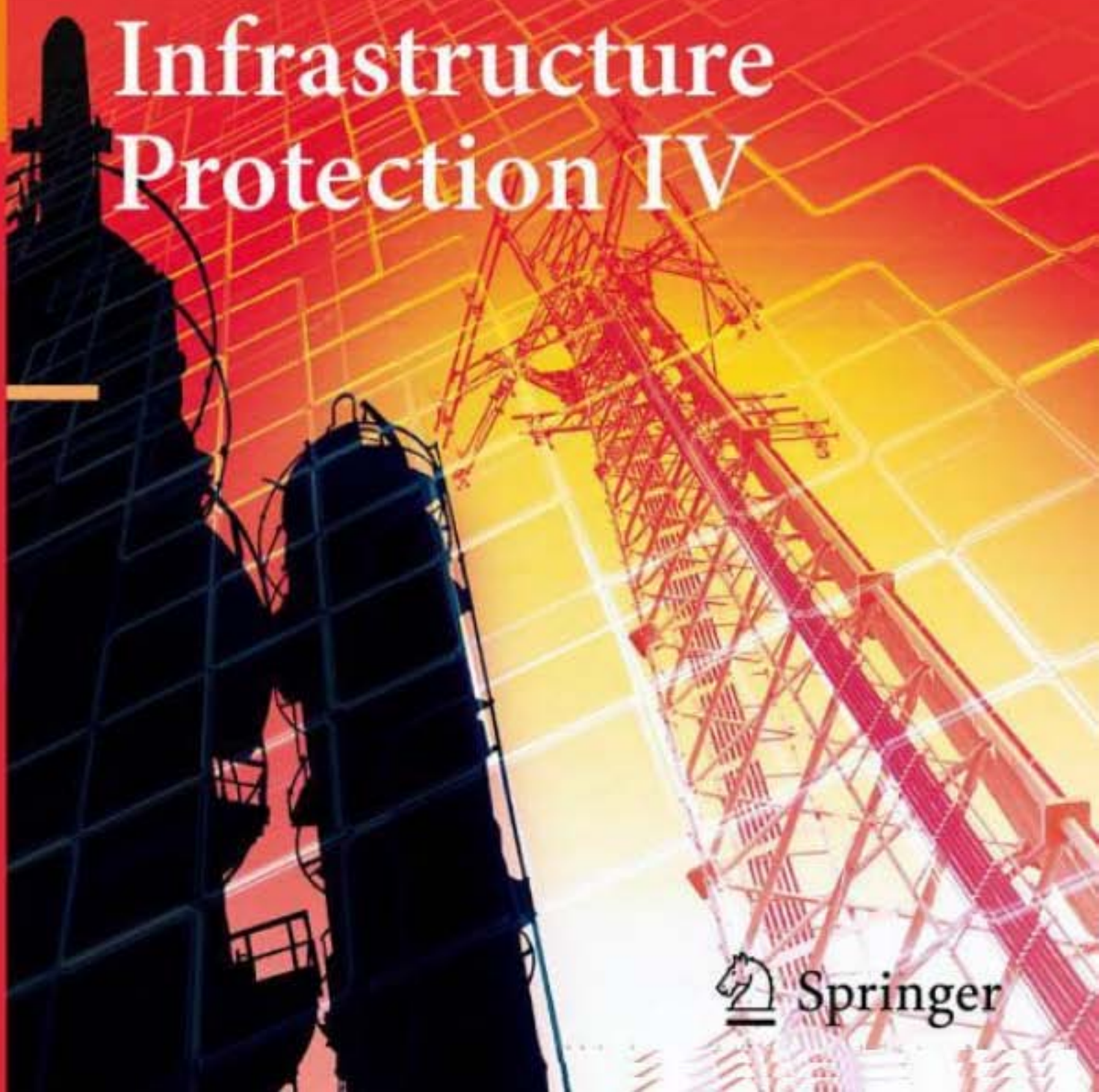
IFIP AICT 342

Tyler Moore
Sujeet Shenoi
(Eds.)

Critical Infrastructure Protection IV



Springer

IFIP Advances in Information and Communication Technology 342

IFIP – The International Federation for Information Processing

IFIP was founded in 1960 under the auspices of UNESCO, following the First World Computer Congress held in Paris the previous year. An umbrella organization for societies working in information processing, IFIP's aim is two-fold: to support information processing within its member countries and to encourage technology transfer to developing nations. As its mission statement clearly states,

> *IFIP's mission is to be the leading, truly international, apolitical organization which encourages and assists in the development, exploitation and application of information technology for the benefit of all people.*

IFIP is a non-profitmaking organization, run almost solely by 2500 volunteers. It operates through a number of technical committees, which organize events and publications. IFIP's events range from an international congress to local seminars, but the most important are:

- The IFIP World Computer Congress, held every second year;
- Open conferences;
- Working conferences.

The flagship event is the IFIP World Computer Congress, at which both invited and contributed papers are presented. Contributed papers are rigorously refereed and the rejection rate is high.

As with the Congress, participation in the open conferences is open to all and papers may be invited or submitted. Again, submitted papers are stringently refereed.

The working conferences are structured differently. They are usually run by a working group and attendance is small and by invitation only. Their purpose is to create an atmosphere conducive to innovation and development. Refereeing is less rigorous and papers are subjected to extensive group discussion.

Publications arising from IFIP events vary. The papers presented at the IFIP World Computer Congress and at open conferences are published as conference proceedings, while the results of the working conferences are often published as collections of selected and edited papers.

Any national society whose primary activity is in information may apply to become a full member of IFIP, although full membership is restricted to one society per country. Full members are entitled to vote at the annual General Assembly, National societies preferring a less committed involvement may apply for associate or corresponding membership. Associate members enjoy the same benefits as full members, but without voting rights. Corresponding members are not represented in IFIP bodies. Affiliated membership is open to non-national societies, and individual and honorary membership schemes are also offered.

Tyler Moore Sujeet Shenoi (Eds.)

Critical Infrastructure Protection IV

Fourth Annual IFIP WG 11.10 International Conference
on Critical Infrastructure Protection, ICCIP 2010
Washington, DC, USA, March 15-17, 2010
Revised Selected Papers

Springer

Volume Editors

Tyler Moore
Harvard University
Cambridge, MA 02138, USA
E-mail: tmoore@seas.harvard.edu

Sujeet Shenoi
University of Tulsa, Department of Computer Science
Tulsa, OK 74104, USA
E-mail: sujeet@utulsa.edu

Library of Congress Control Number: 2010937784

CR Subject Classification (1998): B.8, C.4, B.1.3, B.2.3, B.7.3, C.2, I.6

ISSN 1868-4238
ISBN-10 3-642-16805-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-16805-5 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 219/3180

Contents

PART V RISK MANAGEMENT

Contributing Authors

Craig Allwardt is a Research Scientist at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include knowledge systems, visualization and software architectures.

Ainara Alvarez is a Researcher in the Department of Industrial Organization at the University of Navarra, San Sebastian, Spain. Her research interests include innovation, marketing and risk assessment.

Maria Jesus Alvarez is a Professor of Operations Research at the University of Navarra, San Sebastian, Spain. Her research interests include operations research applied to logistics and systems productivity.

Dan Assaf is a Doctor of Juridical Science (S.J.D.) degree candidate in the Faculty of Law, University of Toronto, Toronto, Canada. His research interests are in the intersection of law, economics and security, in particular, the regulation and governance of information security.

Zaw Zaw Aung is a Ph.D. student in Information Science and Control Engineering at Nagaoka University of Technology, Nagaoka, Japan. His research interests include operational risk management, interdependency analysis and critical infrastructure modeling.

Fabio Bisogni is a Member of the Board of the Formit Foundation, Rome, Italy. His research interests include critical infrastructure protection, critical event management and policy support.

Thomas Brandstetter is a Program Manager at Siemens CERT, Siemens Corporate Research and Technology, Munich, Germany. His research interests include vulnerabilities in critical infrastructures, incident handling methods and economic aspects of IT security.

Peter Burge is an Associate Director at RAND Europe, Cambridge, United Kingdom. His research focuses on modeling choice making behavior, the design and administration of surveys, and the estimation of discrete choice models.

Jonathan Butts is an Assistant Professor of Computer Science at the Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio. His research interests include network, telecommunications and SCADA systems security.

Andrea Carcano is a Ph.D. student in Computer Science at the University of Insubria, Varese, Italy. His research interests include industrial SCADA protocols and architectures.

Simona Cavallini is a Senior Researcher at the Formit Foundation, Rome, Italy. Her research interests include interdependency analysis, economics of security and macroeconomics modeling.

Yousu Chen is a Research Engineer at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include high-performance computing applications, power system operation and decision support, and power system modeling and analysis.

Maria Carla De Maggio is a Researcher at the Complex Systems and Security Laboratory at University Campus Bio-Medico of Rome, Rome, Italy. Her research interests include critical infrastructure protection, risk analysis and risk management.

Ian Ellefsen is a Ph.D. student in Computer Science at the University of Johannesburg, Johannesburg, South Africa. His research interests include critical infrastructure protection and critical information infrastructure protection models for developing nations.

Aaron Engel received his M.S. degree in Computer Science from the University of Tulsa, Tulsa, Oklahoma. His research interests include information assurance, and network and telecommunications system security.

Rajni Goel is an Associate Professor of Information Systems and Decision Sciences at Howard University, Washington, DC. Her research interests include information assurance, digital forensics, control systems security and data mining.

Frank Greitzer is a Chief Scientist at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include situational awareness and decision making in grid operations, cyber security and the insider threat, and applications of cognitive informatics to decision making.

Daniel Guernsey is a Ph.D. student in Computer Science at the University of Tulsa, Tulsa, Oklahoma. His research interests include information assurance, and network and telecommunications system security.

Michele Guglielmi is a Research Trainee at the Joint Research Centre of the European Commission, Ispra, Italy. His research interests include industrial SCADA protocols and architectures.

Mark Hartong is a Senior Electronics Engineer with the Office of Safety, Federal Railroad Administration, U.S. Department of Transportation, Washington, DC. His research interests include information assurance, control systems security, risk analysis and regulatory development.

Zhenyu Huang is a Staff Engineer at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include high performance computing, wide-area measurement technology, information visualization, and power system stability and simulation.

Stig Johnsen is a Senior Research Scientist at SINTEF, Trondheim, Norway. His research interests include information security, SCADA systems, integrated oil and gas operations, and plant safety.

Chong Kim is a Researcher at RAND Europe, Cambridge, United Kingdom. His main research areas are discrete choice modeling and stated preference research in the transportation and health domains.

Ray Klump is an Associate Professor of Mathematics and Computer Science at Lewis University, Romeoville, Illinois; and a Visiting Research Scientist at the Information Trust Institute at the University of Illinois at Urbana-Champaign, Urbana, Illinois. His research interests include electric power system stability and smart grid security.

Konstantin Knorr is a Professor of IT Security in the Computer Science Department at Trier University of Applied Sciences, Trier, Germany. His research interests include SCADA security, authorization models of information and communications systems, and patch management.

Matthew Kwiatkowski is the Cyber Operations Lead in the Cyber Security Program Office at Argonne National Laboratory, Argonne, Illinois; and an Adjunct Professor of Information Security at Lewis University, Romeoville, Illinois. His research focuses on intrusion detection and response mechanisms.

Hans Lund received his B.Sc. degree in Computer Science from Gjovik University College, Gjovik, Norway. His research interests include critical infrastructure protection and communication systems security.

Nils Lunden received his B.Sc. degree in Computer Science from Gjovik University College, Gjovik, Norway. His research interests include security modeling and visualization.

Jian Ma is a Research Engineer at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include power system stability and reliability, renewable integration, wide-area measurement technology and applications of artificial intelligence in power systems.

Patrick Mackey is a Research Scientist at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include visual analytics, data visualization, scientific computation, auditory displays and human-computer interaction.

Marcelo Masera is a Scientific Officer at the Institute for the Protection and Security of the Citizen, Joint Research Center of the European Commission, Ispra, Italy. His research interests include securing networked systems and systems of systems, risk governance and control systems security.

Thomas McEvoy is a Ph.D. student in Mathematics at Royal Holloway, University of London, London, United Kingdom; and a Principal Consultant at Vistorm Ltd., Warrington, United Kingdom. His research interests include the modeling and simulation of critical infrastructures and hybrid systems in relation to security properties.

Igor Nai Fovino is a Scientific Officer at the Institute for the Protection and Security of the Citizen, Joint Research Center of the European Commission, Ispra, Italy; and an Adjunct Professor of Operating Systems at the University of Insubria, Varese, Italy. His research interests include critical infrastructure protection, intrusion detection, secure communication protocols and industrial informatics.

Ainhoa Oses is a Researcher in the Department of Industrial Organization at the University of Navarra, San Sebastian, Spain. Her research interests include manufacturing processes, supply chain management and risk analysis.

Dimitris Potoglou is a Researcher at RAND Europe, Cambridge, United Kingdom. His research involves the design of discrete choice stated preference experiments for understanding individuals' choices and valuations of goods and services.

Mason Rice is a Ph.D. student in Computer Science at the University of Tulsa, Tulsa, Oklahoma. His research interests include network and telecommunications security, and cyberspace deterrence strategies.

Neil Robinson is a Researcher at RAND Europe, Cambridge, United Kingdom. His research interests include critical infrastructure protection, cyber crime and information assurance.

Ute Rosenbaum is a Senior Consultant in IT security at Siemens CERT, Siemens Corporate Research and Technology, Munich, Germany. Her research interests include control systems security, and enhancing development and service processes using security methodologies.

Roberto Setola is the Director of the Complex Systems and Security Laboratory at University Campus Bio-Medico of Rome, Rome, Italy. His research interests include critical infrastructure modeling and analysis, critical infrastructure protection, risk assessment and control strategies for complex systems.

Sujeet Shenoi, Chair, IFIP Working Group 11.10 on Critical Infrastructure Protection, is the F.P. Walter Professor of Computer Science at the University of Tulsa, Tulsa, Oklahoma. His research interests include information assurance, digital forensics, critical infrastructure protection, reverse engineering and intelligent control.

Sean Smith is an Associate Professor of Computer Science at Dartmouth College, Hanover, New Hampshire. His research interests include trusted computing and usable security.

Rouslan Solomakhin is a Software Engineer at Microsoft, Redmond, Washington. His research interests include information assurance, network security and cloud computing.

Robin Sveen received his B.Sc. degree in Computer Science from Gjovik University College, Gjovik, Norway. His research interests include software security and critical infrastructure protection.

Nils Svendsen is an Associate Professor of Computer Science at the Norwegian Information Security Laboratory, Gjovik University College, Gjovik, Norway. His research interests include the modeling and simulation of critical infrastructures, graph theory, cryptography and coding theory.

Alberto Trombetta is an Assistant Professor of Computer Science and Communication at the University of Insubria, Varese, Italy. His research interests include data security and privacy, data integration, query languages, imprecise data management and systems security.

Marcella Trombetta is a Professor of Chemical Fundamentals in Technology at University Campus Bio-Medico of Rome, Rome, Italy. Her research interests include the synthesis and characterization of new materials for biomedical applications, energy and the environment.

Patrick Tsang, who passed away on October 27, 2009, was a Ph.D. student in Computer Science at Dartmouth College, Hanover, New Hampshire. His research interests included cryptography, network security and privacy-enhancing technologies.

Sebastiaan von Solms is a Research Professor in the Academy for Information Technology at the University of Johannesburg, Johannesburg, South Africa. His research interests include information security and critical information infrastructure protection.

Richard Warnes is a Researcher at RAND Europe, Cambridge, United Kingdom. His research interests include counterterrorism, policing and intelligence.

Kenji Watanabe is a Professor in the Graduate School of Social Engineering, Nagoya University of Technology, Nagoya, Japan. His research areas include IT risk management, business continuity and critical infrastructure protection.

Duminda Wijesekera is an Associate Professor of Information and Software Engineering at George Mason University, Fairfax, Virginia. His research interests include information, network, telecommunications and control systems security.

Stephen Wolthusen is a Professor of Information Security at the Norwegian Information Security Laboratory, Gjovik University College, Gjovik, Norway; and a Reader in Mathematics at Royal Holloway, University of London, London, United Kingdom. His research interests include critical infrastructure modeling and simulation, and network and distributed systems security.

Pak-Chung Wong is a Chief Scientist and Project Manager at the Pacific Northwest National Laboratory, Richland, Washington. His research interests include visual analytics, power grid analytics, graph and network analytics, and multimedia analytics.

Preface

The information infrastructure – comprising computers, embedded devices, networks and software systems – is vital to operations in every sector: information technology, telecommunications, energy, banking and finance, transportation systems, chemicals, agriculture and food, defense industrial base, public health and health care, national monuments and icons, drinking water and water treatment systems, commercial facilities, dams, emergency services, commercial nuclear reactors, materials and waste, postal and shipping, and government facilities. Global business and industry, governments, indeed society itself, cannot function if major components of the critical information infrastructure are degraded, disabled or destroyed.

This book, *Critical Infrastructure Protection IV*, is the fourth volume in the annual series produced by IFIP Working Group 11.10 on Critical Infrastructure Protection, an active international community of scientists, engineers, practitioners and policy makers dedicated to advancing research, development and implementation efforts related to critical infrastructure protection. The book presents original research results and innovative applications in the area of infrastructure protection. Also, it highlights the importance of weaving science, technology and policy in crafting sophisticated, yet practical, solutions that will help secure information, computer and network assets in the various critical infrastructure sectors.

This volume contains seventeen edited papers from the Fourth Annual IFIP Working Group 11.10 International Conference on Critical Infrastructure Protection, held at the National Defense University, Washington, DC, March 15–17, 2010. The papers were refereed by members of IFIP Working Group 11.10 and other internationally-recognized experts in critical infrastructure protection.

The chapters are organized into five sections: themes and issues, control systems security, infrastructure security, infrastructure modeling and simulation, and risk management. The coverage of topics showcases the richness and vitality of the discipline, and offers promising avenues for future research in critical infrastructure protection.

This book is the result of the combined efforts of several individuals and organizations. In particular, we thank Daniel Guernsey, Jonathan Butts, Mason Rice, Heather Drinan and Nicole Hall Hewett for their tireless work on behalf

of IFIP Working Group 11.10. We gratefully acknowledge the Institute for Information Infrastructure Protection (I3P), managed by Dartmouth College, for supporting IFIP Working Group 11.10. We also thank the Department of Homeland Security and the National Security Agency for their support of IFIP Working Group 11.10 and its activities. Finally, we wish to note that all opinions, findings, conclusions and recommendations in the chapters of this book are those of the authors and do not necessarily reflect the views of their employers or funding agencies.

Tyler Moore and Sujeet Shenoi

Chapter 1

SECURITY AT WHAT COST?

Neil Robinson, Dimitris Potoglou, Chong Kim, Peter Burge and Richard Warnes

Abstract In the presently heightened security environment in the United Kingdom there are a number of examples of policy that must strike a delicate balance between strengthening security and endangering civil liberties and personal privacy. The introduction of national identity cards and biometric passports, expansion of the National DNA Database and interdepartmental sharing of personal information raise a number of privacy issues. Human rights may also be suspended by the exercise of stop-and-search powers by the police or the detention of suspects prior to trial. However, much of the current debate concerning civil liberties and security is adversarial, and little robust research data informs arguments on both sides. This paper outlines the results of a study that attempts to objectively understand the real privacy, liberty and security trade-offs made by individuals, so that policymakers can be better informed about the preferences of individuals with regard to these important issues.

Keywords: Security measures, stated preferences, trade-offs

1. Introduction

The entities responsible for protecting critical infrastructures such as transportation networks and physical assets often have to make difficult decisions in the face of considerable uncertainty regarding the imposition of security measures to mitigate the risks due to a particular threat. Where individuals are involved in critical infrastructures – as users or consumers of a service or product that the specific sector provides – their civil liberties or privacy may be affected. Contemporary examples of security measures that affect privacy or civil liberties include: (i) new forms of body scanning technologies; (ii) closed-circuit television (CCTV); (iii) fingerprint identification, facial recognition and other biometric identification systems; and (iv) the sharing, mining and use of personal information by government agencies.

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 3–15, 2010.

Most attempts to provide an evidence base for understanding the preferences and views of users of security measures are largely based on opinion polls, surveys or qualitative research. These approaches have limitations because they only permit absolute (Yes/No) responses to questions, and are generally not conducive to instances in which individuals are faced with a series of realistic choices that may have different effects on their privacy, liberty or security. Recent examples include the Westin-Harris privacy surveys [17], a Gallup Flash Eurobarometer survey conducted for the European Commission [31], a British Social Attitudes Survey [15], and tracking research conducted for the Home Office's National Identity Scheme [5]. These approaches suffer from three main limitations: (i) they are generally one-dimensional, unrealistic, and ask abstract, one-off questions that lead to polarized preferences towards absolutes instead of grading choices involving privacy, liberty and security trade-offs; (ii) they do not quantify the extent to which people may be prepared to give up civil liberties or privacy; and (iii) they cannot be integrated easily into an economic appraisal toolkit.

This paper reports on the application of stated preference discrete choice experiments (SPDCEs) for understanding, quantifying and, in some cases, monetizing the privacy, liberty and security trade-offs made by individuals. In particular, the research questions addressed are:

- Given that national security is a non-market public good, does the use of stated preference techniques have merit for gathering data on the willingness of individuals to make trade-offs?
- If so, what drives choice when individuals decide to relinquish or surrender their liberty or privacy in order to obtain security benefits?
- Is it then possible to monetize the impacts of these security measures on liberty and privacy?

2. Research Methodology

Our study used SPDCEs to investigate the importance of specific drivers for the choices made by individuals (see, e.g., [11]). These techniques have been used extensively in marketing, healthcare, environmental and transportation economics [18–20, 29]. In combination with discrete choice analysis, SPDCEs offer the potential to provide empirical evidence for making informed decisions, for example, regarding the importance that individuals attach to advanced CCTV cameras supported by real-time, face recognition technology. As national security and privacy may be considered to be examples of non-market public goods (like healthcare and the environment), there is some validity to the application of these techniques in the domain of interest. Furthermore, the use of a methodology that permits the identification of real choices and the trade-offs that individuals are prepared to make contrasts well with the "top-down" risk-based approach in use by government, which matches vulnerabilities and threats against resource investments (see, e.g., [13]). Finally, the methodology

may assist cost-benefit decision-making processes dealing with the economic evaluation of security measures, since it can determine the threshold at which individuals are prepared to tolerate privacy and civil liberty intrusions in the name of security.

2.1 Case Studies

Based on a review of the literature and semi-structured interviews with representatives from both sides of the national security versus civil liberties debate, we identified three contexts for applying the experimental methodology: (i) applying for a passport, where individuals provide personal information; (ii) traveling on the national rail network, where individuals may be under the surveillance of CCTV networks; and (iii) attending a major public event, where individuals may be subject to identification processes and interact with various security officials.

Attributes describing each case study and their values were derived from information available in the public domain such as estimates of the numbers of terrorist suspects [16], ongoing conspiracies [22] and illegal immigrants [3]. Information about the processing time of passport applications and the personal data collected during the application process was obtained from [7, 32]. The design and specification of the case studies are described in more detail in [28].

2.2 Data Collection

The SPDCEs were conducted over the Internet between September 17 and 19, 2008. The survey was pre-tested and modified in accordance with post-survey cognitive questions by 260 individuals between June 27 and 29, 2008.

Invitations to participate in the survey were emailed to 15,214 individuals registered with Research Now [27], a market research company with the largest panel of Internet users in the United Kingdom. Individuals who did not meet the eligibility criteria (e.g., 18 years or older), provided incomplete information or belonged to sample quotas that were already filled were eliminated. A total of 2,058 participants were recruited.

Table 1 presents the descriptive statistics of the survey sample compared with those from the 2001 U.K. Census [24]. While the survey sample is not representative of the U.K. population, it covers an active segment of the population that matches the demographic profiles (i.e., age and gender) in the 2001 U.K. census.

2.3 Model Development

Following an initial discrete choice model that used only the attribute levels of the experiments, alternative specifications of the model that included socio-demographic characteristics of respondents and their attitudes were employed to test whether certain groups of respondents placed different valuations on any of the attributes. Possible differences were identified by examining cross tables

Table 1. Sample characteristics compared with the 2001 U.K. Census.

Variable	**Sample (%)**	**2001 Census (%)**
Gender (Females)	52	52
Age Group		
18–24	7	16
25–34	13	16
35–44	19	19
45–54	18	16
55–64	21	14
65+	22	20
Education Level		
None	10	29.1
O Level/GCSE	32	59.6
A Level/CSE	26	8.3
Degree	32	19.8
Other		6.9
Occupational Status		
Fulltime	42	59.6
Parttime	16	
Student	4	7.2
Retired	28	13.4
Seeking Work	3	4.5
Other	7	15.3
Income		
Below £30,000	58	
£30,000 to £69,999	26	
£70,000 and Higher	2	
Not Reported	14	

that summarized the in-sample predictive ability of the model. This approach enabled us to address key differences in the choices made by individuals within the sample. The SPDCE method is consistent with utility maximization and demand theory [19, 26]. After the parameter estimates were obtained using the most appropriate model, a willingness-to-pay (WTP) measure for changes across different levels of attributes was computed using the equation [11]:

$$WTP = -\beta_{price}^{-1} \ln \frac{\sum_i \alpha_i e^{V_i^1}}{\sum_i \alpha_i e^{V_i^0}}$$

where β_{price} is the coefficient of the price increase on a ticket to cover security; V_i^0 is the utility of the base level (e.g., no CCTV) for a segment of the sample (e.g., males) with proportion α_i; and V_i^1 is the utility of the same segment

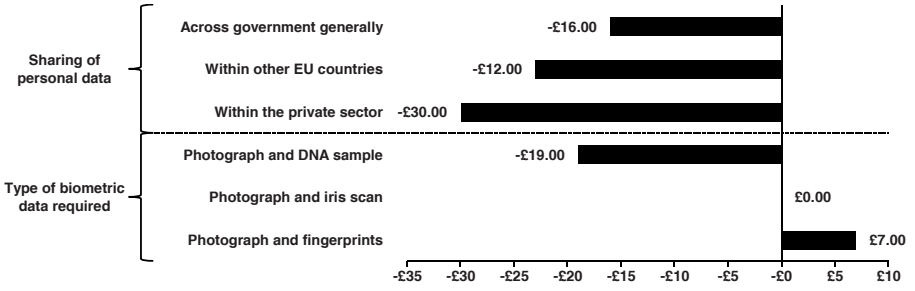


Figure 1. Willingness to pay for security (passport application).

for a security improvement (e.g., CCTV) compared to the base level. Complete details about the results of the model estimation and WTP estimates are provided in [28].

3. Results and Discussion

This section discusses the experimental results for the case studies involving passport applications, national rail travel and public event attendance.

3.1 Passport Applications

Due to heightened concerns about national security and identity theft, there is considerable debate and political pressure to implement ID cards, a National Identity Register (NIR) and biometric passports, all of which will have significant amounts of personal information. It is expected that this information will be shared among government organizations responsible for security, border management and immigration. The current U.K. passport application process is already raising concerns about privacy and civil liberties being relegated in favor of national security. Citizens are required to provide a significant quantity of personal information with their passport applications because the information can help fight against "social bads" such as illegal immigration and terrorism.

The security characteristics of biometric passports may affect privacy and liberty in several ways. For example, personal information collected for the purpose of law enforcement may be shared (mistakenly or deliberately) with other organizations not associated with achieving security objectives, possibly resulting in discrimination or disenfranchisement of individuals based on the identity information stored in their passports. As more organizations are permitted to use this personal information, the risk of abuse and mistakes increases.

Our experimental data indicated a universal degree of discomfort in the provision of advanced forms of biometric information (e.g., DNA) as part of the passport application process (Figure 1). Respondents were only willing to accept (i.e., they derived negative utility from) the collection of DNA and photograph data at the time of a passport application if there was a subsidy of

£19 in the cost of a passport. The respondents preferred to provide personal information in the form of a photograph or fingerprint, and they indicated a willingness to pay £7 for this privilege. This finding is relevant given recent policy statements that indicate that fingerprint biometrics will be collected as part of the passport application process [34]. Note that there is no requirement to submit further biometric information at this time because a facial biometric is compiled from the passport photograph [8].

More worrisome from a privacy perspective were the responses to the question of the sharing of personal information collected during the passport application process with other organizations in the public or private sectors. Indeed, this question provoked universal discomfort in the respondents. All else being equal, the respondents preferred to see their personal information kept within the Identity and Passport Service, and not shared with other government departments, other European nations or the private sector. This has a number of important policy implications – most notably, if the desire by the public sector to use the collected information to achieve efficiencies or help in the fight against organized crime, illegal immigration and terrorism matches the preferences of the general public [25]. Furthermore, there is the question of consent and choice and if this may ever be construed as meaningful given the extent of the demand for passports.

The survey also shows that large incentives (e.g., a discount on the passport fee of as much as £30) would be required to reach a threshold where the respondents would be comfortable sharing their personal information with third parties. Respondents indicated that sharing information with the private sector was the least preferred alternative, and they would be willing to accept this only if the price of a passport was discounted by £30. A subsidy of £23 would have to be provided in order to share information with other European nations, and £16 to share information with other government departments.

Evidence from this case study appears to contradict current government policy, particularly regarding the sharing of NIR information (which may be collected as part of the passport application process) with the private sector or other government departments as part of the "identity assurance" policy agenda. For example, it has been suggested that banks may wish to use the identity information in the NIR as a government-authenticated identity, removing the need for customers to present other credentials when applying for a bank account [4]. Finally, with regard to sharing information with other countries, the European Secure Identity Across Borders Linked (STORK) Project [30] is evaluating methods to do just this – sharing information between EU member states to deliver pan-European services such as the European electronic health insurance card [33]. The existence of such compelling evidence regarding the preferences of the survey participants suggests that policymakers ought to explore and consider the implications of collecting and sharing personal information, whether a subsidy is necessary, or whether to consider (at the very least) the unintended consequences of implementing policies that are contradictory to individual preferences.

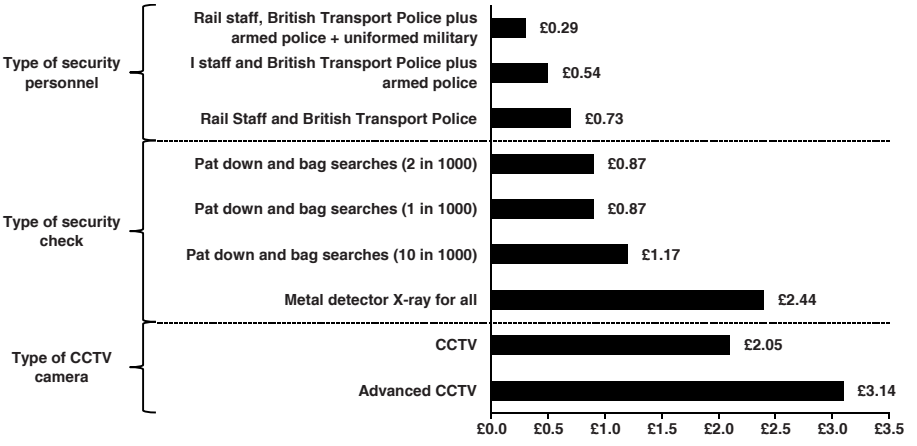


Figure 2. Willingness to pay for security (rail travel).

3.2 National Rail Travel

The terrorist attacks on public transportation systems around the world have made safety and security a top priority in the policy agenda of many countries, particularly the United Kingdom. Security measures for air travel have historically received a great deal of attention, but authorities are now increasingly focusing on land-based mass transit systems. These systems have become targets for terrorist groups due to their vulnerability and ease of access arising from their intrinsically open nature.

In the United Kingdom, measures to address security threats include legislation and regulations as well as campaigns that raise public awareness of the risk of attacks. The Transport Security and Contingencies Team (TRANSEC) of the U.K. Department of Transport [6] plays an important role with regard to security arrangements for multi-modal transportation systems. Its task is complicated by the fact that many transportation systems are privately owned.

Several attributes compete with privacy and liberty in this case study: most notably, the presence of security personnel who may inadvertently detain individuals. The presence of CCTV cameras has an impact on privacy as do other types of security checks, which could be regarded as an invasion of personal space (e.g., security personnel going through bags and personal effects).

Security mechanisms that may affect personal privacy or civil liberties when traveling on the national rail network were viewed more favorably by the survey respondents (Figure 2). This may be due to familiarity: in contrast with sharing personal information in the passport case study, which is relatively abstract and distant, the security mechanisms present in this case, such as CCTV and security arches, are much more physically present and perceptively "closer" to the individual. This is seen in the preferences regarding X-ray machines or physical "pat downs" and bag searches; the latter being considered as more invasive, perhaps due to their physical intrusiveness. Despite this, the poten-

tial to exercise the right to privacy under this security measure may be less restricted than when personal information is collected when passing through an X-ray arch, where data may be recorded, shared with others and stored for a longer period of time with little, if any, self-determination by the individual.

Individuals were comfortable with more intrusive types of security cameras (e.g., face detection systems) as they seemed to outweigh concerns related to personal privacy and civil liberties. Indeed, the extent to which this finding is representative of the oft-discussed "surveillance society" is interesting, since it illustrates a degree of familiarity with privacy-invasive forms of technology such as CCTV cameras [1].

However, there remains the question about the extent to which context plays a role. Many individuals have identified that being monitored by CCTV of any form in the environment of a railway station is an acceptable sacrifice to obtain the security benefits. Similarly, the evidence may illustrate confusion about the perception that CCTV is a tool for detecting low-level street crime such as burglary, mugging and anti-social behavior, rather than for dealing with more complex forms of criminal behavior or terrorism [10].

The findings regarding the degree of comfort attached to different types of security checks are counterintuitive. We anticipated that security checks with an obvious privacy implication would be less preferable than others with which individuals are more familiar. However, the evidence indicates that individuals are much more comfortable with passing through an X-ray arch or scanner than being subjected to a security pat down or bag search. Understandably, these are more privacy-invasive due to their personal and physical nature but, by comparison, the information recorded by a metal detector or X-ray scanner may adversely affect personal privacy in a broader manner as it may be recorded and passed on. There is also the extent to which pat downs and bag searches are more effective from a security perspective. Historical evidence from the Israeli airline El-Al indicates that alert, trained staff who can spot indicative behavior patterns can be a very effective security measure.

Finally, and somewhat unsurprisingly, there was a high degree of comfort expressed for more specialized security personnel, albeit up to a point. Despite the perception in the security community that the deployment of armed police or the military creates an atmosphere of fear, in all cases, the survey respondents were willing to pay for security personnel; in fact, no negative utility was identified. Regarding the visible presence of uniformed military personnel, as was seen, for example, at London's Heathrow Airport in 2003 [2], most survey respondents were willing to pay for these measures, but less so than other "low key" forms of security personnel. Also, the respondents felt that the effectiveness of uniformed military personnel was not correlated with an increasing level of sophistication.

3.3 Public Event Attendance

There is widespread concern regarding security at major sporting and entertainment events, particularly given the terrorist attacks at the 1972 Munich

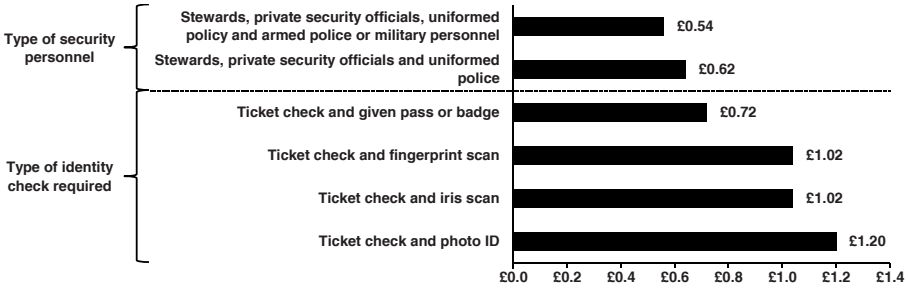


Figure 3. Willingness to pay for security (public event attendance).

Olympics and the 1984 Conservative Party Conference in Brighton. Such events are recognized as prime terrorist targets because they involve large concentrations of members of the general public [12]. This is on top of the challenge of maintaining security given the porous perimeters of most venues. In preparation for the 2012 London Olympics, a number of security measures are being considered, including monitoring, access control, overhead surveillance and CCTV systems [21].

The measures implemented at major public events to deal with security may affect liberty in a number of ways. These include the impact on privacy resulting from the collection of personal information upon entry to an event, various forms of personal information being used to verify the identity of an individual, and the possibility of detention by security authorities.

In the major public event case study, the survey respondents preferred to have some form of identity check. However, all else being equal, they were less likely to pay for checks that would require biometric identification (Figure 3). Based on an expected ticket price of £40 for attending the opening ceremonies of the 2012 Olympic Games, the respondents were willing to pay £1.20 for an identity check based on a picture ID and an examination of the ticket. Biometric checks such as fingerprint and iris scans were less preferable; individuals were prepared to pay £1.02 for these forms of identity checks. This could be explained by the acceptance that it would be necessary to check the identity of the individual presenting the ticket in order to ensure that he/she is a legitimate ticket holder.

The more interesting finding is that, despite media reports about concerns regarding the use of biometric technologies, individuals are willing to pay for the checks and accommodate civil liberty intrusions to achieve the security objectives. This is reinforced by the finding that survey respondents were willing to pay less (£0.72) for a simple ticket check that does not involve identity information than one involving some form of personal or biometric information. This evidence is relevant to the discussions regarding possible security technologies for administering entry to events at the 2012 London Olympics. As such, it is pertinent to note that the Olympic Delivery Authority is considering "facial and palm" biometric identification for workers at Olympic sites [23].

4. Conclusions

The views and preferences of citizens as users of security infrastructures can be quantified and, in some cases, monetized. This information can be used to support security investment decisions that balance the risk of an incident versus the costs and implications of implementing security infrastructures to mitigate the risk.

The methodology used is based on the expectation that individuals act rationally. For example, when presented with a set of alternatives, individuals tend to choose the option that best satisfies their needs. This notion is the cornerstone of neoclassical economics. The diagnostic and evaluative questions asked of the survey respondents facilitated the understanding, measurement and economical quantification of the relative degree of comfort or distaste for security measures. The results provide useful indicators of current concerns about how security measures may affect privacy and liberty.

The rational actor model employed in this work is the basis of many investment decisions in public policy. This study can shed light on where policy and preferences differ and, thus, assist policymakers in making informed, evidence-based decisions as to whether the cost of contravening or ignoring user preferences outweighs the benefits of implementing security measures. Similarly, it might be possible to identify where the measures could be adjusted to take better account of preferences without undermining security gains.

Although the philosophical and moral aspects of the valuation of human life, privacy and civil liberties may be difficult to accept, the real uncertainty is in understanding and quantifying the expected security benefits of certain types of infrastructure. These benefits might be expressed in terms of lives saved or terrorist incidents prevented. Some studies [9] have quantified the overall loss of life and economic damage arising from terrorist incidents, but as of yet there is little or no actuarial data to link the measures to benefits.

This methodology can also support policymaking and security decisions regarding the data to use as input in risk assessments. The approach may have particular relevance in privacy impact assessments, a relatively new policy tool that considers the privacy perceptions of the "users" of policy initiatives when designing security measures [14]. Finally, the application of the methodology can bring a degree of objectivity to the highly charged debate on striking the right balance between civil liberties and security. Ultimately, this study shows that using the metaphor of "balance" is counterproductive without robust measurements of the weights of the factors that are balanced.

References

[1] K. Ball, D. Lyon, D. Wood, C. Norris and C. Raab, A Report on the Surveillance Society, The Surveillance Studies Network, London, United Kingdom, 2006.

[2] P. Barkham, Heathrow show of force after terror alert, *The Times*, February 12, 2003.

[3] BBC News, Illegal immigrant figure revealed, London, United Kingdom (news.bbc.co.uk/2/hi/uk_news/politics/4637273.stm), June 30, 2005.

[4] BBC News, In full: Smith ID card speech, London, United Kingdom (news.bbc.co.uk/2/hi/uk_news/politics/7281368.stm), March 6, 2008.

[5] Central Office of Information, Identity and Passport Service, National Identity Scheme Tracking Research Wave 3, Home Office, London, United Kingdom, 2008.

[6] Department for Transport, Responsibilities of Transport Security's Land Transport Division, London, United Kingdom (www.dft.gov.uk/pgr/security/land/responsibilitiesoftransports4898).

[7] Directgov, Timetable for passport applications, Her Majesty's Government, London, United Kingdom (www.direct.gov.uk/en/TravelAndTransport/Passports/howlongittakesandurgentappplications/DG_174148), 2008.

[8] Directgov, Table of passport fees, Her Majesty's Government, London, United Kingdom (www.direct.gov.uk/en/TravelAndTransport/Passports/howlongittakesandurgentappplications/DG_174 109), 2009.

[9] W. Enders and T. Sandler, Distribution of transnational terrorism among countries by income class and geography after 9/11, *International Studies Quarterly*, vol. 50(2), pp. 367–393, 2006.

[10] D. Farrington, M. Gill, S. Waples and J. Argomaniz, The effects of closed-circuit television on crime: Meta-analysis of an English national quasi-experimental multi-site evaluation, *Journal of Experimental Criminology*, vol. 3(1), pp. 21–28, 2007.

[11] D. Hensher, J. Rose and W. Greene, *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge, United Kingdom, 2005.

[12] Her Majesty's Government, Countering International Terrorism: The United Kingdom's Strategy, London, United Kingdom (www.fco.gov.uk/resources/en/pdf/contest-report), 2006.

[13] Her Majesty's Treasury, The Green Book: Appraisal and Evaluation in Central Government, London, United Kingdom (www.hm-treasury.gov.uk/d/green_book_complete.pdf), 2003.

[14] Information Commissioner's Office, Privacy Impact Assessment Handbook (Version 2.0), Wilmslow, United Kingdom (www.ico.gov.uk/upload/documents/pia_handbook_html_v2/files/PIAhandbookV2.pdf), 2009.

[15] M. Johnson and C. Gearty, *A Price Worth Paying? Changing Public Attitudes to Civil Liberties Under the Threat of Terrorism, British Social Attitudes: The 23rd Report – Perspectives on a Changing Society*, Sage Publications, London, United Kingdom, 2007.

[16] P. Johnston, Yard is watching thousands of terror suspects, *Daily Telegraph*, September 2, 2006.

[17] P. Kumaraguru and L. Cranor, Privacy Indexes: A Survey of Westin's Studies, Technical Report CMU-ISRI-5-138, Institute for Software Research International, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2005.

[18] J. Louviere, Experimental choice analysis: Introduction and overview, *Journal of Business Research*, vol. 23(4), pp. 89–96, 1991.

[19] J. Louviere, D. Hensher and J. Swait, *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, United Kingdom, 2000.

[20] J. Louviere and G. Woodworth, Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregated data, *Journal of Marketing Research*, vol. 20(4), 350–367, 1983.

[21] J. Merrick, Security bill for London's 2012 Olympics to hit £1.5bn – Triple the original estimate, *The Independent*, September 28, 2008.

[22] R. Norton-Taylor, MI5: 30 terror plots being planned in the UK, *The Guardian*, November 10, 2006.

[23] A. O'Connor and J. Sherman, Biometrics screening for Olympic workers, *The Times*, March 5, 2008.

[24] Office for National Statistics, Census 2001, London, United Kingdom (www.statistics.gov.uk/census2001/census2001.asp).

[25] D. Omand, The National Security Strategy: Implications for the UK Intelligence Community, Discussion Paper, Institute of Public Policy Research, London, United Kingdom, 2009.

[26] J. Ortuzar and L. Willumsen, *Modeling Transport*, John Wiley, Chichester, United Kingdom, 2001.

[27] Research Now, Welcome to Research Now, London, United Kingdom (www.researchnow.co.uk).

[28] N. Robinson, D. Potoglou, C. Kim, P. Burge and R. Warnes, Security at What Cost? Quantifying People's Trade-offs Across Liberty, Privacy and Security, Technical Report TR-664, RAND Europe, Cambridge, United Kingdom, 2010.

[29] M. Ryan, A. Bate, C. Eastmond and A. Ludbrook, Use of discrete choice experiments to elicit preferences, *Quality in Health Care*, vol. 10(1), pp. 55–60, 2001.

[30] STORK, The Secure Identity Across Borders Linked (STORK) Project, Madrid, Spain (www.eid-stork.eu).

[31] The Gallup Organization, Data Protection in the European Union – Citizens' Perceptions: Analytical Report, Flash Eurobarometer 225, Brussels, Belgium (ec.europa.eu/public_opinion/flash/fl_225_en.pdf), 2008.

[32] The National Archives, Question the head of the ID card scheme, London, United Kingdom (www.number10.gov.uk/Page10364), November 14, 2006.

[33] The NETC@RDS Project, NETC@RDS: A step towards the electronic European health insurance card, Luxembourg (netcards-project.com/web/frontpage).

[34] ZDNet UK, Government U-turns on passport pledge, London, United Kingdom, October 1, 2009.

Chapter 2

FOREIGN DIRECT INVESTMENT IN AN ERA OF INCREASED THREATS TO CRITICAL INFRASTRUCTURES

Dan Assaf

Abstract The need to maintain national security while deriving the benefits of global economic liberalization presents a significant challenge for governments attempting to privatize critical infrastructure assets. In the post September 11, 2001 world, the notion that foreign direct investment positively contributes to an economy is being tempered by the realization that it can pose a threat to national security. This paper discusses the principal issues that governments must consider when authorizing foreign investment in critical infrastructures. The policies of the United States and Israel are compared to focus and clarify the challenges associated with using a national security rationale to constrain foreign investment.

Keywords: Foreign direct investment, critical infrastructures, United States, Israel

1. Introduction

An important issue in international trade and investment is the control and ownership of critical infrastructure assets by foreign corporations and governments. Although foreign control and ownership may be a viable economic strategy, they can directly jeopardize national security.

Critical infrastructures rely extensively on information and communications technologies that are susceptible to cyber threats. At the same time, critical infrastructure assets are undergoing privatization and deregulation processes that present attractive opportunities for foreign investors.

The conflict between economics and security interests has intensified primarily because of two recent global developments. The first is global economic liberalization and integration. The second is the change in the nature of the global security environment. Both these aspects are evident in the 2006 Dubai

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 17–27, 2010.

Ports World controversy, which demonstrated how tension is exacerbated when the security of critical infrastructures is involved.

This paper discusses the challenges that governments face concerning foreign direct investment in critical infrastructures. On the one hand, governments seek to increase foreign investment and endorse free trade to promote economic growth and prosperity. On the other hand, they must protect their citizens from threats to critical infrastructure assets that underlie national economies. This conflict, which is influenced by competing political and ideological perspectives, can lead to the adoption of policies that upset the delicate balance required to maintain the benefits of open investment and a secure homeland. The policies of the United States and Israel are compared to illustrate the challenges associated with using a national security rationale to constrain foreign investment.

2. Changing National Security Threats

Critical infrastructure protection is a concept *du jour* in many developed countries. Faced with the inherent vulnerabilities of critical infrastructures to physical and cyber attacks, governments around the world have become very preoccupied with their state of security.

The United States Critical Infrastructure Protection Act of 2001 defines critical infrastructures as "those systems and assets, whether physical or virtual, so vital to the U.S. that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters" [14]. The U.S. has identified eighteen critical infrastructure sectors: agriculture and food; defense industrial base; energy; healthcare and public health; banking and finance; water; chemicals; commercial facilities; critical manufacturing; dams; emergency services; nuclear reactors; information technology; communications; postal and shipping; transportation; government facilities; and national monuments and icons.

Nation states and terrorist organizations pose viable physical and cyber threats to critical infrastructure assets; however, the cyber threat has grown in recent years. The heavy reliance on information and communications technologies by critical infrastructures creates new vulnerabilities that can be exploited by "information warfare." Information warfare is considered to be asymmetric in nature because it enables a weaker adversary to counterbalance the military and strategic superiority of a stronger power at relatively low cost. The potential damage from a cyber attack is aggravated by the interdependencies existing between critical infrastructures.

Consider, for example, a scenario where Iran fears that the United States might attack its nuclear facilities. Recognizing that it may not be able to withstand a full-scale American military assault, Iran decides to initiate a series of asymmetric attacks against the United States intended to diminish America's strategic advantages in the conventional military sphere. These attacks combine physical attacks against American facilities around the world (e.g., embassies and military bases) with cyber attacks on American critical infrastructure as-

sets (e.g., power grids, telecommunications networks, oil and gas facilities, and pipelines). The physical attacks result in 500 American casualties. However, the cyber attacks produce blackouts in the Northeastern United States, which adversely affect telecommunications services, energy production, air and land transportation, banking and financial services, emergency services, etc. The cyber attacks bring major industries to a grinding halt and cause enormous economic losses. At the same time, the Iranian military, exploiting the havoc and panic in America, launches a pre-emptive attack against U.S. forces in the Persian Gulf. By combining physical and cyber attacks, Iran may, in fact, be able to inflict significant damage to vastly superior American military forces in the Gulf.

This scenario illustrates how cyber attacks on critical infrastructure assets can be used as part of a military campaign. However, information warfare may be used independently of military action. This was demonstrated by the denial-of-service attacks on Estonia in 2007, which disabled the websites of government agencies, financial institutions and media outlets for several days [9, 12]. While no casualties occurred as a result of these attacks, the damage to the Estonian economy was significant.

3. Foreign Direct Investment

Since the late 1970s, governments have increasingly adopted policies that support global economic liberalization and integration. For example, the Washington Consensus prescribes a set of policies that encourage market liberalization, privatization and deregulation [17]. As a result, essential products and services that were traditionally produced or provided by the state are being produced or provided by private actors. In addition, the adoption of the Washington Consensus has resulted in the removal of barriers to trade and foreign ownership. These factors, no doubt, contribute to the rising presence of foreign companies, including multi-national enterprises, in domestic economies.

Not surprisingly, the United States is both the world's largest foreign direct investor and the world's largest recipient of foreign direct investment. As such, America rigorously promotes policies that enhance free trade and reduce restrictions and barriers on foreign direct investment. Figure 1 shows recent trends in American foreign direct investment, both as a recipient and as an investor.

Foreign direct investment is considered to be a necessary element in the economic policy of a developed country. Graham and Marchick [6] identify several positive effects of foreign direct investment on the American economy. First, because American savings are insufficient to finance domestic investment, the United States depends heavily on the flow of money from foreign investors. Second, foreign investments create more jobs and these jobs often pay higher salaries than jobs in American-based firms. Third, foreign companies tend to invest in research and development and, in some cases, they invest more than their American counterparts. Fourth, foreign investment positively affects

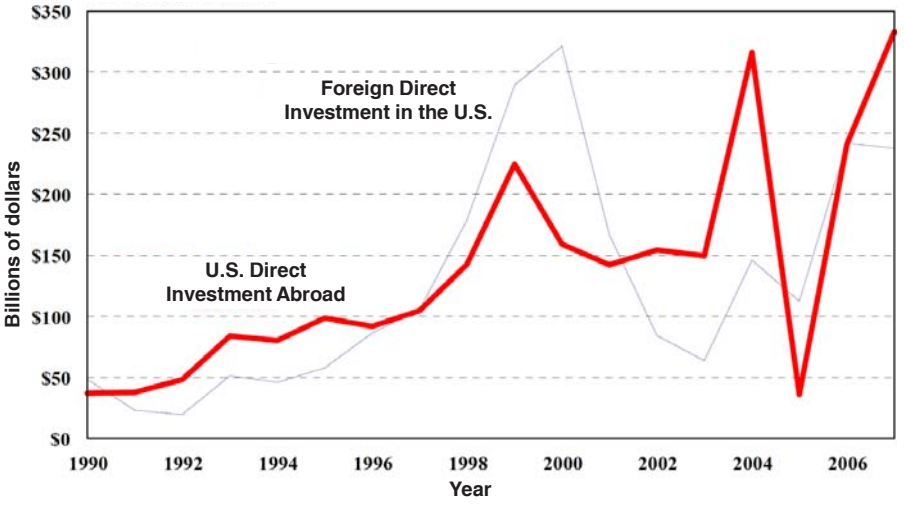


Figure 1. Foreign direct investment in the United States [8].

economic growth. Finally, foreign investment leads to higher productivity and improved product quality.

4. Comparative Analysis

With the seemingly opposing goals of economics and security, the protection of critical infrastructure assets poses a unique challenge. Countries that follow free market principles tend to privatize and deregulate these infrastructures, possibly opening the door to foreign participation. Meanwhile, critical infrastructure assets have become viable targets for asymmetric attacks by adversaries [1].

A decision to privatize, deregulate or allow foreign entities to control critical infrastructure assets has major national security implications. This presents a formidable challenge for countries that place national security at the forefront of their public policy. Liberal policies may well place critical infrastructure assets under foreign control or ownership, potentially providing foreign entities with direct access to the assets that can be exploited in times of conflict.

Gordon and Dion [5] describe restrictions that nation states can place to control the access of foreign entities to infrastructure sectors. These restrictions include blanket limitations (e.g., banning foreign entities from reaching a threshold of ownership and control); sector-specific licensing provisions (e.g., licenses or contractual arrangements between the government and the private entities); and trans-sectoral measures (e.g., investment approval procedures such as those adopted by the Committee on Foreign Investment in the United States).

The United States and Israel provide excellent case studies for a comparative analysis. Both the United States and Israel share a perceived high threat,

resulting in an increased focus on national and homeland security. Both countries have emphasized the threats to critical infrastructures in their national security policies, and this has had an effect on their policies towards foreign ownership of assets in key sectors. In addition, both economies rely on foreign direct investment.

The United States and Israel, however, differ in three main aspects. First, the United States is a developed country; Israel is considered an emerging market that is still undergoing decentralization and privatization processes [11]. Second, the U.S. has a large market economy while Israel has a small market economy [4]. Third, the United States relies primarily on the private sector for critical infrastructure protection. Israel, on the other hand, has adopted a state-centric approach that relies on the government security apparatuses for critical infrastructure protection.

4.1 United States

The United States is cognizant of the adverse national security implications of foreign investment in certain industries and sectors. U.S. foreign investment was initially governed by the Exon-Florio Amendment [13]; currently, however, it is governed by the Foreign Investment and National Security Act of 2007 (FINSA) [15]. The act requires a review by the Committee on Foreign Investment in the United States (CFIUS) to determine if foreign investment proposals threaten national security. CFIUS is chaired by the Department of the Treasury and staffed by representatives from various departments, including the Departments of Defense and Homeland Security. If a 30-day review determines that the transaction indeed poses a potential risk to national security, a 45-day investigation is conducted, upon which CFIUS either specifies the terms for mitigation or prohibits the transaction.

Traditionally, the United States has welcomed foreign investment as part of its open market economic ideology and CFIUS has rarely restricted foreign direct investment. However, this policy shifted following the terrorist attacks of September 11, 2001. In 2003, President Bush added the Department of Homeland Security to CFIUS and provided avenues for other security organizations to submit opinions regarding transactions. The changes have created a balance of power that favors agencies that prioritize security over economic considerations [6].

In 2006, a proposed foreign investment transaction created a major controversy that resulted in greater scrutiny of foreign investments. Towards the end of 2005, Dubai Ports World (a UAE-government-owned company) entered into negotiations for the purchase of the U.K.-based Peninsular and Oriental Steam Navigation Company (P&O) – one of the largest operators of ports worldwide. At the time, P&O operated six major American ports – New York, Philadelphia, Miami, Baltimore, New Jersey and New Orleans.

Dubai Ports World notified CFIUS of the transaction and CFIUS concluded that the transaction would not threaten national security. However, members of the U.S. Congress questioned the transaction and the process that led

Table 1. Trends in the CFIUS process (2005–2008).

Year	Notices (N)	Notices Withdrawn	Invstgns. (I)	Notices Withdrawn	Presidential Decisions
2005	55	1	1	1	0
2006	111	14	7	5	2
2007	138	10	6	5	0
2008	155	19	23	5	0
Total	459	44	37	16	2

CFIUS to its conclusion. Several bills were introduced in Congress to amend the CFIUS legislation and require a process with more emphasis on national security. The proposed changes included moving the chairmanship of CFIUS from the Department of Treasury to the Department of Homeland Security or the Department of Defense, and requiring majority American ownership of critical infrastructure assets. Ultimately, the public controversy led Dubai Ports World to sell the U.S. operations of P&O to an American company.

The Dubai Ports World affair was the major driver for enacting FINSA as a reform to the Exon-Florio Amendment. FINSA emphasizes security at the expense of economic interests (albeit not explicitly). It requires the consideration of critical infrastructure protection and homeland security issues in the CFIUS review process. Also, it requires CFIUS to conduct investigations when foreign investment transactions are initiated by foreign government owned or controlled entities. Previously, the burden of proof was on CFIUS to show that a transaction was a threat to national security. Now, the burden is on the investing entity to demonstrate that the transaction does not pose a national security threat.

An executive order by President Bush in January 2008 [2] and new regulations issued by the Department of Treasury in November 2008 [16] followed the enactment of FINSA. The executive order gave individual members of CFIUS the ability to initiate an inquiry if a transaction is deemed to have national security implications. The new regulations provide CFIUS with stronger enforcement mechanisms (e.g., strict penalties) on parties who fail to act in accordance with FINSA.

As mentioned above, the U.S. has identified eighteen industrial sectors of the economy as critical infrastructures. One impact of this categorization is that the majority of foreign direct investment transactions are required to adhere to reviews by CFIUS. While this does not imply that CFIUS would review and/or investigate every transaction, certain administrative burdens are levied. The potential increase in time and costs may lead to reluctance on the part of foreign entities to invest in U.S. critical infrastructure assets.

Table 1 provides preliminary data relating to the changes enacted in the CFIUS process [3]. Upward trends are evident in the numbers of notices submitted to CFIUS, notices withdrawn during CFIUS review and investigations

Table 2. Investigations following CFIUS reviews.

Year	Notices (N)	Invstgns. (I)	I/N Ratio	Difference in Invstgns. Successive Years
2005	55	1	1.82%	–
2006	111	7	6.31%	600.00%
2007	138	6	4.35%	–14.29%
2008	155	23	14.84%	283.33%
Total	459	37	8.06%	–

initiated by CFIUS following review. Also, the numbers have grown significantly following the Dubai World Ports controversy in early 2006. This clearly demonstrates the increased consideration of national security implications with regard to foreign direct investment in the United States.

The fact that notices were withdrawn during the CFIUS review process does not imply that the parties abandoned the transactions. Instead, the parties likely resubmitted the notices to CFIUS after making changes to reduce the likelihood of an investigation.

Interestingly, the number of presidential decisions (i.e., complete denial of transactions) has remained low. This could mean that, despite the heightened scrutiny, the basic preference for a free market approach persists. Nevertheless, it is clear that the burden (and associated costs) of proving that a transaction may threaten American national security has increased.

Perhaps the most notable change is the steep increase in the number of investigations initiated by CFIUS following a review of notices. Table 2 shows data related to investigations following CFIUS reviews. It is reasonable to conclude that the scrutiny of foreign investment transactions is becoming more strict. While foreign investment still flows into the United States from around the world, the increased emphasis on national security may negatively influence the willingness of foreign entities to direct their investments to the United States in the long term.

4.2 Israel

Israel is often considered to be a developed country, but it is actually an emerging market, which is still undergoing decentralization, deregulation and privatization processes. As such, Israel is highly dependent on foreign investment.

Since the 1980s, Israel has followed the Washington Consensus ideals. Foreign direct investment is encouraged by providing incentives to foreign investors (e.g., tax exemptions and subsidies), reducing trade barriers and eliminating the central bank's intervention in foreign currencies.

The Israeli economy has historically been controlled by the central government. However, in recent years, the government has begun to privatize some

critical infrastructure assets. These include Bezeq (a telecommunications company); Bank of Israel; Oil Refineries Limited; Paz Ashdod Refinery; Israel Electricity Corporation; Eilat-Ashkelon Pipeline Company; Israel Airports Authority; and Tel Aviv Stock Exchange. The specific sectors that are considered to be critical infrastructures are still evolving. It is likely that a list similar to that for the United States will eventually emerge.

The Israeli government limits foreign ownership through a series of laws, government decrees and executive orders. In certain cases, foreign investment in critical infrastructures is banned either wholly or partially. In other cases, foreign investment is implicitly banned by requiring that the directors and key executives of private sector entities pass security screenings. In essence, Israeli citizenship is required, leading to a *de facto* limitation on the degree of control by foreign entities.

Foreign entities wishing to invest in critical infrastructure assets in Israel must apply for permission from the Ministry of Finance and from the Israeli security community. The nature and substance of the deliberations are shrouded in secrecy and the criteria used to determine whether or not a foreign investor poses a national security risk are not publicly known. The procedure clearly lacks transparency and, consequently, accountability; not surprisingly, it is very difficult for a foreign investor to obtain approval.

The case of Bezeq, an Israeli telecommunications company, illustrates the limitations on foreign ownership. Bezeq is designated as a critical infrastructure asset under the Regulation of Security in Public Bodies Law of 1998. It underwent privatization during the 1990s, but the government still owns approximately 16% of the company.

Bezeq is heavily regulated by the Ministry of Communications under the Regulation of Electronic Communications Services in Israel [7]. To limit the privatization of Bezeq, the Ministry of Communications issued a telecommunications order [10] that restricts the ability of foreign entities from controlling 5% or more of the company. Furthermore, 75% of the company directors (including the chairman of the board) must be Israeli citizens and should hold security clearances. The order also gives the Israel Security Agency broad discretion for protecting critical information infrastructures.

The privatization of Oil Refineries Limited is another example where foreign investment in an Israeli critical infrastructure asset was limited because of national security concerns. In 2007, the Israeli government decided to privatize Oil Refineries Limited. One of the bidders for the company was an investment group comprising Israeli companies and the Swiss company Glencore International AG, one of the world's largest suppliers of industrial raw materials. The Israeli government, based on advice from its security apparatuses, refused to grant Glencore a control permit for Oil Refineries Limited. Glencore ultimately had to withdraw from the investment group.

The primacy of national security over economic interests is reflected by the Israeli government's disregard for the adverse effects that foreign investment restrictions could have on economic liberalization. In theory, restricting foreign

ownership in the framework of privatization can have three adverse economic effects. First, barring foreign ownership limits the number of potential bidders in a privatization process. This has an adverse effect on competition and drives the bid amounts down compared with perfect (or close to perfect) competitive processes. Second, in small market economies characterized by a high aggregate concentration, only a small number of entities have sufficient resources to participate in privatization bids [4]. Thus, an already concentrated market becomes even more concentrated. Third, a restriction creates high opportunity costs for the restricting country. Foreign investment, in general, benefits a nation's international economic situation, local employees and research and development efforts, and increases long-term growth. By restricting foreign ownership, a country like Israel forgoes these benefits and incurs an opportunity cost.

4.3 Analysis

The United States and Israel balance economic and security interests when approving foreign direct investment in critical infrastructure assets. However, Israel's security-biased policy limits foreign investment to a much greater extent than U.S. policy.

American and Israeli regulatory policies covering foreign investment and critical infrastructure protection grant the government broad discretion in approving foreign control and ownership. In the United States, the President (in concert with CFIUS) can review the national security consequences of mergers and acquisitions involving foreign entities [13]. In Israel, the review and approval process is mandated through ministerial decrees in addition to privatization documents.

American and Israeli policies also differ in terms of substance and procedure. Although both countries do not strictly prohibit foreign ownership, the Israeli policy is more stringent than the American policy. The American CFIUS process is more clear and transparent than its Israeli counterpart, and, therefore, provides less uncertainty to foreign investors. The differences between the two mechanisms reflect the challenges inherent with conflicting ideologies: the importance of national security in the case of Israel versus the prevailing economic ideology of minimal market intervention in the case of the United States.

5. Conclusions

The need to preserve national security while deriving the benefits of global economic liberalization presents a significant challenge for governments attempting to privatize critical infrastructure assets. In the post September 11, 2001 world, the notion that foreign direct investment positively contributes to an economy is gradually being tempered by the realization that it can pose a threat to national security. Indeed, the threshold of what constitutes a national security risk has lowered considerably.

With regard to foreign direct investment in critical infrastructure assets, the United States now appears to favor security over economic benefits. The result is additional investigations and stricter security conditions for government approval, increasing the risk and the uncertainty for foreign investors. This tendency entails adverse economic effects for countries regardless of whether they have large or small market economies. As a small market economy, Israel is already characterized by a high aggregate concentration. Thus, Israel's restrictions on foreign investment in critical infrastructure assets may result in even higher aggregate concentration with clear adverse effects. It is imperative that policy makers strike the right balance between national security concerns and economic liberalization.

References

[1] J. Arquilla and D. Ronfeldt, *Networks and Netwars: The Future of Terror, Crime and Militancy*, RAND Corporation, Santa Monica, California, 2001.

[2] G. Bush, Executive Order 13456 of January 23, 2008, The White House, Washington, DC (www.fas.org/irp/offdocs/eo/eo-13456.html), 2008.

[3] Department of the Treasury, Committee on Foreign Investment in the United States Annual Report to Congress (Public Version), Washington, DC (www.ustreas.gov/offices/international-affairs/cfius/docs/CFIUS-Annual-Rpt-2008.pdf), 2008.

[4] M. Gal, *Competition Policy for Small Market Economies*, Harvard University Press, Cambridge, Massachusetts, 2003.

[5] K. Gordon and M. Dion, Protection of Critical Infrastructure and the Role of Investment Policies Relating to National Security, Organization for Economic Cooperation and Development, Paris, France (www.oecd.org/dataoecd/2/41/40700392.pdf), 2008.

[6] E. Graham and D. Marchick, *U.S. National Security and Foreign Direct Investment*, Institute for International Economics, Washington, DC, 2006.

[7] D. Ivry-Omer, *Regulation of Electronic Communications Services in Israel*, Israel Democracy Institute Press, Jerusalem, Israel, 2009.

[8] J. Jackson, Foreign Direct Investment in the United States: An Economic Analysis, CRS Report for Congress RS21857, Congressional Research Service, Washington, DC (fpc.state.gov/documents/organization/109490.pdf), 2008.

[9] B. Krebs, Estonia incident demonstrated power of Russia-based cyber networks, *The Washington Post*, October 13, 2007.

[10] Ministry of Communications, The Telecommunications Order, Tel Aviv, Israel (www.moc.gov.il/sip_storage/FILES/1/371.pdf), 2004.

[11] J. Nitzan and S. Bichler, *The Global Political Economy of Israel*, Pluto Press, London, United Kingdom, 2002.

[12] I. Traynor, Russia accused of unleashing cyberwar to disable Estonia, *The Guardian*, May 17, 2007.

[13] United States Government, Authority to review certain mergers, acquisitions and takeovers, Title 50 Appendix, Section 2170, *United States Code Service*, pp. 353–360, 1996.

[14] United States Government, Title 42, Public Health and Welfare, Critical Infrastructure Protection Act of 2001, *United States Code Annotated*, pp. 456–459, 2003.

[15] United States Government, Foreign Investment and National Security Act of 2007, Public Law 110–149, 110th Congress, *U.S. Statutes at Large*, vol. 121, pp. 246–260, 2007.

[16] United States Government, Regulations pertaining to mergers, acquisitions and takeovers by foreign persons, *Federal Register*, vol. 73(226), pp. 70701–70729, 2008.

[17] J. Williamson, The Washington Consensus as Policy Prescription for Development, Institute for International Economics, Washington, DC (www.iie.com/publications/papers/williamson0204.pdf), 2004.

Chapter 3

CRITICAL INFORMATION INFRASTRUCTURE PROTECTION IN THE DEVELOPING WORLD

Ian Ellefsen and Sebastiaan von Solms

Abstract Critical information infrastructure protection (CIIP) has long been an area of concern, from its beginnings with the creation of the Internet to recent high-profile distributed denial-of-service attacks against critical systems. Critical systems rely heavily on information infrastructures; a disruption of the information infrastructure can hinder the operation of critical systems. The developed nations have mature CIIP solutions in place, but these solutions are not always suitable for developing countries, where unique challenges and requirements have to be addressed. Meanwhile, the developing nations are experiencing unprecedented growth of their information infrastructures. However, the lack of national CIIP efforts creates a situation for developing nations to become launch pads for cyber attacks. This paper discusses the need for CIIP in developing nations. It examines the current state and future development of information infrastructures in these nations and outlines a number of CIIP requirements.

Keywords: Critical information infrastructure protection, developing countries

1. Introduction

Critical information infrastructure protection (CIIP) is an area of worldwide concern. Developed and developing countries employ a number of critical systems [9]. These critical systems rely heavily on information infrastructures in order to function.

However, the information infrastructure is a single point of failure, where critical systems can be interrupted, and possibly disabled, by disrupting the underlying information infrastructure. The incidents in Estonia in 2007 [22] and Georgia in 2008 [19] have demonstrated the inability of countries to function effectively in the face of cyber attacks on their information infrastructures.

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 29–40, 2010.

The interconnected nature of systems brought about by the Internet allows cyber attacks to be conducted from anywhere on the globe. Due to advances in technology and growth of their infrastructures, developing nations are being used to launch these attacks. This problem is compounded by ineffective or nonexistent cyber security policies and CIIP solutions.

The development of CIIP structures in developing nations is an issue of vital importance to protect new information infrastructures and to support critical systems. This paper discusses CIIP as it pertains to the developing world. It examines existing protection models and their relevance to developing nations. The current state of affairs in South Africa is presented to set the stage for formulating CIIP requirements in the developing world.

2. CIIP

Critical information infrastructure protection (CIIP) is an issue of vital important to every nation. Developed countries have long had structures in place to protect their critical information infrastructures. Moteff, *et al.* [20] observe that there are a number of different infrastructures that can be considered to be "critical." They define critical infrastructures as those that are "... so vital that their incapacitation or destruction would have a debilitating impact on defense or economic security."

Critical systems, such as electricity distribution, water distribution and financial systems, are of utmost important to the operation of a country [9]. As critical systems become more complex, there is an ever increasing level of interconnection that is required for their operation. Interconnected critical systems heavily rely on information infrastructures. The interconnecting information infrastructures themselves are classified as critical due to the role they play in the operation of other critical systems.

Critical information infrastructures such as the Internet are designed to be fault resistant; however, they can quite easily be affected by events outside the control of a nation's protection structure. A cyber event of sufficient scale can have a detrimental effect on the global operation of the Internet and, thus, critical systems in countries around the world. This section discusses the vulnerability of information infrastructures to cyber attacks that target a nation's critical systems.

2.1 Cyber Attacks

Critical information infrastructures are particularly vulnerable to cyber attacks. For example, large-scale distributed denial-of-service (DDoS) attacks can be initiated quickly using botnets to prevent national systems from operating at full capacity. Such cyber attacks impact information infrastructures and may have significant physical effects.

Cyber attacks can affect countries directly or indirectly. Attacks on infrastructures within one country can have indirect effects on another country; alternatively, a large-scale cyber attack can have global effects. This is largely

due to the interconnected nature of the Internet. Indeed, the world exists in a state of collective vulnerability because of interconnected infrastructures.

The monitoring and management of critical systems that are heavily reliant on information infrastructures are particularly important to mitigate the impact of cyber attacks. The following sections discuss some major cyber attacks, in particular, the Estonian and Georgian incidents, and the DNS root server attacks. These attacks, which impacted the operation of national critical systems, provide insight into the importance of CIIP at the national, regional and global levels.

The Estonian Incident Beginning on April 27, 2007, a series of DDoS attacks were launched against several key computer systems in Estonia. The attacks, which affected the private and public sectors, were executed during a period of civil unrest and increased tension between Estonia and Russia, due to the Estonian Government's decision to move a World War II war memorial. At the time, Estonia blamed Russia for the attacks [5, 6].

The attacks ranged from generic traffic floods to coordinated botnet attacks [22]. Network traffic from the attacks was measured at 90 Mbps for upwards of 10 hours [22]. This had a devastating effect on web access in Estonia.

Even in 2007, Estonia had an extensive information infrastructure structure and relied heavily on Internet services [23]. The attacks disrupted or disabled access to financial institutions, government services and other critical systems, severely impacting the country's ability to function.

The Georgian Incident During the South Ossetia War between Georgia and Russia in August 2008, a number of Georgian governmental and commercial computer systems came under coordinated cyber attacks [10]. These attacks eliminated the ability of Georgian officials to communicate with the outside world [19]. In order to regain the ability to communicate, Georgian officials contracted hosting companies located in other countries, including the United States [10, 19].

Although the attacks on Georgian assets were similar to those that affected Estonia the previous year, they provide insight into the role of the Internet in CIIP. Korns and Kastenberg [19] report that the transfer of key Georgian websites to U.S.-based Internet hosts resulted in portions of the U.S. information infrastructure being affected by the DDoS attacks.

The interconnected nature of the Internet causes other countries to become indirect targets of cyber attacks. While the cyber attacks discussed above were targeted at individual countries, it is conceivable that attacks against the Internet in general could disrupt operations in almost every country around the world.

DNS Root Server Incidents Cyber attacks are not limited to a single country or geographic region; they can have a global impact. This was demonstrated by two DDoS attacks on the Domain Name System (DNS) root servers

that occurred on October 21, 2002 [26] and February 6, 2007 [16]. Although the effects were limited, the attacks demonstrate the ability of malicious actors to cause global disruption of the Internet.

The core of the DNS includes thirteen root servers, with as many as 200 instances in existence around the globe. The root servers translate human-understandable domain names into machine-based IP addresses. Global DNS server disruption can severely impact the operation of the Internet because many critical systems rely on DNS servers to translate domain names into the associated IP addresses. As it turned out, the 2002 and 2007 DDoS attacks did not cause major disruptions due to the over-provisioning of services.

Nevertheless, cyber attacks can have a major impact on the functioning of critical systems in the public and private sectors. These attacks can be organized rapidly and strike without warning. Every country must implement mechanisms to protect the national and global critical information infrastructures. The next section discusses CIIP with regard to developing nations and its current and future impact on the Internet and associated systems.

2.2 Developing Nations and CIIP

The information infrastructure in developing nations is often used to launch or coordinate cyber attacks [12]. According to a 2009 report by Akamai Technologies [2], much of the attack traffic that targets software and hardware vulnerabilities originates in developing countries. This is not to say that users in these countries are actively involved in attacks, only that their computer systems and networks are being utilized for cyber attacks. Indeed, developing countries are often "staging points" for attacks because of their rapidly growing information infrastructures coupled with the lack of coordinated cyber security measures.

Internet Connectivity Developing nations are becoming increasingly dependent on the Internet for communications, e-commerce and e-government services. They are rapidly provisioning their information infrastructures in order to support these services. Countries such as India or China, in particular, are seeing phenomenal growth in Internet-based technologies to support their critical systems [28].

Broadband penetration and Internet connection speeds in developing countries have historically been low, especially for countries in Sub-Saharan Africa [15]. However, several projects are underway to bring massive amounts of bandwidth to these countries [1]. Figure 1 illustrates the current status and future growth of Internet connectivity and bandwidth in the African continent.

The investment in information infrastructures will advance public and private sector efforts, which are essential to economic and social development. In particular, the new infrastructures will increase the resources available in critical areas such as telecommunications, finance, education, health care and social services. Individuals will also benefit from the new infrastructures, with more people having access to the Internet and Internet-based services. However, the

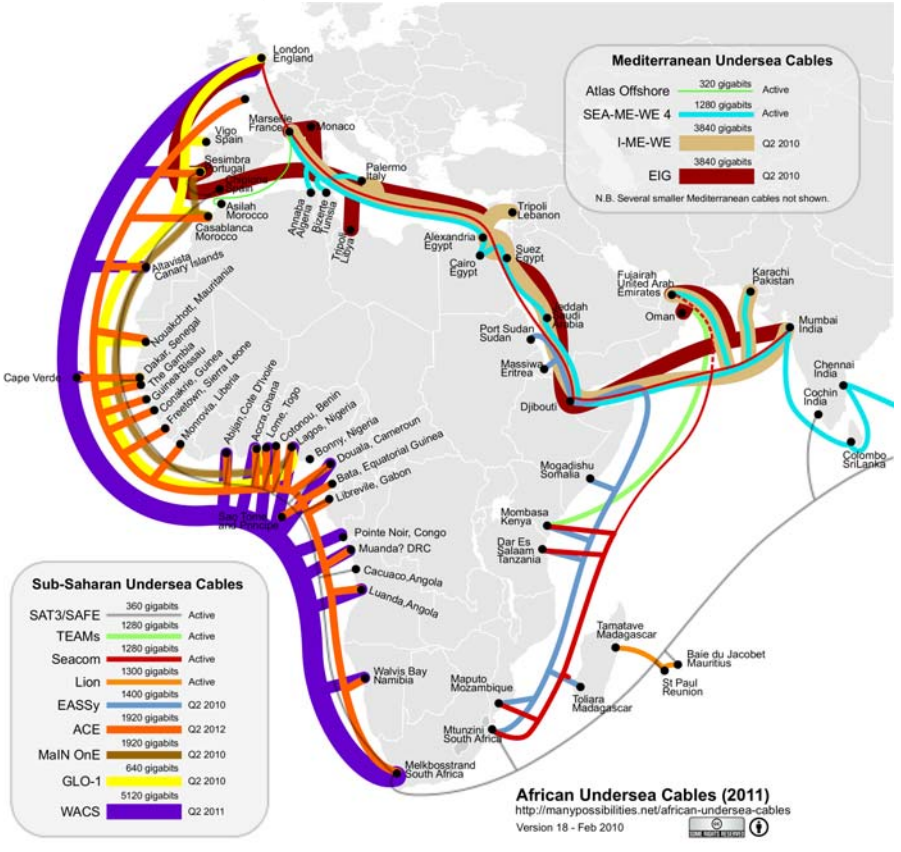


Figure 1. Undersea cables for the African continent (2011 projection) [24].

"always on" Internet culture brings with it its own set of problems such as malware, phishing schemes and botnets.

A 2009 study by Akamai Technologies [2] reveals that a significant percentage of attack traffic originates from developing countries. Table 1 shows that six developing countries or newly industrialized countries (in bold font) are in the top ten list. At the top of the list are Russia and Brazil, most likely due to the prevalence of Conficker-related infections [2].

Table 1 shows an 8% increase in attack traffic from the second quarter to the third quarter of 2009 in the "Other" category, which includes most of the developing countries in the world. This statistic coupled with the growth of their information infrastructures imply that attack traffic from these countries will increase very significantly in the future.

The increased connectivity and bandwidth in developing countries will have the effect of increasing the available pool of users and resources for malware

Table 1. Top ten originators of attack traffic [2].

Rank	Country	2009 Q3	2009 Q2
1	**Russia**	**13.0%**	**1.2%**
2	**Brazil**	**8.6%**	**2.3%**
3	U.S.	6.9%	15.0%
4	**China**	**6.5%**	**31.0%**
5	Italy	5.4%	1.2%
6	Taiwan	5.1%	2.3%
7	Germany	4.8%	1.9%
8	**Argentina**	**3.6%**	**0.8%**
9	**India**	**3.4%**	**0.9%**
10	**Romania**	**3.2%**	**0.6%**
–	Other	39.0%	31.0%

creators and botnet operators. These new pools of users and resources can be leveraged to launch highly destructive DDoS attacks against assets in other countries.

The expansion of the information infrastructure is not limited to investments in fiber optic cables and Internet connectivity. Developing countries are also experiencing unprecedented growth in mobile technologies. According to Cisco Systems [8], developing countries accounted for approximately 75% of the four billion mobile phones in use worldwide in 2009.

Mobile technologies enable developing countries to provide telecommunication services much more effectively than traditional land-based technologies. The MTN Group [21] reported that the "100 million mobile subscriber mark" was attained in developing and emerging markets during the middle of 2009. Mobile devices are being used increasingly as entry points into critical systems, a fact that is often overlooked in existing security policies [4].

The extensive use of mobile technologies also exposes more users to cyber attacks. Cisco Systems [8] predicts that large numbers of users in developing nations will fall victim to cyber attacks that leverage mobile technologies.

The information infrastructures in developing countries must be secured, managed and monitored to prevent them from being used as staging points for attacks. The countries will have to invest in legislation, education and CIIP mechanisms to prevent their new infrastructures from being abused. CIIP will have to be accomplished without limiting the functionality of the infrastructures. This is a particularly challenging aspect of "bridging the digital divide."

Cyber Security Policies Cyber security policies are essential to the management and operation of information infrastructures. Developed countries have created extensive security mechanisms and policies over time, which enables them to identify threats and mitigate the effects of attacks on their critical systems.

Until recently, developing countries have had little need for complex security mechanisms and policies due to their limited infrastructures. The lack of adequate protection for their information infrastructures creates a situation where criminals can utilize them for malicious purposes without fear of attribution or reprisal [7].

Despite the paucity of security mechanisms and policies, most developing countries do have some structures in place to deal with cyber crime. These normally take the form of incident response teams in large companies and government agencies, and digital forensic units in law enforcement agencies [11]. These entities are essential to identifying and prosecuting cyber criminals, but they do not provide monitoring and reporting capabilities for the national information infrastructures.

The next section discusses the structures that are in place for CIIP. These structures are loosely hierarchical in nature and are designed to monitor and report cyber incidents, enabling the relevant parties to react quickly and efficiently to incidents.

3. Protection Structures

Several structures exist for protecting critical information infrastructures. Some of these structures are specifically designed to provide incident handling and monitoring functions. The primary goal is to enable the relevant authorities to take quick, decisive steps to prevent cyber incidents and mitigate their adverse effects.

This section discusses two key structures: (i) computer security incident response teams (CSIRTs) that are used in large organizations; and (ii) warning, advice and reporting points (WARPs) that cater to smaller organizations and individuals.

3.1 CSIRTs

Computer security incident response teams (CSIRTs) (or computer emergency response teams (CERTs)) are commonly used by large corporations and government agencies, as well as for local, regional and national CIIP efforts. CSIRTs provide incident handling services [27], responding to cyber events and providing information and support to their stakeholders. For example, a CSIRT may monitor vulnerability reports from software and security appliance vendors and report information about threats, vulnerabilities and security controls to its stakeholders, enabling them to take the appropriate steps to protect their critical systems.

Organization CSIRTs are normally loosely hierarchical in nature. A tiered approach allows for the coordination of many, possibly diverse, stakeholders. The CSIRT hierarchy typically includes coordinating CSIRTs, regional CSIRTs and private CSIRTs. Each of these CSIRTs operates as a security team that is responsible for a specific constituency [17].

A coordinating CSIRT spearheads national information infrastructure protection efforts. It coordinates regional and private CSIRTs, and communicates with its counterparts in other countries. Its constituency includes regional and private CSIRTs, and other international CSIRTs.

Regional CSIRTs operate in a specific geographic region, providing support to organizations and the general population. They prevent the coordinating CSIRT from becoming overwhelmed by a large constituency. Regional CSIRTs also serve as the regional contact points for CIIP efforts.

Private CSIRTs (or private security teams) are created for large companies, academic institutions, government and law enforcement agencies, and military entities. They are normally responsible for managing incidents directly related to their particular organizations. Private CSIRTs are a vital entity in the CSIRT hierarchy as they experience the direct effects of cyber incidents and serve an first responders in their organizations.

The CSIRT hierarchy is presented in a generic manner. Koivunen [18] observes that each CSIRT and CSIRT hierarchy are unique, depending on the requirements imposed by the organizations and stakeholders, and their operating environments.

Analysis The establishment of a national CSIRT hierarchy is a proven method for implementing CIIP. Killcrece [17] notes that individual CSIRTs serve as trusted points of contact for cyber incidents. The coordinating CSIRT can help establish national best practices for cyber security, and provide advanced support for security incidents. However, Harrison and Townsend [14] argue that a CSIRT hierarchy is expensive to set up and maintain in terms of personnel and technology costs. Moreover, CSIRTs are primarily reactive as opposed to proactive.

3.2 WARPs

CSIRTs are large structures that are not designed to support smaller organizations and individuals. Warning, advice and reporting points (WARPs) fill the gap by serving as informal providers of cyber security information and expertise to small, focused constituencies.

Organization WARPs were first created in the United Kingdom as part of its CIIP efforts [3]. They are informal in nature and are focused on small member communities, to whom they provide computer security advice and limited incident handling services [14]. The members of a WARP typically number between 20 and 50, enabling the WARP to remain community-driven and focused on its member needs. The informal and focused nature of WARPs makes them very cost effective [14].

Analysis WARPs are very effective at providing CSIRT-like services to small communities that may not be adequately served by a larger CSIRT. Their obvious benefit, specifically for developing countries, is their low cost.

WARPs cannot provide adequate protection at the national level and are, therefore, not a replacement for CSIRTs. However, WAPRs can operate very effectively in conjunction with traditional CSIRTs.

4. South African Case Study

The creation and implementation of effective cyber security policies in developing countries are vitally important to national, regional and international CIIP efforts. Many of these countries are using hastily-created policies to cope with the dramatic expansion of their information infrastructures and the associated vulnerabilities in their critical systems. But these policies are only adequate for the short term; sustained efforts are necessary to provide long-term CIIP solutions.

In February 2010, the South African Department of Communications released a draft cyber security policy for South Africa [25]. This document outlines various structures for protecting the South African information infrastructure and associated critical systems.

The document notes that South Africa neither has a coordinated cyber security effort nor a broad legal framework for dealing with cyber crime. It goes on to stress that these technological and legal deficiencies must be addressed. The document also highlights the need for international cooperation in the area of CIIP. However, South Africa does not currently have adequate international relationships for effective information infrastructure protection.

The document makes a number of recommendations regarding the development of a CIIP framework. The recommendations are aimed at securing South African cyber space as well as reducing threats and vulnerabilities.

A key recommendation is the creation of a National Cybersecurity Advisory Council (NCAC). The NCAC will be responsible for advising governmental entities on issues related to cyber security. It will also be responsible for coordinating cyber security across the South African Government.

The document also recommends the creation of a CSIRT structure for managing threats and vulnerabilities, and to serve as point of contact for cyber security information. The proposed structure will consist of a national CSIRT, a governmental CSIRT and a number of sector-specific CSIRTs. Finally, the document highlights the need for local and international partnerships for effective CIIP.

5. CIIP Requirements for the Developing World

In order for developing nations to implement effective CIIP solutions, there are a number of requirements that should be satisfied. These requirements are diverse and depend on the goals mandated by governmental policies.

The nature of a CIIP solution would clearly depend on the specific country and infrastructure that needs to be protected. However, any solution that is developed will have to be cost effective.

Because developing nations are experiencing phenomenal growth in their information infrastructures, CIIP solutions have to be extensible to support future development without incurring excessive costs.

CIIP solutions in developing countries will require the support of international entities. To this end, the CIIP structures must support information exchange and knowledge transfer, both locally and internationally.

Special care must be taken with regard to mobile technologies. The growth of mobile technologies in the developing world enables millions of individuals to access information and services that were previously unavailable. However, mobile technologies dramatically increase the size of the user pool for exploitation by malicious actors. CIIP solutions in developing countries must take this into consideration.

Developing countries will have to embrace technology in order to supplement traditional methods of communication. This will allow vital information to be communicated in the event that traditional modes are unavailable. For instance, should communication via email not be possible, information could be exchanged via SMS messages or fax.

Finally, developing nations have to invest in broad-based awareness programs to ensure that new users are aware of the risks associated with their activities in cyber space. Such programs benefit users as well as the nation as a whole. Critical systems are connected by the same networks that are used by the general public; reducing threats and vulnerabilities at the user end helps protect critical systems as well as the underlying information infrastructure.

6. Conclusions

Modern critical systems rely heavily on information infrastructures in order to operate efficiently; this is true for developed countries as well as developing countries. However, due to the lack of adequate CIIP structures and security policies, developing countries are often used as launch pads for cyber attack. Much of the world's attack traffic already originates in these countries and the proportion of attack traffic will only increase with the dramatic growth of their information infrastructures.

Traditional CIIP solutions, such as CSIRTs and WARPs, support the monitoring and reporting of cyber incidents. Such solutions hold promise for the developing world, but issues such as cost-effectiveness, information exchange and international cooperation must be addressed.

Our future research will investigate a number of models that will satisfy CIIP requirements for developing countries. It will also examine the relationships between successful CIIP solutions across the developing world, and articulate best practices for national and regional CIIP efforts.

References

[1] Akamai Technologies, State of the Internet, vol. 2(2), Cambridge, Massachusetts (www.akamai.com/stateoftheinternet), 2009.

[2] Akamai Technologies, State of the Internet, vol. 2(3), Cambridge, Massachusetts (www.akamai.com/stateoftheinternet), 2009.

[3] B. Askwith, WARP case study – Experience setting up a WARP, Center for the Protection of the National Infrastructure, London, United Kingdom (www.warp.gov.uk/Index/indexarticles.htm), 2006.

[4] S. Baker, S. Waterman and G. Ivanov, In the Crossfire: Critical Infrastructure in the Age of Cyber War, Technical Report, McAfee, Santa Clara, California, 2010.

[5] BBC News, The cyber raiders hitting Estonia, London, United Kingdom (news.bbc.co.uk/2/hi/europe/6665195.stm), May 17, 2007.

[6] BBC News, Estonia fines man for "cyber war," London, United Kingdom (news.bbc.co.uk/2/hi/technology/7208511.stm), January 25, 2008.

[7] BBC News, What makes a cyber criminal? London, United Kingdom (news.bbc.co.uk/2/hi/americas/7403472.stm), May 19, 2008.

[8] Cisco Systems, Cisco 2009 Annual Security Report, San Jose, California (www.cisco.com/en/US/prod/collateral/vpndevc/cisco_2009_asr.pdf), 2009.

[9] R. Dacey, Critical Infrastructure Protection: Challenges and Efforts to Secure Control Systems, Testimony before the Subcommittee on Technology Information Policy, Intergovernmental Relations and the Census, House Committee on Government Reform, GAO-04-628T, General Accounting Office, Washington, DC (www.gao.gov/new.items/d04628t.pdf), 2004.

[10] D. Danchev, Coordinated Russia vs Georgia cyber attack in progress, ZDNet, San Francisco, California (blogs.zdnet.com/security/?p=1670), August 11, 2008.

[11] J. Fick, Cyber Crime in South Africa: Investigating and Prosecuting Cyber Crime and the Benefits of Public-Private Partnerships, Technical Report, PriceWaterhouseCoopers, Sunninghill, South Africa (www.coe.int/t/dghl/cooperation/economiccrime/cybercrime/Documents/Reports-Presentations/2079if09pres_JaquiFick_report.pdf), 2009.

[12] Georgia Tech Information Security Center, Emerging Cyber Threats Report for 2009, Georgia Institute of Technology, Atlanta, Georgia (www.gtisc.gatech.edu/pdf/CyberThreatsReport2009.pdf), 2008.

[13] M. Handley and E. Rescorla, Internet Denial-of-Service Considerations, RFC 4732, Internet Engineering Task Force, Fremont, California (www.ietf.org/rfc/rfc4732.txt), 2006.

[14] J. Harrison and K. Townsend, An update on WARPs, *ENISA Quarterly Review*, vol. 4(4), pp. 13–15, 2008.

[15] R. Heacock, Internet filtering in Sub-Saharan Africa, Technical Report, OpenNet Initiative, Harvard University, Cambridge, Massachusetts (opennet.net/sites/opennet.net/files/ONI_SSAfrica_2009.pdf), 2009.

[16] ICANN, Factsheet: Root server attack on 6 February 2007, Marina del Rey, California (www.icann.org/announcements/factsheet-dns-attack-08mar 07.pdf), 2007.

[17] G. Killcrece, Steps for Creating National CSIRTs, CERT Coordination Center, Carnegie Mellon University, Pittsburgh, Pennsylvania (www.cert.org/archive/pdf/NationalCSIRTs.pdf), 2004.

[18] E. Koivunen, Survey on Certain European CSIRT Teams' Administration, Operations, Cooperation and Communications, Technical Report, CERT-FI, Helsinki, Finland, 2009.

[19] S. Korns and J. Kastenberg, Georgia's cyber left hook, *Parameters*, vol. XXXVIII, pp. 60–76, 2008.

[20] J. Moteff, C. Copeland and J. Fischer, Critical Infrastructures: What Makes an Infrastructure Critical? Report for Congress RL31556, Congressional Research Service, Library of Congress, Washington, DC (www.fas.org/irp/crs/RL31556.pdf), 2003.

[21] MTN Group, MTN reaches the 100 million subscriber milestone, Press Release, Johannesburg, South Africa (www.mtn.com/media/overviewdetail.aspx?pk=381), May 2009.

[22] J. Nazario, Estonian DDoS attacks – A summary to date, Arbor Networks, Chelmsford, Massachusetts (asert.arbornetworks.com/2007/05/estonian-ddos-attacks-a-summary-to-date), May 17, 2007.

[23] J. Richards, Denial-of-service: The Estonian cyberwar and its implications for U.S. national security, *International Affairs Review*, vol. XVIII(1) (www.iar-gwu.org/node/65), 2009.

[24] S. Song, African undersea cables, Many Possibilities, Durbanville, South Africa (manypossibilities.net/african-undersea-cables), 2010.

[25] South African Department of Communications, Draft Cybersecurity Policy of South Africa, Government Gazette No. 32963, Pretoria, South Africa, 2010.

[26] P. Vixie, G. Sneeringer and M. Schleifer, Events of 21-Oct-2002 (c.root-servers.org/october21.txt), November 24, 2002.

[27] M. West-Brown, D. Stikvoort, K. Kossakowski, G. Killcrece, R. Ruefle and M. Zajicek, Handbook for Computer Security Response Teams (CSIRTs), Handbook CMU/SEI-2003-HB-002, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003.

[28] P. Wolcott, The provision of Internet services in India, in *Information Systems in Developing Countries: Theory and Practice*, R. Davison, R. Harris, S. Qureshi, D. Vogel and G. de Vreede (Eds.), City University of Hong Kong Press, Hong Kong, China, pp. 253–267, 2005.

Chapter 4

MODELING CONTROL SYSTEM FAILURES AND ATTACKS – THE WATERLOO CAMPAIGN TO OIL PIPELINES

Jonathan Butts, Mason Rice and Sujeet Shenoi

Abstract This paper presents a model for expressing control system failures and attacks on control protocols that involve the exchange of messages. Control failures and attacks are modeled using the notion of an attacker who can block and/or fabricate messages. These two attack mechanisms can cover a variety of scenarios ranging from control failures in the Waterloo Campaign to cyber attacks on oil pipelines. The model helps provide a comprehensive understanding of control system failures and attacks, which supports the development of strategies for attack as well as defense.

Keywords: Control systems, failure modeling, attack modeling

1. Introduction

Mankind's first conflicts were waged on land. As military technology advanced, battles were fought on sea and in the air. The earliest recorded naval battle occurred in 1210 BC when the Hittites led by Suppiluliumas II defeated a fleet from Cyprus. Aerial warfare was pioneered by the ancient Chinese who launched fire arrow attacks from "war kites" [1]; the first airplane bombing occurred in 1911 during the Libyan War between Italy and Turkey [1]; the first dogfights came soon after during World I. The first attack in space was a 1985 test that involved a U.S. F-15 shooting down a P78-1 communications satellite in a 345 mile orbit [15].

The 21st century brings a new dimension to the field of battle – cyberspace. Cyberspace, through its inextricable connection with the critical infrastructure, pervades all aspects of human endeavor – business, government and military

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 43–62, 2010.

operations, and societal functions. It is certain that future warfare will involve both cyberspace and the critical infrastructure.

Control is a vital component of any battle plan. A commander is responsible for controlling military operations. The commander must maintain constant situational awareness of the battlespace – allied forces, opposing forces, time and terrain. The commander uses various control techniques to maneuver forces to accomplish the mission within the battlespace parameters. Historians believe that Napoleon lost the Battle of Waterloo because of control failures made two days earlier during the Battles of Ligny and Quatre Bras [2, 11]. Sometimes, control failures are accidental; at other times, they are the result of the enemy compromising the control protocol. But however they occur, battles are won and lost because of control successes and failures.

This paper presents a model for expressing control system failures and attacks on control systems. Control system failures as well as attacks are modeled using the notion of an attacker who blocks and/or fabricates messages. In fact, these two types of attacks on control protocols can be used to express scenarios ranging from control failures during the Waterloo Campaign to cyber attacks on critical infrastructure assets such as oil pipelines.

The model, which is readily defined using graph theory, helps conceptualize attacks and failures in one control protocol and translate them to similar attacks and failures in another protocol. It also assists in targeting specific control protocol and system implementations. In particular, the model helps identify the information requirements, articulate possible outcomes and examine the feasibility of attacks based on the available information. The comprehensive understanding of attacks can facilitate risk analysis and risk management, the implementation of defensive postures and the design of robust control protocols.

2. The Waterloo Campaign

On June 16, 1815, two days before the pivotal Battle of Waterloo, Napoleon's troops were arrayed south of the road between the Belgian towns of Ligny and Quatre Bras (Figure 1). His 125,000 troops were divided into two commands [5]. Napoleon himself commanded the force on the east, just south of Ligny. Marshall Ney led the force to the west, a few miles south of Quatre Bras.

Napoleon's troops were opposed by two allied forces [5]. One force of 90,000 British, Dutch, Belgians and Germans commanded by the Duke of Wellington was positioned just north of Quatre Bras. The other force, a 115,000-man Prussian army led by Marshal Blucher, was positioned on the northern outskirts of Ligny.

The Battles of Ligny and Quatre Bras were fought that day. Napoleon won the Battle of Ligny; the Battle of Quatre Bras was a standoff [2]. Historians believe that the French forces could have crushed the opposition were it not for certain "control" failures [2, 11]. As a result, Wellington and Blucher were able to move north and reconstitute their forces. On June 18, 1815, the combined armies led by Wellington routed Napoleon's forces near the village of Waterloo.

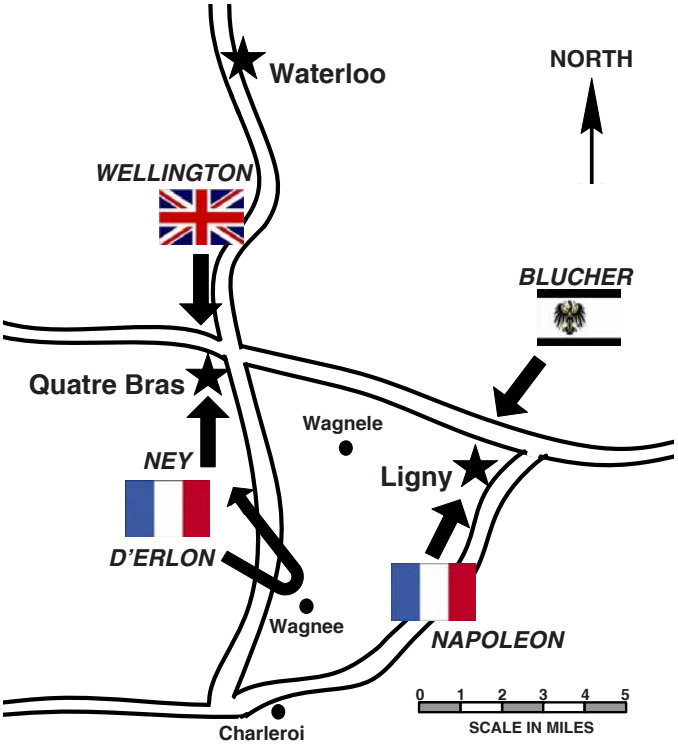


Figure 1. Battles of Ligny and Quatre Bras.

Before the Battles of Ligny and Quatre Bras, Napoleon had split his forces into two groups intending to drive a wedge between Wellington and Blucher (Figure 1). Napoleon wanted Ney to seize the crossroads at Quatre Bras while he destroyed Blucher's forces at Ligny [11]. The control failures occurred in the following chronological order:

- **Delay in Attacking Quatre Bras (Control Failure 1):** On the morning of June 16, 1815, Ney was in good position to take the strategic crossroads at Quatre Bras. Ney claimed that he did not receive the order to attack that morning. The Battle of Quatre Bras did not begin until 2 p.m. The delay gave Wellington time to place his troops in strong defensive positions.

- **Failure to Mobilize a Blocking Force (Control Failure 2):** Napoleon assumed that Ney would take the crossroads at Quatre Bras with ease. Shortly after 1 p.m., he ordered Ney to send a force towards Ligny to block the retreat of the Prussian forces involved in the Battle of Ligny. Ney did not receive the order. In any case, the Battle of Quatre Bras was still underway and Ney could not spare a blocking force.

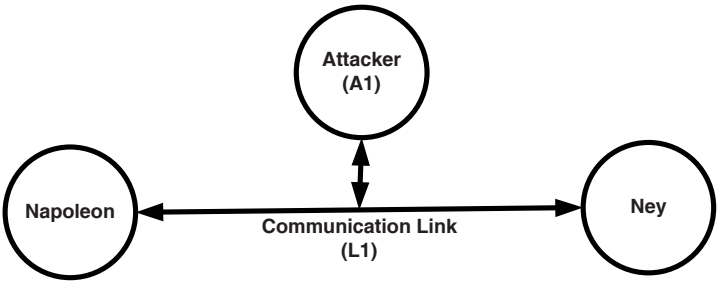


Figure 2. Communication pathway for Control Failures 1 and 2.

- **Failure to Engage a Reserve Force (Control Failure 3):** While the battle raged on at Quatre Bras, Napoleon's forces inflicted significant damage to the Prussians at Ligny. Sensing an opportunity for a devastating blow, Napoleon sent two messages: one to General d'Erlon ordering him to move his troops to Wagnele to attack the Prussian flank, and the other to Ney (d'Erlon's commander) informing him about the order. d'Erlon received Napoleon's message, but misread the message and marched towards the town of Wagnee instead of Wagnele.

  The message from Napoleon to Ney about moving d'Erlon's forces never arrived. When Ney learned that his reserve force under d'Erlon was moving away, he ordered it to turn back and support his forces at Quatre Bras. But it was too late and d'Erlon's reserve force neither joined the Battle of Ligny nor the Battle of Quatre Bras.

3. Modeling Napoleon's Control Failures

The control failures during the Battles of Ligny and Quatre Bras occurred as a result of delayed, lost and misinterpreted messages. The failures are attributed to the fog of war. However, we can formally model the failures using the notion of an attacker who blocks and/or fabricates messages. Indeed, these two types of attacks on messaging protocols can be used to express scenarios ranging from the control failures in the Waterloo Campaign to cyber attacks on critical infrastructure assets.

3.1 Modeling Control Failure 1

Figure 2 illustrates the communication pathway for Control Failure 1. Three nodes are involved: (i) Napoleon ($\in$ *ControlNodes*), (ii) Ney ($\in$ *EdgeNodes*) and (iii) A1 ($\in$ *AttackNodes*). Communications occur over L1 ($\in$ *Links*). In general, a link supports bidirectional message transfer, although each message is unidirectional from sender to receiver.

Control Failure 1 is consistent with a man-in-the-middle attack involving Attacker A1. In particular, Ney's delay in attacking Quatre Bras is modeled

System
Napoleon ∈ *ControlNodes*
Ney ∈ *EdgeNodes*
A1 ∈ *AttackNodes*
L1 ∈ *Links*
L1 = (A1: (Napoleon, Ney))
Request ∈ *MsgTypes*

Capabilities
A1 = {¬, +}

Initial Node States
Ney := Ney^0
Napoleon := Ney^0

Possible Node States
Ney := Ney^0 | Ney^1
Ney^0 ≡ Attack on Quatre Bras is FALSE
Ney^1 ≡ Attack on Quatre Bras is TRUE
Napoleon := Ney^0 | Ney^1

Control Failure 1
1. Napoleon $\xrightarrow[A1\ \neg Request]{L1}$ Ney[Attack]
Ney := Ney^0
Napoleon := Ney^1
2. Napoleon $\xrightarrow[A1\ +Request]{L1}$ Ney[Attack]
Ney := Ney^1
Napoleon := Ney^1

Figure 3. Formal specification of Control Failure 1.

by A1 blocking Napoleon's message to Ney, then fabricating the same message and transmitting it to Ney some time later. This first step, a block (¬) of Napoleon's message, is represented as:

$$\text{Napoleon} \xrightarrow[A1\ \neg Request]{L1} \text{Ney[Attack]}.$$

Since Napoleon ∈ *ControlNodes*, he can issue request messages (Request ∈ *MsgTypes*) that are received and acted on by *EdgeNodes* like Ney. For a message to be valid, the nodes must have access to a common link (L1). In the man-in-the-middle attack, A1 ∈ *AttackNodes* blocks (¬) the message sent from Napoleon to Ney along L1. At this point, Napoleon expects Ney to engage; however, Ney did not receive the message.

The second message transfer step in Control Failure 1 involves A1 fabricating a copy of Napoleon's original message (+) and sending it to Ney:

$$\text{Napoleon} \xrightarrow[A1\ +Request]{L1} \text{Ney[Attack]}.$$

Figure 3 shows the complete representation of Control Failure 1. The model includes the various nodes, links and message types. Attacker A1 has the capability to block (¬) and fabricate (+) messages. Note that L1 = (A1: (Napoleon, Ney)) expresses the fact that A1 has compromised Link L1 to perpetrate a man-in-the-middle attack between Napoleon and Ney. Ney's status is represented as one of two possible states: (i) Ney is not attacking Quatre Bras (Ney^0) or (ii) Ney is attacking Quatre Bras (Ney^1). Similarly, Napoleon maintains his perception of Ney's status (Ney^0 or Ney^1). The initial states for Ney and Napoleon are both Ney^0, i.e., Ney is not attacking Quatre Bras.

Figure 3 also shows the two-step message sequence for Control Failure 1. After Step 1 (message block), Napoleon assumes that Ney is attacking Quatre Bras (Napoleon := Ney^1) when, in fact, Ney is not attacking Quatre Bras (Ney := Ney^0). It is only after Step 2 (message fabrication) that Napoleon's and Ney's states match (Napoleon := Ney^1 and Ney := Ney^1).

System
Napoleon ∈ *ControlNodes*
Ney ∈ *EdgeNodes*
A1 ∈ *AttackNodes*
L1 ∈ *Links*
L1 = (A1: (Napoleon, Ney))
Request ∈ *MsgTypes*

Capabilities
A1 = {¬}

Initial System State
Ney := Ney^0
Napoleon := Ney^0

Possible Node States
Ney := Ney^0 | Ney^1
Ney^0 ≡ Blocking Force is FALSE
Ney^1 ≡ Blocking Force is TRUE
Napoleon := Ney^0 | Ney^1

Control Failure 2
1. Napoleon $\xrightarrow[A1\ \neg Request]{L1}$ Ney[Attack]
Ney := Ney^0
Napoleon := Ney^1

Figure 4. Formal specification of Control Failure 2.

3.2 Modeling Control Failure 2

The failure of Napoleon to mobilize a blocking force against the retreating Prussians is modeled by Attacker A1 blocking Napoleon's order to Ney. The communication pathway is the same as that for Control Failure 1 (Figure 2).

Figure 4 shows the complete representation of Control Failure 2. The nodes, links and message types are the same as for Control Failure 1 (Figure 3). However, Attacker A1 only requires the capability to block (¬) messages. Ney's status is represented as one of two possible states: (i) Blocking force is not engaged (Ney^0) or (ii) Blocking force is engaged (Ney^1). Similarly, Napoleon maintains his perception of Ney's status (Ney^0 or Ney^1). The initial states for Ney and Napoleon are both Ney^0, i.e., Blocking force is not engaged.

The bottom half of Figure 4 shows the one-step message sequence corresponding to Control Failure 2. After Step 1 (message block), Napoleon assumes that Ney's blocking force is engaged (Napoleon := Ney^1) when, in fact, Ney's blocking force is not engaged (Ney := Ney^0).

3.3 Modeling Control Failure 3

Figure 5 shows the communication pathways involved in Control Failure 3. Note that Napoleon, Ney ∈ *ControlNodes*; d'Elron ∈ *EdgeNodes*; and A1, A2 ∈ *AttackNodes*. Communication occurs over three links: L1, L2, L3 ∈ *Links*. Links L1 and L2 are compromised by Attackers A1 and A2, respectively. Link L3 supports communication between Ney and d'Erlon, and is not compromised by an attacker.

Figure 6 shows the complete representation of Control Failure 3. Attacker A1 has the capability to block (¬) messages while Attacker A2 can block (¬) and fabricate (+) messages. d'Erlon's status is represented as one of three possible states: (i) d'Erlon is at Quatre Bras ($d'Erlon^0$), (ii) d'Erlon is at Wagnele ($d'Erlon^1$) or (iii) d'Erlon is at Wagnee ($d'Erlon^2$). Similarly, Napoleon and Ney each maintain their own perceptions of d'Erlon's status ($d'Erlon^0$, $d'Erlon^1$

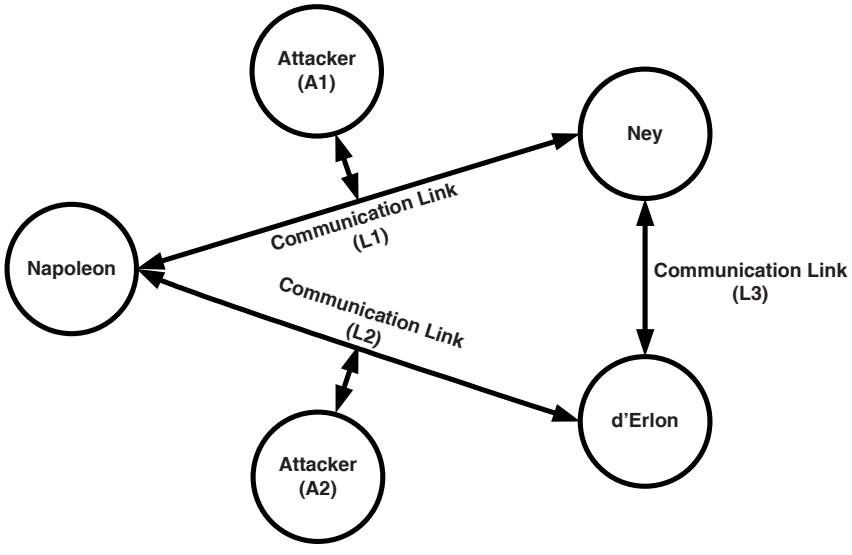


Figure 5. Communication pathways for Control Failure 3.

or d'Erlon^2). The initial states for d'Erlon, Ney and Napoleon are all d'Erlon^0, i.e., d'Erlon is staged at Quatre Bras.

The message sequences that result in Control Failure 3 are broken down into three processes. Processes 1 and 2 are independent and can occur in parallel; however, Process 3 must occur after Process 1.

The first step in Process 1 involves Attacker A2 blocking Napoleon's message to d'Erlon that orders his troops to Wagnele. After Step 1, Napoleon believes that d'Erlon is at Wagnele (Napoleon := d'Erlon^1); however, d'Erlon is still at Quatre Bras (d'Erlon := d'Erlon^0). Since Ney is not involved in the message exchange, his perception of d'Erlon's status is unchanged (Ney := d'Erlon^0). In Step 2, Attacker A2 sends a fabricated message to d'Erlon ordering him to move his troops to Wagnee (d'Erlon := d'Erlon^2). Napoleon still believes d'Erlon to be at Wagnele (Napoleon := d'Erlon^1) and Ney's perception of d'Erlon's status is unchanged (Ney := d'Erlon^0).

In Process 2, A1 blocks Napoleon's message to Ney that would have informed Ney of d'Erlon's movement. Process 2 may run concurrently with Process 1, implying that d'Erlon is either at Quatre Bras (d'Erlon := d'Erlon^0) or Wagnee (d'Erlon := d'Erlon^2) depending on the order of execution. The superscript $(*)$ denotes that d'Erlon is in one of multiple possible states (d'Erlon := d'Erlon*). Note that the specific state is not relevant because d'Erlon would still be out of position.

Process 3 is conditional on Ney's observation that d'Erlon is moving away from Quatre Bras (Ney := d'Erlon^1 or Ney := d'Erlon^2). At this point, Ney sends a message to d'Erlon to reverse course and engage at Quatre Bras. Upon receiving this message, d'Erlon moves to Quatre Bras (d'Erlon := d'Erlon^0),

System
Napoleon, Ney $\in$ *ControlNodes*
d'Erlon $\in$ *EdgeNodes*
A1, A2 $\in$ *AttackNodes*
L1, L2, L3 $\in$ *Links*
L1 = (A1: (Napoleon, Ney))
L2 = (A2: (Napoleon, d'Erlon))
L3 = (Ney, d'Erlon)
Request $\in$ *MsgTypes*

Capabilities
A1 = $\{\neg\}$
A2 = $\{\neg, +\}$

Initial System State
d'Erlon := d'Erlon^0
Ney := d'Erlon^0
Napoleon := d'Erlon^0

Possible Node States
d'Erlon :=
d'Erlon^0 | d'Erlon^1 | d'Erlon^2
d'Erlon^0 $\equiv$ Quatre Bras is TRUE
d'Erlon^1 $\equiv$ Wagnele is TRUE
d'Erlon^2 $\equiv$ Wagnee is TRUE
Ney :=
d'Erlon^0 | d'Erlon^1 | d'Erlon^2
Napoleon :=
d'Erlon^0 | d'Erlon^1 | d'Erlon^2

Process 1
1. Napoleon $\xrightarrow[A2\ \neg Request]{L2}$ d'Erlon[Wagnele]
d'Erlon := d'Erlon^0
Ney := d'Erlon^0
Napoleon := d'Erlon^1
2. Napoleon $\xrightarrow[A2\ +Request]{L2}$ d'Erlon[Wagnee]
d'Erlon := d'Erlon^2
Ney := d'Erlon^0
Napoleon := d'Erlon^1

Process 2
1. Napoleon $\xrightarrow[A1\ \neg Request]{L1}$ Ney[d'Erlon Wagnele]
d'Erlon := d'Erlon*
Ney := d'Erlon^0
Napoleon := d'Erlon^1

Process 3
CONDITIONAL: IF (d'Erlon != d'Erlon^0)
1. Ney $\xrightarrow[Request]{L3}$ d'Erlon[Quatre Bras]
d'Erlon := d'Erlon^0
Ney = d'Erlon^0
Napoleon = d'Erlon^1

Figure 6. Formal specification of Control Failure 3.

Ney believes that d'Erlon is moving to Quatre Bras (Ney := d'Erlon^0), but Napoleon believes that d'Erlon is at Wagnele (Napoleon := d'Erlon^1).

4. Formal Model

This section formalizes the approach used to express the control failures involved in the Battles of Ligny and Quatre Bras. The formal model is intended to express and reason about failures and attacks on SCADA systems used to control critical infrastructure assets.

Control protocols use messages to direct actions and provide feedback using a hierarchical, request-reply paradigm. Figure 7 shows a generic process diagram. *ControlNodes* send request messages to subordinate *EdgeNodes* or other control nodes (sub-control devices) to specify control actions and/or obtain data. *EdgeNodes* translate request messages into physical actions and/or physical actions into reply messages that are transmitted to their *ControlNodes*. In general, a message may involve a request-reply sequence, a request without a reply or an unsolicited reply. Request and reply messages are transmitted along bi-directional communication links that connect two or more nodes.

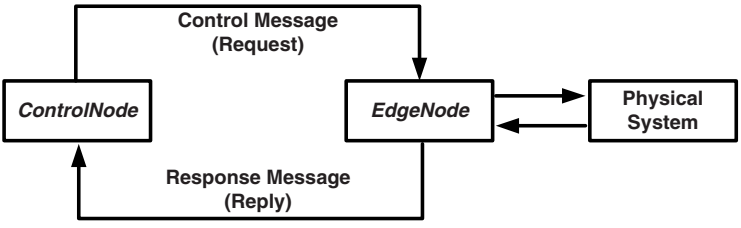


Figure 7. Generic process diagram.

A communication involves a sender transmitting a message to a receiver along a communication link:

$$MsgSource \underset{MsgTypes}{\overset{Link}{\longrightarrow}} MsgDest[Payload].$$

MsgSource and *MsgDest* are *ControlNodes* or *EdgeNodes* that communicate over the specified *Link*. *MsgType* is the type of message as defined by the control protocol (e.g., Request or Reply). *Payload* is the data contained in the message.

An attack on a control protocol occurs when an attacker (represented as an *AttackNode*) blocks legitimate messages or sends fabricated messages along a communication link. Formally, a message used in an attack is specified as:

$$MsgSource \underset{AttackNodes,Capabilities,MsgTypes}{\overset{Link}{\longrightarrow}} MsgDest[Payload].$$

MsgSource is the original sender or the spoofed sender of the attack message ($\in$ *ControlNodes* $\cup$ *EdgeNodes*). *MsgDest* is the intended target of the message ($\in$ *ControlNodes* $\cup$ *EdgeNodes*). The *AttackNode*, the perpetrator of the attack, has the *Capabilities* to block and/or fabricate messages along the *Link*. Note that message blocking and fabrication enable an attacker to launch a variety of messaging attacks. Message modification is implemented by blocking a legitimate message followed by sending a fabricated message. Message replay is implemented by sending a fabricated message with the same payload as an earlier message. Likewise, message delay is implemented by blocking a legitimate message followed by sending the original message some time later.

In general, an attacker has two attack avenues. The first is to compromise a *ControlNode* or *EdgeNode* and convert it into an *AttackNode*; this enables the attacker to block messages sent to the compromised node and to send fabricated messages from the compromised node. The second attack avenue is to compromise a link, which enables the attacker to perpetrate man-in-the-middle attacks. As shown in Figures 3, 4 and 6, the situation where Attacker A1 compromises Link L1 between Nodes N1 and N2 is expressed as L1 = (A1: (N1, N2)).

A node N has an initial state ($N := N^p$) and may change its state ($N := N^q$) upon receiving a message or as a result of a change in the physical state of the

device expressed by the node (e.g., closed valve). Note that a superscript $(*)$ is used to denote that a node is in one of multiple possible states. A *ControlNode* also maintains information about the status of its subordinate *EdgeNodes*. For example, Z := (N1^p, N2^q) denotes that *ControlNode* Z perceives the states of subordinate nodes N1 and N2 to be N1^p and N2^q, respectively.

The formal model expresses temporal and causal properties using sequential steps, conditional statements and independent processes. A "process" is a sequence of messages that occur in a specific order; the process may be executed ψ number of times. A "conditional process" only executes when a Boolean condition holds.

5. Modeling an Attack on an Oil Pipeline

Oil pipelines often rely on SCADA systems to manage, direct and monitor large-scale, distributed operations. Similar to the Waterloo Campaign, control failures in these systems can result in devastating consequences.

This section describes and models a pipeline rupture incident that occurred at Fork Shoals, South Carolina on June 26, 1996. The rupture released 957,600 gallons of fuel oil and caused damage estimated at $20.5 million. According to the National Transportation Safety Board (NTSB) Pipeline Accident Report [8], the incident occurred as a result of failures in system components and improper operator actions. However, as we show in this section, the same results can be produced by targeted cyber attacks. For brevity, only the critical events that led to the pipeline rupture are discussed.

5.1 Pipeline Rupture Incident

The Fork Shoals pipeline transports fuel oil from Atlanta to Greensboro (North Carolina). Figure 8 shows the pipeline section of interest, which contains four pump stations (A–D), a delivery facility (F) with breakout tankage and a control center (Z) located in a central office north of the pipeline. Control Center Z houses operators that remotely monitor and control the pipeline using communication links (L1–L4). The remote pump stations (A–D) each have one RTU that controls actuators and reads pipeline sensors. Delivery Facility F is monitored by the control center, but the communication link is not shown because it is not pertinent to the analysis.

The pipeline rupture was due to two primary factors: (i) increase in pressure flow beyond the maximum allowable pipeline pressure, and (ii) failure of operator to realize and correct the conditions before structural failure occurred. The control failures that resulted in the pressure increase occurred in the following chronological order:

- **Increase in Pumping Capacity (Control Failure 1):** Pumping capacity is increased by starting additional pumps and/or turning on larger pumps and shutting down smaller ones. After a transfer to Delivery Facility F was completed, pumping capacity at downstream pumping facilities was sequentially increased to accommodate the additional fuel oil in the

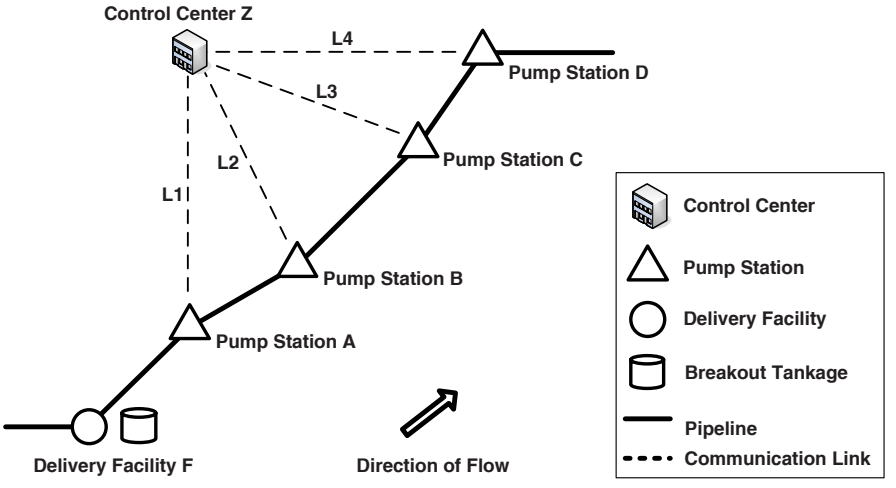


Figure 8. Pipeline layout.

pipeline. In particular, Pump Stations A and B each started a second pump. Also, Pump Station C started a larger pump and shut down a smaller one.

- **Failure to Start a Second Pump at Pump Station D (Control Failure 2):** The operator attempted to start a larger pump at Pump Station D to increase capacity. The operator noted that a green light appeared on the console to indicate that the pump had started, but, for some reason, the pump did not start.

- **Stoppage of the Active Pump at Pump Station D (Control Failure 3):** Believing that two pumps at Pump Station D were running, the operator stopped the smaller (and only operating) pump. Shutting down the only pump at Pump Station D created a pressure surge that traveled upstream to Pump Station C, causing its only operating pump to shut down due to high discharge pressure. The resulting second pressure surge caused the two pumps at Pump Station B to shut down. The continued high fuel oil flow rate caused the pipeline pressure to grow rapidly, resulting in a rupture between Pump Stations A and B.

Meanwhile, the pipeline operator at Control Center Z either ignored, misinterpreted or did not receive alarm notifications. The following control failure hindered the pipeline operator's situational awareness:

- **Failure to Receive and React to Alert Notifications (Control Failure 4):** A pressure alarm was triggered shortly after the pump at Pump Station B shut down; however, the operator took no action. Additionally, low suction pressure alarms were triggered intermittently for readings at Pump Station B, but the operator did not react because the

alarms were behaving erratically and he assumed that the pressure readings were inaccurate. Moreover, the SCADA system did not report the failure of the pump to start at Pump Station D.

5.2 Cyber Attack Scenario

This section uses a cyber attack scenario to recreate the control system failures that led to the pipeline rupture. As shown in Figure 8, the system incorporates five nodes: Z ($\in$ *ControlNodes*) and A–D ($\in$ *EdgeNodes*), and four communication links L1–L4 ($\in$ *Links*).

The attacks described below involve compromising *ControlNode* Z and using it as the *AttackNode*. Essentially, the attacker has "root" access to Z, and can block ($\neg$) and fabricate (+) messages.

Control Failure 1: This control failure occurred because of a series of messages that started and stopped various pumps along the pipeline. The attacker begins by fabricating (+) a message from Z to A to start the second pump (Pump #2):

$$\text{Z} \underset{Z\ +Request}{\overset{L1}{\longrightarrow}} \text{A[Start Pump \#2]}.$$

Pump Station A then generates an acknowledgment message to Z confirming that Pump #2 has started. In pipeline control protocols (e.g., Modbus), field (slave) devices typically send acknowledgements in response to requests from the control center (master); these acknowledgment messages must be blocked to mask the attack from the operator. Thus, the next step is to block ($\neg$) the acknowledgment message from A to Z:

$$\text{A} \underset{Z\ \neg Reply}{\overset{L1}{\longrightarrow}} \text{Z[ACK]}.$$

Next, the attacker sends a fabricated message from Z to B to start Pump #2 and blocks the acknowledgement message:

$$\begin{aligned} \text{Z} &\underset{Z\ +Request}{\overset{L2}{\longrightarrow}} \text{B[Start Pump \#2]}, \\ \text{B} &\underset{Z\ \neg Reply}{\overset{L2}{\longrightarrow}} \text{Z[ACK]}. \end{aligned}$$

The attacker then starts a larger pump (Pump #3) at Pump Station C and blocks the acknowledgement message:

$$\begin{aligned} \text{Z} &\underset{Z\ +Request}{\overset{L3}{\longrightarrow}} \text{C[Start Pump \#3]}, \\ \text{C} &\underset{Z\ \neg Reply}{\overset{L3}{\longrightarrow}} \text{Z[ACK]}. \end{aligned}$$

Finally, the attacker stops the smaller pump (Pump #1) at Pump Station C and blocks the acknowledgement message:

$$\mathrm{Z} \xrightarrow[Z\ +Request]{L3} \mathrm{C[Stop\ Pump\ \#1]},$$
$$\mathrm{C} \xrightarrow[Z\ \neg Reply]{L3} \mathrm{Z[ACK]}.$$

Control Failure 2: This control failure occurred because the operator did not realize that the larger pump (Pump #3) at Pump Station D had not started despite directing it to start. The attacker implements this failure by ensuring that Pump #3 does not start and that the operator is unaware of this situation. Consequently, the attacker blocks a message from the Control Center Z to Pump Station D to start Pump #3 and fabricates an acknowledgement from D to Z that Pump #3 has started:

$$\mathrm{Z} \xrightarrow[Z\ \neg Request]{L4} \mathrm{D[Start\ Pump\ \#3]},$$
$$\mathrm{D} \xrightarrow[Z\ +Request]{L4} \mathrm{Z[ACK]}.$$

Control Failure 3: This control failure occurred because the operator stopped the only pump (Pump #1) at Pump Station D. The attacker implements this failure by fabricating a message to stop Pump #1 at Pump Station D and then blocking the acknowledgment message from D to Z:

$$\mathrm{Z} \xrightarrow[Z\ +Request]{L4} \mathrm{D[Stop\ Pump\ \#1]},$$
$$\mathrm{D} \xrightarrow[Z\ \neg Reply]{L4} \mathrm{Z[ACK]}.$$

Control Failure 4: This control failure occurred because the operator did not receive and react to alert notifications. Pressure fluctuations and pump shutdowns trigger alarms in SCADA systems. Alarms in typical SCADA systems are sent in the form of response messages from a field device to the master during normal polling cycles. The control failure is implemented by having the attacker block polling messages from Z to A–D and fabricate response messages to reflect normal operating conditions:

$$\mathrm{Z} \xrightarrow[Z\ \neg Request]{L1} \mathrm{A[Poll]},\ \mathrm{A} \xrightarrow[Z\ +Reply]{L1} \mathrm{Z[ACK]},$$
$$\mathrm{Z} \xrightarrow[Z\ \neg Request]{L2} \mathrm{B[Poll]},\ \mathrm{B} \xrightarrow[Z\ +Reply]{L2} \mathrm{Z[ACK]},$$
$$\mathrm{Z} \xrightarrow[Z\ \neg Request]{L3} \mathrm{C[Poll]},\ \mathrm{C} \xrightarrow[Z\ +Reply]{L3} \mathrm{Z[ACK]},$$
$$\mathrm{Z} \xrightarrow[Z\ \neg Request]{L4} \mathrm{D[Poll]},\ \mathrm{D} \xrightarrow[Z\ +Reply]{L4} \mathrm{Z[ACK]}.$$

5.3 Modeling the Cyber Attack Scenario

Figures 9 and 10 present the formal specification of the cyber attacks described above. Figure 9 specifies the nodes, attacker capabilities, and the possible and initial node states. The system has five nodes: Control Center Z,

System	**Possible Node States**
Z $\in$ *ControlNodes*	A := A^0 | A^1 | A^2
A, B, C, D $\in$ *EdgeNodes*	$A^0 \equiv$ (1, 0, 0, 0, 0)
Z $\in$ *AttackNodes*	$A^1 \equiv$ (1, 1, 0, 0, 0)
L1, L2, L3, L4 $\in$ *Links*	$A^2 \equiv$ (*, *, *, *, 1)
L1 = (Z, A)	B := B^0 | B^1 | B^2
L2 = (Z, B)	$B^0 \equiv$ (1, 0, 0, 0, 0)
L3 = (Z, C)	$B^1 \equiv$ (1, 1, 0, 0, 0)
L4 = (Z, D)	$B^2 \equiv$ (*, *, *, *, 1)
Request, Reply $\in$ *MsgTypes*	C := C^0 | C^1 | C^2 | C^3
	$C^0 \equiv$ (1, 0, 0, 0, 0)
Capabilities	$C^1 \equiv$ (1, 0, 1, 0, 0)
Z = {¬, +}	$C^2 \equiv$ (0, 0, 1, 0, 0)
	$C^3 \equiv$ (*, *, *, *, 1)
Initial Node States	D := D^0 | D^1 | D^2 | D^3
A := A^0	$D^0 \equiv$ (1, 0, 0, 0, 0)
B := B^0	$D^1 \equiv$ (0, 0, 1, 0, 0)
C := C^0	$D^2 \equiv$ (0, 0, 0, *, *)
D := D^0	$D^3 \equiv$ (*, 0, 1, 0, 0)
Z := (A^0, B^0, C^0, D^0)	Z := (A^0 | A^1 | A^2,
	B^0 | B^1 | B^2,
	C^0 | C^1 | C^2 | C^3,
	D^0 | D^1 | D^2 | D^3)

Figure 9. Formal specification of the cyber attacks (nodes, capabilities and states).

which is both a *ControlNode* and an *AttackNode*, and Pump Stations A–D. The attacker has the capability to block (¬) and fabricate (+) messages.

Table 1 shows the possible states of Pump Stations A–D. The first three columns list the status of (small) Pump #1, (small) Pump #2 and (large) Pump #3. The fourth column shows whether or not the pressure reading is within acceptable limits. The fifth column indicates if the node is an alarm state. The status of a pump is represented as On (1) or Off (0). The pressure status is Acceptable (0) (i.e., within acceptable limits) or Not Acceptable (1). The alarm status is True (1) or False (0). Note that a "*" entry signifies that the specific binary value does not matter for that particular state.

The initial states for all four pump stations are identical:

$$A^0, B^0, C^0, D^0 \equiv (1, 0, 0, 0, 0).$$

This means that (small) Pump #1 is on, the other two pumps are off, the pressure readings are within acceptable limits and no alarms are triggered. The node states vary as different actions are performed along the pipeline.

The four control failures can be broken down into four processes: Processes 1, 2, 3 and 4. Process 1 (Figure 10) occurs only once, so $\psi = 1$. The second pump (Pump # 2) at Pump Stations A and B are activated. At Pump Station C, the small pump (Pump #1) is deactivated and the large pump (Pump #3) is activated unbeknownst to the operator.

Process 2 (Figure 10) corresponds to Control Failure 2. It is conditional on a message being sent to activate the large pump (Pump #3) at Pump Station

Table 1. Possible states of *EdgeNodes* (Pump Stations A–D).

State	Pump #1	Pump #2	Pump #3	Pressure	Alarm
A^0	On	Off	Off	Acceptable	F
A^1	On	On	Off	Acceptable	F
A^2	*	*	*	*	T
B^0	On	Off	Off	Acceptable	F
B^1	On	On	Off	Acceptable	F
B^2	*	*	*	*	T
C^0	On	Off	Off	Acceptable	F
C^1	On	Off	On	Acceptable	F
C^2	Off	Off	On	Acceptable	F
C^3	*	*	*	*	T
D^0	On	Off	Off	Acceptable	F
D^1	Off	Off	On	Acceptable	F
D^2	Off	Off	Off	*	*
D^3	*	Off	On	Acceptable	F

D. This condition always holds after the operator sends an activation message, so $\psi = \infty$.

Process 3 (Figure 10) corresponds to Control Failure 3. The steps deactivate Pump #1 at Pump Station D. Process 3, like Process 1, is executed once, so $\psi = 1$.

Process 4 (Figure 10) corresponds to Control Failure 4. Since polling is a continuous process, $\psi = \infty$. As discussed above, the attacker blocks polling messages from Z to A–D and fabricates response messages to reflect normal operating conditions despite the build-up of pressure that eventually causes the pipeline to rupture.

Note that a targeted cyber attack would simply turn off the pumps and mask the actions. However, the additional steps are incorporated in the example above to model the events described in the NTSB report.

6. Model Evaluation

This section discusses key applications of the model and its relationship to other work in the field.

6.1 Model Applications

The model was created specifically to express attacks on critical infrastructure assets. However, as demonstrated in the examples involving the Waterloo Campaign and pipeline rupture incident, the model can also be used to express failures in control protocols. In fact, the model is capable of expressing attacks and failures in diverse protocols that involve the exchange of messages.

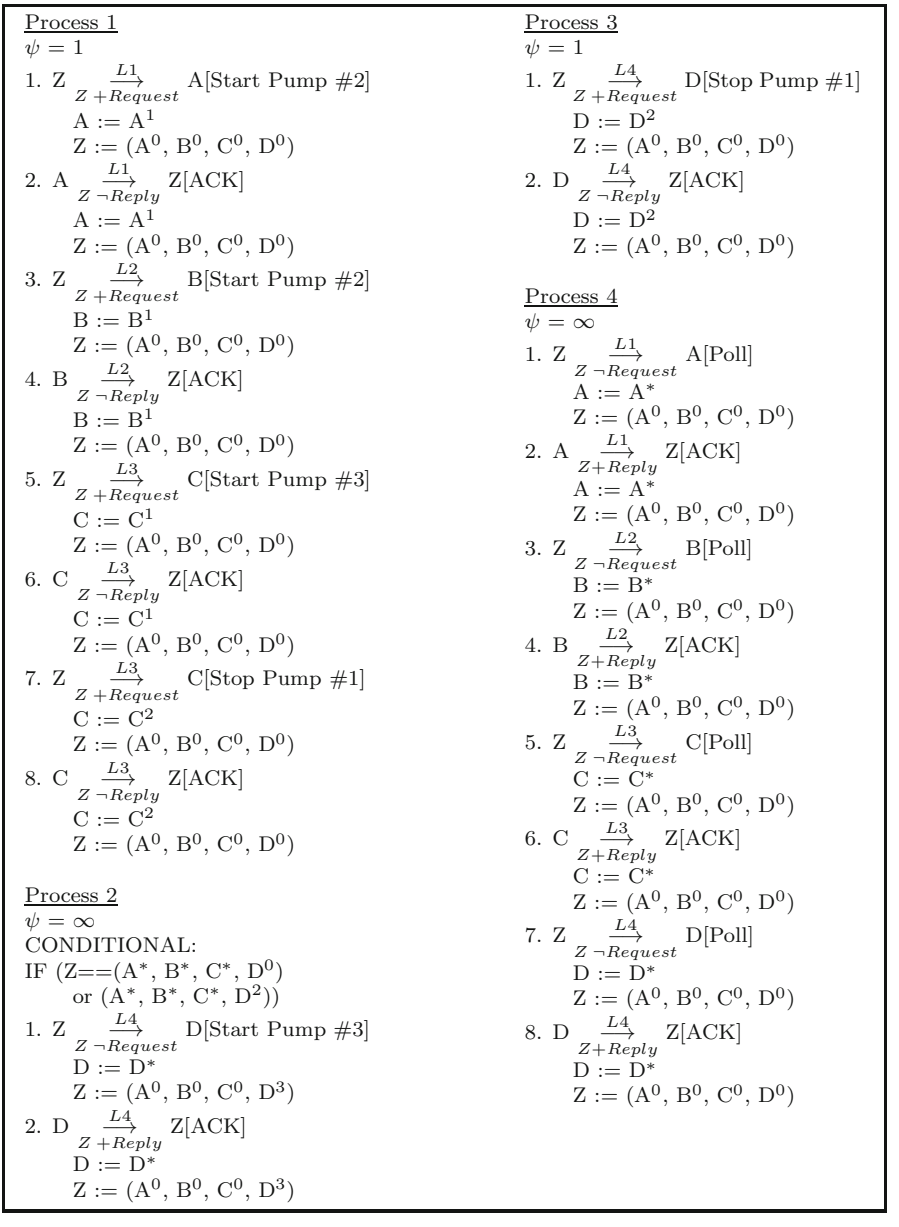


Figure 10. Formal specification of Control Failures 1, 2, 3 and 4.

The model provides a powerful mechanism for conceptualizing attacks and failures in one control protocol and translating them to similar attacks and failures in another protocol. For example, the sequence of attacks involved in rupturing a pipeline is very similar to the attack sequences that turn off a section

of the electric power grid or shut down telephone service. The nodes (pumping stations, generators and service switching/transfer points) and protocols (Modbus, DNP3 and SS7) for pipelines, power grid and telecommunications infrastructures are different, but the attack strategies are practically identical and merely involve different messages with different payloads.

Techniques for masking attacks are just as similar across critical infrastructures. For example, the attack on the polling mechanism in the oil pipeline example is applicable to numerous protocols in the oil and gas sector (e.g., Modbus and Fisher ROC) and to protocols in other sectors such as the electric power grid (DNP3) and manufacturing (Profibus). The underlying strategy is to block polling messages and fabricate normal responses to mask alert conditions. Some critical infrastructure protocols (e.g., DNP3) support additional communication modes (e.g., unsolicited replies), but these modes are readily accommodated by the model.

In addition to conceptualizing common attacks and attack strategies for different protocols, the formal model assists in targeting specific protocol and system implementations. Developing attacks requires detailed information about the control system as well as the underlying cyber-physical systems. The model helps identify the information requirements, articulate possible outcomes and examine the feasibility of attacks based on the available information.

The formal model provides a framework for infrastructure asset owners and operators to evaluate system implementations and configurations for possible weaknesses. Moreover, the comprehensive understanding of attacks supports risk analysis and risk management, and the implementation of defensive postures. In particular, common vulnerabilities derived via attack analysis can be grouped into general security dimensions to aid the development and deployment of mitigation strategies. The formal model also enables researchers to analyze existing control protocols and to evaluate the implications of design decisions in new protocols.

6.2 Comparison With Other Work

Numerous models have been developed for expressing and reasoning about the security properties of computer systems and protocols. This section compares the proposed model with some of the established approaches for modeling attacks on computer and control systems.

The majority of attack models for computer systems are founded on the notion of an attack tree [12]. The root node of an attack tree denotes the goal of the attacker. The steps for completing the attack are decomposed using parent-child relations such that a path from a leaf node to the root node expresses one instance of the attack. The possible attacks correspond to the different branches of the attack tree. Attack trees have been shown to capture a variety of computer system attack scenarios [7, 17].

Attack tree models are attractive because of their simplicity and ease of analysis. However, certain ambiguities and the lack of expressiveness of attack tree models hinder quantitative reasoning and comparison. Precise analysis is

difficult because an attack tree node represents both the current system state and the specific attack action [19]. Also, attack trees are susceptible to state explosion, they do not represent the temporal aspects of dependent attacks, and lack the ability to generalize beyond the modeled scenario [16].

An alternative model [16] views an attack as a series of capabilities instead of a sequence of events. It provides constructs for defining intrusion signatures and automating the discovery of attacks. Other researchers have proposed taxonomies for classifying attacks [9] and ontology-based models [18]. These models describe complex scenarios involving multiple attacks; however, they lack formalisms for analyzing causality and sequential events. Additionally, in these models, attacks focus on peer-to-peer communications and do not arise from a holistic view of the system. Attack models for control systems require the ability to express hierarchical communications, logical sequences of events and system-wide situational awareness.

Models based on finite state machines allow the formal representation of discrete-event, dynamic systems. Finite state machines facilitate the logical analysis of deterministic and non-deterministic attributes. Finite state machine models based on Petri nets have been used to express common computer attacks [6, 19]. Petri nets use graphs and set theory to model concurrency, synchronization, resource allocation and randomness. Our formal model specifies attack sequences and events that may be expressed and analyzed using Petri nets or other graph theoretical constructs.

Models for failures in control systems have been developed primarily to analyze reliability, resilience, functionality and risk. The U.S. Department of Energy [10], National Institute of Standards and Technology [14] and other entities [13] have developed models for predicting, reacting to and understanding failures in control systems. Several failure analysis models (e.g., [4]) demonstrate the conditions that can lead to failures in control systems. Our model does not directly focus on failure analysis. However, a failure analysis model can be used to determine the conditions under which physical damage can occur. The conditions can then be analyzed using our model to determine how a control protocol may be attacked to cause physical damage.

Little, if any, research has specifically focused on modeling attacks on control system protocols. Current work has tended to focus on modeling intrusion and anomaly signatures, identifying attributes for system resilience and evaluating system vulnerabilities for risk analysis. Cheung, *et al.* [3] describe a language for modeling attacks on process control systems. Because the systems are relatively static, models can be constructed to characterize the expected system behavior. Traffic that does not conform to normal traffic patterns is identified as a potential attack. The concept focuses on a purely defensive posture that assumes attackers will not use legitimate traffic in an attack. Our model is attack-centered in that it allows legitimate (as well as non-standard) messages in attack scenarios. Moreover, our model expresses temporal aspects and causal effects while lending itself to formal analysis.

7. Conclusions

The model presented in this paper can express control system failures and attacks on control protocols that involve the exchange of messages. The model helps provide a comprehensive understanding of control system failures and attacks, which supports the development and analysis of attack and defense strategies.

Our future research will concentrate on developing a graph-theoretic model augmented with temporal and belief attributes. The extended model will permit the specification of common modes of attack (e.g., control node compromise, edge node compromise, malicious control of edge nodes and polling mechanism manipulation). The model will also facilitate the development of attack metrics and will support formal reasoning about attacks and defensive strategies.

References

[1] J. Buckley, *Air Power in the Age of Total War*, Indiana University Press, Bloomington, Indiana, 1999.

[2] D. Chandler, *Waterloo – The Hundred Days*, Macmillan, New York, 1980.

[3] S. Cheung, U. Lindqvist and M. Fong, Modeling multistep cyber attacks for scenario recognition, *Proceedings of the Third DARPA Information Survivability Conference and Exposition*, pp. 284–292, 2003.

[4] L. Decker, A risk assessment model for pipeline facility operations, *Pipeline and Gas Journal*, vol. 236(3), pp. 38–44, 2009.

[5] HowStuffWorks.com, Battle of Waterloo, Atlanta, Georgia (history.howstuffworks.com/european-history/battle-of-waterloo.htm), 2008.

[6] J. McDermott, Attack net penetration testing, *Proceedings of the New Security Paradigms Workshop*, pp. 15–21, 2000.

[7] A. Moore, R. Ellison and R. Linger, Attack Modeling for Information Security and Survivability, Technical Note CMU/SEI-2001-TN-001, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2001.

[8] National Transportation Safety Board, Pipeline Rupture and Release of Fuel Oil into the Reedy River at Fork Shoals, South Carolina, Pipeline Accident Report PB98-916502/NTSB/PAR-98/01, Washington, DC, 1996.

[9] P. Neumann and D. Parker, A summary of computer misuse techniques, *Proceedings of the Twelfth National Computer Security Conference*, pp. 396–407, 1989.

[10] North American Electric Reliability Corporation, Reliability Functional Model, Function Definitions and Functional Entities, Version 5, Princeton, New Jersey (www.nerc.com/fileUploads/File/Standards/Functional_Model_V5_Clean_2009Sept24.pdf), 2009.

[11] A. Roberts, *Waterloo – June 18, 1815: The Battle for Modern Europe*, HarperCollins, New York, 2005.

[12] B. Schneier, Attack trees, *Dr. Dobb's Journal*, vol. 24(12), pp. 21–29, 1999.

[13] J. Stamp, M. Berg and M. Baca, Reference Model for Control and Automation Systems in Electrical Power, Version 1.2, Sandia National Laboratories, Albuquerque, New Mexico (www.oe.energy.gov/DocumentsandMedia/Reference_Model_for_Control_and_Auto_Systems_in_Elec_Ind.pdf), 2005.

[14] K. Stouffer, J. Falco and K. Scarfone, Guide to Industrial Control Systems Security, Final Public Draft, NIST Special Publication 800-82, National Institute of Standards and Technology, Gaithersburg, Maryland, 2008.

[15] A. Tan, G. Badhwar, F. Allahdadi and D. Medina, Analysis of Solwind fragmentation event using theory and computations, *Journal of Spacecraft and Rockets*, vol. 33(1), pp. 79–85, 1996.

[16] S. Templeton and K. Levitt, A requires/provides model for computer attacks *Proceedings of the New Security Paradigms Workshop*, pp. 31–38, 2000.

[17] C. Ten, C. Liu and M. Govindarasu, Vulnerability assessment of cybersecurity for SCADA systems using attack trees, *Proceedings of the IEEE Power Engineering Society General Meeting*, pp. 1–8, 2007.

[18] J. Undercoffer, J. Pinkston, A. Joshi and T. Finin, A target-centric ontology for intrusion detection, *Proceedings of the IJCAI Workshop on Ontologies and Distributed Systems*, pp. 47–58, 2004.

[19] R. Wu, W. Li and H. Huang, An attack modeling based on hierarchical colored Petri nets, *Proceedings of the Second International Conference on Computer and Electrical Engineering*, pp. 918–921, 2008.

Chapter 5

HIGH SECURITY WITH LOW LATENCY IN LEGACY SCADA SYSTEMS

Rouslan Solomakhin, Patrick Tsang and Sean Smith

Abstract Message authentication with low latency is necessary to ensure secure operations in legacy industrial control networks such as those in the power grid. Previous authentication solutions that examine single messages incur noticeable latency. This paper describes Predictive YASIR, a bump-in-the-wire device that reduces the latency by considering broader patterns of messages. The device predicts the incoming plaintext based on previous observations; compresses, encrypts and authenticates data online; and pre-sends a portion of the ciphertext before receiving the entire plaintext. The performance of Predictive YASIR is evaluated using a simulation involving the Modbus/ASCII protocol. By considering broader message patterns and using predictive analysis, improvements in latency of $15.48 \pm 0.35\%$ are obtained.

Keywords: Legacy SCADA systems, security, low latency

1. Introduction

The United States power grid was built half a century ago, when network-based attacks were practically non-existent. New threats warrant retrofitting security in legacy networks in the power grid. Protecting a legacy network is difficult, however, because critical infrastructure components must communicate rapidly, but security slows the communications. This paper presents a predictive approach that optimizes the performance of the previous fastest security solutions.

2. Background

A power utility typically monitors and controls operations using a partially-unsecured, slow legacy network that connects substations and control centers. In a control center, human operators ensure safe and continuous operation of the grid by monitoring data terminals. A terminal provides a visual represen-

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 63–79, 2010.

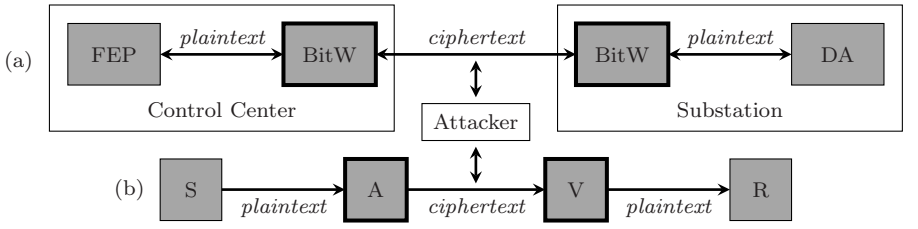


Figure 1. Typical BitW device setup.

tation of the data received from a front end processor (FEP), which exchanges messages with data aggregators (DAs) in substations. A FEP and DA connect via a slow legacy point-to-point network, which is often unsecured [4].

When communications are unsecured, an adversary can insert messages in network traffic and impersonate any device on the network. For example, an adversary could impersonate a FEP and command a DA to perform tasks that are not appropriate under normal operations. The adversary may replay an "increase power output" message from the FEP multiple times or increase the value in the message "set power output to 10 MW," which would overload the substation, possibly causing a rolling blackout in the power grid.

An adversary who impersonates a DA can replay old DA status messages to the FEP, which would forward these messages to the operator's terminal. Since the terminal would be receiving old status messages, it would not reflect the abnormalities in the power grid, and the operator would likely not detect the attack. Even if the operator discovers abnormalities through an alternative channel, understanding the scope of the problem is difficult in the absence of correct data; at the very least, this would slow the operator's response to the attack, giving the adversary time to subvert other substations.

Such attacks violate the message authenticity assumption made by the FEP and DA. Although FEPs and DAs can be upgraded to authenticate messages, the upgrades are prohibitively expensive and the upgraded devices would still have to communicate over the slow legacy network. Also, the required network upgrade is expensive. The cheaper and faster option is a bump-in-the-wire (BitW) device [10, 14, 15, 17] that secures all messages in a legacy network.

Two BitWs work in concert to authenticate a message (Figure 1(a)). Before a message from the FEP leaves the control center, the authenticator BitW reformats the original plaintext message into a ciphertext message with an added counter and a redundancy check (or message digest). The counter prevents an attacker from replaying old messages. The digest depends on the message, counter and a digest key shared by the pair of BitWs. Due to the cryptographic properties of the mechanism used to generate the digest, it is intractable for an adversary without the digest key to construct the correct digest for an altered ciphertext. When the ciphertext arrives at the substation, the verifier BitW compares its own calculation of the ciphertext digest with what it has received. If the two digests match, the verifier reformats the ciphertext into a plaintext

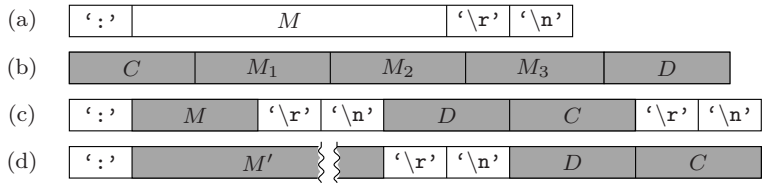


Figure 2. Message formats.

message and forwards it to the DA. Note that a BitW is still susceptible to attacks if the adversary gains access to an FEP or DA before the BitW.

When a DA sends a message to a FEP, the authenticator (A) and the verifier (V) switch roles. Due to this symmetry, we prefer to refer to the entities as sender (S) and receiver (R) (Figure 1(b)).

Figure 2 presents the Modbus/ASCII protocol message formats, which are used in this paper. Figure 2(a) shows the plaintext representation of a message while Figure 2(b) shows the position embedding (PE) ciphertext in blocks (described later). Figures 2(c) and 2(d) present the YASIR and Predictive YASIR ciphertext representations, respectively (described later). Note that the C and D denote the counter and digest, respectively. Also, the shaded areas in a message correspond to portions that are modified or generated by a BitW.

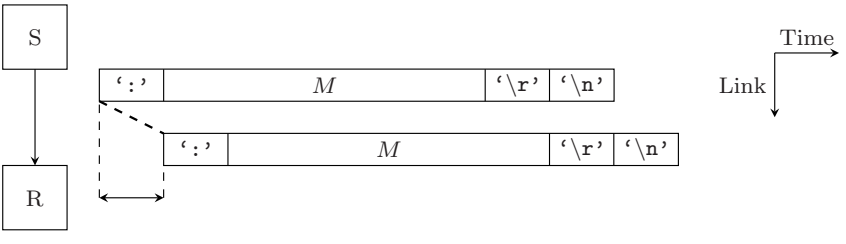


Figure 3. Transmission latency without authentication.

In a slow legacy power grid network, the end-to-end latency of a message typically should not exceed a certain value. Suppose this bound is 300 ms (Figure 3, where we use the diagram style from the YASIR paper [14]). A millisecond equals the time in which a device transmits 0.9 bytes (or 0.9 "byte-times") on a network with a bandwidth of 9,600 baud if a byte has 10 bits. Because of the low bandwidth, a BitW should not wait to receive the entire message before processing it, a practice known as "hold-back." If both BitWs in a pair hold-back a message that is longer than 144 bytes, the delay would exceed 300 ms. Instead, a BitW should forward each byte quickly or process the message online. The online processing must thwart an attacker who attempts to replay or modify ciphertext.

Figure 4 illustrates the notion of latency with hold-back, where both BitWs delay the entire message before forwarding it. As before, the shaded areas are modified or generated by a BitW. In the figure, the hold-back delays a message

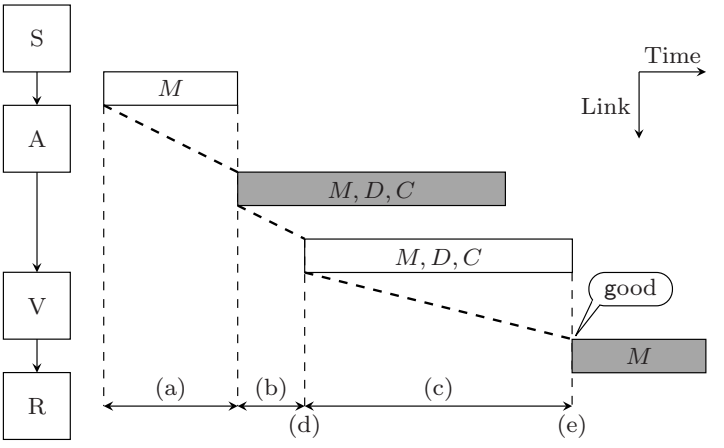


Figure 4. Latency with hold-back.

by time intervals (a) and (c), and the transmission delay is (b). The verifier starts receiving the message at time (d), but begins to forward it to the receiver only at time (e), when it has checked that the digest is correct. If an attacker modifies the message, the verifier drops it.

Next, we consider how online processing addresses each type of attack.

Tsang and Smith [14] have demonstrated that an online BitW can stop an attacker who attempts to replay an old message (replay attack) without message delay. The delay is absent because, although the authenticator transmits the counter at the end of the ciphertext, the verifier forwards the entire message to the receiver before checking the counter. Since the counter affects the digest, the verifier increments the counter from the previous message to calculate the new value. If the calculated value is larger than what is received by the verifier, the received counter is ignored. If the calculated value is smaller than what is received, the verifier sets its own counter to the received value to synchronize with the authenticator. Before the counter overflows, the pair of BitWs reset their values to zero and change the digest key.

To prevent an attacker from modifying a message, the BitW pair appends a digest after the message but before the counter. This digest depends on the message, counter and digest key. The verifier compares the received digest with the value it calculates. If the two digests match, then the verifier forwards the message to the receiver. Intuitively, one might think that the verifier cannot process the ciphertext online – that it must wait until it receives the entire digest before it can forward the message to the receiver, thus incurring a byte-time of latency for each byte in the message. Although this is done in some earlier approaches, Wright, *et al.* [17] have suggested that the message be forwarded as soon as the verifier receives it, but to introduce a CRC error if an attacker modifies the message. This exploits the receiver's ability to detect random errors. We use a similar technique (described later) to enable the verifier to process the message online.

When an authenticator appends the digest to a message, it must ensure that the verifier can distinguish between them. Two options are available. The first option is to prepend the message length to the ciphertext. However, a common protocol like Modbus/ASCII (Figure 2(a)) has variable length messages and does not specify the length in a message. Consequently, to identify the length of a message in this protocol, an authenticator must hold-back the entire message. The second option is do it online by delimiting the ciphertext portions using a message digest separator. If the separator appears in the message data, then the authenticator marks it with a special symbol to avoid confusing the verifier, i.e., the authenticator "escapes" the separator within the message. Modbus/ASCII has a message end indicator that can be used as the message digest separator. In general, new separators and escapes delay a message, but do not enhance its authenticity. Indeed, they constitute encoding inefficiencies required for online authentication. One of our goals is to eliminate these inefficiencies in online processing.

Note that we do not attempt to eliminate the overhead due to the digest by compressing or pre-sending it. A BitW cannot compress the digest because a strong digest does not have a pattern. Also, a BitW cannot pre-send a part of the digest before the device receives the entire plaintext message; this would weaken the strength of the digest. If an authenticator pre-sends a part of the digest, the pre-sent part would contain no information about the part of the message that the authenticator has not yet received.

3. Related Work

In most SCADA environments, message integrity is more important than confidentiality. An attacker can always learn the state of the system from the physical world. For example, an attacker may see the open flood gates of a dam and deduce that the control station has sent an "open flood gates" message and the substation has sent a "flood gates open" message. On the other hand, if no measures are in place to preserve message integrity, an attacker can cause actions such as opening the flood gates or shutting down a generator with significant negative repercussions. If a utility also needs to hide message content, a BitW device can provide confidentiality with zero latency using a stream cipher (e.g., AES in counter mode), which encrypts a plaintext stream by XORing it with a pseudo-random stream. Therefore, we consider related BitW solutions only if they authenticate messages, namely, SEL-3021-2, AGA SCM and YASIR. Also, we ignore solutions such as SEL-3021-1 because it does not protect message integrity and PNNL's SSCP embedded device because it does not represent a bump-in-the-wire solution.

3.1 SEL-3021-2

SEL-3021-2 [10] is a commercial off-the-shelf BitW device from Schweitzer Engineering Laboratories. The device uses the Message Authentication Protocol (MAP) [11] to provide integrity with HMAC-SHA-1 or HMAC-SHA-256

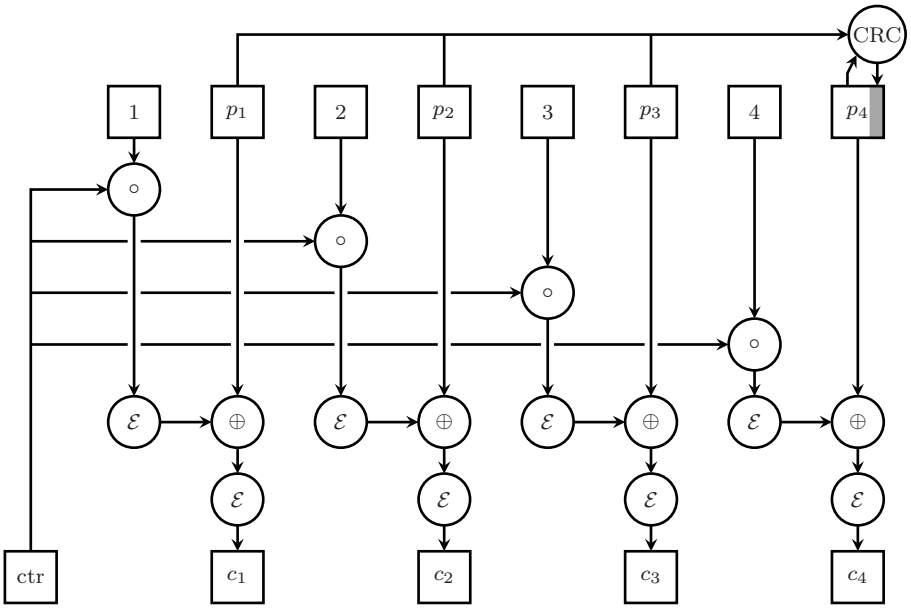


Figure 5. Position embedding (PE) mode.

digests. The specifications do not provide latency information for this device; however, Schweitzer does not recommend the use of SEL-3021-2 if low latency is desired.

3.2 AGA SCM

The American Gas Association SCADA Cryptographic Module (AGA SCM) [15] is a BitW device proposed by the AGA 12 Task Group. A reference implementation is available that provides several hold-back modes and one online mode [16]. The hold-back modes buffer the entire message before checking the digest and sending the message, slowing the message by a time interval in linear proportion to its size (see Figure 3). The online position embedding (PE) mode [17] is a modified version of AES in counter mode (AES-CTR) followed by a standard version of AES in electronic code book mode (AES-ECB) (Figure 5).

As shown in Figure 5, the PE mode is AES-CTR followed by AES-ECB with the same key. The mode relies on CRC to authenticate messages. The symbols $\circ$ and $\oplus$ denote concatenation and XOR, respectively. The symbol $\mathcal{E}$ is the encryption function. The plaintext is $p_1 \circ p_2 \circ p_3 \circ p_4$ and the ciphertext is $\mathrm{ctr} \circ c_1 \circ c_2 \circ c_3 \circ c_4$. Blocks p_i and c_i are 16 bytes long; the message counter *ctr* is 14 bytes long and the block counters (1–4 in Figure 5) are 2 bytes long. Depending on the protocol, the CRC is 2 to 4 bytes long.

Figure 6 demonstrates the latency in the PE mode. The PE mode delays a message by 32 byte-times, because both BitWs in a pair buffer 16-byte blocks

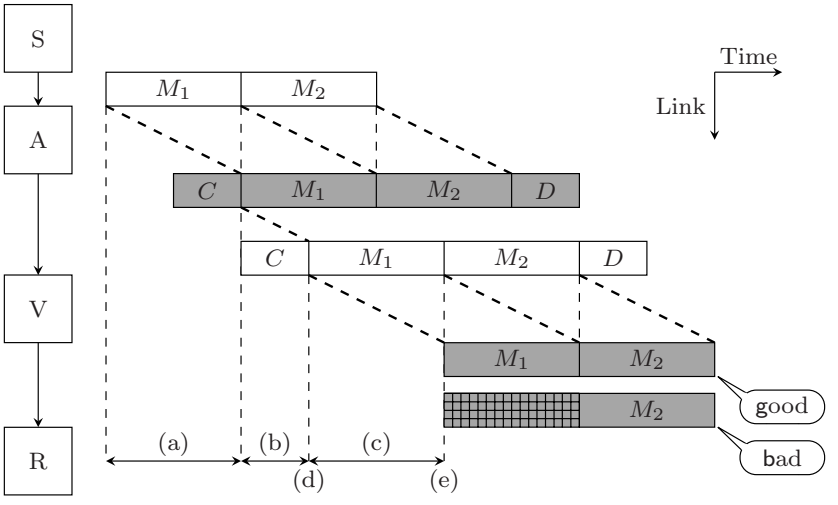


Figure 6. Latency in the PE mode.

of data to apply AES-ECB. M_1 and M_2 are blocks of message M. Both BitWs buffer a 16-byte block of a message before forwarding it, delaying the message by 32 byte-times, denoted by (a) and (c). The verifier starts receiving the message at time (d) and begins to forward it to the receiver at time (e). If an attacker attempts to modify a block in the message, the verifier unconsciously scrambles the block (crossed out), which the receiver detects by checking the CRC at the end of the message.

AGA 12 modifies the manner in which AES-CTR generates counters. First, the authenticator increments a 14-byte session clock by one every r microseconds, where r is the "counter resolution." Next, the authenticator concatenates the session clock with a 2-byte block counter. The authenticator sets the block counter to zero at the beginning of each message and increments it by one for each block in the message. Finally, the authenticator encrypts the resulting 16-byte counter value and XORs the encrypted counter with a plaintext block like the standard AES-CTR. To avoid using the same counter for two messages, AGA 12 requires the counter resolution to be set so that an authenticator can send at most one message in a single session clock tick.

The PE mode relies on CRC for message integrity. Note that using CRC for plaintext protects against random errors, but not from malicious attacks on message integrity. Similarly, a BitW device cannot protect message integrity using AES-CTR or AES-ECB alone. The counter mode is malleable [3, 7], i.e., an adversary can modify the ciphertext with predictable changes to the CRC even without the encryption key.

The ECB mode is vulnerable to a known-plaintext attack, where an adversary who knows the plaintext of two messages can splice message portions to create a third message such that the CRC of the new message equals the CRC of one of the original messages. The PE mode prevents splicing and predictable

changes to the ciphertext by combining the CTR and ECB modes. However, using such a combination with CRC to ensure message integrity is not recommended because of potential security problems. One example is the cipher block chaining (CBC) mode, which may not provide message integrity protection with CRC. This result was demonstrated by Stubblebine and Gligor [13], who exploited the predictability of CRC to create undetectable bogus messages for Kerberos and remote procedure calls (RPCs).

Thus, the PE mode depends on the nonmalleability of its ciphertext: if an adversary changes the ciphertext, it is impossible to predict what happens to the CRC. If an adversary inserts, removes or reorders blocks, then the verifier BitW scrambles the plaintext in the CTR step. If an adversary modifies a message in the PE mode, the verifier BitW scrambles the plaintext in the ECB step. Because of such scrambling, the receiver will detect a CRC error. The probability of this error is 2^{-h} where h is the length of the CRC. CRC variants range in length from 8 bits to 32 bits, but AGA 12 specifies that this mode be used when the CRC is at least 16 bits.

3.3 YASIR

YASIR [14] is a BitW device that authenticates messages with at most 18 byte-times of overhead (Figure 2(c)). The actual overhead depends on the underlying protocol. With Modbus/ASCII, YASIR delays a message by about 16 byte-times. The delay comprises the 12 bytes of HMAC-SHA-1-96 digest for data integrity, 2 bytes for the authenticator to detect the end of the message, and 2 bytes for the verifier to have an opportunity to control the message CRC. Building on the ideas from the PE mode [15], YASIR turns malicious errors into random errors by sending an incorrect CRC to the receiver if the digest is invalid. Compared with the PE mode, YASIR delays a message by fewer byte-times and authenticates a message with a fully standard and accepted cryptographic technique [9].

Figure 7 illustrates the latency obtained with YASIR. As mentioned above, the authenticator buffers two bytes of the message to detect its end, and the verifier buffers 14 bytes of the message to verify its digest. The total latency is 16 byte-times, denoted (a) and (c). The verifier starts receiving the message at time (d) and starts forwarding the message to the receiver at time (e). At time (f), the verifier knows if the digest is correct. If an attacker attempts to modify a message, the verifier sends the wrong CRC (crossed out) to the receiver.

4. Predictive YASIR Approach

Previous work [10, 14, 15, 17] focusing on the authentication of individual messages with a digest has tended to delay messages because of encoding inefficiencies (i.e., searching for special symbols in the plaintext and escaping special symbols in the ciphertext). Predictive YASIR attempts to eliminate these inefficiencies by examining broader message patterns.

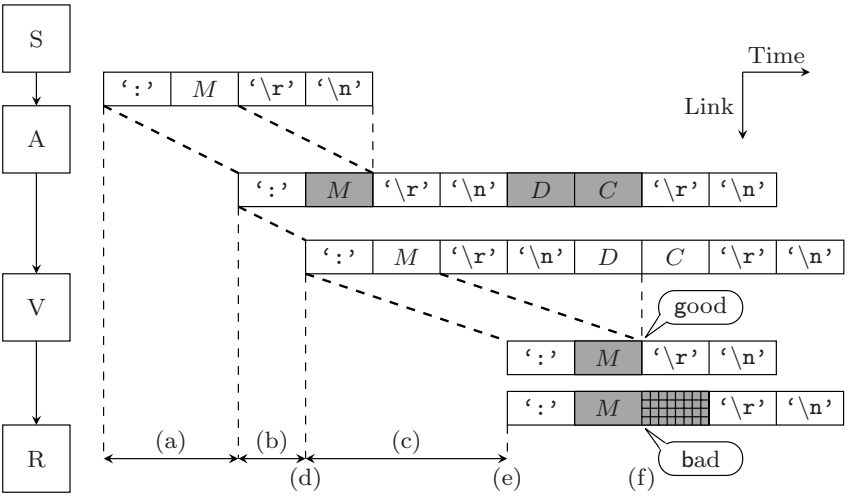


Figure 7. Latency with YASIR.

The approach employs a Bayesian network to predict the incoming plaintext and pre-send the prediction. As each byte reaches the authenticator, the device predicts the rest of the message based on its previous observations. It compresses and encrypts its hypothesis and pre-sends as much ciphertext as possible. Prediction is used to take advantage of the higher bandwidth for ciphertext that is provided by this optimistic *a priori* compression of plaintext. Intuitively, the solution augments YASIR by predicting plaintext messages to eliminate encoding inefficiencies.

Figure 8 illustrates latency with Predictive YASIR. The authenticator does not buffer the message, but must delay it by one byte-time (denoted by (a)). The verifier also does not buffer the message and must delay it by one byte-time (denoted by (c)). In addition, the verifier delays the message by $|D| - 1$ byte-times (denoted by (d)). When prediction works well, the overall delay is $|D| + 1$, which is 13 byte-times. The verifier starts to receive the message at time (e) and starts forwarding it to the receiver almost immediately at time (f). At time (g), the verifier knows if the digest is correct. If an attacker attempts to modify a message, the verifier resets the receiver with `':'` instead of forwarding the entire message.

A BitW device can use compression to avoid overloading a channel that is close to its capacity. This is done by overlapping compressed messages with digests and counters. As shown in Figure 9, messages M_2 and M_3 are compressed and overlapped with digests and counters for messages M_1 and M_2.

Similar to YASIR, Predictive YASIR causes the receiver to drop the message if an attacker modifies the ciphertext. The verifier forwards the message without the last byte to the receiver, which must have the last byte before it acts on the message. When the verifier receives the digest, it calculates its own digest for comparison. If the two digests match, the verifier forwards the last

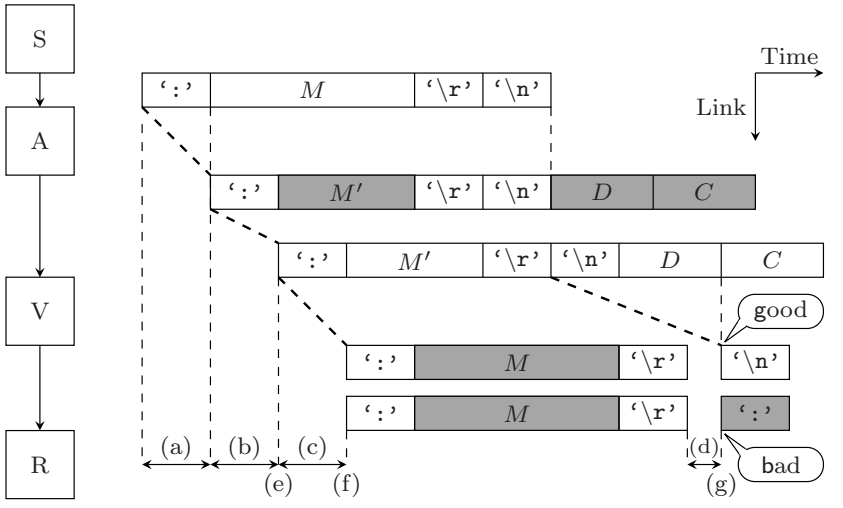


Figure 8. Latency with Predictive YASIR.

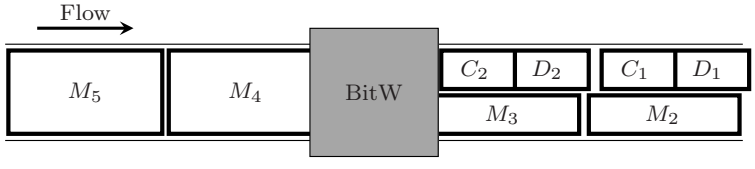


Figure 9. Overlapping compressed messages.

byte of the digest to the receiver, who now has the entire valid message. On the other hand, if the two digests differ, the verifier forwards the reset symbol to the receiver, who has to drop the incomplete message to adhere to the Modbus/ASCII protocol specifications.

Figure 10 illustrates the operation of Predictive YASIR. The sender begins message transmission (Figure 10(a)). The authenticator receives prefix a, predicts that the message is $abcd$, compresses and encrypts the prediction into ciphertext xy and transmits xy (Figure 10(b)). The authenticator receives prefix abe, changes its prediction to $abel$, compresses and encrypts the prediction to xz and transmits the back-away to replace y with z (Figure 10(c)). The authenticator receives the entire message from the sender and transmits the digest τ to the verifier (Figure 10(d)). The authenticator transmits the counter υ to the verifier (Figure 10(e)). The verifier compares the received digest τ to its own calculation. If the two digests match, the verifier forwards the last byte of the message to the receiver; otherwise, the verifier resets the receiver.

Note that if the authenticator changes its hypothesis, the device sends a back-away signal to the verifier to indicate how much of the prediction is incorrect plus the increment for the correct ciphertext (Figure 10(c)). When the authenticator receives the entire message, it sends the digest and counter for

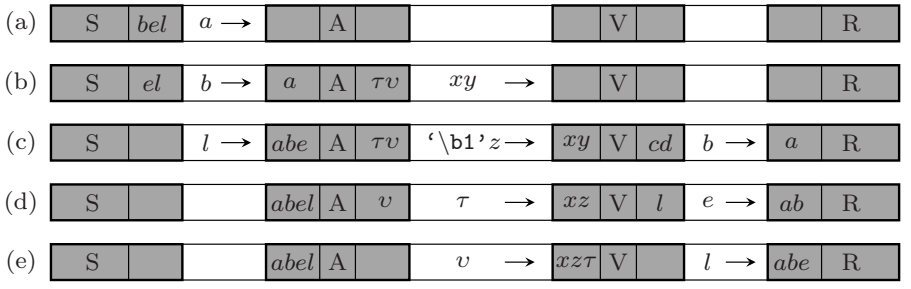


Figure 10. Example of Predictive YASIR operation.

the message (Figures 10(d) and 10(e)). The device then updates the weights in the Bayesian network (described in the next section).

Our solution is a non-intrusive way to "steal" bandwidth for security needs via data coding techniques and to utilize this bandwidth with the help of message prediction. The coding is effective because the data that is sent has sufficiently low entropy and can, therefore, be compressed and predicted to some extent. Note that if a BitW device compresses without predicting, it would have to wait for a larger portion of the message, incurring more latency.

5. Experimental Evaluation

This section describes the experimental methods used to evaluate the performance of Predictive YASIR and presents the experimental results. Interested readers are referred to [12] for additional details.

5.1 Modbus Protocol

Control centers often communicate with substations using Modbus/ASCII [8]. The sender transmits a message starting with the reset symbol `':'` (colon); when a device receives this reset symbol, it drops any incompletely received message (i.e., it resets itself). Every byte of a Modbus/ASCII message is encoded in ASCII. The sender appends a CRC and the terminating symbols `'\r\n'` (carriage return and a newline) at the end of the message.

Figure 11 shows an example Modbus/ASCII message. If the CRC of `0xABCD` is `0xEF`, then the sender encodes the hex message `0xABCD` in a Modbus/ASCII message `':ABCDEF\r\n'` which is `0x3A4142434445460D0A` in hex.

Note that ASCII encoding is inefficient because every byte of the message is two bytes in Modbus/ASCII. This inherent inefficiency allows for greater debugging capabilities in the field; we leverage it to compress messages.

5.2 Scalable Simulation Framework

We use the Scalable Simulation Framework (SSF) [1] to conduct experiments with Predictive YASIR and measure the overhead of the approach. SSF sim-

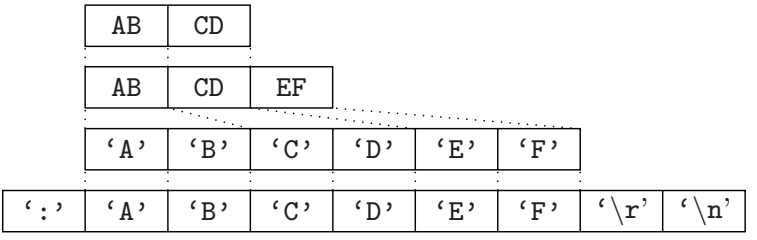


Figure 11. Modbus/ASCII message.

ulates networked entities that exchange events. The framework automates the collection of various statistics related to a simulation. If the simulation is large and runs slowly on a single computer, it can be scaled up with minimal effort by distributing the workload over a set of machines. The device entities exchange single byte events to ensure that they can process one byte at a time. To synchronize the timing, a BitW device outputs at most one byte for each byte that it receives, except after it has received the entire message.

5.3 BitW Devices

A BitW device has two ports: one for plaintext and the other for ciphertext. The device continuously listens for input on both ports. The machinery for processing data on these ports is independent. If a device receives data on the ciphertext port while processing plaintext input, the device deals with the two inputs independently in order.

5.4 Bayesian Network

In order to predict the incoming plaintext, the authenticator models the network traffic using a Bayesian network. A Bayesian network can be represented as a labeled directed acyclic graph. A vertex label is either a message prefix or an entire message and its frequency. All edges are directed from the prefixes to the full-length messages. A prefix vertex may have multiple outgoing edges. For example, the prefix `':'` has an edge to all observed messages because all Modbus/ASCII messages begin with this symbol. Note that a message vertex has inbound edges from all of its prefixes.

We implement the Bayesian network with a hash table of prefixes and a table of tuples (m, f) corresponding to messages and their frequencies.

Figure 12 uses the symbol $\mathcal{H}$ to denote hashing. Each prefix object has a list of the message-frequency tuples. When a plaintext message passes through an authenticator, the frequency of this message increases by one. To predict the rest of the message from its prefix, the authenticator looks up the prefix hash in the Bayesian network. This prefix may have edges to multiple messages, from which the authenticator predicts the most frequent message.

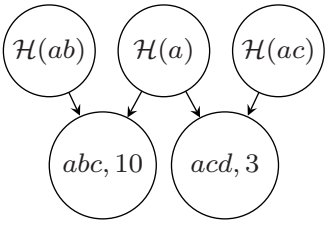


Figure 12. Bayesian network as a bipartite graph.

To prevent incorrect predictions, the authenticator uses Bayes' theorem to calculate the probability $\Pr(H|D)$ of the correctness of a hypothesis for the current data observation. A hypothesis expresses the message prediction while a data observation expresses the prefix. If a hypothesis is less than 50% likely, then the device falls back to its non-predictive mode, which is similar to YASIR. According to Bayes' theorem, $\Pr(H|D) = \Pr(D|H) \cdot \Pr(H) \,/\, \Pr(D)$.

- $\Pr(D|H)$ is the conditional probability of the current data observation given the hypothesis. If the predicted message is correct, then the prefix must occur. Therefore, $\Pr(D|H) = 1$.

- $\Pr(H)$ is the prior probability of a hypothesis. It is the ratio of the number h of occurrences of the hypothesis to the total number t of messages of the same or greater length that passed through the device. Therefore, $\Pr(H) = h/t$.

- $\Pr(D)$ is the prior probability of data occurrence. It is the ratio of the number d of occurrences of the data over the total number o of all the pieces of data of the same length that passed through the device. Therefore, $\Pr(D) = d/o$.

Substituting these terms into the Bayes' equation yields $\Pr(H|D) = (h \cdot o)/(t \cdot d)$.

5.5 Back-Away

As an authenticator pre-sends a predicted message, it monitors the incoming plaintext to verify that the prediction is correct. If the authenticator discovers an error in its prediction, it sends the back-away signal `'\b'` (backspace) to the verifier, followed by the number of bytes to discard from the predicted message, and transmits the corrected part of the message (Figure 10(c)). The discarded bytes are always the last bytes that the device sends because Predictive YASIR uses stream compression and encryption algorithms. For example, if the authenticator has to discard the last byte and replace it with z, then it sends the back-away signal `'\b1'`z. The verifier computes the digest for the final version of the message after it discards all the incorrect predictions.

5.6 Cipher Format

Modbus/ASCII uses only half the available bandwidth and our compression reclaims this space. The authenticator converts each ASCII character (‘0’ to ‘9’ and ‘A’ to ‘F’) into its equivalent four-bit representation: 0x0 to 0xF. The authenticator appends the digest and the counter after the terminating symbol ‘\r\n’. Thus the entire encrypted and authenticated message comprises the ‘:’ symbol, followed by message data, followed by ‘\r\n’, followed by the digest and the counter (Figure 2(d)).

5.7 Simulation Experiments

The simulation contains four components: one FEP, two BitWs and one DA (Figure 1(a)). The FEP connects to the plaintext port of the first BitW. The two BitWs connect to each other via their ciphertext ports. The plaintext port on the second BitW connects to the DA.

The FEP has a set of messages that it sends to the DA in random order. It sends each byte of a message individually, but without delays. Therefore, the authenticator can only act on information from a single byte, which simulates a slow legacy network. The authenticator sends at most one byte of ciphertext for every byte of plaintext it receives, except after it has received the entire message. This simulates enough computational power to query the Bayesian network on every byte of plaintext and enough silence on the wire to avoid congestion due to the ciphertext being longer than the plaintext.

The data used in the experiments was a trace from a GE XA/21 SCADA/Energy Management System that communicated with a GE D400 Substation Data Manager in a laboratory setting. The devices used the DNP3 protocol to communicate and record the trace. Before the simulation, the trace was converted to the Modbus/ASCII format for input into the SSF. The conversion was done because it is difficult to obtain real-world Modbus/ASCII traces.

YASIR and Predictive YASIR were run 30 times each. During the i^{th} run of the simulation, the FEP had $10i$ unique messages to send to the DA. The number of unique messages was varied because the prediction ability can deteriorate with many unique messages. Each run lasted for 200,000 SSF ticks, enough to send every message more than once. The Bayesian network was reset after every run.

The simulation assumes that the BitWs have sufficient computational power so that prediction, compression, encryption and authentication operations do not affect latency. The average byte-time latency was measured in each test and the improvement percentage from YASIR to Predictive YASIR was computed. No comparisons were made with other approaches because they exhibit higher latency than YASIR.

The results of the simulation are presented in Figure 13. The byte-time latency for perfect prediction with a 12-byte HMAC digest is presented as a reference.

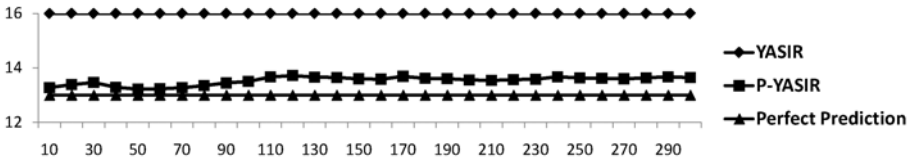


Figure 13. Average end-to-end latency of YASIR and Predictive YASIR.

These results demonstrate that Predictive YASIR has 15.48% less average latency than YASIR with a 95% confidence interval of 0.35 percentage points. Recall that Predictive YASIR is not compared with other bump-in-the-wire devices that provide message authenticity, because they have higher latency than the original YASIR.

Note that prediction does not degrade when the number of unique messages increases. The latency for Predictive YASIR is 13.52 byte-times with a 95% confidence interval of 0.06. In contrast, original YASIR always has a latency of 16 byte-times. When the authenticator makes a prediction mistake, successful recovery occurs with a back-away. The verifier determines all the messages to be valid because errors are not introduced in the ciphertext stream.

6. Future Work

Our future research will focus on several enhancements to Predictive YASIR. Also, it will attempt to validate its performance in real-world settings.

- **Historical Data:** An authenticator can use historical data to predict plaintext, similar to a branch predictor in an instruction pipeline. Historical data can be useful for predicting a natural phenomenon such as temperature. For example, consider a sensor that measures the temperature of water in a river. The temperature is usually 10°C, but assume that it recently increased to 11°C and will remain at this level. An authenticator that only uses statistics will continue to mistakenly predict the temperature as 10°C. On the other hand, a historical system would adjust its predictions even though the long-term majority of the temperature reports is 10°C.

- **SCADA Protocols:** This work has focused on the Modbus/ASCII protocol, but we believe that the technique should scale to other industrial control protocols such as DNP3 [2]. While it is easy to compress a Modbus/ASCII message, all sensors should have a finite and small number of states. For instance, the outdoor water temperature has only 100 integer states in Celsius and varies little.

- **Space:** Predictive YASIR computes statistics about the data stream to predict the next message. Our implementation uses space that is linear in the number of unique messages in the stream. Of course, more efficient stream statistics algorithms (e.g., [5, 6]) may be used.

- **Key Management:** We do not address key distribution, but concentrate on the BitW algorithm. Other researchers have addressed key distribution. For example, the AGA SCM design [15] specifies how devices negotiate the keys and ScadaSafe [16] implements these specifications. Key management implementations are easy to misconfigure or ignore, creating a false sense of security. Therefore, it is important to consider security policies that can be configured easily and correctly.

- **Validation:** Finally, collecting real industrial network data traces from substations and control centers is an important task to verify the correctness of the simulation and test future hypotheses. Unfortunately, vendors hesitate to share data because it may reveal proprietary information or trade secrets.

7. Conclusions

Message prediction and coding can be used to decrease latency arising from encoding inefficiencies, supporting the implementation of message authentication in slow legacy power grid networks. This method, which we call Predictive YASIR, is effective because message data in these networks has low enough entropy to enable BitW devices to predict and compress the data. Simulation results demonstrate a $15.48 \pm 0.35\%$ improvement in byte-time latency without compromising security. These savings can be significant in congested networks that require fast response as well as in other applications that suffer from encoding inefficiencies.

Acknowledgements

This research is based on work supported by the Department of Energy under Award No. DE-OE0000097. We are also grateful to Paul Myrda of Electric Power Research Institute for providing data used in this research.

References

[1] J. Banks, J. Carson, B. Nelson and D. Nicol, *Discrete-Event System Simulation*, Prentice Hall, Upper Saddle River, New Jersey, 2005.

[2] DNP Users Group, Overview of the DNP3 Protocol, Pasadena, California (www.dnp.org/About), 2010.

[3] D. Dolev, C. Dwork and M. Naor, Non-malleable cryptography, *Proceedings of the Twenty-Third ACM Symposium on the Theory of Computing*, pp. 542–552, 1991.

[4] T. Fleury, H. Khurana and V. Welch, Towards a taxonomy of attacks against energy control systems, in *Critical Infrastructure Protection II*, M. Papa and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 71–85, 2008.

[5] S. Ganguly, A. Singh and S. Shankar, Finding frequent items over general update streams, *Proceedings of the Twentieth International Conference on Scientific and Statistical Database Management*, pp. 204–221, 2008.

[6] P. Indyk and D. Woodruff, Optimal approximations of the frequency moments of data streams, *Proceedings of the Thirty-Seventh ACM Symposium on the Theory of Computing*, pp. 202–208, 2009.

[7] A. Menezes, P. van Oorschot and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, Florida, 2001.

[8] Modbus IDA, MODBUS Application Protocol Specification v1.1b, North Grafton, Massachusetts (www.modbus.org/specs.php), 2006.

[9] National Institute of Standards and Technology, Secure Hash Standard, FIPS Publication 180-3, Gaithersburg, Maryland (csrc.nist.gov/publications/fips/fips180-3/fips180-3_final.pdf), 2008.

[10] Schweitzer Engineering Laboratories, SEL-3021-2 Serial Encrypting Transceiver, Pullman, Washington (www.selinc.com/SEL-3021-2), 2007.

[11] Schweitzer Engineering Laboratories, SEL-3021-2 Serial Encrypting Transceiver Data Sheet, Pullman, Washington (www.selinc.com/WorkArea/DownloadAsset.aspx?id=2855), 2007.

[12] R. Solomakhin, Predictive YASIR: High Security with Lower Latency in Legacy SCADA, Technical Report TR2010-665, Department of Computer Science, Dartmouth College, Hanover, New Hampshire, 2010.

[13] S. Stubblebine and V. Gligor, On message integrity in cryptographic protocols, *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pp. 85–104, 1992.

[14] P. Tsang and S. Smith, YASIR: A low-latency, high-integrity security retrofit for legacy SCADA systems, *Proceedings of the Twenty-Third IFIP TC 11 International Information Security Conference*, pp. 445–459, 2008.

[15] A. Wright, AGA 12 Part 2-AKW Proposed SCADA Encryption Protocol (scadasafe.sourceforge.net/Protocol), 2006.

[16] A. Wright, ScadaSafe (scadasafe.sourceforge.net).

[17] A. Wright, J. Kinast and J. McCarty, Low-latency cryptographic protection for SCADA communications, *Proceedings of the Second International Conference on Applied Cryptography and Network Security*, pp. 263–277, 2004.

Chapter 6

DETECTING SENSOR SIGNAL MANIPULATIONS IN NON-LINEAR CHEMICAL PROCESSES

Thomas McEvoy and Stephen Wolthusen

Abstract Modern process control systems are increasingly vulnerable to subversion. Attacks that directly target production processes are difficult to detect because signature-based approaches are not well-suited to the unique requirements of process control systems. Also, anomaly detection mechanisms have difficulty coping with the non-linearity of industrial processes.

This paper focuses on the problem where attackers gain supervisory control of systems and hide their manipulations in signal noise or conceal computational states. To detect these attacks, we identify suitable proxy measurements for the output of a control system. Utilizing control laws, we compare the estimated system output using real-time numerical simulation along with the actual output to detect attacker manipulations. This approach also helps determine the intervention required to return the process to a safe state.

The approach is demonstrated using a heat exchange process as a case study. By employing an explicit control model rather than a learning system or anomaly detection approach, the minimal requirements on proxy sensors and the need for additional sensors can be characterized. This significantly improves resilience while minimizing cost.

Keywords: Process control systems, attack detection, proxy measurements

1. Introduction

Supervisory control and data acquisition (SCADA) systems are vital components in critical infrastructures. Advanced technologies have significantly improved the operation and management of these systems, but they increase the vulnerability to attack [12]. Control systems are often generic computing hosts with complete operating systems and network stacks [7] as opposed to isolated, proprietary systems. This increases the potential for manipulation

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 81–94, 2010.

of the computational states in a SCADA network [10, 21] and process control signals at the sensors and remote terminal units (RTUs) [4].

In previous work [20], we showed that simple statistical anomaly detection can be bypassed by a knowledgeable attacker, underscoring the need for a multi-sensor approach to detection. We, therefore, proposed a novel approach that analyzes process correlations using additional sensors [15]. This paper presents a formal model of the approach, which utilizes process control laws to directly achieve the goals. A beer pasteurizer is considered as a case study because it is a simple, but realistic, example of heat exchange in a production environment.

Pasteurization involves a series of heat exchanges that are controlled to ensure specific target temperatures required for production. The relations between heat exchange inputs and outputs are captured using material and energy balance equations [1]. An attack may be defined as a "concealed" manipulation of these relations [15], which implies that a process degradation cannot be detected by conventional fault analysis. We assume that the attacker has supervisory access to the plant and can alter setpoint values or sensor values, while hiding these manipulations from plant operators [4, 21].

This paper shows how to identify suitable (composite) proxy measurements for making a determination of the current process behavior. The approach relies on the presence of non-linear relations, so that small alterations in proxy values may be correlated with significant process changes. These proxy measurements support the comparison of the actual behavior of the pasteurizer with its estimated behavior, making direct use of material and energy balances and utilizing real-time numerical simulation techniques. This enables the efficient detection of process signal inconsistencies that indicate the presence of manipulated states. In contrast, non-linearity in industrial processes renders conventional anomaly detection approaches (e.g., statistical analysis or learning systems) computationally infeasible for real-time applications.

The proxy measurements also provide a basis for implementing intervention strategies. In a practical implementation, the sensors used for proxy measurements may reside in an out-of-band network with data being pushed to the network via data diodes for detection purposes.

2. Related Work

SCADA systems are increasingly vulnerable to cyber attack due to their modernization and exposure to untrusted networks [3, 12, 16, 23]. This has led to increased interest in intrusion detection [18]. As in the case of conventional computer systems, intrusion detection research has focused on signature-based approaches (generally) at the perimeter and anomaly-based approaches that address the insider threat and direct attacks on control processes [9, 13, 24].

The predictable nature of SCADA traffic can be leveraged to detect system anomalies [5, 21, 24]. However, a knowledgeable attacker can seize the advantage by manipulating computational states or utilizing signal noise to obfuscate attacks that would otherwise be recognized [4, 20, 21].

We argue that this adversary capability highlights a requirement for additional sensors [2, 6, 20] to provide different points of view in order to detect anomalies [21]. This requirement is underscored by the introduction of sophisticated control processes that rely on multivariate controls and, hence, require more complex forms of supervision [19]. Note that this approach would also apply to traditional control systems. The approach has strong parallels with fault detection strategies in SCADA environments [22], but unlike fault detection approaches, the sensors cannot always be assumed to be reliable.

We use functional models to map systems [17] and identify suitable redundant characteristics for evaluating process behavior. We combine these readings with a process simulation to detect signal manipulation [1], extending the invariant model proposed in [15]. This approach obviates the need for linear approximations as used in explicit control models (see, e.g., Lin, *et al.* [11]).

3. Problem Definition

We assume that an attacker is capable of remote penetration attacks on a process control system and understands the industrial process under control. Furthermore, the attacker may gain unauthorized supervisory access to the system and be able to alter setpoints and sensor readings while disguising this from plant operators [4, 21]. For a complex process, which requires multivariate analysis to ensure that production values are achieved, the attacker may not be able to disguise an attack simply by manipulating signals to hide the attack in signal noise [20]. However, as discussed in the next section, it is possible to conceal the attack by falsifying a subset of control signals and relying on the behavior of other parts of the system to normalize the anomalous signals.

Our focus is to identify a conservative number of additional sensors that could provide an inexpensive, practical and computationally-feasible means of uncovering attacks in real time. It is also desirable to be able to use the detection approach as a basis for intervention, although meeting this requirement is beyond the scope of this paper.

4. Detection Model

Heat exchange is a common industrial process. In our case study, we consider heat exchange in the operation of a flash pasteurizer and model the identification of potential proxy measurements. Subsequently, we use a simulation of the pasteurizer to develop a profile of proxy behavior under different operating conditions to determine information about the state of the pasteurizer by observing the proxy behavior.

4.1 Pasteurizer Simulation

Pasteurization is achieved by a series of heat exchanges between hot and cold fluids, where the hot side setpoint temperature is determined by the flow rate and the cold side temperature by the packaging requirements. In flash pasteur-

ization, the interaction of the pasteurized (hot) and unpasteurized (cold) product is used as part of the temperature cycle with target values being achieved through secondary heat exchanges using steam-heated water on the hot side and a glycol refrigerant on the cold side [15]. The basic equations for heat exchange are:

$$\dot{q} = UA(T_{in,s} - T_{out,c}) \tag{1}$$

$$wCp\frac{dT}{dt} = \dot{w}Cp(T_{in,c} - T_{out,c}) + \dot{q} \tag{2}$$

$$uCp\frac{dT_s}{dt} = \dot{u}CP(T_{in,s} - T_{out,s}) - \dot{q} \tag{3}$$

where $\dot{q}$ is the rate of heat exchange; U is the coefficient of heat exchange for the construction material; A is the heat exchange area; $T_{in,s}$ is the initial hot side temperature; $T_{out,s}$ is the final hot side temperature; $T_{in,c}$ is the initial product temperature; and $T_{out,c}$ is the final product temperature. Equations (2) and (3) represent the respective energy balances of the cold and hot sides where $\dot{w}$ and $\dot{u}$ represent the hot side and cold side flow rates, respectively; w and u are the corresponding liquid volumes; and Cp is the specific heat capacity of the product.

We use a diagrammatic form of these equations in Matlab/Simulink (Figure 1(a)), which we subsequently link together in larger blocks to simulate the main plate heat exchanger called the splitter/regeneration unit. This unit uses the hot and cold beer flows as a primary counterflow heat exchange mechanism (Figure 1(b)).

The secondary heating and cooling actions of steam-heated water and glycol refrigerant at the hot and cold sides may be modeled in separate heat exchange sections using a proportional-integral-derivative (PID) block to mimic valve action. The cooling section is presented in Figure 1(c).

The pasteurization unit (PU) is derived from the flow rate and temperature values using the equation:

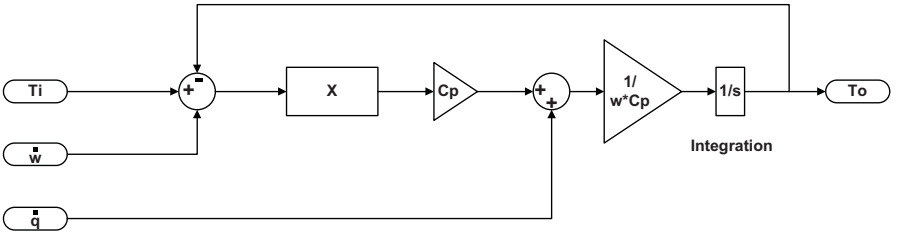
$$PU = \frac{w}{\dot{w}}(60)1.393^{(T-60)} \tag{4}$$

where w is the holding volume; $\dot{w}$ is the flow rate; and T is the temperature. This is an oddly dimensioned measure, which was derived by Dayharsh, *et al.* [8]. Note the non-linear relationship between flow rates, temperature and pasteurization values.

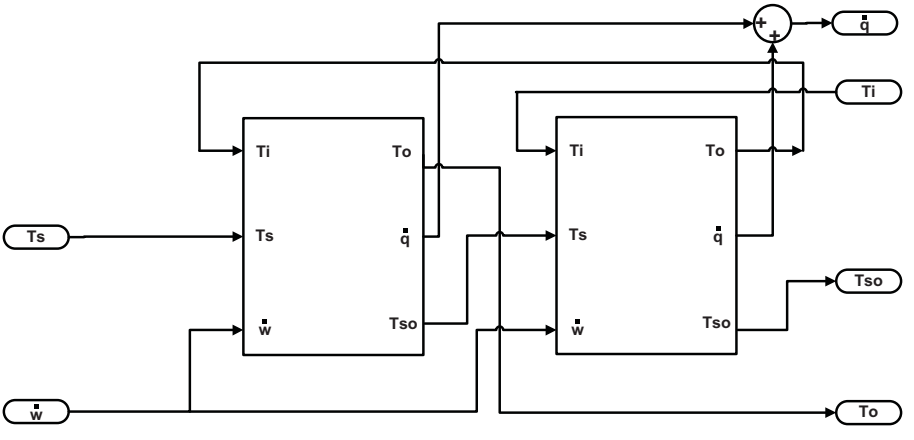
The flow rate $\dot{w}$ is given by:

$$\dot{w} = \dot{w}_{max} - \dot{w}_{min}\frac{L_{actual} - L_{min}}{L_{max} - L_{min}} \tag{5}$$

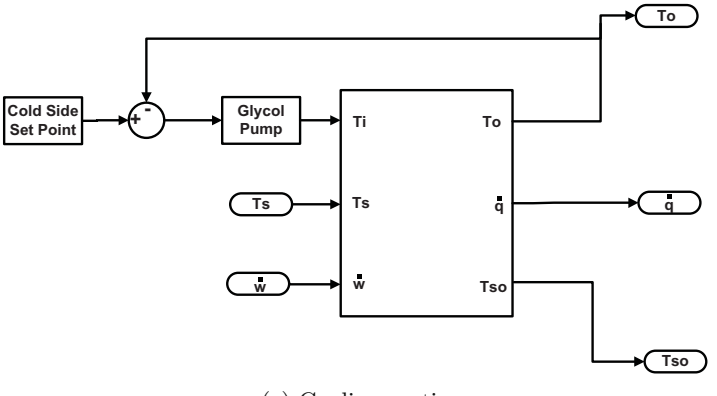
where L is the tank level; and $\dot{w}$ is the flow rate as before. The minimum and maximum tank levels for this specimen are 100 and 220 (cm), respectively; and the minimum and maximum flow rates are 250 and 500 (l/hr), respectively. Flow rates are clamped to their extrema when tank level values exceed their



(a) Block diagram (Equation (1)).



(b) Counterflow heat exchange (splitter/regeneration unit).



(c) Cooling section.

Figure 1. Heat exchange process system.

minimum or maximum values. The tank levels change almost continuously during pasteurization due to alterations in packaging (called "kegging") rates. We add the appropriate calculations to the model to simulate these setpoints.

4.2 Proxy Discovery

We use an adapted functional causal model of the pasteurizer to visualize significant relations and sensor values under specific attack conditions [17] in order to uncover potential proxy nodes. This is a technique that we have found to be suitable for analyzing small-scale configurations. An algebraic transform of the same technique may be used for large-scale configurations.

Let $\overrightarrow{G} = V(\overrightarrow{E})$ be a graph. Let each $v \in V$ represent a characteristic aspect of pasteurizer functionality (e.g., water temperature). Let each $\overrightarrow{e} \in \overrightarrow{E}$ be a causal relationship between distinct pasteurizer characteristics. We assume that conditional questions may be asked about the state of a node $v \in V$ where it is directly or transitively linked to a node $u \in V$, except where v is also linked to a dominant node w. Where a dominant node exists, its state determines the value of all slave (or invariant) nodes that are tail adjacent. All other relations to invariant nodes are represented as dotted lines. We assume, but do not explicitly show in graph form, that the state of each node is subject to unmeasured disturbances that create a probability distribution over node values. These values are perturbed under an attack, but the perturbation may be concealed by the attacker who manipulates computational states or uses process noise. Clearly, the attacker would attempt to conceal the values of all nodes that are directly implicated in determining the success of pasteurization. Supervisory access also allows the manipulation of certain nodes.

Figure 2 shows the pasteurizer under attack. Potentially manipulated nodes have a shaded ring, while probable concealed nodes have an unshaded ring. All the ringed nodes belong to the set of "covered" nodes whose values may not be known during an attack. Note that PU_{SP} is the pasteurization unit setpoint; TL is the tank level; FR is the flow rate; T_{SP} is the hot side temperature setpoint; S is the steam temperature; W is the water temperature; T_{in} is the initial product temperature; HX_S is the product temperature after heating in the splitter; HX_W is the product temperature after being heated by the water; HX_H is the product temperature at the end of the holding pipe and pre-regeneration; PU_{est} is the estimated PU value; HX_R is the product temperature post regeneration; T_{out} is the product temperature at the end of the process after glycol cooling; C_{SP} is the cold side setpoint; and G is the glycol temperature.

According to Equation (4), the temperature and flow rate determine the pasteurization value; and these values are central to the material and energy balance equations in the heat exchange process (Equations (1-3)). These observations suggest that it may be possible to estimate the pasteurization values from the heat exchange performance and vice versa. Finally, note that the heat

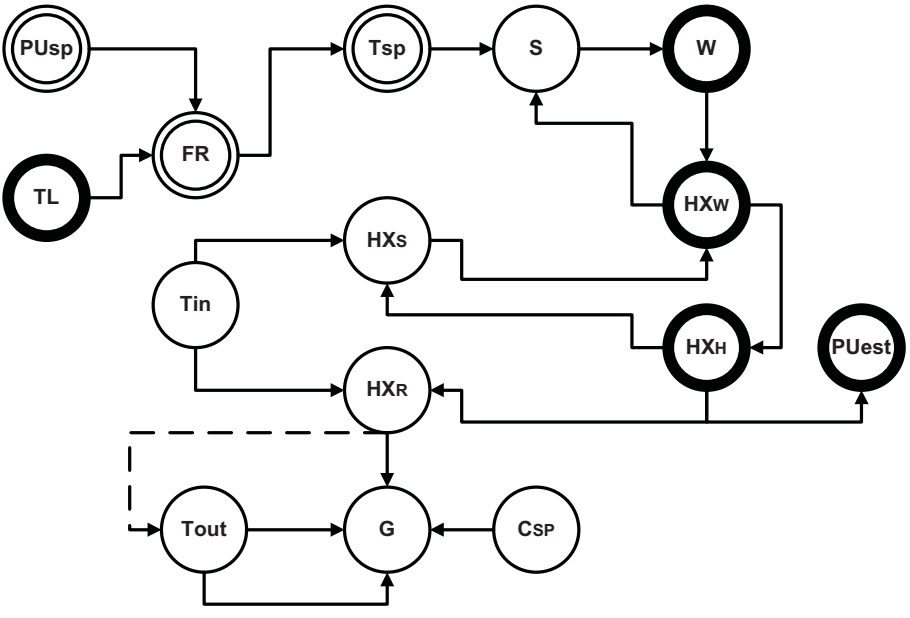


Figure 2. Adapted functional causal model.

exchange output values are uncovered at HX_S and HX_R, potentially enabling their use as proxy measurements for determining the success of pasteurization.

5. Analysis of Heat Exchange Profiles

Based on the assumptions about attacker capabilities and assuming no insider collusion, three attack strategies are possible:

- Lower the *PU* by spoofing a lower tank level, thus increasing the flow rates relative to temperature using negative error values.
- Lower the *PU* by lowering the water temperature and, hence, product temperature relative to the flow rates using positive error values, or equivalently by resetting the *PU* setpoint.
- Combine the above two strategies in a single attack.

In the first two attack strategies, the adversary has to cover all the relevant sensors so that they appear to show consistent values. In the third strategy, the attacker may omit to cover the hot side temperature as a stepped approach to jointly lower the water temperatures and raise the relative flow rates. This can be concealed in process noise, where the product temperature remains constant, but the pasteurization process is still degraded. Therefore, it is necessary to show that the proxy measurements identified in Section 4.2 can be used to estimate the flow rate and the temperature and, hence, the pasteurization unit value.

5.1 Pasteurization Profiles

As a preliminary, we show that there are distinct temperature to flow rate ratios for each PU. Upon solving Equation (4) for temperature [15], we plot the temperature values against flow rates for PU values of 40 (nominal value) and 20 (fail or "divert" value) and determine the PU value for a 1°C alteration downwards for the nominal value of $PU = 40$, which represents a significant loss of quality (i.e., we also solve for $T - 1$ with $PU = 28.7$). The results are shown in Figure 3(a).

5.2 Establishing Flow Rate

The temperature of the cooled (pasteurized) product leaving the regenerator tank can be used to estimate the flow rates. To show this, we plot this temperature for four distinct tank levels $T = 120$, 140, 160 and 180 (representing different flow rates) against the nominal pasteurization rate of 40, the divert value of 20 and the quality loss value of 28.7. The distinct banding of temperatures for the cooled product (color coded by flow rate in Figure 3(b)) shows that the flow rate and, hence, the tank level can be estimated. Confidence intervals are estimated at ±10 l/hr. It follows that it should be possible to detect the first attack strategy that alters the flow rate by spoofing false tank level readings.

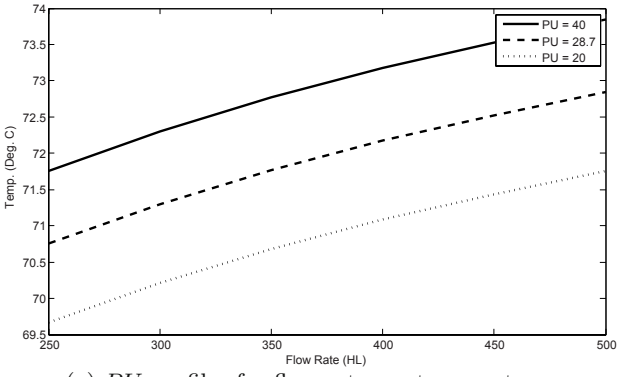
5.3 Establishing a Temperature Profile

Next, we show how a concealed alteration in temperature may be detected. Setting the tank level at TL =180 to lock in the flow rate, we present the results of three runs with distinct pasteurization profiles as before (Figure 3(c)). Given the non-linear relationship between pasteurization levels, these temperature differences are on the average sufficiently significant so that, in combination with the flow rate, it is possible to estimate the process success with a confidence level greater than 3σ. Similar results are obtained for the other flow rates. Hence, it is possible to determine the current PU level modulo approximately four units.

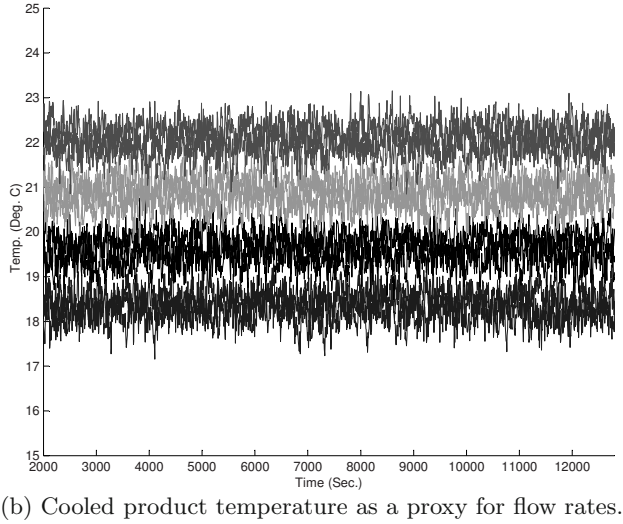
5.4 Uncovering Concealed Attacks

Finally, we show how transitions concealed by signal noise can be detected. We assume that an attack causes the PU value to drop gradually by lowering the water temperatures and raising the flow rates in an effort to hide the manipulations in the process signal. We set the tank level to 220 and dropped it in stages to 172, altering the flow rate upwards. We dropped the pasteurization target from 40 to 28.7 (equivalent to quality loss) by altering the water temperature error signal or PU setpoint.

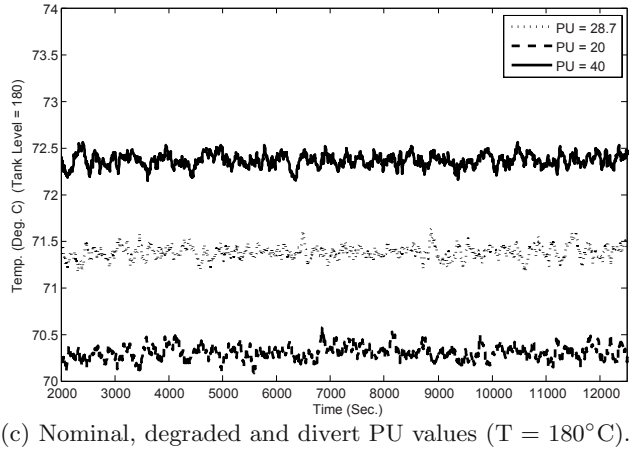
Figures 4(a) and 4(b) show that the product temperatures do not vary from their mean values, while the PU rates drop significantly. Obviously, the attacker could continue this process until the divert or abort values are achieved.



(a) *PU* profiles for flow rate vs. temperature.

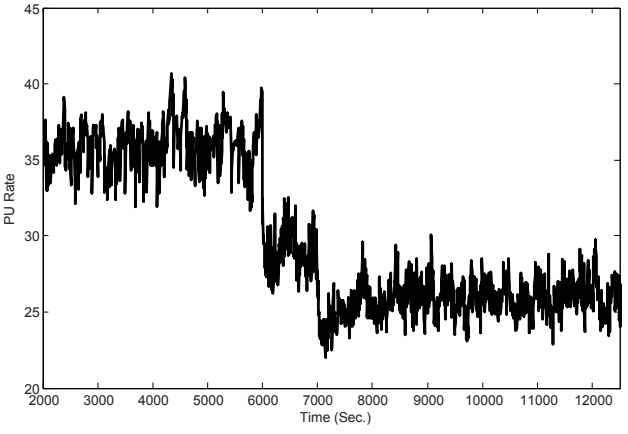


(b) Cooled product temperature as a proxy for flow rates.

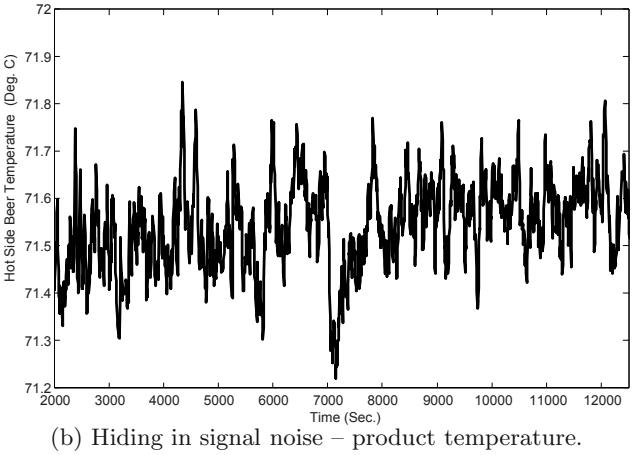


(c) Nominal, degraded and divert PU values (T = 180°C).

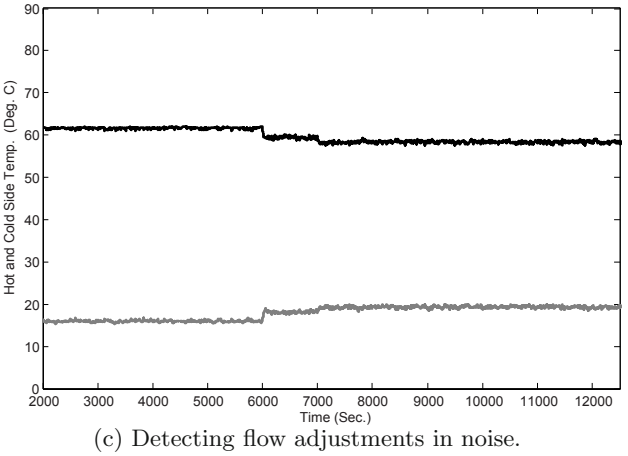
Figure 3. Heat exchange profiles.



(a) Hiding in signal noise – pasteurization rate alteration.



(b) Hiding in signal noise – product temperature.



(c) Detecting flow adjustments in noise.

Figure 4. Disguising and detecting alterations in signal noise.

Figure 4(c) shows the difference in the heat exchange profile as a result of the concealed adjustments of the hot side and cold side temperatures. Note that the attack produces a greater contrast in the heat exchange profile compared with the other attacks because both the lower and upper product temperatures are altered simultaneously to create anomalous steps in the heat exchange profile. It follows that it is possible to estimate (and even control) the actual pasteurization rates using this profile. We estimate that confidence levels of $\pm 4PU$ can be achieved assuming a precision of ± 0.5°C in the hot side temperature values and ± 10 l/hr in the flow rate estimations.

6. Discussion

In Section 5, we discussed three possible attacks on the pasteurization process. The first attack lowers the water temperature setpoint so that the product is not heated to the requisite temperature to achieve the target pasteurization value. The second attack raises the flow rate by spoofing lower tank levels. The third attack combines the previous two attacks while keeping the beer temperature invariant to hide the attack manifestations in process noise while degrading the pasteurizer value.

Using well-chosen proxy measurements, it is possible (in combination with numerical simulation) to capture inconsistent and anomalous behavior in a manner that minimizes the computational cost of detection. This approach also supports attack intervention strategies. While this is not strictly relevant for a batch process such as pasteurization (which can restart if there is a fault), there are other processes involving heat exchange where the ability to adjust to attacker actions "on the fly" is necessary because the processes are not easily restarted. Intrusion prevention thus becomes a dynamic process of defending process integrity.

Clearly, the ideal situation is to have a fully redundant set of sensors, but physical and cost constraints along with certification and accreditation requirements make this approach infeasible. Our approach, therefore, seeks to minimize the effort involved in implementing a signal-based anomaly detection mechanism, which is important when dealing with large-scale industrial processes. Although this paper focuses on heat exchange in the context of pasteurization, the approach is applicable to a variety of non-linear engineering processes.

In many cases, there exists the potential to identify output values that are redundant for process control, but that can be used as proxy measurements in combination with real-time control system simulation for intrusion detection. We have previously shown that anomaly detection techniques based on univariate statistical techniques may be unable to distinguish signal noise from an attack, but in the case of a non-linear process, even a difference of σ in the average performance can radically alter the result [20]. Classical intrusion techniques, in general, do not consider process signals. Even if these approaches were to be applied, their reliance on signature-based detection has no validity in the application. Moreover, anomaly detection techniques that rely on

complex statistical analyses (e.g., Markov models) have limited applicability to non-linear systems because they do not accommodate the sharp disparities in process state that result from small alterations in such systems. Learning systems face similar problems because considerable data is required to create a training set that accommodates non-linearity. In case of the pasteurizer, a learning system would have to learn the pasteurization profile of each product that is processed by a pasteurizer, track the performance degradation and estimate state changes. In contrast, a control model of the system encapsulates these aspects in a straightforward manner and provides a computationally inexpensive numerical simulation of process behavior.

Nevertheless, our approach has certain limitations. Some limitations may be introduced by physical constraints such as sensor placement [1] that can reduce the confidence in the results. Distinct processes are associated with different perturbation levels; this can reduce (or increase) confidence levels. However, in most cases, even the process models used to set up the control systems are limited in precision and tuned based on experience rather than physical or chemical models. Thus, the approach is ultimately limited by the precision of these models.

7. Conclusions

Attacks on industrial control systems that involve signal manipulations are often invisible to traditional intrusion detection systems. A promising solution is to use proxy measurements to determine anomalous readings in key process characteristics in a computationally efficient manner while minimizing the need for additional sensors, thereby reducing the accompanying costs. This approach permits the continued safe operation of a process when shutdown is not feasible.

The primary limitations of the approach are process specific and plant specific in nature. Different processes are associated with distinct perturbations. In addition, variations in plant design may not permit the satisfactory placement of supplementary sensors. These factors result in lower confidence levels.

Our future work will attempt to characterize processes that are amenable to this approach. We will also develop a more rigorous adversary capability model. Finally, we hope to combine this approach with other anomaly detection mechanisms (e.g., [14]) to eliminate some of the assumptions imposed on sensor and actuator integrity.

Acknowledgements

This research was partially supported by Vistorm, an HP Company. The authors also wish to acknowledge the assistance of Diageo and, in particular, Mr. Brian Furey for helping validate the model and providing data sets for analysis.

References

[1] B. Bequette, *Process Control: Modeling, Design and Simulation*, Prentice-Hall, Upper Saddle River, New Jersey, 2002.

[2] J. Bigham, D. Gamez and N. Lu, Safeguarding SCADA systems with anomaly detection, *Proceedings of the Second International Workshop on Mathematical Methods, Models and Architectures for Computer Network Security*, pp. 171–182, 2003.

[3] E. Byres and D. Hoffman, The Myths and Facts behind Cyber Security Risks for Industrial Control Systems, Technical Report, Department of Computer Science, University of Victoria, Victoria, Canada, 2004.

[4] A. Cardenas, T. Roosta and S. Sastry, Rethinking security properties, threat models and the design space in sensor networks: A case study in SCADA systems, *Ad Hoc Networks*, vol. 7(8), pp. 1434–1447, 2009.

[5] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner and A. Valdes, Using model-based intrusion detection for SCADA networks, *Proceedings of the SCADA Security Scientific Symposium*, pp. 127–134, 2007.

[6] M. Coutinho, G. Lambert-Torres, L. da Silva, J. da Silva, J. Neto, E. da Costa Bortoni and H. Lazarek, Attack and fault identification in electric power control systems: An approach to improve the security, *Proceedings of the PowerTech Conference*, pp. 103–107, 2007.

[7] A. Creery and E. Byres, Industrial cybersecurity for power systems and SCADA networks, *Proceedings of the Fifty-Second Annual Petroleum and Chemical Industry Conference*, pp. 303–309, 2005.

[8] C. Dayharsh and H. Del Vecchio, Thermal death time studies on beer spoilage organisms, *Proceedings of the American Society of Brewing*, vol. II, pp. 48–52, 1952.

[9] D. Gamez, S. Nadjm-Tehrani, J. Bigham, C. Balducelli, K. Burbeck and T. Chyssler, Safeguarding critical infrastructures, in *Dependable Computing Systems: Paradigms, Performance Issues and Applications*, H. Diab and A. Zomaya (Eds.), Wiley-Interscience, Hoboken, New Jersey, 2005.

[10] G. Hoglund and J. Butler, *Rootkits: Subverting the Windows Kernel*, Addison-Wesley, Reading, Massachusetts, 2005.

[11] Y. Huang, A. Cardenas, S. Amin, Z. Lin, H. Tsai and S. Sastry, Understanding the physical and economic consequences of attacks on control systems, *International Journal of Critical Infrastructure Protection*, vol. 2(3), pp. 73–83, 2009.

[12] R. Krutz, *Securing SCADA Systems*, Wiley, Indianapolis, Indiana, 2006.

[13] O. Linda, T. Vollmer and M. Manic, Neural network based intrusion detection system for critical infrastructures, *Proceedings of the International Joint Conference on Neural Networks*, pp. 1827–1834, 2009.

[14] T. McEvoy and S. Wolthusen, Using observations of invariant behavior to detect malicious agency in distributed environments, *Proceedings of the Fourth International Conference on IT Incident Management and IT Forensics*, pp. 55–72, 2008.

[15] T. McEvoy and S. Wolthusen, Using observations of invariant behavior to detect malicious agency in distributed control systems, presented at the *Fourth International Workshop on Critical Information Infrastructures Security*, 2009.

[16] P. Motta Pires and L. Oliveira, Security aspects of SCADA and corporate network interconnections: An overview, *Proceedings of the International Conference on Dependability of Computer Systems*, pp. 127–134, 2006.

[17] J. Pearl, *Causality: Models, Reasoning and Inference*, Cambridge University Press, Cambridge, United Kingdom, 2009.

[18] J. Rrushi and K. Kang, Detecting anomalies in process control networks, in *Critical Infrastructure Protection III*, C. Palmer and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 151–165, 2009.

[19] J. Schlesser, D. Armstrong, A. Cinar, P. Ramanauskas and A. Negiz, Automated control and monitoring of thermal processing using high temperature, short time pasteurization, *Journal of Dairy Science*, vol. 80(10), pp. 2291–2296, 1997.

[20] N. Svendsen and S. Wolthusen, Modeling and detecting anomalies in SCADA systems, in *Critical Infrastructure Protection II*, M. Papa and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 101–113, 2008.

[21] J. Verba and M. Milvich, Idaho National Laboratory supervisory control and data acquisition intrusion detection system, *Proceedings of the IEEE Conference on Technologies for Homeland Security*, pp. 469–473, 2008.

[22] X. Wang, J. Lizier, O. Obst, M. Prokopenko and P. Wang, Spatiotemporal anomaly detection in gas monitoring sensor networks, *Proceedings of the Fifth European Conference on Wireless Sensor Networks*, pp. 90–105, 2008.

[23] D. Watts, Security and vulnerability in electric power systems, *Proceedings of the Thirty-Fifth North American Power Symposium*, pp. 559–566, 2003.

[24] D. Yang, A. Usynin and J. Hines, Anomaly-based intrusion detection for SCADA systems, *Proceedings of the Fifth International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*, 2006.

Chapter 7

DISTRIBUTED INTRUSION DETECTION SYSTEM FOR SCADA PROTOCOLS

Igor Nai Fovino, Marcelo Masera, Michele Guglielmi, Andrea Carcano and Alberto Trombetta

Abstract This paper presents an innovative, distributed, multilayer approach for detecting known and unknown attacks on industrial control systems. The approach employs process event correlation, critical state detection and critical state aggregation. The paper also describes a prototype implementation and provides experimental results that validate the intrusion detection approach.

Keywords: Industrial control systems, SCADA protocols, intrusion detection

1. Introduction

Critical infrastructures rely very heavily on information and communications technologies (ICT). These technologies provide features and services such as remote monitoring, remote management, intra-system coordination, inter-system communication and self-orchestration. Unfortunately, critical infrastructure assets are susceptible to a large number of ICT attacks [5, 10]. These attacks can be categorized into two classes. The first class includes traditional ICT attacks that leverage vulnerabilities in general purpose ICT systems; these attacks can be mitigated by adopting ICT countermeasures such as software patches, antivirus software and firewalls. The second class includes industrial system attacks that exploit vulnerabilities specific to industrial ICT systems, e.g., attacks that leverage the lack of authentication and integrity checks in SCADA communication protocols [1].

Due to the peculiarities of industrial systems, ICT countermeasures cannot be deployed efficiently in all environments. Indeed, the countermeasures are inadequate for dealing with attacks of the second class that exploit SCADA protocols. Also, even when countermeasures (e.g., signature-based intrusion detection) are successfully deployed, they may not protect against unknown

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 95–110, 2010.

attacks. Such exposure is unacceptable as far as critical infrastructure assets are concerned.

This paper presents a novel approach for detecting attacks on industrial control systems based on the concepts of event analysis, event aggregation and correlation, and critical state detection. A prototype distributed intrusion detection system for monitoring SCADA systems is described. It uses event correlation to identify race conditions in critical states induced by malicious actions.

2. Background

Intrusion detection is a well-established field of research. The basic idea, presented in the mid-1980s (see, e.g., [3]), is to search for evidence indicating that a malicious attack is underway during a certain period of time. Intrusion detection systems (IDSs) can be classified according to (i) the source of the information used to detect intrusions; and (ii) the technique used to discriminate between licit behavior and malicious behavior. Discriminating IDSs based on their information source leads to their classification as network-based IDSs that analyze network traffic in search of malicious packets; and host-based IDSs that analyze host behavior for suspicious activities. Categorizing IDSs based on their discrimination technique gives rise to signature-based IDSs that detect attacks based on known attack patterns; and anomaly-based IDSs that detect attacks based on deviations from normal system behavior.

This paper focuses on network-based intrusion detection system (NIDS) architectures. Typical NIDS architectures incorporate a number of distributed sensors that analyze network traffic in search of attack signatures and anomalies. In the case of a SCADA system, a NIDS must be able to understand and analyze an industrial communication protocol such as Modbus, DNP3 or Profibus. These protocols, which were originally designed for serial communication, are currently embedded in TCP packets and ported over TCP/IP. Traditional NIDSs are unable to understand such "application level" protocols.

Digital Bond [4] has released a set of *ad hoc* rules for detecting certain attacks on the Modbus protocol. These rules specify unauthorized uses of the Modbus protocol, protocol errors and network scans. A traditional NIDS that incorporates these rules could identify primitive, single-packet-based attacks, in which the attacker sends a malicious packet to a Modbus device or uses a rare command. However, as shown in [10], SCADA attacks can be extremely complex and rarely involve a single step (i.e., the exploitation of a single vulnerability). Consequently, it is necessary to identify complex and dangerous attacks based on the analysis of different, low-risk atomic operations.

Several researchers have used such "attack correlation" techniques for intrusion detection in traditional ICT networks. Gross, *et al.* [6] have proposed a "selecticast" mechanism for collaborative intrusion detection that uses a centralized server to dispatch information about suspicious activities to intrusion detection sensors. Yegneswaran, *et al.* [16] employ a distributed overlay technique to monitor attacks. These approaches provide a broad picture regarding

suspicious events, but they are unable to identify complex malicious actions that are strategic as opposed to tactical.

Ning, *et al.* [15] have developed a model for identifying causal relationships between alerts on the basis of prerequisites and consequences. Likewise, Cuppens and Miege [2] engage pre-conditions and post-conditions in multiple analysis phases such as alert clustering, alert merging and intention recognition. This approach facilitates the automatic generation of correlation rules, but it can produce a large number of spurious rules that significantly increase noise in intrusion detection.

In summary, the correlation techniques described in the literature attempt to identify "malicious actions." However, in an industrial control system, routine actions can be used to implement devastating attacks. Consequently, an effective IDS for process control networks should be able to correlate licit actions and events in its search for malicious attacks.

3. Event Correlation and Anomaly Detection

As mentioned above, traditional IDSs often fail to detect complex attacks on process control networks. Most IDSs are unable to analyze SCADA communication protocols and detect attacks that exploit protocol vulnerabilities. The few IDSs that are able to analyze SCADA protocols (e.g., Snort) employ single-packet signatures; they cannot detect complex attacks where sequences of licit packets put the system in a critical state. Classical anomaly detection techniques also have limited effectiveness in industrial control environments. One of the major challenges is to specify and detect anomalies in process control networks that may have hundreds of PLCs.

Thus, we argue that a different intrusion detection approach must be devised for industrial control systems. Our approach is described as follows:

- A SCADA system is at the core of every process network in an industrial process system. A SCADA system controls every process in the industrial system. Therefore, the key to detecting intrusions is to monitor the activity of SCADA systems.

- Most industrial process systems are analyzed carefully and the possible "critical states" (or dangerous states) are usually identified,

- The data flowing between master and slave devices in a SCADA system can be used to construct a virtual image of the monitored system. The virtual states can be compared with the critical states that must be avoided. Upon tracing the evolution of the virtual states, it is possible to predict if the system is moving to a critical state.

- The industrial system is modeled in a modular fashion. A set of critical states for each subsystem comprising the industrial system is identified. The dependencies between subsystems are described so that the state of the system can be monitored. This enables the detection of many types

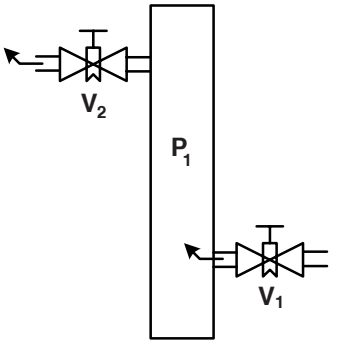


Figure 1. Example of a critical state.

of attacks. The effectiveness of this approach depends on the granularity of the virtual state representation and on the effects that attacks have on the evolution of the virtual states.

In general, anomaly detection defines the normal behavior of a system using a mathematical model and flags any deviation beyond a given threshold as a potential attack. We use a complementary method that identifies the critical states that are to be avoided (i.e., anomalous configurations that might put the system at risk) and flag them as the possible results of an attack. This approach is rarely employed in large, open ICT systems because it is almost impossible to describe all the possible combinations of behaviors that can drive a system into a critical state. However, industrial systems operate in tightly-controlled environments comprising electromechanical devices that react to a limited set of well-defined commands in relation to physical conditions that are known *a priori*. In these systems, the critical states are limited in number and are well-known; also, the task of describing them is more manageable than in traditional ICT systems.

The problem posed by false positives is addressed if attention is focused on identifying the sequences of events that could drive a system into a critical state. Moreover, focusing on system state evolution and on critical states (and not on specific attacks), makes it possible to detect new, unknown attacks that traditional signature-based IDSs would not be able to detect.

Figure 1 clarifies the notion of a critical state. The example system is a pipe P_1 through which high-pressure steam flows. The pressure is regulated by two valves V_1 and V_2.

An attacker with the ability to inject command packets in the process control network could direct the programmable logic controller (PLC) to close valve V_2 and open valve V_1. These two operations are perfectly licit when executed separately. However, the two operations performed in sequence put the system in a critical state because the pressure in P_1 could rise to a high enough level to cause the pipe to burst.

Table 1. Internal structure of PLCs.

Name	Object	Type	Comments
Discrete Inputs	1 bit	R	Provided by I/O system
Coils	1 bit	R/W	Alterable by application
Input Registers	16-bit word	R	Provided by I/O system
Holding Registers	16-bit word	R/W	Alterable by application

This simple example describes a scenario involving a two-command sequence. However, in general, the command sequences are long and complex, making it very difficult to specify their signatures for traditional IDSs. For this reason, we concentrate on the results of command sequences (i.e., the resulting states).

We represent critical states using our Industrial Critical State Modeling Language (ICSML). The current version of the language supports SCADA systems that use the Modbus protocol, but it is easily extended to accommodate other protocols.

The system under consideration is modeled in a modular manner as a collection of subsystems. Each subsystem is, in turn, modeled as a set of subsubsystems, and so on. Thus, the overall system can be decomposed to the desired level of granularity.

At the most basic level, a system is composed of a set of PLCs, whose internal structure is essentially a sequence of registers and their corresponding values (Table 1).

ICSML is specified as follows:

```
<Critical State> ::= <term> | <Critical State><op><Critical State>
                     | NOT<Critical State> | (<Critical State>)
<op>             ::= AND | OR
```

As mentioned above, a system is decomposed in terms of subsystems, which are, in turn, decomposed in terms of sub-subsystems, and so on.

```
<system>         ::= <sysName>
                     | <sysName>[<component>.<componentList>]
                     | <system>.<system>
<componentList>  ::= <component>.<componentList> | e
<sysName>        ::= valid system name over the ASCII character set
```

Note that PLCs constitute the basic building blocks of a system. A component in ICSML represents a specific PLC (and its status) in a given subsystem. Each PLC in a Modbus network is identified by an IP address, port and identification number.

```
<component>      ::= PLC[<address>:<port>:<id>].<comp_status>
<address>        ::= <byte>.<byte>.<byte>.<byte>
<byte>           ::= 0 | 1 | ... | 255
<port>           ::= 0 | 1 | ... | 65535
```

```
<id>              ::= <byte>
<comp_status>     ::= CO[<reg_index>]<rel><bit>
                    | DI[<reg_index>]<rel><bit>
                    | IR[<reg_index>]<rel><word>
                    | HR[<reg_index>]<rel><word>
<reg_index>       ::= 0 | 1 | ... | 65535
<bit>             ::= 0 | 1
<word>            ::= 0 | 1 | ... | 65535
<rel>             ::= <= | >= | < | > | =
```

Using ICSML, it is possible to formally describe the critical states of a system and subsystems. For example, suppose that the following facts hold in the example above: (i) the output stream of valve V_1 is connected to PLC 192.168.0.1 port 502 id 1 and the holding register 10 contains 100 if V_1 is open and 0 if V_1 is closed; (ii) the input stream of valve V_2 is connected to PLC 192.168.0.2 port 502 id 1 and the holding register 20 contains 100 if V_2 is open and 0 if V_2 is closed; and (iii) the system is in a critical state if valve V_1 is open less than 50% and if valve V_2 is open more than 50%. Then, the critical state can be formalized in the following manner using ICSML:

```
PLC[192.168.0.1 : 502 : 1].HR[10] < 50 AND
PLC[192.168.0.2 : 502 : 1].HR[20] > 50
```

The relationships between subsystems are modeled using transition rules that specify what happens (in terms of changes in the values of components) in one or more subsystems given that something has changed in some other subsystem. The syntax of a transition rule is very simple. Given two components C_1 and C_2, a transition rule is an expression of the form $C_1 \rightarrow C_2$. The transition rule states that if the status of the first component is described by the expression C_1, then the status of the second component changes to that described by expression C_2.

ICSML permits the description of a set of critical states that represent – at the desired level of detail – unwanted occurrences of the subsystems contained in the industrial system under scrutiny. Also, ICSML permits the modeling of data flows between SCADA masters and slaves to create a virtual representation of the state of the entire system. The set of production rules should faithfully represent the industrial system and its subsystems along with the relationships between the various subsystems.

Of course, production rules alone are inadequate to ensure that changes in the virtual system states accurately model those in the industrial system over time. For this reason, it is necessary to periodically poll the SCADA network components to map changes related to external events (e.g., variations in sensor readings) to virtual states. Also, as we discuss below, it is important to recognize as early as possible that a state is evolving into a critical state.

3.1 High-Level Event Correlation

It would be useful to correlate all the licit low-level events to discover critical patterns that are potentially caused by malicious attacks. However, it is practically impossible to enumerate all the critical states of a complex system with hundreds of devices, let alone correlate all the low-level events.

Masera and Nai Fovino [9] have presented a methodology for modeling system features that are relevant to detection and correlation. The methodology expresses a complex system and its subsystems as a "system-of-systems" – a set of interconnected collaborative systems. A system is a set of independently-managed subsystems that provide services in a producer/consumer environment with a locality property. Thus, a system (which is also a subsystem) is recursively defined in terms of subsystems. Each subsystem has four attributes:

- **Name:** Uniquely identifies the subsystem.
- **Description:** Describes the purpose of the subsystem.
- **Service List:** Specifies the services provided by the subsystem.
- **Data Flow List:** Specifies data flows as tuples of the form $< S_1, S_2 >$, which represents a flow of information from subsystem S_1 to subsystem S_2.

The notion of a service provides the glue for describing a system-of-systems in terms of subsystems. Every action that a subsystem performs can be viewed as a service, and every subsystem communicates with other subsystems by providing services. Damage to a subsystem usually manifests itself as a corresponding lack in some service. Indeed, the modeling of a service has a central role in our critical state event correlation approach. Each service has four attributes:

- **ID:** Uniquely identifies the service.
- **Service Dependency List:** Specifies the service dependencies as tuples of the form $< Service, DamageThreshold >$ where $Service$ is the ID of a service and $DamageThreshold$ is the maximum level of damage to the service that can be tolerated.
- **Dependencies:** Describe the relations between the service and the services in the service dependency list using first-order logical expressions.
- **Service Value:** Specifies the value of the service (essential, valuable or expendable).

Thus, we represent a complex system in terms of subsystems, dependencies and services. Each subsystem is a black box that produces and consumes services (from which the data flows are derived). The subsystems, services and dependencies yield a system graph that represents the intrinsic interdependencies between the elements of the industrial system of interest.

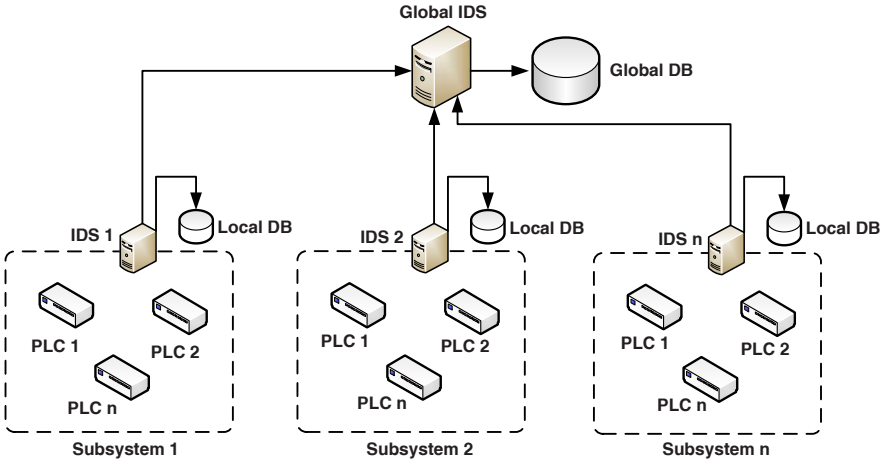


Figure 2. Global intrusion detection system architectures.

Using ICSML, it is possible to define high-level critical states in which each critical state represents a particular scenario where certain services provided by a collection of subsystems are partially or completely damaged, potentially moving the entire system into an unsafe state. The description of the critical states at this level is compact because it only involves the description of the failure of the higher-level services. In other words, a high-level critical state is reached when certain high-level services fail partially or totally, causing the system to move to a dangerous state.

The high-level and low-level event correlations can be merged using the following three-step procedure:

- When a low-level critical state is identified, information about the impacted subsystem services is delivered as an event to the high-level event correlation engine (HLEC).
- The HLEC propagates subsystem service failures across the whole system based on an exploration of the system graph.
- After the HLEC completes the propagation of service failures, it analyzes the entire system by comparing its general status with the set of high-level critical states in search of a match.

4. Architectural Overview

The distributed IDS is presented in Figure 2. The SCADA system under consideration is decomposed into component subsystems, each of which is monitored by a local IDS (figure 3). Each local IDS implements low-level correlation and detection, and raises specific alerts. The entire system is monitored by a

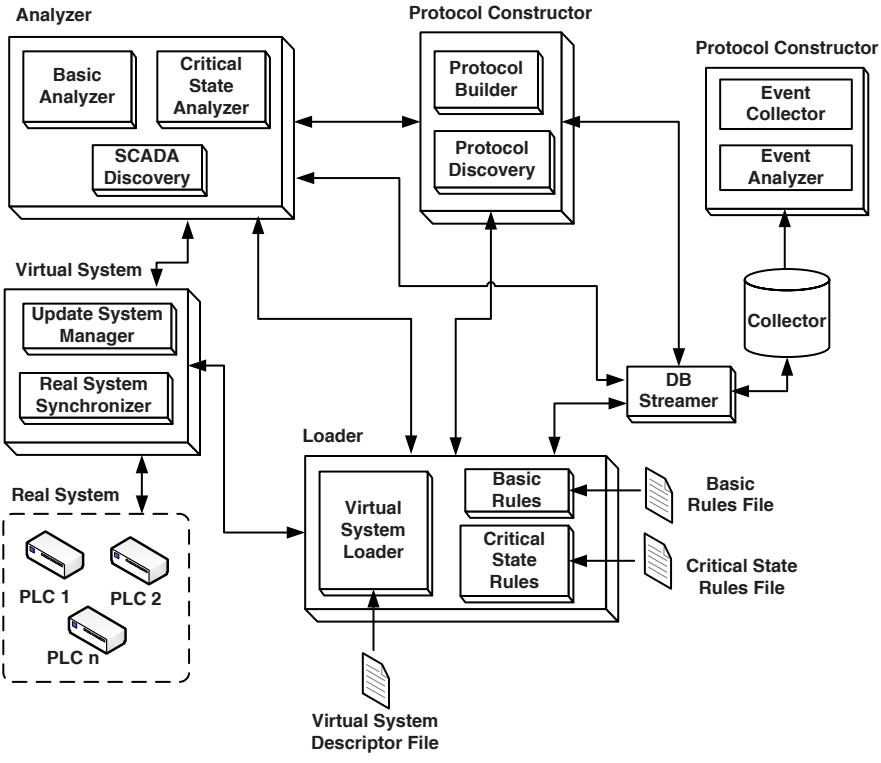


Figure 3. Local intrusion detection system architecture.

global IDS, which receives information from the local IDSs and implements high-level correlation and detection.

A local IDS has two information sources: network traffic flowing through the monitored subsystem, and direct queries that are periodically sent by the IDS to PLCs for information about their status. A local IDS contains a virtual representation of the monitored subsystem, and uses the two information sources to keep the local virtual system (LVS) in line with the real subsystem. The IDS identifies local unsafe states by comparing the state of the LVS with the set of critical states of the subsystem described using ICSML. Once a critical state is detected, the local IDS raises an alert and sends the list of subsystem services impacted by the critical state to the global IDS. The global IDS maintains a high-level virtual system, which is updated with the information received from the local IDSs. Then, the global state of the system is compared with the set of high-level critical states in search of a match.

4.1 Local IDS Prototype

The local IDS prototype incorporates five modules and fourteen functional components implemented in C# (Figure 3).

- **Loader (LS):** This module is responsible for initializing the system. It uses three XML files. The XML virtual system descriptor file contains the information used to create the virtual system. The basic rules file, commonly used in IDSs such as Snort, specifies malicious commands. The critical state rules file contains ICSML descriptions of rules related to critical states. The two rule files are imported and loaded into memory by separate functional components.
- **Protocol Constructor (PC):** This module contains the functional components that manage the construction and interpretation of SCADA protocols (currently Modbus and DNP3).
- **Analyzer (AZ):** This module monitors SCADA communications and performs single-packet detection, event correlation and local critical state detection.
- **Virtual System Manager (VSM):** This module stores a virtual representation of the SCADA system in memory. It updates the virtual system using the command flows captured by other components as input and by periodically querying the real system.
- **Database:** This MySQL database stores the alerts received from the analyzer.

4.2 Global IDS Prototype

The global IDS prototype receives critical state alerts from the local IDSs and correlates them to identify global critical states. The structure is simpler than that of the local IDS because it does not need modules to handle specific SCADA protocols. The global IDS has three modules:

- **Loader (LS):** This module is responsible for initializing the system. It uses two XML files, one to create the global virtual system image and the other to store the global critical states.
- **Global Virtual System Manager (GVSM):** This module manages the global virtual system, keeping track of its evolution and using as input the alerts received from the local IDSs.
- **Critical State Discovery (CSD):** This module analyzes the global virtual system to identify global critical states.

5. Experimental Tests

Several experiments were conducted to verify the performance of the distributed IDS. The SCADA testbed employed for the experiments reproduces the network, hardware and software used in a typical gas power plant [10]. In addition to evaluating IDS performance, the experiments analyzed the delays introduced by single-signature analysis, packet capture, virtual system updates and critical state analysis.

Data Rate Kbps	Alerts Expected	Alerts Detected
2.24	800	800
22.50	8,000	8,000
44.79	16,000	16,000
89.58	32,000	32,000
156.77	56,000	56,000
179.17	64.000	64,000
223.96	80.000	80,000
313.64	112,000	112,000
358.33	128,000	128,000
403.12	144,000	144,000
492.71	176,000	176,000

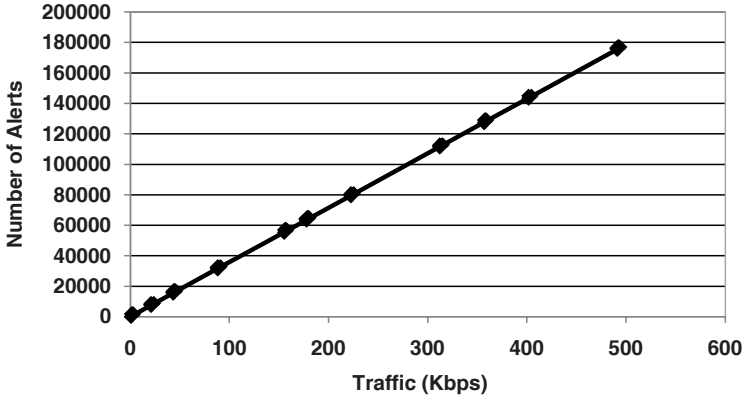


Figure 4. Alerts expected and alerts detected.

5.1 Single Signature Analysis

The amount of bandwidth that can be analyzed is a key critical issue in a NIDS. To evaluate this aspect, we implemented standard master/slave communications involving 100 request/response messages comprising 40 read requests, 50 write requests and 10 special functions. The IDS was configured to monitor Modbus and DNP3 traffic using a set of 2,000 *ad hoc* rules.

Figure 4 shows the results obtained when SCADA packets were sent simultaneously to a group of PLCs. The first column of the table shows the data rate (for an average SCADA packet size of 253 Bytes); the second and third columns show the numbers of expected alerts and detected alerts, respectively. Note that the IDS can analyze a large number of packets per second without packet loss.

Figure 4 also compares the number of expected alerts with the number of alerts raised by the IDS. Note that the IDS was able to raise all the alerts expected up to a data rate of 500 Kbps, which is very satisfactory given the

Table 2. Packet capture performance.

Requests Sent	100,000
Responses Sent	100,000
Request Size	315 Bytes
Response Size	315 Bytes
Request Rate	1 Request/ms
Traffic Rate	615.2 Kbps
Packet Loss	0

number of rules inspected and the low bandwidth available in industrial networks.

5.2 State-Based IDS

The state-based portion of the IDS has the largest impact on the real-time performance and, therefore, needs accurate control. For this reason, we carried out several tests, one for each step necessary to check the critical states of the system.

Packet Capture Packet loss is rare because "thread programming" was used to implement the IDS. Such losses can occur in the case of network congestion, but this is unlikely in a SCADA network. We tested the packet capture performance by sending a large number of packets at a very high bit rate.

The request/response transaction used in the packet capture experiment involved the master sending the slave a large request packet of 260 Bytes (maximum size allowed in Modbus); and the slave then responding with a large response packet of 260 Bytes. The request and response packets were both captured by the IDS.

The experiment to measure packet loss repeated this request/response transaction 100,000 times in order to generate a large amount of network traffic. The results are shown in Table 2. Note that the packet size is 315 Bytes (TCP header: 260 + 20 Bytes; IP header: 20 Bytes; Ethernet header: 15 Bytes).

No packet loss occurred for a burst of 100,000 packets at a rate of 615.2 Kbps (which is extremely high for a SCADA network). The experiment clearly demonstrates the reliability of the IDS.

Virtual System Updates The IDS updates the virtual system image in two steps: (i) it finds the PLC related to the content of the packet; and (ii) it updates the virtual object that represents the PLC.

The first step has no impact on IDS performance because the list of PLCs is stored in a hash table. The time required to find a PLC is the same for tables with 1 or 1,000 PLCs – around 0.0042 ms in our test environment.

Number of Coils	Average Time (ms)
1	0.0012168
50	0.0030485
100	0.0044824
500	0.0173109
1,000	0.0334344
2,000	0.0624535

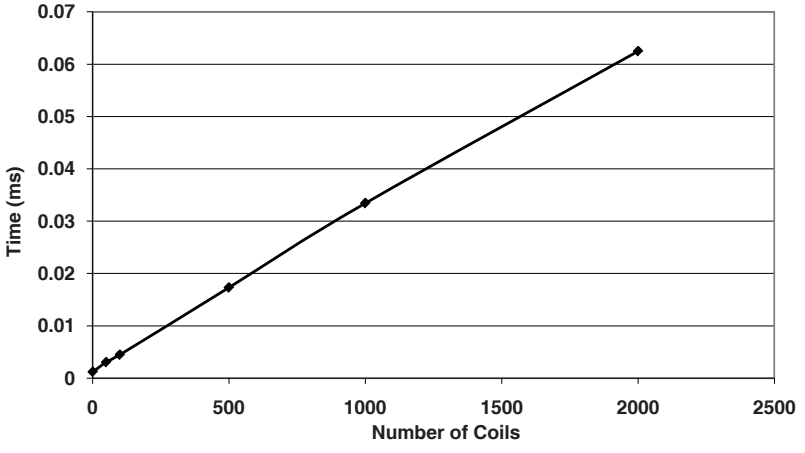


Figure 5. Virtual system performance test.

The second step takes more time, especially if it involves many registers or coils. Consequently, we conducted an analysis of PLC update times in a worst-case scenario. This worst-case scenario occurs when the IDS receives a packet with the function code 01 (read coils) with 2,000 coils to be read (maximum value in the Modbus specification [11–13]). The scenario requires the IDS to update the values of 2,000 coils.

Experiments were conducted for 1, 50, 100, 500, 1,000 and 2,000 coils to be updated. Figure 5 shows the average time taken to update values in the virtual PLC. The request/response transaction was repeated 1,000 times in order to obtain the average update time. As expected, the average time increases with the number of coils to be updated, but the increase is linear.

Critical State Analysis The performance of the critical state analyzer depends on two factors: (i) the number of conditions in each rule; and (ii) the number of rules.

To analyze the impact of rule size, we employed a request/response transaction with IDS capture and rule checking. The transaction involved the master sending a generic request to the slave; and the slave then sending the appropriate response. The IDS captured the request/response transaction and checked

Number of Conditions	Average Time (ms)
2	0.0204746
4	0.0217611
8	0.0244149
16	0.0301169
32	0.0370071
64	0.0550301
128	0.1206957
256	0.2127598
512	0.4226185
1,024	1.0706136

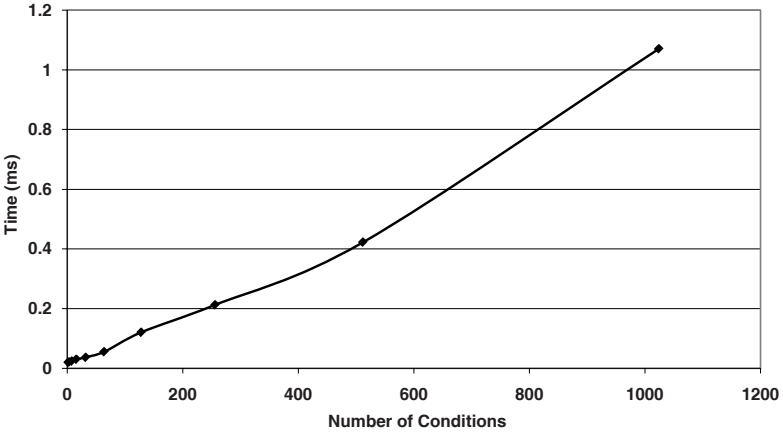


Figure 6. Critical state analyzer performance (Test 1).

if the virtual system entered into a critical state based on only one rule with a certain number of conditions.

Experiments were performed using rules with 2, 4, 8, 16, 32, 64, 128, 256, 512 and 1,024 conditions. In each case, the request/response transaction was repeated 1,000 times to obtain the average time for checking a rule.

Figure 6 shows the results of the experiments. Note that the elapsed time increases with the number of rule conditions and that the growth is linear.

Similar experiments were conducted to evaluate the impact of the number of rules. However, in this case, each rule had two conditions. The experiments used 10, 50, 100, 500, 1,000 and 2,000 rules. The results are shown in Figure 7. Note that the elapsed time increases with the number of rules and that the increase is linear. The results also demonstrate that critical state rules analysis is the performance bottleneck because it requires the most time of all the IDS operations.

Number of Rules	Average Time (ms)
10	0.1123061
50	0.5153591
100	1.0248889
500	2.6010271
1,000	5.0175991
2,000	9.9285867

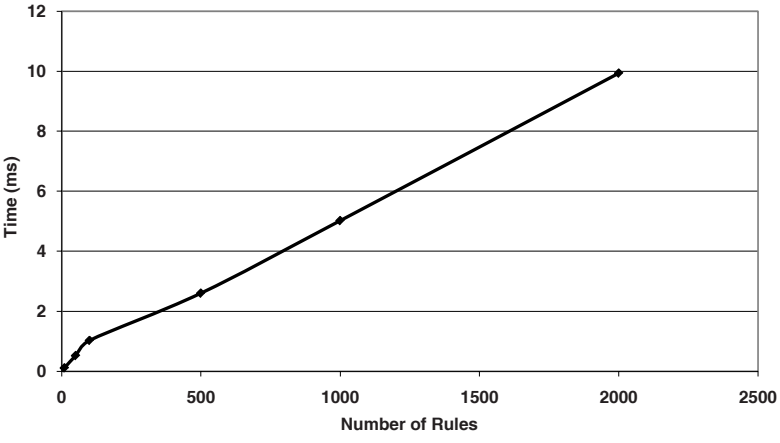


Figure 7. Critical state analyzer performance (Test 2).

6. Conclusions

The distributed intrusion detection approach developed for industrial control environments takes into account the state of the system of interest instead of attack signatures and anomaly heuristics. The approach rests on the assumption that the ultimate goal of an attacker is to put the system into a critical state. Consequently, instead of searching for the evolution of an attack, the approach tracks the evolution of the system. This approach addresses problems posed by false positives and permits the detection of unknown attacks.

Experimental results indicate that the IDS prototype exhibits good performance with respect to packet capture, virtual system updates and critical state analysis. Our future research will extend the prototype for application in a real-world industrial control environment. In addition, we plan to incorporate a critical state prediction feature, which will anticipate the evolution of the system into a known critical state on the basis of local sensor information.

References

[1] A. Carcano, I. Nai Fovino, M. Masera and A. Trombetta, SCADA malware: A proof of concept, presented at the *Third International Workshop on Critical Information Infrastructure Security*, 2008.

[2] F. Cuppens and A. Miege, Alert correlation in a cooperative intrusion detection framework, *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 202–215, 2002.

[3] D. Denning, An intrusion-detection model, *IEEE Transactions on Software Engineering*, vol. 13(2), pp. 222–232, 1987.

[4] Digital Bond, Modbus TCP IDS signatures, Sunrise, Florida (www.digitalbond.com/index.php/research/ids-signatures/modbus-tcp-ids-signatures).

[5] G. Dondossola, J. Szanto, M. Masera and I. Nai Fovino, Effects of intentional threats to power substation control systems, *International Journal of Critical Infrastructures*, vol. 4(1/2), pp. 129–143, 2008.

[6] P. Gross, J. Parekh and G. Kaiser, Secure selecticast for collaborative intrusion detection systems, *Proceedings of the International Workshop on Distributed Event-Based Systems*, 2004.

[7] M. Masera and I. Nai Fovino, Modeling information assets for security risk assessment in industrial settings, *Proceedings of the Fifteenth EICAR Annual Conference*, 2006.

[8] M. Masera and I. Nai Fovino, Models for security assessment and management, *Proceedings of the International Workshop on Complex Network and Infrastructure Protection*, 2006.

[9] M. Masera and I. Nai Fovino, A service-oriented approach for assessing infrastructure security, in *Critical Infrastructure Protection*, E. Goetz and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 367–379, 2007.

[10] M. Masera, I. Nai Fovino and R. Leszczyna, Security assessment of a turbogas power plant, in *Critical Infrastructure Protection II*, M. Papa and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 31–40, 2008.

[11] Modbus IDA, MODBUS Application Protocol Specification v1.1a, North Grafton, Massachusetts (www.modbus.org/specs.php), June 4, 2004.

[12] Modbus IDA, MODBUS Messaging on TCP/IP Implementation Guide v1.0a, North Grafton, Massachusetts (www.modbus.org/specs.php), June 4, 2004.

[13] Modbus.org, MODBUS over Serial Line Specification and Implementation Guide v1.0, North Grafton, Massachusetts (www. modbus.org/specs.php), February 12, 2002.

[14] I. Nai Fovino and M. Masera, Emergent disservices in interdependent systems and system-of-systems, *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, vol. 1, pp. 590–595, 2006.

[15] P. Ning, Y. Cui and D. Reeves, Constructing attack scenarios through correlation of intrusion alerts, *Proceedings of the Ninth ACM Conference on Computer and Communications Security*, pp. 245–254, 2002.

[16] V. Yegneswaran, P. Barford and S. Jha, Global intrusion detection in the DOMINO overlay system, *Proceedings of the Network and Distributed System Security Symposium*, 2004.

Chapter 8

DISTRIBUTED IP WATCHLIST GENERATION FOR INTRUSION DETECTION IN THE ELECTRICAL SMART GRID

Ray Klump and Matthew Kwiatkowski

Abstract The electric power infrastructure in the United States is undergoing a significant transformation. To enhance the ability of the grid to support the use of diverse and renewable energy resources and to respond to problems more quickly, the infrastructure is being redesigned to include greater options for automation, measurement and control. An enormous communications system will underlie the network of smart grid sensors and actuators. Devices will send messages to each other to coordinate control activity and formulate corrective strategies. The diversity and scale of this network will pose significant security challenges, especially since the number of entities charged with managing the grid will be large. A means for sharing information about cyber risks within the smart grid communications infrastructure is sorely needed. This paper proposes a strategy for sharing cyber security risks among smart grid stakeholders to enable them to identify attacks and mitigate their effects. The approach is inspired by the federated model, a cyber risk communications strategy employed by several U.S. national laboratories.

Keywords: Smart grid, federated model, intrusion detection

1. Introduction

The electric power grid is a complex interconnected control system of enormous scale and diversity. Disturbances in one part of the grid can profoundly impact conditions far away, despite control actions that are designed to isolate their impact. Different loads exhibit different dynamic response characteristics; different energy sources exhibit different availability profiles and rates of response to fluctuations in demand; and different units of protection equipment

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 113–126, 2010.

respond at different rates to different signals. Furthermore, measurements of the system are reported at vastly different rates and are required by different applications running on varied computing platforms [14]. Reporting rates for synchrophasor measurement units, a key component for enhanced wide-area monitoring and control [17], now occur at 30 to 60 measurements per second. Moreover, because of its geographical expanse, the grid is operated by multiple entities. Despite a universal mandate to keep the system operationally reliable in the face of the loss of any one credible contingency [2, 13], these operating entities adhere to different policies and procedures to meet the reliability mandate.

The diverse enterprise that is the electric power grid operates in an increasingly threatened environment. The period from 2000 to 2004 saw a tenfold increase in successful cyber attacks on the supervisory control and data acquisition (SCADA) systems that comprise the bulk of its communications, monitoring and control infrastructure [4]. Furthermore, it is believed that many (if not most) SCADA systems are inadequately protected against cyber attack [24]. While SCADA systems monitor and control the bulk of the grid infrastructure, they increasingly operate alongside new devices that use standard networking protocols like IP to provide what is described as an "end-to-end smart grid communications architecture" [20]. In fact, the adoption of networking equipment in the emerging smart grid is expected to create a network that will eclipse the size of the Internet [8]. The deployment of various smart-grid-related enhancements is currently well underway.

Despite the challenges, the creation of the smart grid promises several benefits. Modernization of the electrical grid is central to the nation's push for greater energy efficiency, the incorporation of renewable and cleaner resources, and the creation of more energy-sector jobs. Although there is no single model for the smart grid, all the various visions call for the expanded use of computing and networking technologies to support the two-way communication and control of power system devices [12]. This complies with the Energy Independence and Security Act (EISA) of 2007, which calls for the increased use of information and control technology to improve efficiency, reliability and security [23]; implementing this will involve the integration of a vast number of smart devices [22]. However, meeting the EISA mandate will require the collaboration of all the grid stakeholders to keep the system secure in the face of growing threats.

One way to increase the effectiveness of the collaboration is to capitalize on the fact that cyber attackers often prey on similar organizations, so that an incident at one location can be a precursor to an attack at another similar location [1]. Indeed, at various levels of detail, the electric power grid can be considered to be a network of related organizations. If the cyber security experiences of one organization can be broadcast securely in real time to its peers, then the threat awareness of the entire system can be greatly enhanced. While threat awareness involves a variety of considerations, the analysis generally begins with the identification of the source and destination IP addresses and port numbers associated with communications in a monitored network.

This paper proposes a distributed approach to generating watchlists and warning lists of IP addresses for intrusion detection and prevention. The concept is quite simple – it merely globalizes what local intrusion detection systems (IDSs) already do. This approach is based on the federated model, a technique used at a number of U.S. national laboratories [1, 11, 18]. Security and scale issues brought about as intrusion detection reports from increasing numbers stakeholders and devices contribute to the global watchlist are addressed using techniques implemented in the Worminator Project [21] and elsewhere [9, 10]. This paper also offers recommendations for sharing intrusion detection data in a variety of current and future grid architectures. Note that the framework for sharing IP address and port information from individual intrusion detection systems is just one component of a comprehensive cyber defense strategy for the power grid – one that formalizes the exchange of IDS data to strengthen the security vision of grid operators. It will be up to the individual entities to act on the shared data as they see fit.

2. Distributed Intrusion Detection

Several efforts have focused on techniques for sharing intrusion detection data among peers. Many of these efforts, including the popular online DShield tool [7], are cited in [9]. The federated model instituted at Argonne National Laboratory [1, 11, 18] is a recent implementation of distributed intrusion detection data sharing with centralized storage in a domain similar in scale and mission to the electric utility industry. The federated model grew from a "grass roots effort" to combat cyber security incidents at U.S. Department of Energy facilities [1]. The intent was to capitalize on the notion that attackers attempt to compromise related organizations and that, by working together, the related organizations can benefit from shared experiences as they combat attacks. By sharing potentially dangerous communication sources with each other, the organizations can contain the damage to just the first participant that received the communication.

Figure 1 illustrates the benefits of the model. Although Participant 1 is impacted by the attack, Participants 2, 3 and 4 benefit from the data reported via the federated repository and establish the appropriate defenses in a timely manner.

Argonne's implementation uses two mirrored repositories to store IP address and port combinations reported by the participating organizations. The additional repository provides a backup in case one goes down. The organizations report suspicious IP addresses identified by their intrusion detection systems to the repositories in an XML file based on the intrusion detection message exchange format (IDMEF) [6]. The organizations encrypt these files using their private PGP keys. The federated repositories collect this data and correlate it within a single IDMEF-formatted XML file, which is then passed to each organization encrypted under its public PGP key. All the organizations must register the IP addresses that are eligible to send and receive files.

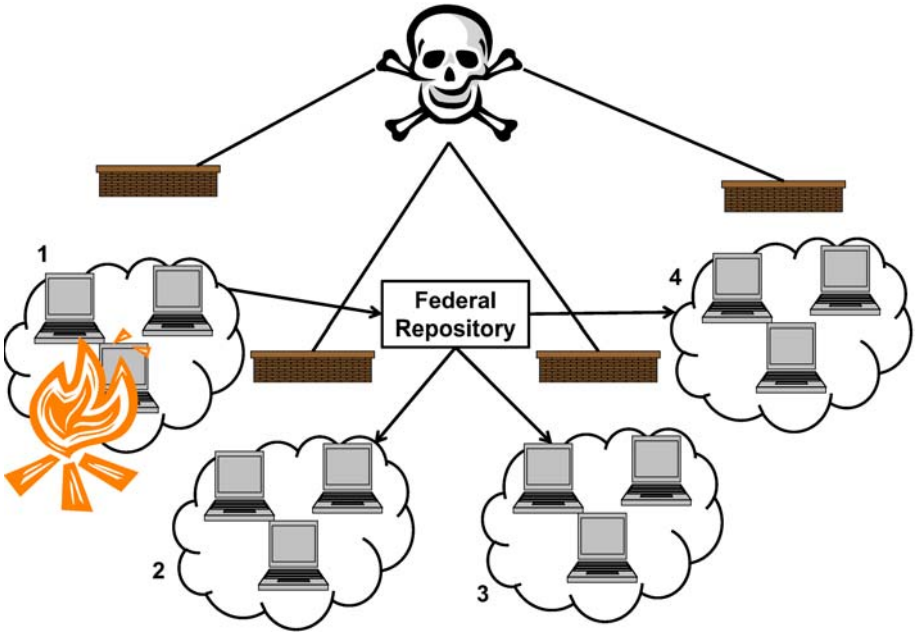


Figure 1. Attack limitation using the federated model.

Currently, more than twenty member organizations participate in the federated model. Nearly 1,000 events are communicated to the repository each day. Each organization has complete control over what it shares with other members through the central repository and the events to which it responds. Furthermore, each member is free to respond to the information it downloads from the repository as appropriate. This demonstrates respect for individual practices and an appreciation of the political pressures that may cause some organizations to be reluctant to share security-related data. Argonne personnel see this as a way to refine the organizations' observe-orient-decide-act (OODA) loop, because all but the directly-impacted organization will have more time and more intelligence to handle attacks. Martin [11] provides an example in which information conveyed via the federated model could have blocked a malicious site full two weeks before it was blocked manually. The federated model is germane to the electrical power grid because of its similarity in scale and mission, particularly when the grid is viewed from a regulatory framework.

3. Distributed IP Watchlist Generation

Whenever a hierarchical structure is to be controlled and monitored, it is necessary to determine the level at which most of the tasks will be performed. Only then can an appropriate strategy for sharing data and decisions be chosen. Deciding where to assign responsibilities requires the consideration of various operational models of the electrical grid. Details on how smart grid devices will

be integrated within the electrical and communications networks of the grid are still taking shape. Therefore, this section describes the electrical grid as it is currently managed and how it might be organized in the future. It is important to understand the models, because they affect the intrusion detection strategies that can be employed.

Like the Internet, the electric power grid can be viewed as a network of networks. Administratively, though, it can also be viewed as a hierarchy of managing entities. The organizational management perspective is most germane to how the grid operates today and reflects the current regulatory environment. For example, the North American grid has three interconnections: the Eastern Interconnection, the Western Interconnection and the Texas Interconnection. Each interconnection has one or more reliability councils, each of which monitors the operations of balancing authorities that usually deliver power to geographically contiguous areas. Within each balancing authority are generation sources, loads and transmission facilities that deliver power from generators to loads. Balancing authorities are responsible for ensuring that their generation matches their load and power exchange demands so that a constant system frequency can be maintained. The reliability councils help coordinate activities when their constituents are out of balance. The entire grid is monitored by the North American Electric Reliability Corporation (NERC).

At present, before the widespread adoption of smart grid technologies and the decentralized control strategies they may afford, the control of grid assets is centralized in the owning balancing authorities. Thus, the operation of the grid currently adheres to a regulatory model, which is shaped by the grid management and accountability structures.

Each balancing authority has its own information technology (IT) staff and each has a portion of the communications infrastructure for which it is responsible. The communications infrastructure for a balancing authority supports corporate systems as well as power monitoring and control systems. While a barrier typically exists between the two systems, it is not always secure (see, e.g., [5]).

Given this model, each balancing authority is required to monitor and respond to cyber attacks against the equipment within its jurisdiction. Viewed from a regulatory perspective, the structure that results is quite similar to the network of national laboratories that currently share intrusion detection data via the federated model. In this analogy, the grid's balancing authorities have jurisdictions similar to those of the national laboratories. The balancing authorities and national laboratories operate independently; both kinds of organizations answer to a supervisory body: reliability councils in the case of balancing authorities and federal agencies in the case of national laboratories. The balancing authorities share a common mission to operate their portions of the grid in accordance with the requirements set by the reliability councils, just as the national laboratories collaborate to achieve the larger research agendas.

To satisfy its own operating responsibilities and meet the requirements prescribed by its reliability council, each balancing authority has to share data

with the reliability council and with its neighbors. To increase the cyber security awareness of the reliability council, the data should include intrusion detection information. A reasonable data set consists of the IP addresses and ports of suspicious communications. Using the federated model as a template, each balancing authority within a reliability council could watch for possible intrusions on its network according to its IT department's policies and procedures. At a minimum, each balancing authority would have to maintain two lists of IP address and port combinations: a watchlist and a warning list [9]. The IP watchlist would contain potentially rogue address/port combinations encountered by the balancing authority during the monitoring period (set by the council to be a certain number of days). For example, a reliability council might require the watchlist to keep track of new IP address and port combinations for the past thirty days. IP address and port combinations that are deemed by the balancing authority to pose a particular threat, perhaps because they have appeared with a worrisome frequency during the monitoring period, would be transferred to the balancing authority's warning list, a permanent record of source or destination points that should be regarded as rogue or potentially dangerous by balancing authorities in the reliability council. How these lists are populated depends on the policies of the individual balancing authority. This approach would preserve local control over cyber security monitoring while enhancing cyber security awareness throughout the council.

In real time, or at some interval defined by the reliability council, the balancing authorities would be required to communicate the updates to their watch and warning lists to the reliability council. The reliability council would consolidate the watchlist updates from the balancing authorities into a single council-wide watchlist containing unique IP address and port combinations along with the number of times that each combination was reported system-wide. Using criteria established by the reliability council, the global warning list would be augmented with the warning list updates provided by the member balancing authorities along with additions to the watchlist that exceeded the council's frequency threshold. For example, if the council's frequency threshold dictates that IP address and port combinations be moved from the watchlist to the warning list when the report count exceeds five, then a combination may be deemed dangerous if five different balancing authorities reported it recently, or if a single balancing authority reported the address more than five times during the monitoring period, or when some other combination of reporting parties and address detections arises. Updates to the global warning list would then be distributed to the balancing authorities to enable them to augment their firewall rules as appropriate.

Figure 2 illustrates the communications that would take place between the reliability council and its member balancing authorities. A similar set of communications could occur to manage global watch and warning lists for sets of reliability councils. In this case, each reliability council could publish its watch and warning list updates to a coordinating body, perhaps a NERC designee. This entity would be responsible for updating and disseminating global warning

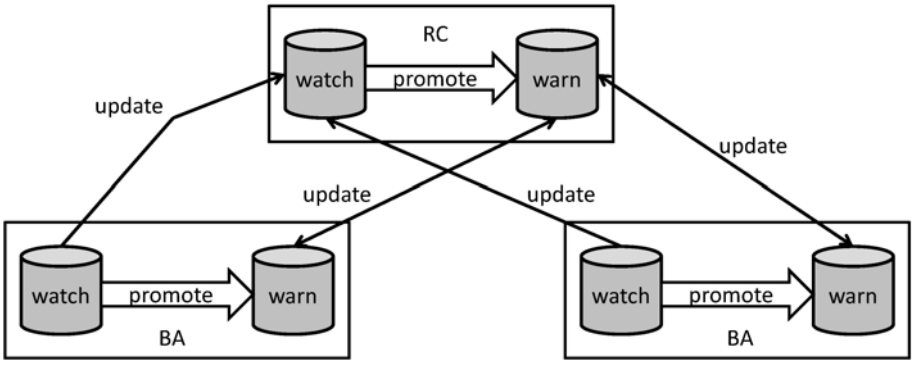


Figure 2. Distributed IP watch and warning list generation.

list updates to the member reliability councils, which in turn would pass these updates to their constituent balancing authorities.

An alternative to this approach is for each balancing authority to maintain a "white list" of IP addresses with which it may communicate. The proposed system could support this choice, since it merely provides a formal system for communicating events and threats and leaves it to the participants to decide how to use the data. Depending solely on a white list, however, may prove problematic for the balancing authority's operations and business centers. Because the facilities of balancing authorities host corporate and command and control systems, a comprehensive white list that supports both types of applications would be cumbersome to manage.

Based on the experiences of the national laboratories with the federated model, a collaboratively-generated black list that disallows communications based on data shared through the proposed system would be both proactive and flexible. Also, as individual devices evolve into smart components that can act autonomously, it is conceivable that a device that was within the safety zone defined by the white list could be compromised and become a bad actor. This requires the white list to be changed for the owning utility as well as for any other entity (e.g., an aggregator) allowed to communicate with it. In this case, the proposed system could be used to coordinate updates to either a white list or a black list database, depending on the policies in place at the participating entity.

Although it can be claimed that each balancing authority has a competitive impetus to act in an adversarial manner toward its peers, there is considerable motivation to act cooperatively because of the dire public consequences of a grid security failure. Furthermore, each participant has the freedom to choose the entities with which it will share intrusion detection data and on whose data it will act. If trust in particular entities is compromised, the participants of the federation can decide how to respond to mend the relationships. The flexibility comes about because of the voluntary nature of the federation.

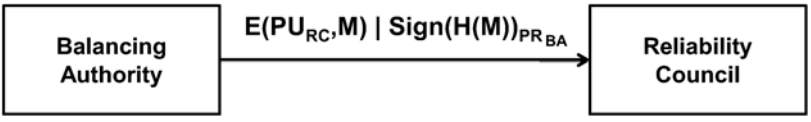


Figure 3. Public key approach for confidentiality, authenticity and integrity

4. Security and Scalability

This section discusses the security of the update messages passed between entities and the scalability of the architecture.

The update messages passed between balancing authorities must be confidential and authentic, and have integrity. The messages must be confidential because one balancing authority may not wish its peers to know the parties with which it is communicating. This is true even if the parties are adversarial – unwanted contact may cause the cyber security readiness of one of the parties to come under the scrutiny of its peers. The messages must be authentic in terms of source and destination, because the recipient, whether it is the reliability council receiving an update from a balancing authority or a balancing authority receiving an update from the reliability council, must be confident that the sender is identified correctly. Finally, update messages must be received as sent. It should not be possible to modify the contents of updates via a man-in-the-middle attack.

There are several ways to achieve the requirements of confidentiality, authenticity and integrity. For example, a public key infrastructure could be used to provide public and private keys to each balancing authority and reliability council. A balancing authority would encrypt a watchlist or warning list update to the reliability council using the reliability council's public key. The update would also be signed by the balancing authority by computing and encrypting the hash of the update using its private key. The encrypted update and the signature are then communicated to the reliability council. Upon receipt, the reliability council can decrypt the update using its private key. It then deciphers the signature by decrypting the hash with the sender's public key. Next, it computes the hash of the decrypted update and compares it with the decrypted signature; if the two match, the sender is authenticated and the received update matches what was sent.

Figure 3 illustrates the public key approach. M denotes a watch or warning list update message sent from the balancing authority to the associated reliability council. To prevent replay, it may also be necessary to send a timestamp that the recipient can check against a list of previously received timestamps. The public key approach works in a similar (albeit reverse) manner for messages sent from the reliability council to a balancing authority. The reliability council could send each balancing authority an update encrypted with the balancing authority's public key. Alternatively, it could send multiple balancing authorities within its jurisdiction the same update encrypted using a public

key shared by the group. Key distribution would be manageable in both cases because the number of communicating entities would be small.

In determining how well this approach scales, it is necessary to consider the number of participants and the sizes of the watch and warning lists and the update messages. In the regulatory model, the number of communicating entities would be small as the number of balancing authorities, which generally coincide with electric utilities, is unlikely to grow much beyond the hundred or so that exist today. Therefore, the number of participants engaged in communications between the reliability council and balancing authority would not contribute to a problem of scale; in fact, the number would be approximately the same as the number of entities participating in the federated model.

However, the sizes of the updates and the watch and warning lists may be a concern. Individual balancing authorities control how much detail they provide to the reliability council. The filter used by a balancing authority for the set of IP addresses and ports it collects and passes to the reliability council may be more stringent than the criteria it uses internally to add the address-port combinations to its watchlist. Also, it may make the reasonable choice to omit attempts to access non-existent services in its reports to the repository. However, if balancing authorities choose not to be as selective in what they report, additional steps would have to be taken to maintain system performance at acceptable levels.

One approach for managing message volume is to use a Bloom Filter to represent messages more compactly [9]. In this approach, the watch and warning lists could each be represented as a large array of bits initialized to zero. When a new IP address and port combination is reported to the reliability council, authenticated and integrity-checked, it is stored as sent and it is also hashed. Portions of that hash are used to calculate the indices of entries in the bit array that should be set to one. Thus, whenever an IP address and port combination is sent in an update, the indices in the Bloom Filter array are checked to see if they are all already turned on. If all the bits are not equal to one, then the combination has been reported for the first time. If all the bits are set to one, then the combination may be a repeat of a earlier combination, in which case the list of hashes recorded as sent have to be checked to determine if it is indeed a duplicate. If it is a duplicate, then its occurrence count is increased, possibly making it a candidate for promotion to the warning list.

5. Decentralized IP Watchlist Generation

Section 3 described the implementation of a distributed scheme for identifying problematic IP addresses that was centralized at the balancing authority. As more smart grid devices capable of two-way communication are deployed in the future, it may become necessary to adopt a more decentralized scheme in which localized clusters of devices share information.

Consider the more device-focused architecture shown in Figure 4. In this model, the end loads L_i at the lowest level of the architecture are controlled to achieve the operating objectives. For example, in [19], the end loads are

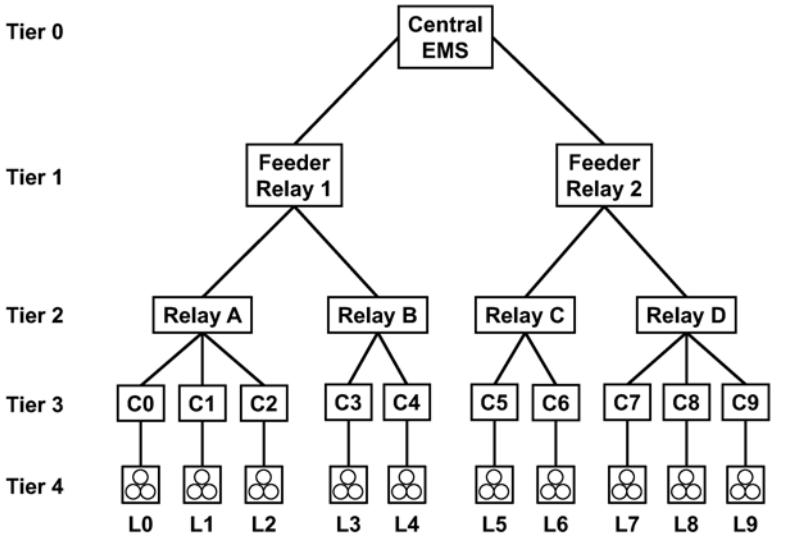


Figure 4. Delegation model for smart grid communications and control.

regulated to provide more reactive power support to regions experiencing depressed voltage. Given the proper equipment, the problem may be detected and addressed locally instead of by the central energy management system (EMS) housed at the balancing authority. If this is possible, then the load L_i is regulated by its corresponding controller C_i to address the problem. If the problem cannot be handled locally, but requires the assistance of peer devices in nearby regions, then the responsibility for the problem may be assigned to the next higher tier. Again, if properly equipped, the device at the next higher tier can formulate a response to mitigate the problem that calls for support from a broader pool of devices than just those in the affected region. Messages passed among tiers of this model must be authenticated and checked for integrity, and the devices in each tier must be "smart" in that they have the processing power to assess the electrical characteristics and formulate a control response.

This architecture manages the grid through delegation: each tier can communicate only with the tier directly above it or directly below it. For example, if a problem at load L_1 in Tier 4 cannot be handled by the controller C_1, Relay A or Feeder Relay 1 in the intervening lower tiers, then the Central EMS in Tier 0 will formulate a strategy that calls for supportive action and communicate it to its two children in Tier 1, Feeder Relay 1 and Feeder Relay 2. These, in turn, will pass instructions contained in the just-received directive from the Central EMS to the appropriate relays in Tier 2, the next lower tier. The relays repeat the parse-and-pass procedure to forward required instructions to the appropriate load controllers C_i in Tier 3.

A distributed approach to collecting and correlating intrusion detection data in this case might involve establishing a separate repository for each tier. The

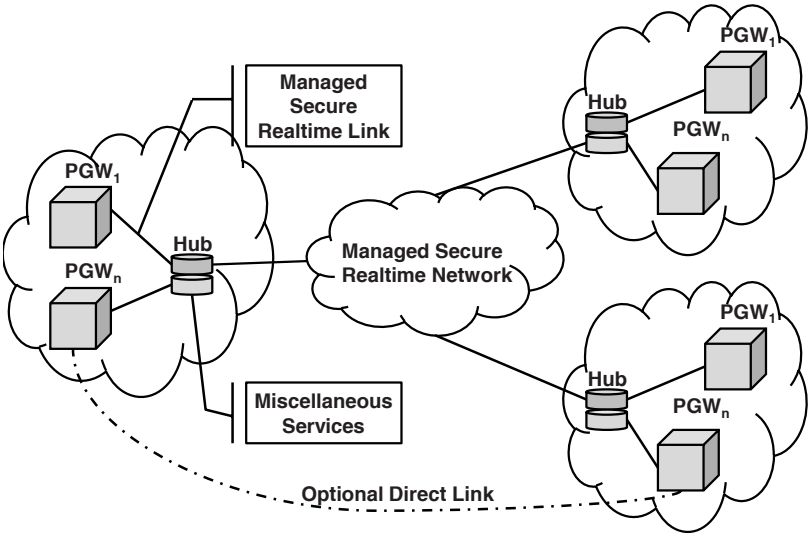


Figure 5. Hub-based NASPInet architecture (from [3]).

repository at Tier n would maintain the watch and warning lists for the devices in Tier $n + 1$. Since control requests never pass beyond the immediately next tier, such a short-range approach to compiling the watch and warning lists would support the needs of the smart grid architecture. By defining the repositories by tier, the scaling problem that would otherwise be encountered if the entire system communicated with a single repository is avoided. Furthermore, in the event of a multi-tier attack, the suspicious activities recorded in each tier could provide the data necessary to interfere with the progress of the attack.

Bobba, *et al.* [3] describe another example of a tiered architecture, motivated by a specific application, that exhibits elements of the regulatory and delegation models. One of the aims of the emerging smart grid is to increase wide-area situational awareness. A tool for achieving this is the synchrophasor measurement unit (PMU), a GPS-time-synchronized meter capable of measuring voltage and current magnitudes, phase angles and frequencies between 30 to 60 times per second. The North American Synchrophasor Initiative (NASPI) is planning the deployment of PMUs throughout the grid. The current plan, documented in [15, 16], assigns monitoring and control of each PMU to its owning utility through devices called phasor gateways (PGWs).

An alternative NASPInet architecture is proposed in [3] to address a potential bottleneck in reporting large quantities of data to the owning authority. In this design, which is illustrated in Figure 5, the phasor gateways report to hubs that share information with each other using a secure realtime network. The hubs manage requests for data as well as the collection and correlation of phasor measurement data. They could also serve as hosts for the distributed intrusion detection effort. Each hub could maintain watch and warning lists

for its constituent phasor gateways and share the lists with its peer hubs. This approach should scale well because the number of hubs is much more than the number of individual PMUs and PGWs. Regardless of whether the hub communications are regulated by a white-list-based or black-list-based approach, by sharing intrusion detection intelligence with each other, the hubs can achieve a more comprehensive view of security threats.

6. Conclusions

The distributed intrusion detection architecture presented in this paper gathers threat data from multiple sources and disseminates consolidated updates to participating entities, helping improve the wide-area security awareness of the electrical power grid. The architecture is applicable to the current grid that operates according to a regulatory model as well as various future smart grid designs that operate in a distributed, device-oriented manner. Also, the architecture supports secure messaging and is scalable.

Acknowledgements

This research was partially supported by the National Science Foundation under Grant No. CNS-0524695 and by the Department of Energy under Award No. DE-OE0000097. This article was created by UChicago Argonne, LLC, Operator of Argonne National Laboratory, a U.S. Department of Energy Office of Science Laboratory, which is operated under Contract No. DE-AC02-06CH11357. Note that the U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license to reproduce this article, prepare derivative works, distribute copies to the public, perform publicly and display publicly by or on behalf of the U.S. Government.

References

[1] Argonne National Laboratory, Federated model for cyber security: Collaborative effort to combat Internet attackers, Argonne, Illinois (webapps .anl.gov/federated), 2009.

[2] N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M. Lauby, B. Wollenberg and J. Wrubel, On-line power system security analysis, *Proceedings of the IEEE*, vol. 80(2), pp. 262–282, 1992.

[3] R. Bobba, E. Heine, H. Khurana and T. Yardley, Exploring a tiered architecture for NASPInet, presented at the *First IEEE PES Conference on Innovative Smart Grid Technologies*, 2010.

[4] E. Byres and D. Hoffman, The Myths and Facts behind Cyber Security Risks for Industrial Control Systems, Technical Report, Department of Computer Science, University of Victoria, Victoria, Canada, 2004.

[5] E. Byres, A. Paller and B. Geraldo, Special webcast: Cyber attacks against SCADA and control systems, SANS Institute, Bethesda, Maryland, 2009.

[6] H. Debar, D. Curry and B. Feinstein, The Intrusion Detection Message Exchange Format (IDMEF) (www.ietf.org/rfc/rfc4765.txt), 2007.

[7] DShield, DShield Cooperative Network Security Community (www.dshield .org).

[8] M. LaMonica, Smart grid will eclipse size of Internet, CNET News (news .cnet.com/8301-11128_3-10241102-54.html), May, 18, 2009.

[9] M. Locasto, J. Parekh, A. Keromytis and S. Stolfo, Towards collaborative security and P2P intrusion detection, *Proceedings of the IEEE Workshop on Information Assurance and Security*, pp. 30–36, 2005.

[10] M. Locasto, J. Parekh, S. Stolfo, A. Keromytis, T. Malkin and V. Misra, Collaborative Distributed Intrusion Detection, Technical Report CUCS-012-04, Department of Computer Science, Columbia University, New York, 2004.

[11] T. Martin, Federated model for cyber security: Sharing intrusion detection results, Argonne National Laboratory, Argonne, Illinois (webapps.anl.gov /federated/site_media/docs/Presentations/DOETechSummit.pdf), 2008.

[12] National Institute for Standards and Technology, NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0, NIST Special Publication 1108, Gaithersburg, Maryland, 2010.

[13] North American Electric Reliability Corporation, Reliability Standards for the Bulk Electric Power Systems of North America, Princeton, New Jersey, 2010.

[14] North American Synchrophasor Initiative, Phasor Applications Taxonomy, Pacific Northwest National Laboratory, Richland, Washington, 2007.

[15] North American Synchrophasor Initiative, Data Bus Technical Specifications for North American Synchrophasor Initiative Network, Pacific Northwest National Laboratory, Richland, Washington, 2009.

[16] North American Synchrophasor Initiative, Phasor Gateway Technical Specifications for North American Synchrophasor Initiative Network, Pacific Northwest National Laboratory, Richland, Washington, 2009.

[17] North American Synchrophasor Initiative, Synchrophasor Technology Roadmap, Pacific Northwest National Laboratory, Richland, Washington, 2009.

[18] S. Pinkerton, A federated model for cyber security, presented at the *Cyberspace Research Workshop*, 2007.

[19] K. Rogers, R. Klump, H. Khurana and T. Overbye, Smart-grid-enabled load and distributed generation as a reactive resource, presented at the *First IEEE PES Conference on Innovative Smart Grid Technologies*, 2010.

[20] J. St. John, Duke Energy enlists Cisco in smart grid efforts, Greentech Media, Cambridge, Massachusetts (www.greentechmedia.com/articles/read /duke-energy-enlists-cisco-in-smart-grid-efforts), June 9, 2009.

[21] S. Stolfo, Worm and attack early warning, *IEEE Security and Privacy*, vol. 2(3), pp. 73–75, 2004.

[22] U.S. Department of Energy, Recovery Act – Smart Grid Investment Grant Program, DE-FOA-0000058, Washington, DC, 2009.

[23] U.S. Government, Energy Independence and Security Act of 2007, Public Law 110–140, *United States Statutes at Large*, vol. 121, pp. 1492–1801, 2007.

[24] C. Wilson, Computer Attack and Cyber Terrorism: Vulnerabilities and Policy Issues for Congress, CRS Report for Congress, RL32114, Congressional Research Service, Washington, DC (www.fas.org/irp/crs/RL32114.pdf), 2003.

Chapter 9

SECURITY ANALYSIS OF THE MPLS LABEL DISTRIBUTION PROTOCOL

Daniel Guernsey, Aaron Engel, Jonathan Butts and Sujeet Shenoi

Abstract Since its inception more than a decade ago, multiprotocol label switching (MPLS) has become one of the fastest-growing telecommunications infrastructure technologies. The speed, flexibility, sophisticated traffic management and cost savings offered by MPLS have prompted service providers to converge existing and new technologies onto common MPLS backbones. Indeed, much of the world's data, voice communications, video traffic and military applications traverse an MPLS core at some point.

The rapid adoption of MPLS raises significant concerns – primarily because of the dependence of critical communication services on a technology that has yet to undergo significant security testing. This paper examines security issues associated with the Label Distribution Protocol (LDP), which is the primary route construction protocol in MPLS networks. Our analysis has identified ten attacks that exploit weaknesses in the LDP specification: six attacks that disrupt service and four that divert traffic from intended routes. Details of the attacks are presented along with suggested mitigation strategies and security postures.

Keywords: Multiprotocol label switching, Label Distribution Protocol, security

1. Introduction

Multiprotocol label switching (MPLS) is quickly becoming the *de facto* protocol for transporting traffic in modern telecommunications networks. MPLS networks leverage the performance and availability of circuit-switched networks with the robustness and flexibility of packet-switched networks. Traffic entering an MPLS network is tagged with labels based on customer quality of service (QoS) and class of service (CoS) requirements. This allows traffic to be classified and then routed according to provisioned services (e.g., data type, message source, message destination and bandwidth requirements) instead of destination-only methods employed in traditional IP networks.

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 127–139, 2010.

In December 2005, the United States Department of Defense (DoD) achieved full operational capability of the Global Information Grid Bandwidth Expansion (GIG-BE) Program. The GIG-BE is designed to deliver global, high-speed classified and unclassified services to meet national security intelligence, surveillance and reconnaissance; and command and control requirements [10]. MPLS was chosen as the network transport backbone primarily due to its efficiency, simplicity and popularity in commercial environments [5, 9]. The DoD's use of MPLS for critical data is by no means unique. Many major telecommunications service providers around the world have invested massively in MPLS technology [2, 4, 16]. In fact, according to one source [13], 84% of enterprises have already transitioned their wide area networks to MPLS.

Despite the massive growth of MPLS networks, very little research has focused on the security aspects of core protocols such as the Label Distribution Protocol (LDP). LDP is the primary mechanism for transforming IP routes into high-speed "autobahns" within the MPLS paradigm. Weaknesses in LDP can be exploited by an attacker to achieve a wide range of strategic effects, including disrupting voice, global data and emergency communications.

This paper examines the security issues related to LDP. In particular, it discusses how LDP can be exploited to isolate network segments, reroute network traffic, disable the routing of network traffic and perform targeted attacks. Ten exploits are discussed: six denial-of-service attacks and four route modification attacks. Denial-of-service attacks target weaknesses in LDP to degrade or deny legitimate traffic delivery. Route modification attacks alter the path of targeted MPLS traffic traversing the network. The paper concludes by outlining mitigation strategies and security postures.

2. Multiprotocol Label Switching Networks

Connection-oriented and connectionless protocols are the two principal paradigms for transporting traffic across large networks [12]. ATM (OSI Layer 2) is an example of a connection-oriented technology that provides low latency and high quality of service (QoS). IP (OSI Layer 3) is a connectionless protocol that supports a multitude of underlying heterogeneous network technologies.

Service providers are eager to leverage the flexibility of IP and the speed of ATM without sacrificing efficiency [8]. In traditional implementations, an overlay model is used to create an ATM virtual circuit between each pair of IP routers. The IP routers are unaware of the ATM infrastructure and the ATM switches are unaware of IP routing. The end result is relatively inefficient: the ATM network must construct a complete mesh of virtual circuits among the IP routers.

MPLS offers an alternative solution that enables connection-oriented nodes to peer directly with connectionless technologies by transforming ATM switches into IP routers. ATM switches participate directly in IP routing protocols (e.g., RIP and OSPF) to construct label switched paths (LSPs). LSPs are implemented in ATM switches as virtual circuits, enabling existing ATM technology to support the MPLS forwarding mechanism. Conversely, MPLS enables

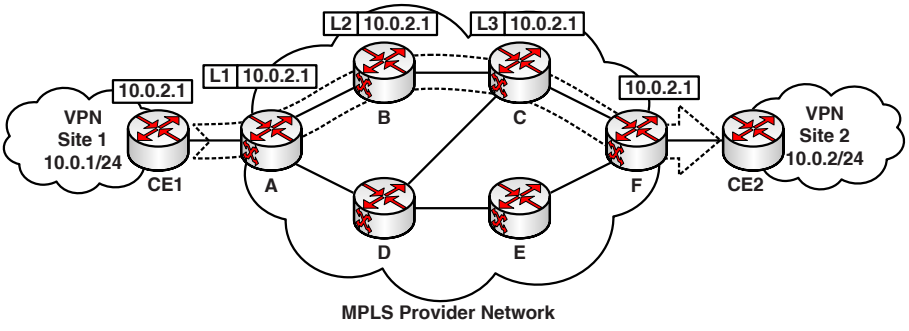


Figure 1. MPLS packet forwarding.

connectionless technologies to behave in a connection-oriented manner by augmenting IP addresses and routing protocols with relatively short, fixed-length labels.

Each label is a 32-bit (fixed length) tag, which is inserted in the Layer 2 header (e.g., for ATM VCI and Frame Relay DLCI) or in a separate "shim" between Layers 2 and 3 [14]. A label works much like an IP address; it dictates the path the router uses to forward the packet. Unlike an IP address, however, an MPLS label only has local significance. When a router receives a labeled packet, the label informs the router (and that router only) about the operations to be performed on the packet. Typically, a router pops the label on an incoming packet and pushes a new label for the router at the next hop in the MPLS network; the network address in Layer 3 is unchanged.

Figure 1 illustrates a typical MPLS architecture that interconnects two customer VPN sites. Routers A through F in the MPLS network are called label switched routers (LSRs). Customer edge routers, CE1 and CE2, sit at the edge of the customer network and provide connectivity to the MPLS core.

Consider the LSP from VPN Site 1 to VPN Site 2 (Routers A, B, C and F). Router A is designated as the "ingress node" and Router F is designated as the "egress node." The ingress and egress nodes are often called label edge routers (LERs) because they are at the edge of the MPLS network [14].

When an IP packet reaches the ingress of the MPLS network, LER A consults a forwarding information base (FIB) and assigns the packet to a forwarding equivalence class (FEC). The FEC maps to a designated label that supports QoS and CoS requirements based on IP parameters in the packet (e.g., source IP address, destination IP address, application).

In this example, LER A pushes Label L1 onto the packet and forwards it to LSR B. LSR B reads the label, consults its local label information base (LIB) to identify the next hop, pops the previous label and pushes a new label (L2), and forwards the packet to LSR C. LSR C behaves similarly, forwarding the packet to LER F. LER F then pops Label L3, examines the destination IP address and forwards the packet to VPN Site 2.

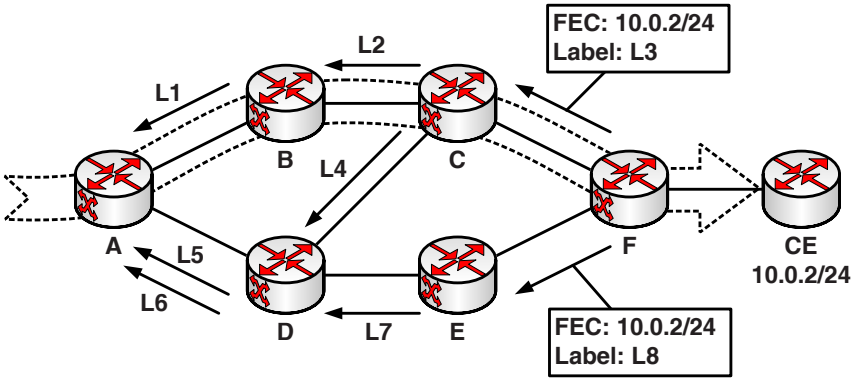


Figure 2. LDP routing information flow.

3. MPLS Routing Information

MPLS defines a forwarding mechanism designed to emulate IP routes using labels and paths. IP networks rely on routing protocols such as RIP and OSPF to populate the IP forwarding table [12]. Similarly, MPLS networks engage label distribution protocols to populate the FIB and LIB and establish end-to-end LSPs.

The Label Distribution Protocol (LDP) is the primary MPLS protocol for exchanging label mapping information [7]. LDP relies on underlying IP routing information to construct a set of LSPs using best-effort routes [6]. LSPs, in turn, can be optimized by employing traffic engineering protocols. MPLS traffic engineering protocols (e.g., RSVP-TE and MP-BGP) use topology information, constraints, specialized algorithms and signaling protocols to create LSPs to match customer QoS and CoS requirements [3, 15]. Traffic engineering protocols rely on LSPs constructed by LDP to discover the underlying routing structure. As such, exploiting a weakness in LDP can be leveraged to affect LSPs generated through traffic engineering.

4. Label Distribution Protocol (LDP)

LDP is designed to distribute information about available routes within an MPLS network. The edge routers begin the process by distributing label information about their adjacent external networks. FECs are created for each network based on IP addresses or prefixes [1].

Consider the example in Figure 2. LER F defines an FEC F1 for 10.0.2/24 and binds it to Labels L3 and L8. Next, it distributes the mappings (L3, F1) and (L8, F1) to its upstream peers (LSR C and LSR E, respectively) to update their LIBs. Upon receiving the mapping, LSR C binds a label to FEC F1 for each of its upstream interfaces and distributes these labels to LSR B and LSR D. Similarly, LSR E distributes the mapping (L7, F1) to LSR D. The process terminates when the information reaches an ingress router (LER A).

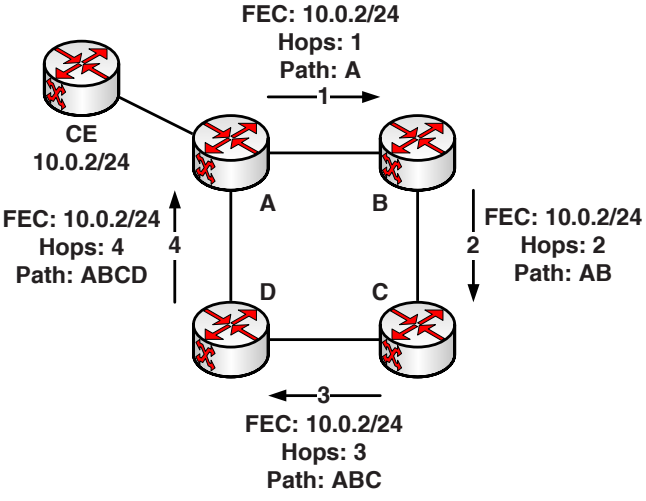


Figure 3. Loop detection using path vectors and hop counts.

The mapping distribution provides LER A with three distinct paths for 10.0.2/24. For example, if LER A receives an IP packet addressed to 10.0.2.1, it consults its FIB for an FEC with the longest matching prefix. Because three paths exist for 10.0.2/24, LER A selects the least cost path determined by its IP table. To meet customer requirements, FECs can be generated through traffic engineering for distinct destinations or applications to ensure that specific bandwidth, latency and other services are adequately provided.

4.1 Label Merging

It is common for two LSPs to converge prior to reaching a common egress [1]. To save memory and label space, LSRs may merge the LSPs at the point of convergence. When a merge-capable LSR receives a label request for an existing FEC and label mapping, it does not forward the request. Rather, it distributes the existing mapping to its upstream neighbors, effectively merging the two requested LSPs.

4.2 Loop Detection

The recursive nature of label request and label mapping messages creates the potential for message loops [1]. LDP uses hop counts and path vectors to detect loops. When a mapping request is forwarded, the LSR increments the message hop count and appends its own ID to the path vector. If the hop count exceeds a configured limit or an LSR discovers its ID in the path vector, the LSR sends a notification to the sender that a loop has been detected. In Figure 3, LSR A detects ID A in the path vector, implying that a loop exists. LSR A

stops forwarding the message and sends a notification to LSR D to prevent the construction of an LSP that contains a loop.

4.3 LDP Messages

Four message classes in LDP are used to facilitate session management and label distribution [1]: (i) Discovery messages that establish network adjacencies; (ii) Session messages that initialize and maintain LDP connections; (iii) Advertise messages that establish and remove LSPs; and (iv) Notification messages that specify advisories and errors.

Discovery Class Messages

- **Hello:** Hello messages are exchanged among LSRs during the discovery process using UDP. There are two types of messages: (i) Link Hello messages and (ii) Extended Hello messages. Link Hello messages are sent between directly-linked LSRs by addressing the messages to the subnet broadcast address. Extended Hello messages are exchanged between non-directly-linked LSRs by addressing the messages directly to peers.

Session Class Messages

- **Initialization:** Once an adjacency is discovered, the LSR peers establish a TCP connection. Initialization messages are then used to exchange session parameters (e.g., retention mode or label distribution mode) between the LSRs.

- **KeepAlive:** KeepAlive messages facilitate the detection of network errors. LSRs periodically transmit these messages to indicate that a link is still working. An error condition is assumed to have occurred when an LSR does not receive a message from a peer within an allotted timeout period; this results in the termination of the established session and the removal of associated labels.

Advertise Class Messages

- **Address:** Address messages provide neighboring LSRs with mapping information about LSR IDs to interface IP addresses. This information is used to identify the label mappings that correspond to the least cost path.

- **Address Withdraw:** Address Withdraw messages notify neighboring LSRs of disabled interfaces or broken links. Receipt of this message causes an LSR to remove the withdrawn address from its LIB mapping.

- **Label Mapping:** Label Mapping messages are used to distribute FEC-to-label bindings from a downstream LSR to an upstream peer. This message is the primary mechanism for constructing LSPs.

- **Label Withdraw:** Label Withdraw messages are used to notify peers that a particular FEC-to-label mapping is no longer valid (e.g., an egress interface goes offline or the network topology changes). When an LSR receives this message, it removes the label from its LIB and sends subsequent Label Withdraw messages to upstream peers.

- **Label Release:** Label Release messages notify downstream peers that an LSR has removed a particular label mapping. An LSR may remove bindings, for example, when an IP table changes or a Label Withdraw message is received.

Notification Class Messages

- **Notification:** Notification messages convey errors and advisories among peer LSRs. If the message indicates a fatal error, the sending and receiving LSRs terminate the LDP session and remove all associated label bindings.

5. LDP Vulnerabilities

In general, attacks may exploit weaknesses in: (i) the LDP specification; (ii) service provider implementations; and (iii) underlying infrastructure. Attacks on the LDP specification leverage inherent weaknesses in the design of the protocol. Any network that conforms with the protocol standard is susceptible to this class of attacks.

Attacks on service provider implementations exploit configuration errors or code flaws. LDP includes several undefined and reserved fields that can be exploited in attacks [1]. LDP also uses a nested structure of Type-Length-Value fields, which offers numerous opportunities for buffer overflow attacks. Our analysis does not focus on implementation vulnerabilities; nevertheless, we note that all implementations should undergo extensive fuzz testing.

Attacks on the underlying infrastructure exploit vulnerabilities in information technology and network assets or weak security policies. For example, LDP relies on IP to provide session communication and routing information. An attack on the underlying IP protocols may be used to reroute a target LSP or hijack a session. Because these attacks do not explicitly exploit LDP messages, they are not considered in this paper.

Our analysis focuses primarily on how an attacker can use LDP messages to exploit MPLS networks. Given only link access, we discuss several vulnerabilities in the LDP specification that could enable an attacker to deny service to various network assets or to reroute traffic.

5.1 Denial-of-Service Attacks

Denial-of-service (DoS) attacks target network resources or capabilities in order to degrade performance or prevent a provider from delivering services to its customers. Our analysis has uncovered six DoS attacks.

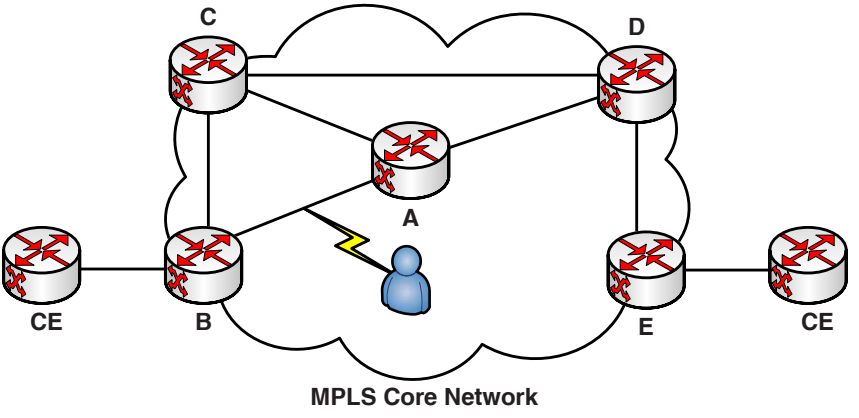


Figure 4. A portion of an MPLS network under attack.

- **Fabricating Notification Messages:** Fabricated Notification messages can be used to target network links. The attack requires read and write access to the target link. In Figure 4, an attacker with access to Link AB fabricates a fatal Notification message from LSR B to LSR A. In response, LSR A and LSR B close the LDP session and remove labels received from the peer. In turn, each router sends Label Withdraw messages to its upstream neighbors to reflect the removed label bindings. Additionally, an attacker with read access to Links AC and AD can intercept TCP sequence numbers and send Notification messages targeting these links via Link AB, which result in the isolation of LSR A.

- **Blocking KeepAlive Messages:** This attack disables a target link. The attacker selectively blocks LDP KeepAlive messages on the target link, which causes the LSRs at either end to terminate the LDP session. The LSRs then remove all the labels associated with the target link as well as the labels from their upstream peers.

- **Fabricating Address Withdraw Messages:** This attack targets three LSRs within an LSP. In Figure 4, an attacker with access to Link AB targets LSPs containing BAC or BAD. To attack BAD, the attacker fabricates an Address Withdraw message from LSR A to LSR B, which withdraws the address associated with the interface for LSR D. LSR B now believes LSR D cannot be reached via LSR A. Subsequently, LSR B tears down any LSPs containing BAD and constructs replacement LSPs.

- **Fabricating Label Withdraw Messages:** This attack targets a specific LSP and requires access to a link along the target path. If the network employs label merging, then the attack also affects all upstream portions of paths merged with the target LSP. Suppose LSR B in Figure 4 binds Label L1 to the LSP EDAB and distributes the binding to LSR A. To tear down EDAB, the attacker fabricates a Label Withdraw message

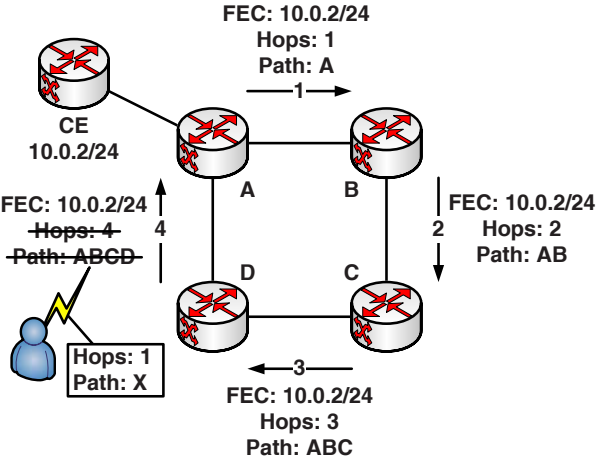


Figure 5. Avoiding loop detection mechanisms.

for L1 from LSR B to LSR A. This causes LSR A to remove L1 from its LIB, delete the label binding for EDAB and send a Label Withdraw message to LSR D. Similarly, LSR D sends a Label Withdraw to LSR E, which completes the destruction of the target path.

- **Exhausting Label Memory:** This attack targets an LSR and requires access to an adjacent link. The attacker floods the target with Label Mapping messages containing random FECs and labels. If the target LSR is configured for the liberal retention mode, it maintains all mappings in its LIB until the memory is exhausted [1]. The target LSR must drop older mappings to replace them with incoming mappings or must refuse all new mappings. In either case, legitimate paths are affected.

- **Creating Loops:** The goal of this attack is to degrade performance within a portion of the network by constructing an LSP loop. The attacker (Figure 5) listens on Link AD for Label Request or Label Mapping messages from LSR D to LSR A. The path vector is modified to reflect one LSR that is not contained within the loop (say LSR X). Any LSR along the path that supports label merging will combine the label request with the existing LSP to create an infinite loop. If no LSR supports label merging, the request completes a full loop, requiring the attacker to perform the modification again. In the absence of label merging, the process continues until the maximum 255 hops exhaust the TTL allowed by MPLS; thus, an infinite loop cannot be created.

5.2 Route Modification Attacks

Route modification attacks change the path of targeted traffic. These attacks enable an attacker to gain access to certain traffic (e.g., maneuver traffic

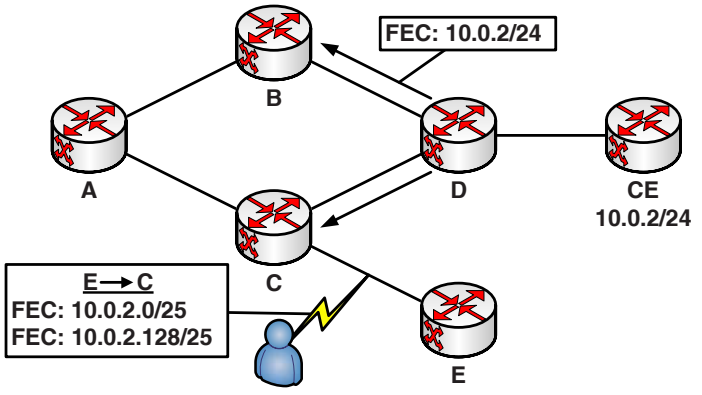


Figure 6. Route modification by creating more specific FECs.

through a compromised link); affect accounting (e.g., trigger automatic financial transactions among cooperating providers); or route traffic across domains (e.g., send one customer's traffic to another customer's network). Our analysis has revealed four route modification attacks.

- **Exploiting FEC Specificity:** This attack takes advantage of the "most specific" or "longest match" rule applied by ingress routers to incoming IP packets. An attacker needs access to a link or a connection to an interface to establish an LDP session. The attacker identifies a target FEC and advertises label bindings for more specific FECs. LSRs that receive the label mappings distribute them throughout the network, thereby building new LSPs toward the compromised link.

  For example, in Figure 6 the attacker targets FEC 10.0.2/24 by distributing mappings for 10.0.2.0/25 and 10.0.2.128/25. When ingress LER A sees a packet for 10.0.2.1, it selects FEC 10.0.2.0/25 and forwards the traffic to the compromised link. The attacker may now read, modify and/or forward this packet to its original destination. The attacker may also be very specific by sending label bindings for a single host such as FEC 10.0.2.1.

- **Fabricating Label Mapping Messages:** This attack reroutes traffic or creates loops by modifying the labels in Label Mapping messages. The attacker needs knowledge of downstream labels, which can be obtained by listening on the compromised link. The attacker may either modify a message in transit or fabricate a Label Mapping message. The message is sent to the upstream router causing it to adjust its LIB. When the upstream LSR receives a packet for the target FEC, it applies the incorrect label, which causes the downstream router to mistakenly recognize the packet as belonging to a different FEC. The packet is then forwarded along the desired LSP. A nefarious attacker can exploit this vulnerability to forward all targeted traffic to a different domain.

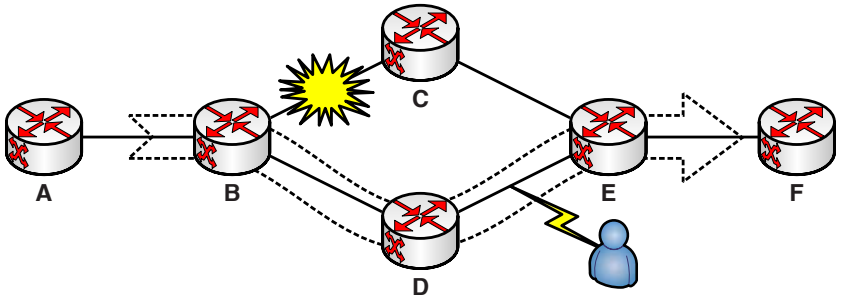


Figure 7. DoS-based route modification attacks.

- **Fabricating Address Messages:** This attack reroutes traffic or creates loops by manipulating the "least cost" mechanism used to select the next hop. Traffic can be redirected using access to a compromised link adjacent to an LSR along a selected LSP (the attacker does not require access to the link carrying the targeted traffic). The attacker crafts an LDP Address message that spoofs the address of the IP next hop. The fabricated message causes the LSR to adjust its LIB and generate a Label Request message. Thus, a new LSP is constructed that forces the targeted traffic along the compromised link.

- **Strategic Placement of DoS Attacks:** As shown in Figure 7, an attacker may execute DoS attacks that force the network to reroute traffic. These attacks change traffic flow within an MPLS network; however, they lack the varying degrees of granularity provided by the other route modification attacks. Nevertheless, the attacks are quite effective and their strategic placement can disable large portions of the network and force traffic through desired paths.

6. Mitigation Strategies

As in the case of traditional networks, most security mechanisms are applied at the perimeter of MPLS networks. However, many of the attacks discussed above occur from within administrative domains. Therefore, it is essential to apply security mechanisms that protect the internal operations of MPLS networks.

Many vulnerabilities in LDP stem from the lack of authentication, integrity and confidentiality mechanisms. LDP messages are sent in the clear, which enables an attacker to gather valuable network information, identify important targets and perform insidious attacks. Without integrity or authentication checks, LSRs are unable to discern the source of a message or verify that a message has not been modified or replayed.

Adequate authentication and integrity mechanisms would mitigate the majority of attacks discussed above. However, implementing these mechanisms

requires significant effort and overhead for key management. According to RFC 3562 [11], keys should be changed at least every 90 days. Additionally, the Internet Engineering Task Force (IETF) suggests strict guidelines for key distribution. Unfortunately, a manageable implementation scheme has yet to be demonstrated. Similar problems surface when using pre-shared keys to encrypt traffic for protecting messaging confidentiality.

In addition to authentication, integrity and confidentiality, simple filtering techniques can be applied to protect LDP from exploitation. For example, an LSR should not accept a Link Hello (used in direct peer discovery) unless the packet is addressed to the link multicast address and the source address is on the same subnet [1]. Without this restriction, it may be possible for an attacker to create LDP adjacencies by addressing Link Hellos directly to a target LSR. To prevent the abuse of Extended Hellos (used in extended peer discovery), each LSR should be configured with an access control list that specifies authorized remote peers. Extended Hello messages should also be filtered at the ingress; unless the source and destination addresses identify an authorized external LDP adjacency, the message should be discarded.

To mitigate memory exhaustion attacks, LSRs should favor existing label bindings over new label bindings. LSRs in the liberal retention mode are susceptible to memory exhaustion because they maintain all label bindings advertised by their peers. LSRs in the conservative retention mode, however, are not susceptible because they release bindings that do not correspond to the IP next hop. Unfortunately, configuring all LSRs for the conservative mode comes at the cost of increased time required to recover from network failures. Alternatively, LSRs in the liberal retention mode should not discard bindings corresponding to the IP next hop when limited by memory constraints. LSRs may also prioritize label bindings based on recent use to protect the most common alternate routes.

7. Conclusions

MPLS has emerged as a mainstay for transporting large volumes of traffic over a wide array of networks. Indeed, much of the world's enterprise traffic already depends on MPLS-based infrastructures to deliver reliable voice, video and application services. A persistent attack on the MPLS infrastructure could cripple corporate, national and even global operations.

LDP, a critical component for discovering and constructing MPLS routes, is vulnerable to several types of attacks. An attacker with internal link access can disable portions of a network or modify traffic flow. Therefore, mitigation strategies should focus on internal operations as well as external operations. We hope that this work prompts a more thorough analysis of security for LDP and related MPLS protocols.

References

[1] L. Anderson, P. Doolan, N. Feldman, A. Fredette and B. Thomas, LDP Specification, RFC 3036, 2001.

[2] AT&T, AT&T wins Frost & Sullivan 2009 North American Market Leadership Award in MPLS/IP VPN services, Dallas, Texas (att.centralcast.net /rss/feeds.aspx), July 20, 2009.

[3] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell and J. McManus, Requirements for Traffic Engineering over MPLS, RFC 2702, 1999.

[4] BT, Delivering the future: BT's 21 Century Network, London, United Kingdom (www.btplc.com/21CN/index.htm).

[5] Congressional Budget Office, Issues Associated with the Global Information Grid Bandwidth Expansion, Washington, DC (www.cbo.gov/doc .cfm?index=6132&type=0), February 28, 2005.

[6] B. Davie and Y. Rekhter, *MPLS: Technology and Applications*, Morgan Kaufmann, San Francisco, California, 2000.

[7] L. Ghein, *MPLS Fundamentals*, Cisco Press, Indianapolis, Indiana, 2007.

[8] E. Gray, *MPLS: Implementing the Technology*, Addison-Wesley, Reading, Massachusetts, 2001.

[9] D. Grayson, D. Guernsey, J. Butts, M. Spainhower and S. Shenoi, Analysis of security threats to MPLS virtual private networks, *International Journal of Critical Infrastructure Protection*, vol. 2(4), pp. 146–153, 2009.

[10] Joint Interoperability Test Command, JITC DISN OT&E Support, Fort Huachuca, Arizona (jitc.fhu.disa.mil/ot&e/gigbe.htm), June 26, 2002.

[11] M. Leech, Key Management Considerations for the TCP MD5 Signature Option, RFC 3562, 2003.

[12] L. Peterson and B. Davie, *Computer Networks: A Systems Approach*, Morgan Kaufmann, San Francisco, California, 2007.

[13] B. Reed, What's next for MPLS? *Network World*, December 21, 2009.

[14] E. Rosen, A. Viswanathan and R. Callon, Multiprotocol Label Switching Architecture, RFC 3031, 2001.

[15] M. Spainhower, J. Butts, D. Guernsey and S. Shenoi, Security analysis of RSVP-TE signaling in MPLS networks, *International Journal of Critical Infrastructure Protection*, vol. 1, pp. 68–74, 2008.

[16] TelecomWeb, Verizon Business Promises Aggressive 2009 Network Investment, Parsippany, New Jersey (www.telecomweb.com/international/262 500.html), February 10, 2009.

Chapter 10

U.S. FEDERAL OVERSIGHT OF RAIL TRANSPORTATION OF TOXIC BY INHALATION MATERIALS

Mark Hartong, Rajni Goel and Duminda Wijesekera

Abstract The 9/11 Commission created as a consequence of the terrorist attacks on New York City and Washington had two goals. The first goal was to study the incidents to determine the specific security failures; the second was to provide recommendations for preventing future incidents. In August 2007, President Bush signed U.S. Public Law 110-53 that implemented the 9/11 Commission recommendations. Section 1551 of the law requires every railroad carrier that transports security-sensitive materials in commerce to provide a written analysis of the safety and security risks for every calendar year. This paper discusses the background behind the current regulatory requirements, the nature of the security-sensitive materials involved, the rail industry and its role in the movement of security-sensitive materials, and the new U.S. federal regulatory requirements associated with the shipment of toxic by inhalation (TIH) materials.

Keywords: Rail transportation, toxic by inhalation materials, regulations

1. Introduction

Following the terrorist attacks of September 11, 2001, the U.S. federal government established a bipartisan commission to study the incidents and to report on the lessons learned. The 9/11 Commission, officially known as the National Commission on Terrorist Attacks upon the United States, conducted an almost two-year study into the circumstances surrounding the events of September 11, 2001 and the associated security failures, and made several recommendations for preventing similar attacks [13]. One of the recommendations dealt with the protection of critical infrastructures — security systems should be integrated into a larger network of screening points that includes the transportation system and access to vital facilities. Based on the commission's

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 141–155, 2010.

recommendations, the U.S. Congress enacted Public Law (PL) 110-53: Implementing Recommendations of the 9/11 Commission Act of 2007 [18]. This law established statutory requirements for the improvement of all facets of transportation system security, in general, and rail transportation, in particular.

This paper provides the background of the current regulatory requirements related to the transportation of security-sensitive materials. It also examines the railroad industry and its role in shipping security-sensitive materials, and the new regulatory requirements associated with the shipment of toxic by inhalation (TIH) materials, in particular.

2. U.S. Rail System and TIH Materials

The rail system is a critical component of the U.S. economy. A total of 563 freight railroads operate on approximately 171,000 miles of track [6], hauling more than 1.85 trillion ton-miles of freight [8] – roughly 40% of all inter-city freight volume. The cargo carried is diverse and supports all facets of the U.S. industrial base. Between 1.7 to 1.8 million carloads comprise hazardous materials [3]. A small percentage (0.3%) of the cargo includes toxic by inhalation (TIH) or poison by inhalation (PIH) materials. However, because of their potential for use in weapons of mass destruction, PL 110-53 specifically mandates the protection of these materials.

TIH materials are defined and regulated by the U.S. Department of Transportation (DOT) under Section 5103 of the Federal Hazardous Materials Transportation Law (49 U.S.C. §5103). These materials include gases or liquids that are known or presumed to be toxic to humans and pose significant health hazards in the event of release during transportation.

The primary DOT hazardous material regulations are issued by the Pipeline and Hazardous Materials Safety Administration (PHMSA) and govern the transportation of hazardous materials by all modes (road, rail, sea and air). The generic transportation regulations address hazardous materials classification, packaging, hazard communication and emergency response. Regulations specific to carriage by rail are in Title 49 of the Code of Federal Regulations (CFR) Parts 172–174 and 179. Part 172 defines the hazardous material classes. Part 173 address the general packaging and shipping requirements for hazardous and TIH materials. Part 174 addresses the minimum specific requirements for loading, placards and special handling requirements for Class 1 (explosive), Class 2 (gaseous), Class 3 (flammable), Class 6.1 (poisonous) and Class 7 (radioactive) materials moved by rail. Part 179 addresses the regulatory weight, marking and design and manufacturing requirements for tank cars.

While the federal government and the rail industry are concerned with the safe and secure shipment of all hazardous materials, the safety and security of certain shipments of explosive (Class 1), toxic by inhalation (Class 6.1) and radioactive materials (Class 7) are of special concern because of their potential for use in weapons of mass destruction and their extreme impact on the human body. TIH materials of concern are categorized according to their biological effects: nerve agents, blister agents, choking agents and blood agents [9].

Nerve agents are man-made chemicals, mostly organophosphates that are used in insecticides. These chemicals affect the nervous system, causing the over-stimulation of muscles. Victims typically suffer from nausea and weakness, and possibly convulsions and spasms. At high enough concentrations, loss of muscle control and nervous system irregularities result in death.

Blister agents or vesicants cause the blistering of tissues. They can enter the body through the lungs or by contact with the skin or eyes. Vaporized blister agents are extremely dangerous even in low concentrations. Victims may have symptoms ranging from mild bronchitis to the blistering of the lungs.

Choking agents act on the lungs, causing breathing difficulty and potentially permanent lung damage. Examples include chlorine, ammonia and phosgene. Exposure to low concentrations causes chest discomfort, shortness of breath and irritation of the nose and throat. High concentrations quickly result in the swelling of the lungs, respiratory failure and death.

Blood agents interfere with oxygen utilization at the cellular level, potentially causing death through oxygen starvation of brain cells. Examples include hydrogen cyanide and cyanide salts used in the chemical, electroplating and mining industries. Exposure to very high concentrations of blood agents leads to violent convulsions and cardiac failure within a few minutes.

Two incidents demonstrate the adverse consequences of the loss of containment of TIH materials during their transportation by rail. The first incident was the January 18, 2002 derailment of a Canadian Pacific freight train in Minot, North Dakota. The derailment and subsequent loss of tank car integrity resulted in the release of anhydrous ammonia that killed one person, injured 333 others and required the evacuation of 11,600 inhabitants for more than one week [14].

The second incident was the January 6, 2005 collision of Norfolk Southern freight trains in Graniteville, South Carolina [15]. In the ensuing derailment, the loss of tank car integrity resulted in the release of chlorine gas that killed nine people and injured 554 others. The gas release rendered the town of Graniteville uninhabitable for two weeks, necessitating the evacuation of 5,400 people. The total damage as a result of the incident exceeded $40 million.

While the consequences of the accidents were severe, they were mitigated by the fact that neither of the accidents occurred in a highly-populated area. Worst-case scenarios evaluated by the Naval Research Laboratory [12] indicate that the release of chlorine gas from a 90-ton car in the center of Washington, DC could kill or injure 100,000 people and render large portions of the city uninhabitable for an extended period of time.

Although TIH materials constitute only 0.3% of all hazardous material shipments by rail, this still equates to more than 21.6 million ton-miles of TIH material movement each year [10]. Consequently, railroads are a critical and sensitive component of the U.S. infrastructure, and they are strictly regulated. While the consequences of a TIH material release can be catastrophic due to the volume of material carried in a freight car, it must be noted that such incidents are very rare. Rail transportation is by far the safest way of ship-

Table 1. TIH shipments (source: U.S. Census Bureau).

Year	Tons (thousands)	Ton-Miles (millions)	Length of Haul (miles)
1997	8,868	6,736	764
2002	6,090	3,226	549
2007	4,005	2,551	580

ping TIH materials. In 2007, 99.996% of hazardous material shipments by rail reached their destination without a release caused by a train accident [5]. The railroads and trucking industries carry roughly the same amount of ton-mileage of hazardous materials, but the trucking industry has sixteen times the amount of hazardous material release of railroads [2].

The Commodity Flow Survey (CFS) conducted by the U.S. Census Bureau is the primary source of national and state-level data on domestic freight shipments in the mining, manufacturing, wholesale, auxiliary and selected retail industries. CFS is a shipper-based survey that is conducted every five years as part of the Economic Census. It provides data on the types of commodities, their origins and destinations, value, weight, modes of transport, distance and ton-miles shipped; and presents a modal picture of national freight flow.

The CFS was conducted in 1997, 2002 and, most recently, in 2007 [8]. Table 1 presents the volumes of national TIH material shipments for these years. Note that the volumes moved have decreased since 1997 as a result of product substitution; the distance hauled has decreased due to greater co-location of suppliers and consumers.

While the DOT maintains records for individual shipments of commercially-transported commodities, these records are deemed proprietary by the individual firms and, consequently, the information is not available to the public. Federal data involving rail operations is suppressed at all levels apart from the national level. The commercial TRANSEARCH database provides estimates for smaller geographic units, but supporting information about the flows is proprietary and is not available to the public.

3. Statutory Obligations and Regulations

U.S. railroads have a statutory common carrier obligation under 49 U.S.C. §11101 to provide transportation for commodities that are not exempted from regulations pursuant to 49 U.S.C. §10502. This obligation creates two interrelated requirements: (i) railroads must provide, in writing, common carrier rates to any person requesting them (49 U.S.C. §11101(b)); and (ii) railroads must provide rail service pursuant to the common carrier rates upon reasonable request (49 U.S.C. §11101(a)).

These statutory requirements place the railroads in a difficult position as they are exposed, by law, to the risk of catastrophic liability when transporting

TIH materials. Railroad companies cannot decline to transport hazardous materials merely because it is inconvenient or unprofitable to do so; nor can they refuse to transport a commodity based on its dangerous characteristics. Unlike accidents involving nuclear materials, for which the Price-Anderson Act limits liability, accidents involving TIH materials have no liability limits. However, recent federal court decisions (e.g., [17]) have found that the Federal Rail Safety Act preempts individual state tort law, which may serve to limit railroad liability from punitive damages in cases where railroad companies are in compliance with federal law.

The railroads, of course, purchase insurance to mitigate the financial risk of carrying hazardous materials, but this coverage is both expensive and limited in availability. According to the Association of American Railroads (AAR), highly hazardous commodities constitute only 0.3% of the total carload, but account for 50% of the insurance costs of railroad companies. Due to the expense and lack of coverage, most railroads can ensure only a fraction of their net worth. A single hazardous materials accident can bankrupt a small carrier. The situation is further complicated by the fact that insurance coverage is regulated by state law instead of federal law, and that state insurance statutes override most federal laws (see the McCarran-Ferguson Act (15 U.S.C. §1011)).

Legal ramifications aside, in order to mitigate the risk of catastrophic liability, AAR, DOT and the U.S. Department of Homeland Security (DHS) have instituted strict protocols for the movement of TIH materials that are intended to minimize hazards. The AAR protocols are included in the United States Hazardous Materials Instructions for Rail [7], OT-55 [4] and Casualty Prevention Circular 1187 (CPC-1187) [1]. The DOT and DHS security protocols are specified in the Hazardous Materials Regulations (49 CFR §171; Parts 105–180); Rail Safety Act (49 CFR Parts 200–244); and Rail Transportation Security Regulations (49 CFR Part 1580).

3.1 Rail Industry Voluntary Requirements

In addition to railroad-specific security plans that provide for variations in the actual movement of hazardous materials corresponding to the different DHS security threat levels, the railroad industry has developed several handling and routing requirements [1, 4]. These requirements specify the list of hazardous and TIH materials, the main technical and handling requirements for trains moving TIH materials, the main rail routes over which TIH materials are moved, along with railroad operating practices and facilities when TIH or other hazardous materials are being transported or stored en route. The requirements also include the type of tracks over which TIH materials may be hauled, the maximum train operating speeds when hauling TIH materials, the positioning of TIH cars in train consists, the placement of placards identifying the TIH materials being transported, and the movement and storage requirements of TIH cars in marshalling yards and customer facilities.

Under OT-55, AAR member railroads are responsible for tracking the locations of hazardous and TIH material shipments from shipper to consignee, and

for ensuring the timely delivery of the materials in accordance with DOT guidelines. OT-55 also establishes mechanisms for the railroads to provide, upon request by public safety officials in a jurisdiction, the list of the top 25 hazardous materials transported through the jurisdiction. The railroad industry considers this information to be restricted information of a security-sensitive nature and that the recipient of the information must agree to release the information only to *bona fide* emergency response planning and response organizations and not to distribute the information publicly in whole or in part without the express written permission of the individual railroads.

The reporting mechanism used in OT-55 to keep local authorities appraised of the nature of the shipments is called TRANSCAER. TRANSCAER (short for Transportation Community Awareness and Emergency Response) is an outreach program initiated by the railroads and shippers. The program provides assistance to emergency response and planning groups in assessing local risks based on the hazardous materials being shipped through their areas of responsibility and in developing response plans in the event of material release. OT-55 also requires that railroads and shippers develop emergency response plans that allow railroads to report the release of materials. The program, known as CHEMTREC, allows a railroad to initiate the shipper's emergency response capability in the event of a derailment, tank shell damage or product release.

The OT-55 requirements, while very successful in mitigating the unplanned release of TIH materials, have a significant shortcoming. Unlike the regulatory requirements issued by DOT or DHS, the OT-55 requirements are merely recommended practices. As such, they are not strictly enforceable should a railroad elect not to comply. As a practical matter, however, the railroads are self-policing, where railroads that do not comply with the recommended practices are embargoed by other railroads that comply with the practices.

CPC-1187 implements industry standards for the shell, head and top fittings of TIH tank cars based on the conditional probability of release (i.e., the probability of release in the event of an accident). CPC-1187 requires tank cars used to transport TIH materials to be equipped with top fittings protection systems designed to withstand, without loss of lading, a rollover with a linear velocity of 9 mph, and the top fittings protection systems to be attached to the tanks by welding. As currently written, tank cars designed and built to the CPC-1187 specifications suffer from a significant drawback. CPC-1187 requirements can be met by using DOT specification tank cars of higher tank classes than required by minimum DOT standards. However, tank cars built to the CPC-1187 standard do not meet the existing minimum DOT standards.

3.2 Federal Safety and Security Regulations

Significantly greater adverse consequences associated with TIH materials are indicated by their special labeling requirements, both those adopted voluntarily by the railroad industry as well as those required by the government. DOT regulations require the words "Poison Inhalation Hazard" to be entered on TIH material shipping papers. Tank cars transporting TIH materials require special

placards (in addition to normal hazardous material placards) that indicate "Poison Inhalation Hazard" or "Poison Gas" (49 CFR 172.504).

These requirements are further enhanced by recent federal regulatory efforts. The new requirements are codified in changes to 49 CFR Parts 172 and 174 – Hazardous Materials: Enhancing Rail Transportation Safety and Security for Hazardous Materials Shipments; 49 CFR Part 209 – Railroad Safety Enforcement Procedures, Enforcement, Appeal and Hearing Procedures for Rail Routing Decisions; 49 CFR Part 236, Subpart I – Positive Train Control; and 49 CFR Part 1580 – Rail Transportation Security Regulations. The Federal Railroad Administration (FRA), PHMSA and Transportation Security Administration (TSA) have developed these new regulations in concert.

The new requirements of 49 CFR 172 and 174 satisfy the requirements in Section 1551 of the Implementing Recommendations of the 9/11 Commission Act of 2007. Section 1551 requires a rail carrier of security-sensitive materials to select the safest and most secure routes when transporting the materials, based on the rail carrier's analysis of the safety and security risks on primary and alternate transportation routes over which the carrier has authority to operate. In summary, these new regulations establish risk-based protocols for evaluating the safety and security of TIH material shipments. All rail carriers are now required to:

- Compile annual data on shipments of explosive, TIH and radioactive materials.
- Use the data to analyze safety and security risks along rail routes where the materials are transported.
- Assess alternative routing options.
- Make routing decisions based on the assessments.

These new regulations also require rail carrier security plans to address en route storage and in transit delays. Also, rail carriers must inspect placarded hazardous material rail cars for signs of tampering and the presence of suspicious items.

Railroads are required to compile annual data on shipments by route. This could be a line segment or series of line segments. The railroads can choose to define what constitutes a line segment and how to aggregate the line segments into a route. However, railroads must translate the routes into geographical locations and identify the materials shipped by their UN identification numbers [16]. The four-digit UN identification numbers are used in international commerce and transportation to identify hazardous chemicals or classes of hazardous materials. The numbers generally range between 0000 and 3500 and are preceded by the letters "UN" (e.g., "UN1005") to avoid confusion with other number codes.

The route analysis requires that railroads identify all practical alternatives and involve state, local and tribal officials in identifying security risks along

Table 2. Rail risk routing factors [19, 20, 22].

Rail Risk Routing Factor	**Risk Reduction Strategy**
Volume of hazardous materials	Minimize volume
Rail traffic density	Minimize density
Trip length	Minimize trip length
Railroad facilities	Maximize availability
Track type and class	Maximize type and class
Track grade and curvature	Minimize grade and curvature
Signal and train control systems	Maximize presence
Wayside detectors	Maximize number
Number and type of grade crossings	Minimize number
Single vs. double track	Maximize double tracks
Frequency and locations of of track turnouts	Minimize number
Proximity to iconic targets	Minimize proximity
Proximity to environmentally sensitive areas	Minimize proximity
Population density	Minimize population
Venues of route	Minimize proximity
Emergency response capability along route	Maximize response
Areas of high consequence	Minimize high consequence areas
Passenger traffic	Minimize volume
Speed of train operations	Minimize speed
Proximity to en route storage and repair facilities	Maximize proximity
Known threats	Minimize threats
Security measures in place	Maximize security
Availability of alternative routes	Maximize alternatives
Past incidents	Minimize incidents
Overall time in transit	Minimize time
Crew training and skill level	Maximize skill and training
Impact on rail network traffic and operations	Minimize impact

proposed routes. Route alternatives must be prepared in writing and must consider all the safety and security risks associated with the critical factors listed in Table 2. Also, they should always consider the possibility of catastrophic release of the shipment. The analysis must also identify remediation or mitigation acts that can be adopted. The route identifying procedure requires a railroad to consider if interchanging the TIH shipment with another railroad will result in an overall lower societal risk and costs, regardless of the financial gain or loss to the railroads. The analysis and the supporting information are

considered to constitute sensitive security information (SSI) and their release is restricted to persons with a need to know. Generally, this means federal, state, local and tribal officials responsible for transportation safety and security, not the general public.

While interchange must be considered, it is not mandated. In order to encourage interchange, the regulations provide an exemption from anti-trust regulations (49 U.S.C. §333) so that railroads and shippers can share cost and route information to facilitate the system-wide optimization of safety and security. Normally, the exchange of such information is deemed to be "anti-competitive," but immunity from prosecution is granted if the discussions are moderated by the FRA and the agreements are approved by the FRA.

The changes to 49 CFR Parts 171–174 and 179 also establish new structural requirements for tank cars, especially those handling TIH materials. In 2004, the National Transportation Safety Board (NTSB) found that more than one-half of the 60,000 rail tank cars used to transport hazardous materials were not built according to current standards and were susceptible to rupture in the event of an accident [24]. The NTSB also reported that the 1989 requirement for tougher steel has made all new tank cars safer, but about 60% of pressurized tank cars currently in use were built before 1989.

Issued pending validation and implementation of new crashworthy designs, the NTSB requirement imposes interim technical rules for tank car design and operation to protect against the release of TIH materials in the event of a collision or derailment. The required technical modifications to a particular tank car are based on the specific TIH materials being shipped. The mandatory functional requirements for tank cars require: (i) blunting the load impacting the tank to prevent tank puncture; (ii) absorbing the kinetic energy associated with a crash without loss of containment; (iii) reinforcing the commodity tank; and (iv) removing in-train forces from the commodity tank. Changes to improve the top fittings performance (where the material is loaded into tank cars) were made along with the steel used in the shells of tank cars. In addition to the functional performance requirements, the new regulation established a 50 mph speed restriction for loaded rail tank cars transporting TIH materials. This codified the speed restrictions established by AAR in OT-55.

The new requirements of 49 CFR Part 209 are more administrative in nature, establishing procedures to enable railroad carriers to challenge rail routing decisions made by the FRA that carry out the new requirements of 49 CFR Part 172 discussed above. The procedures in Part 209 require the FRA to provide written notification if a railroad carrier's route selection, analysis and documentation are deficient and the carrier fails to establish that the route chosen poses the least overall safety and security risks. Once a railroad has been notified, the FRA works with the railroad, the Surface Transportation Board (STB), PHMSA and DHS to address the issues identified by the FRA. After this process is completed, if the railroad still does not address the issues, then the FRA transmits a final written order identifying the unresolved issues and orders the use of a route that the FRA determines to be the safest. The

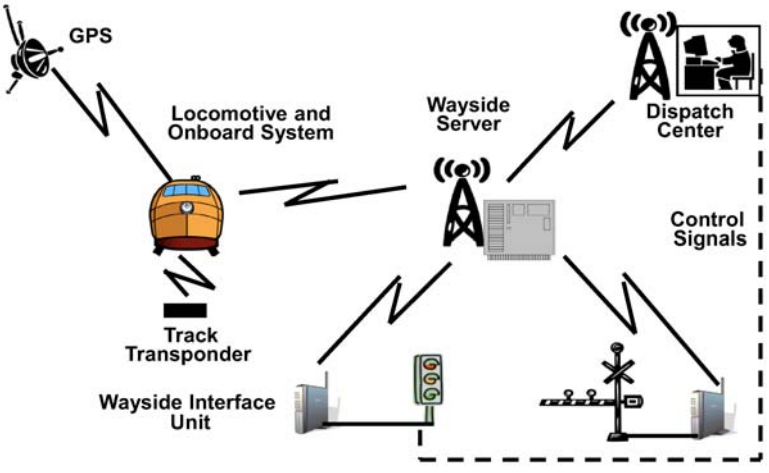


Figure 1. Positive train control system.

railroad may petition for review of the final decision in the appropriate United States Court of Appeals, but compliance with the FRA order is not stayed unless ordered by the appellate court.

The requirements of 49 CFR Part 236, Subpart I [21] implement the Rail Safety Improvement Act (RSIA) of 2008. Among the many provisions of the RSIA is the requirement for Class 1 railroads to install positive train control (PTC) systems on their route segments that transport more than 5 million gross tons annually and carry TIH materials. These supervisory control and data acquisition (SCADA) systems communicate using wireless links and are utilized by railroads to provide positive train separation, over-speed protection and protection for roadway workers working within the limits of their authority [11]. As illustrated in Figure 1, a PTC system consists of four subsystems: office system, wayside system, onboard system and communications network.

In the process of ensuring positive train separation and preventing derailment, PTC safety mechanisms provide some degree of protection against the release of TIH materials due to collision or derailment. When installed, PTC systems will cover approximately 70,000 miles of the US rail system. The new implementation regulations of 49 CFR Part 236, Subpart I for PTC SCADA systems recognize the vulnerability of the systems to wireless attacks, and require the systems to incorporate cryptographically-based message integrity and non-repudiation mechanisms to prevent misuse.

The last significant set of regulations associated with securing TIH materials is found in the TSA Rail Transportation Security Regulations of 49 CFR Part 1580. Published in November 2008, the TSA regulations require that bulk shipments of TIH materials (along with certain explosive and highly radioactive materials) be handled through a continuous chain of custody, including physical delivery to a connecting railroad at a point of interchange where personnel of the receiving railroad are available to take physical control.

Table 3. High threat urban areas [23].

Phoenix, AZ	San Diego, CA	Miami, FL
Anaheim, CA	Santa Barbara, CA	Denver, CO
Orlando, FL	San Francisco, CA	Washington, DC
Tampa, FL	Long Beach, CA	Los Angeles, CA
Fort Lauderdale, FL	Atlanta, GA	Sacramento, CA
Jacksonville, FL	Honolulu, HI	Chicago, IL
Indianapolis, IN	Baton Rouge, LA	New Orleans, LA
Louisville, KY	Boston, MA	Baltimore, MD
Detroit, MI	Twin Cities, MN	Kansas City, KS
St Louis, MO	Charlotte, NC	Omaha, NE
Newark, NJ	Jersey City, NJ	Las Vegas, NV
Buffalo, NY	New York City, NY	Cincinnati, OH
Cleveland, OH	Columbus, OH	Toledo, OH
Oklahoma City, OK	Portland, OR	Philadelphia, PA
Pittsburgh, PA	Memphis, TN	Dallas, TX
Fort Worth, TX	Houston, TX	San Antonio, TX
Seattle, WA	Milwaukee, WI	

These regulations improve security in several ways. First, the regulations require rail carriers and facilities that handle specified hazardous materials to report location and shipping information to the TSA upon request. The reporting criteria are very strict. Class I freight railroad carriers must provide the location and content information to the TSA no later than five minutes (for one car) or 30 minutes (for two or more cars) after receiving the request. To facilitate this, each railroad must identify a Rail Security Coordinator (RSC) who is available at all times to serve as the primary liaison with the TSA on security matters.

Second, the regulations require railroads and shippers to ensure a chain of custody when exchanging extremely high-risk hazardous materials (e.g., explosive, TIH and radioactive materials) when they pass through a high threat urban area (HTUA) (Table 3). Chain of custody is relatively straightforward. The shipment must be under positive control from the time the hazardous material is accepted by the railroad to the time the shipment is delivered. Positive control has three elements: (i) the physical location of a responsible party in close proximity to the car; (ii) the ability to respond promptly to an unauthorized access; and (iii) the ability to contact the appropriate security officials. In conjunction with physical control, a designated responsible party must sign for the materials.

Third, as in the case of rail safety regulations that permit FRA inspectors to conduct announced and unannounced inspections for compliance, the TSA can conduct security inspections. While the carriers would prefer not to have this sort of regulatory oversight, it provides a mechanism for ensuring the implementation of a common level of security throughout the rail system.

The annual TIH material routing analysis fillings by the railroads also require FRA review and approval. In carrying out a review, the FRA can obtain an estimate of the total TIH material tonnage shipped on a particular rail carrier via an analysis of waybill sample data for comparison with the filed routing analysis. The annual rail waybill sample contains shipment data from a stratified sample of confidential rail waybills submitted by freight railroads to the STB in support of rail carrier rate filings. Discrepancies between waybill data and route analysis data, or between route analysis data and observed shipments by government field inspectors may trigger further investigations. The penalty for non-compliance is high. The regulations provide for civil penalties of up to $100,000 per day levied against railroads found to be not in compliance, along with the assignment of individual liability, which results in the assessment of civil penalties to individuals and possible disbarment from employment in the transportation services industry. In extreme cases, criminal felony charges may be filed for non-compliance.

4. Conclusions

The current federal regulations have certain shortcomings that must be acknowledged. First, the regulations leave out other hazardous materials that could also cause considerable damage or that could be used as catalysts to release other toxic materials (e.g., highly volatile liquefied petroleum gas and flammable liquids). Second, the regulations are limited to loaded cars; residue cars containing smaller quantities of hazardous materials are excluded. Because of the way "residual" is defined (i.e., cars that have been unloaded to the maximum extent practicable), a car that has had only half of its contents unloaded could be considered to be a residue car and is, therefore, not subject to the regulations. Third, the regulations only cover a limited number of high threat urban areas – many U.S. cities (e.g., Tulsa, OK) are not classified as high threat urban areas.

The imposition of federal regulations for TIH material security has been, and continues to be, a very divisive topic, with the various stakeholders promoting contradictory agendas. The railroads generally object to the new regulations as being unfunded mandates that are arbitrary and capricious. They point to the fact that the shipment of TIH materials by rail has a proven record of being extremely safe and that there is no credible evidence that the security risks are any higher than the safety risks.

Shippers and other customers are concerned that the railroads will utilize the new regulations to condense TIH material traffic. By gaming the regulations, railroads could eliminate service and/or pass their safety and security costs to customers. Such actions would adversely impact the ability of railroad customers to provide goods and services.

State, local, and tribal entities are concerned that the imposition of federal regulations is preemptive. In preempting state and local laws, the federal government limits the ability of these entities to adequately protect their constituents. These entities believe that regulatory routing and TIH material

handling requirements do not make adequate provisions for state, local and tribal oversight and that the rejection of routes may impose unwanted and unacceptable exposure to their constituents.

All the stakeholders are greatly concerned by the performance-based nature of the regulations. Generally, the requirements specify in broad terms what must be accomplished but are silent on the how. This situation provides for a large solution space. However, because the regulations are not prescriptive, the stakeholders are never entirely sure what the regulators will consider to be acceptable or unacceptable solutions to implementing the requirements.

The three regulatory federal agencies responsible for creating and enforcing security rules for TIH material shipments (FRA, PHMSA and TSA) have been mindful of stakeholder concerns and have worked to make the development of the regulations as transparent as possible. The proposed regulations were made available for public review and comment before their enforcement. After the comments were received, the government regulatory bodies carefully weighed each comment, deciding on a specific course of action and, where appropriate, modifying the proposed rule text based on the comments.

The current regulations governing the movement of TIH materials by rail must deal with significant uncertainties because the associated probability and consequence data are often sparse and of questionable quality. The uncertainties arise because the adverse events have very small probabilities and have rarely, if ever, occurred. Nevertheless, despite regulatory efforts, the release of TIH materials as the result of a train accident or a terrorist incident, while very unlikely, is still possible.

Regardless of the significant level of carrier, labor, vendor, government and public participation in the formulation of the new regulations to address the safety and security of TIH materials, public perception will be the driving motivation for the suitability of these regulations. Anecdotal evidence, despite statistics to the contrary, may result in the creation of additional regulations to address perceived problems. Over regulation of rail shipments could have the unintended effect of forcing TIH material shipments to roadways – a much more risky operating environment. Such policy could be extremely harmful to public health, safety and welfare, and to the economy as a whole.

The views and opinions expressed in this paper are those of the authors and do not necessarily state or reflect the views of the U.S. Government, the U.S. Department of Transportation or the Federal Railroad Administration, and shall not be used for advertising or product endorsement purposes.

References

[1] Association of American Railroads, CPC-1187: Specification for Tank Cars – Manual of Standards and Recommended Practices, Washington, DC, 2008.

[2] Association of American Railroads, Hazmat and the Railroad Industry, Washington, DC, 2009.

[3] Association of American Railroads, Hazmat Transportation by Rail, Washington, DC, 2009.

[4] Association of American Railroads, OT-55: Recommended Railroad Practice for Transportation of Hazardous Materials, Washington, DC, 2009.

[5] Association of American Railroads, Railroads and Chemicals, Washington, DC, 2009.

[6] Association of American Railroads, Railroad Facts 2009, Washington, DC, 2009.

[7] Bureau of Explosives/Association of American Railroads, United States Hazardous Materials Instructions for Rail, Washington, DC, 2008.

[8] Bureau of Transportation Statistics, BTS Special Report: U.S. Freight on the Move – Highlights from the 2007 Commodity Flow Data Survey, Preliminary Data SR 018, Washington, DC, 2009.

[9] Departments of the Army, Navy and the Air Force, and Commandant, Marine Corps, Field Manual: Treatment of Chemical Agent Casualties and Conventional Military Chemical Injuries, Field Manual FM 8-285, U.S. Department of Defense, Washington, DC, 1995.

[10] Federal Highway Administration, Freight Facts and Figures 2009, Washington, DC, 2009.

[11] M. Hartong, R. Goel and D. Wijesekera, Securing positive train control systems, in *Critical Infrastructure Protection*, E. Goetz and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 57–72, 2007.

[12] National Capital Planning Commission, Rail Realignment Feasibility Study Securing Freight Transportation in the National Capital Region, Washington, DC, 2007.

[13] National Commission on Terrorist Attacks upon the United States, *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States*, Government Printing Office, Washington, DC, 2004.

[14] National Transportation Safety Board, Derailment of Canadian Pacific Railway Freight Train 292-16 and Subsequent Release of Anhydrous Ammonia near Minot, North Dakota, January 18, 2002, Railroad Accident Report NTSB/RAR-04/01, Washington, DC, 2004.

[15] National Transportation Safety Board, Collision of Norfolk Southern Freight Train 192 with Standing Norfolk Southern Local Train P22 with Subsequent Hazardous Materials Release at Graniteville, SC, January 6, 2005, Railroad Accident Report NTSB/RAR-05/04, Washington, DC, 2005.

[16] United Nations, U.N. Recommendations on the Transport of Dangerous Goods – Model Regulations, Geneva, Switzerland, 2007.

[17] U.S. District Court (District of North Dakota, Northwestern Division), Mehl v. Canadian Pacific Railway, *Federal Supplement (Second Series)*, vol. 417, pp. 1104–1121, 2006.

[18] U.S. Government, Implementing Recommendations of the 9/11 Commission Act of 2007, Public Law 110–53, *United States Statutes at Large*, vol. 121(1), pp. 266–550, 2007.

[19] U.S. Government, Title 49, Transportation, Part 172, Hazardous Materials Table, Special Provisions, Hazardous Materials Communications, Emergency Response Information, Training Requirements and Security Plans, *Code of Federal Regulations*, Washington, DC, pp. 123–402, 2009.

[20] U.S. Government, Title 49, Transportation, Part 174, Carriage by Rail, *Code of Federal Regulations*, Washington, DC, pp. 680–706, 2009.

[21] U.S. Government, Title 49, Transportation, Parts 179–244, *Code of Federal Regulations*, Washington, DC, pp. 7–891, 2009.

[22] U.S. Government, Title 49, Transportation, Part 209, Railroad Safety Enforcement Procedures, *Code of Federal Regulations*, Washington, DC, pp. 10–63, 2009.

[23] U.S. Government, Title 49, Transportation, Part 1580, Rail Transportation Security, *Code of Federal Regulations*, Washington, DC, pp. 448–461, 2009.

[24] U.S. Senate, Senate Report 108-278: Rail Security Act of 2004, Calendar No. 536, Washington, DC, 2004.

Chapter 11

PROTECTING THE FOOD SUPPLY CHAIN FROM TERRORIST ATTACK

Maria Jesus Alvarez, Ainara Alvarez, Maria Carla De Maggio, Ainhoa Oses, Marcella Trombetta and Roberto Setola

Abstract The food supply chain is a critical infrastructure that is an attractive target for terrorist attacks. Despite its importance, relatively little research has focused on improving the security of the food supply chain infrastructure. This is largely due to a lack of awareness on the part of food supply chain stakeholders and authorities about the threats. This paper describes a methodology for assessing the risk associated with threats to the food supply chain, with the goal of enhancing awareness and helping develop appropriate security measures.

Keywords: Food supply chain, threats, food defense, risk assessment

1. Introduction

The food supply chain is an attractive target for terrorist attacks. In the aftermath of the attacks of September 11, 2001, the World Health Organization (WHO) stressed the risks due to food terrorism. Of particular concern is "an act or threat of deliberate contamination of food for human consumption with biological, chemical or physical agents or radionuclear materials for the purpose of causing injury or death to civilian populations and/or disrupting social, economic or political stability" [8]. The need to protect the food supply chain was also underscored by Resolution WHA 55.16 [24] at the *Fifty-Fifth World Health Assembly*, which stressed that food is a likely and highly effective way to disseminate biological, chemical or radionuclear agents and materials.

The protection of the food supply chain – termed "food defense" – has attracted considerable attention in the United States [15]. Agriculture and food is recognized as one of the seventeen national critical sectors [4, 5] and a specific work plan [20] was released in 2007. Despite the efforts, many instances of *salmonella* and *E. coli* contamination have been reported in the United States. These outbreaks – large and small – mostly led to hospitalization and, in some

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 157–167, 2010.

cases, death. Interestingly, the authorities were unable to determine the causes of the outbreaks in the majority of cases [3].

The U.S. incidents demonstrate that compromises of the food supply chain can have a significant impact on public health. The food infrastructure is massive and highly distributed. As emphasized by the U.K. Centre for the Protection of National Infrastructure (CPNI) [6] and the Asia-Pacific Economic Cooperation (APEC) Counter Terrorism Task Force (CTTF) [2], every country and geographic region is exposed to a wide range of threats.

The European Commission's Green Paper on Bio-Preparedness [9] highlights efforts for reducing biological risks and enhancing preparedness and response with regard to the food supply chain. Nevertheless, few comprehensive initiatives are underway to secure the European food supply chain from attack. One example is the Rapid Alert System on Food and Feed (RASFF) [10], but it focuses on food safety warnings, not on preventing malicious contamination.

The U.K. CPNI and British Standard Institute (BSI) define food defense as "the security of food and drink and their supply chains from all forms of malicious attack including ideologically motivated attacks leading to contamination or supply failure" [6]. As explained in [8], the potential effects of a terrorist attack on the food supply chain are many, the most significant of which are human disease and death. Terrorist acts are also designed to create fear and anxiety in the population and reduce confidence in the government, which can lead to political instability.

Dalziel [7] has conducted a systematic examination of incidents involving the intentional and malicious contamination of food from 1950 to 2008. The analysis reveals that almost 98% of the incidents occurred downstream in the food supply chain (e.g., at retail outlets, food service establishments, homes and the workplace). Typically, the incidents involved commonly-available household, agricultural or industrial chemicals. When more esoteric chemicals were used, the perpetrators often had access to these agents at work and also possessed the knowledge to use them. Incidents involving biological or radiological agents typically occurred at the retailer or at the consumer and had little impact on public health.

Analysis of the data indicates that the most common reason for the deliberate contamination of food was to disrupt business or tourism and cause economic loss rather than injure people. Thus, a distinction should be made between actions aimed at spreading pathogens in large populations and "symbolic" attacks designed to provoke social anxiety and economic loss. Contaminated food products often spread panic in the population. The mad cow disease and avian flu scares modified consumer behavior in a very significant manner, creating negative effects on the market and massive losses for producers.

Symbolic attacks on the food supply are both efficient and effective. These attacks are easy to perpetrate, and can target any aspect of the food supply chain, especially the least controlled and protected portions of the chain. Widespread monitoring of contamination is complicated by food imports. Most countries import significant quantities of food; the figure for the United States

is about 15%. Illegally-imported food poses additional problems because it bypasses government testing.

This paper presents an approach for analyzing the risk to the food supply chain in terms of potential threats, system vulnerabilities and countermeasures. The research, which has been performed under the SecuFood Project [17], has considered a broad sampling of foods consumed in Europe (e.g., milk, yoghurt, juice, bread, oil, salads, fish and baby food). However, this paper specifically examines the major issues related to securing the European milk supply chain.

2. SecuFood Methodology

Ezell and von Winterfeldt [11] have noted that estimating the probabilities of an attack on the food supply chain is a hard task, requiring knowledge about the motivation, intent and capabilities of attackers. In addition, these probabilities change with the defensive measures that are implemented. For these reasons, we focus our attention on food supply chain vulnerabilities with the goal of identifying them in order to implement preventive measures.

To estimate the risk posed by terrorist attacks, and more generally, criminal attacks, we consider the threats posed by the availability of various biological and chemical agents and their potential consequences. This is because any attack on the food supply chain requires the introduction of a dangerous agent. The agent can be added during harvest, storage, processing, preparation, retail or food service.

To conduct a more effective analysis, we decomposed the food supply chain into its main macroscopic steps, taking into account the peculiarities of each step in terms of vulnerabilities and countermeasures. To this end, we assume that a generic food supply chain can be decomposed into the five macroscopic steps shown in Figure 1. A typical workflow starts with a large set of production sites that supply one or more industries. The food is processed, transformed and packaged at these sites, and is then sent, via a logistic system, to wholesalers. The wholesalers distribute the food items to retailers and food service establishments who pass them on to consumers. Note that the decomposition in Figure 1 represents an abstraction; the actual process is very complex and includes numerous sources, processes and exchanges of raw materials between various entities.

We identified specific threats at each macroscopic step for each food type in terms of contamination by chemical and biological agents and by other instruments [18]. Our analysis revealed that the types of threats at the different steps are essentially the same, although the impact and the ability to detect and neutralize the threats can be very different. In fact, the impact of a contaminant is greater when the agent is introduced early in the supply chain. This complicates and delays the localization of the contamination, especially when the adverse effects are not immediate. Also, a contaminant that is introduced in an early step of the food supply chain is difficult to identify and isolate, especially if the problem is discovered after processing and delivery. However, some agents can be detected by quality control testing and neutralized during

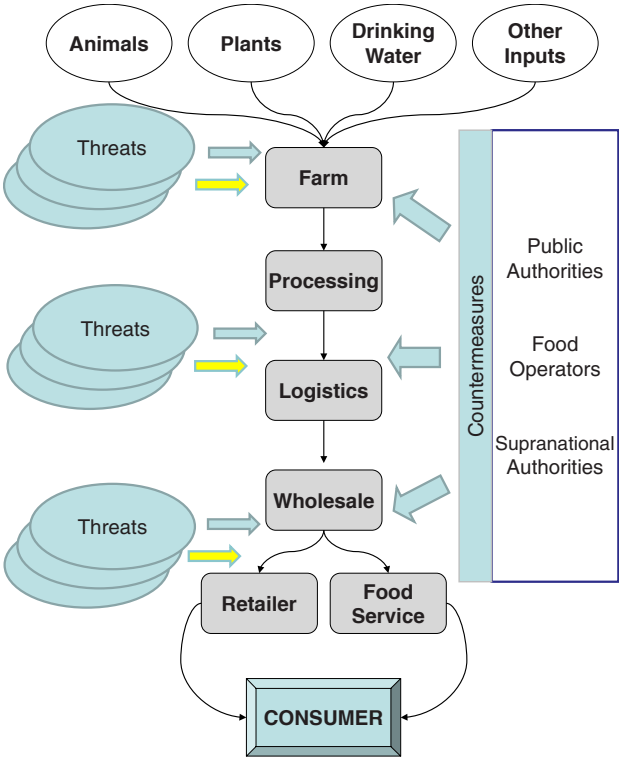


Figure 1. Food supply chain decomposition.

processing. On the other hand, as reported by Lee, *et al.* [14], the most probable targets in the supply chain are food vendors, which includes food producers, retailers, restaurants and other food service establishments. This is because, even if the overall impact is limited in terms of the concrete consequences, the attacker would obtain a large "return on investment."

We also considered the "likelihood" of attacks. The likelihood takes into account the availability and manageability of the agents, the vulnerability of the specific product supply chain, and the possible effects in terms of causalities, economic loss and psychological impact. Specifically we considered:

- Processes in terms of their ability to neutralize agents and product accessibility.
- Company policies regarding employees and visits (e.g., monitoring and access control).
- Security measures adopted (e.g., alarms, cameras and guards).
- Quality control mechanisms implemented (e.g., number and types of controls and hazard analysis and critical control points (HACCP)).

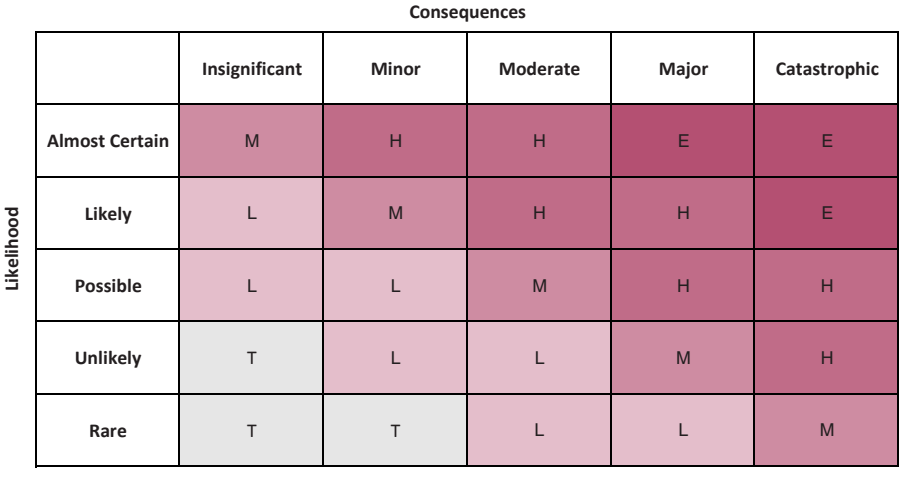
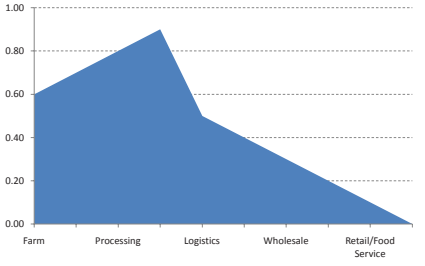


Figure 2. Risk assessment matrix [21].

Next, for each of the eight types of food items (milk, yoghurt, juice, bread, oil, salads, fish and baby food), we performed a risk analysis based on the research literature and interviews with principal stakeholders and public authorities. We collected about 40 questionnaires and performed inspections of several food processing facilities. Also, we analyzed all the incidents reported by Dalziel [7] and others, amounting to more than 450 cases of malicious contamination of food. Finally, we evaluated and classified about 50 biological and chemical agents in terms of their availability, manageability and possible pathological effects.

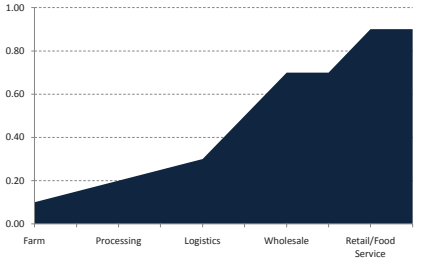
These activities enabled us to collect a large quantity of qualitative and quantitative data about threats to the food supply chain. The data was analyzed with the help of experts from a specialized police corps [1]. An operational risk management (ORM) approach [21] was used to classify the attacks from extreme to tolerable. We also identified the degree of likelihood for each agent with respect to each food item and step in the supply chain. The likelihood was evaluated in terms of the availability of the agent and the vulnerability of the corresponding supply chain step. We created a risk matrix taking into consideration the ability to detect the attack and the possible consequences (Figure 2). The risk matrix employs the following risk categories:

- **Extreme (E):** Causes a large number of injuries, several deaths and catastrophic economic consequences.
- **High (H):** Causes severe injuries, some deaths and severe economic consequences.
- **Moderate (M):** Causes some injuries that may require medical attention, and significant economic consequences.



(a) Significant contamination;

(b) Effectiveness of surveillance.



(c) Symbolic contamination;

(d) Effectiveness of countermeasures.

Figure 3. Risks related to different steps in the food supply chain.

- **Limited (L):** Causes no injuries, but some economic consequences.
- **Tolerable (T):** Causes no health effects, but has a limited impact on reputation and some economic consequences.

Figure 3 summarizes the results obtained by averaging the behavior corresponding to each of the 50 contaminating agents with respect to the eight classes of food considered in our analysis. Figure 3(a) illustrates how the risk due to an exposure to significant contamination increases from the farm to the processing phase, reaching a maximum just before the packaging operation. After this, it reduces monotonically due to the decreasing sizes of the lots in the subsequent steps.

In contrast, the graph in Figure 3(c) shows that the risk due to symbolic contamination reaches its maximum close to the consumer. This happens because, as seen in Figures 3(b) (effectiveness of surveillance) and 3(d) (effectiveness of countermeasures), the last steps in the food supply chain are less controlled and less secure. Indeed, most of the controls and countermeasures in the food supply chain are intended to guarantee the safe production of food. Therefore, they are largely concentrated in the production step, where tests are conducted on raw materials, semi-processed goods and final products. After the production step, security-related activities are mostly focused on preventing theft and only minimally on preventing food tampering.

Table 1. Biological and chemical agents [13, 16, 19].

Agent	Lethality	Availability
Biological Agents		
E. coli	3–5%	Easy
Yersinia	100% (pneumonic) 50% (bubonic)	Easy
Salmonella	<5% (*S. enteritidis*) 12–30% (*S. typhi*)	Easy
Staphylococcus aureus	< 5%	Easy
Brucella	Low	Easy
Francisella tularensis	30–40%	Difficult
Coxiella burnetii	<5%	Difficult
Chemical Agents		
Abrin	Fatal (no antidote)	Very Easy
Aflatoxin	Fatal (no antidote)	Easy
Tetrahydrocannabinoids	Toxic at high levels	Easy
Safrol	Carcinogen	Moderately Easy
Diphosgene Lewisite	Fatal at high levels	
Nicotine	Fatal	
Ricin	Fatal (no antidote)	Very Easy
Tetrodotoxin	Fatal at low levels	Difficult
Saxitoxin	Fatal at low levels	Moderately Easy
Shigatoxin	Fatal at low levels	Easy
Nitrogen Mustard Gas	Fatal at high levels	Difficult
Cadmium	Fatal	Moderately Difficult
Chromium VI	Fatal	Easy
Mercury	Fatal at high levels	Easy
Red Phosphorus	Fatal	Easy
Thallium	Fatal at high levels	Difficult
Titanium	Fatal at high levels	
White phosphorus	Fatal	Easy
Arsenic	Fatal	Very Easy

3. Milk Case Study

This section focuses on a case study of the milk sector. Milk was selected because it is a basic component of the European diet; as such, it is consumed in large quantities either directly or indirectly in other food products. Moreover, it has been the target of malicious attacks [23].

Table 1 lists the main biological and chemical agents that can be used to contaminate milk. Information is also provided about the lethality and ease of availability of these agents.

In the case of milk, it is important to distinguish between biological and chemical agents. Most processed milk goes through a pasteurization (thermal) process that kills biological agents. The subsequent cooling of milk to 4°C

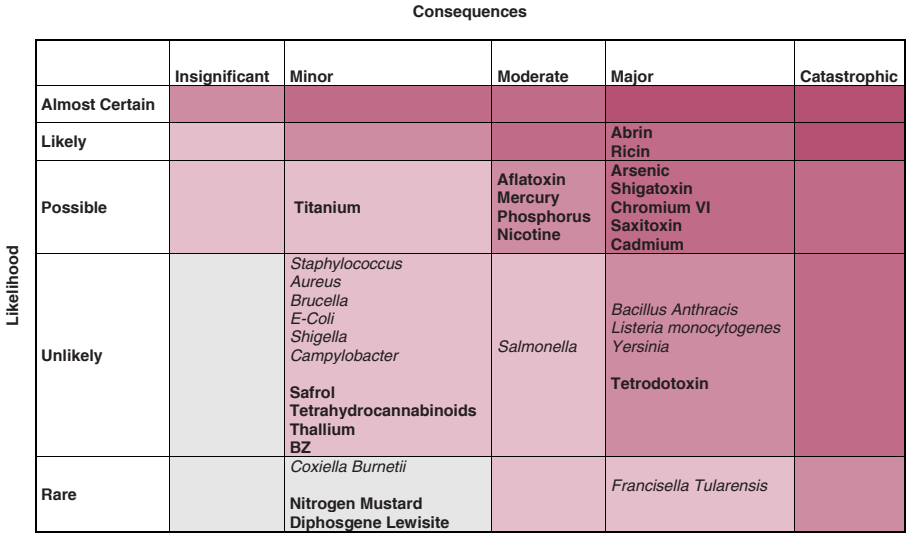


Figure 4. ORM matrix for the milk supply chain.

makes it very difficult for most biological agents to grow. On the other hand, the heating and cooling process does not affect chemical agents, enabling them to be added at any time during milk production.

Milk producers perform several tests on raw milk to check its quality and detect the presence of biological agents. However, these tests are not comprehensive and tests for dangerous agents such as *botulinum* are not performed.

In general, the deliberate contamination of milk at the output stage is much more complicated than at the input stage because the product is packaged in small lots. However, Blasco and Bledsoe [16] observe that with the appropriate technical knowledge and access, any product can be tampered with during the distribution or retail steps. Indeed, packaged food can be sabotaged by terrorists or criminals with a relatively low degree of sophistication.

The ORM matrix in Figure 4 demonstrates that, in the case of milk, the adverse consequences of chemical agents (bold) are much higher than those due to biological agents (italics). This is because few, if any, controls are in place for chemical agents. Furthermore, detecting some chemical agents is very difficult because they are colorless and odorless. However, the most important factor is that chemical agents, unlike biological agents, are not destroyed by the heating and cooling processes involved in milk production.

In summary, the milk sector is prepared to prevent spontaneous contamination via the implementation of controls against zoonosis and other health risks of a microbiological origin. However, it is woefully unprepared to deal with malicious contamination using chemical agents.

4. Conclusions

Food is an unconventional weapon in the hands of terrorists. Despite the worldwide attention paid to the malicious tampering of food products, the majority of the stakeholders in the food supply sector have little understanding of the risks related to deliberate contamination. In general, they believe that their production processes are secure and that their controls and countermeasures are adequate. However, they concede that malicious entities can target food products almost anywhere in the supply chain. This means that they admit that many vulnerabilities exist in food production and distribution.

The consequences of contamination vary according to the specific step in the supply chain that is targeted. An attack that targets a step closer to the consumer has a greater probability of success but affects fewer people. On the other hand, an attack in the early steps of the supply chain affects many more people, but has to evade many controls and countermeasures to be successful.

The transportation and storage steps are, in general, more vulnerable that the manufacturing step. Raw materials are more vulnerable than packaged products, but it is difficult to successfully target raw materials because of strong quality controls. Packaged products are more susceptible to contamination during transportation and storage. The risk is high and the probability of detection is very low – until consumers are affected.

With regard to the milk supply chain, pasteurization and quality control processes reduce the likelihood of a successful attack involving biological agents. However, because of the absence of controls and countermeasures, attacks using chemical agents have a high probability of success.

The absence of major food contamination events leads us to believe that the food supply is relatively safe, but we cannot afford to be complacent. All the entities in the food supply chain should develop security plans for managing the risk. The hazard analysis and critical control points (HACCP) approach is an effective technique as it focuses on proactive (preventive) measures instead of reactive measures, which is prudent in any critical infrastructure sector.

Acknowledgements

This research was partially supported by the European Commission Directorate General for Justice, Freedom and Security under the SecuFood – Security of European Food Supply Chain Project (Grant No. JLS/2008/CIPS/022).

References

[1] Arma dei Carabinieri, Comando Carabinieri per la Tutela della Salute, Rome, Italy (www.carabinieri.it/Internet/Cittadino/Informazioni/Tutela/Salute).

[2] Asia-Pacific Economic Cooperation, APEC to increase protection of the food supply from terrorist attack, News Release, APEC Secretariat, Da Nang, Vietnam, September 15, 2006.

[3] W. Boddie and L. Kun, Health care, public health and the food and agriculture critical infrastructures, *IEEE Engineering in Medicine and Biology*, vol. 27(6), pp. 54–58, 2008.

[4] G. Bush, Critical Infrastructure Identification, Prioritization and Protection, Homeland Security Presidential Directive 7 (HSPD-7), The White House, Washington, DC, December 17, 2003.

[5] G. Bush, Defense of United States Agriculture and Food, Homeland Security Presidential Directive 9 (HSPD-9), The White House, Washington, DC, January 30, 2004.

[6] Centre for the Protection of National Infrastructure and British Standards Institute, Defending Food and Drink: Guidance for the Deterrence, Detection and Defeat of Ideologically Motivated and Other Forms of Malicious Attack on Food and Drink and their Supply Arrangements, Report PAS 96:2010, London, United Kingdom (www.cpni.gov.uk/Docs/PAS96_vis14.pdf), 2010.

[7] G. Dalziel, Food Defense Incidents 1950-2008: A Chronology and Analysis of Incidents Involving the Malicious Contamination of the Food Supply Chain, Technical Report, Centre of Excellence for National Security, S. Rajaratnam School of International Studies, Nanyang Technological University, Singapore, 2009.

[8] Department of Food Safety, Terrorist Threats to Food: Guidance for Establishing and Strengthening Prevention and Response Systems, World Health Organization, Geneva, Switzerland, 2008.

[9] European Commission, Green Paper on Bio-Preparedness, COM (2007) 399 Final, Brussels, Belgium (ec.europa.eu/food/resources/gp_bio_preparedness_en.pdf), 2007.

[10] European Commission, Rapid Alert System for Food and Feed (RASFF), Brussels, Belgium (ec.europa.eu/food/food/rapidalert/index_en.htm).

[11] B. Ezell and D. von Winterfeldt, Probabilistic risk analysis and bioterrorism risk, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice and Science*, vol. 7(1), pp. 108–110, 2009.

[12] A. Khan, D. Swerdlow and D. Juranek, Precautions against biological and chemical terrorism directed at food and water supplies, *Public Health Reports*, vol. 116(1), pp. 3–14, 2001.

[13] R. Lawley, L. Curtis and J. Davis, *Food Safety Hazard Guidebook*, Royal Society of Chemistry, London, United Kingdom, 2008.

[14] R. Lee, R. Harbison and F. Draughon, Food as a weapon, *Food Protection Trends*, vol. 23(8), pp. 664–674, 2003.

[15] J. Monke, Agroterrorism: Threats and Preparedness, CRS Report for Congress RL32521, Congressional Research Service, Washington, DC (www.fas.org/irp/crs/RL32521.pdf), 2004.

[16] B. Rasco and G. Bledsoe, *Bioterrorism and Food Safety*, CRC Press, Boca Raton, Florida, 2005.

[17] SecuFood Project Secretariat, SecuFood: Security of the European Food Supply Chain, University Campus Bio-Medico, Rome, Italy (secufood.unicampus.it).

[18] M. Sekheta, A. Sahtout, F. Sekheta, N. Pantovic and A. Al Omari, Terrorist threats to food and water supplies and the role of HACCP implementation as one of the major effective and preventive measures, *Internet Journal of Food Safety*, vol. 8, pp. 30–34, 2006.

[19] D. Shea and F. Gottron, Small-Scale Terrorist Attacks Using Chemical and Biological Agents: An Assessment Framework and Preliminary Comparisons, CRS Report for Congress RL32391, Congressional Research Service, Washington, DC (www.fas.org/irp/crs/RL32391.pdf), 2004.

[20] U.S. Department of Homeland Security, U.S. Department of Agriculture and U.S. Food and Drug Administration, Agriculture and Food: Critical Infrastructure and Key Resources Sector-Specific Plan as Input to the National Infrastructure Protection Plan, Washington, DC (www.dhs.gov/xlibrary/assets/nipp-ssp-ag-food.pdf), 2007.

[21] U.S. Food and Drug Administration, CARVER Software, Washington, DC (www.fda.gov/Food/FoodDefense/CARVER/default.htm#whatis).

[22] P. Valle, A. Girard and O. Saldate, Defensa alimentaria, *Mundo Lácteo y Cárnico*, pp. 12–19, September/October 2007.

[23] L. Wein and Y. Liu, Analyzing a bioterror attack on the food supply: The case of botulinum toxin in milk, *Proceedings of the National Academy of Sciences*, vol. 102(28), pp. 9984–9989, 2005.

[24] World Health Organization, Global Public Health Response to Natural Occurrence, Accidental Release or Deliberate Use of Biological and Chemical Agents or Radionuclear Material that Affect Health, Resolution WHA 55.16, *Fifty-Fifth World Health Assembly*, Geneva, Switzerland (apps .who.int/gb/archive/pdf_files/WHA55/ewha5516.pdf), 2002.

Chapter 12

INTERACTIVE VISUALIZATION OF INTERDEPENDENCIES AND VULNERABILITIES IN CONSTRAINED ENVIRONMENTS

Nils Lunden, Robin Sveen, Hans Lund, Nils Svendsen and Stephen Wolthusen

Abstract Many critical infrastructure assets from hospitals to industrial facilities rely on multiple infrastructure services whose close proximity can result in the failure of one component causing cascading failures in other assets. Ths effective analysis and mitigation of risks requires the consideration of numerous scenarios and input from domain experts.

This paper describes a distributed interactive visualization and analysis mechanism for constrained environments such as buildings and industrial facilities in which the physical and logical components and dependencies are considered in a three-dimensional model. Physical effects modeled include flooding, fire, smoke and gas explosions; the logical dependencies considered include telecommunications and electrical power. This allows the creation of scenarios by interactively creating events whose effects are subsequently captured by physical and logical submodels that can be viewed and replayed from multiple angles to permit efficient analysis and review. The capabilities of the system are illustrated using a complex scenario of cascading physical and logical failures resulting from a water leakage in a hospital environment.

Keywords: Visualization, concurrent models, distributed simulation

1. Introduction

Most critical infrastructure facilities are dependent on other critical infrastructure components and sectors. Dependencies and interdependencies between components and sectors can be captured and analyzed at the logical level [19]. However, this view is often imperfect as physical dependencies and hazards must also be considered in order to obtain an understanding of the risks to

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 171–183, 2010.

an infrastructure element or cluster and possible mitigation approaches. The simultaneous consideration of logical dependencies and physical effects permits the identification of potential hazards due to unexpected (inter-)dependencies in logically separate but physically proximate infrastructure elements. Also, it facilitates the development of dynamic scenarios for assessing the existence and severity of risks as well as the time available to take mitigation measures [18].

While such models provide valuable insights in larger, sparsely populated geospatial domains and must have limited complexity to permit effective use, their utility in confined environments such as buildings and industrial facilities is limited. Elaborate models and visualization environments exist for some types of hazards [7, 11], but these are concentrated on a single incident or hazard category. They are often unsuitable for identifying the cascading failures of interest when analyzing the robustness of a facility to disruptions and its ability to provide critical services under degraded circumstances. However, they may be utilized to assess risks and mitigation strategies. Nevertheless, given the intractable number of possible simultaneous and cascading contingencies, automated analysis is unlikely to yield significant insights. It is therefore critical to allow domain experts from different realms to create and analyze scenarios interactively while providing physical and logical models that track the evolution of the scenarios and allow for their visualization. The resulting interactive visualization paradigm, which would facilitate the viewing of event traces from different perspectives and with varying levels of detail and emphasis, can yield important insights.

2. Model Components

The model used for visualization incorporates multiple modules centered around the geospatial representation of the area under consideration. This is achieved using a "scene graph" representation found in most vector-based graphical editing systems and in some distributed gaming environments. The scene graph representation permits the efficient and intuitive representation of logical and spatial groupings of entities in the form of a tree.

The visualization is controlled by a "master node" that retains the full model of the scenes to be rendered. The "rendering nodes" that participate in a simulation register with the master node, download the initial scene graphs and forward any local changes that are made to the master node. A tree-structured change propagation hierarchy can be employed reduce the overhead on the master node. Communication between nodes is realized via IPSec-secured channels to minimize the latency of individual update bursts. Note that the master node may or may not have a rendering engine running on it.

Distributed operations on scene graphs may be grouped by sub-trees that, for example, contain the contents of individual rooms; the operations must be transmitted once for the basic configuration when a rendering (visualization) node joins a simulation, adds or changes rooms. Any changes made must be transmitted to all nodes. When a node performs modifications locally, these must be communicated first to the master node, which then transmits the

modifications to all affected visualization nodes. To minimize communications traffic, rendering is performed on visualization nodes.

All objects and entities in the model are characterized by three-dimensional geo-referenced coordinates. The precise representation may range from a simple point cloud description to arbitrary polygon mesh forms for complex objects, where the latter may result from tessellations of other types of object or object surface descriptions (e.g., non-uniform rational B-spline surfaces) because polygon meshes are the natural representation used by the visualization engine.

The overall architecture is modular in that it supports an extensible set of sub-models for different types of events. This also allows the replacement of sub-models with others of higher or lower fidelity depending on resource availability and other requirements. To minimize the complexity of sub-models, components may be shared and reused. This includes material properties that describe the behavior of objects with regard to flammability, water-related effects and response to pressure, in addition to visualization properties such as color and transparency. In particular, the model captures both bulk material and surface properties.

The second major component shared across models is the basic physics model. Given the choice of visualization platform (see Section 3), this is based on the jME Physics 2 interface, which, in turn, provides an adaptation layer for several physics engines, including the Open Dynamics Engine (ODE) [16] and the proprietary PhysX engine by Nvidia. The two engines run on a variety of platforms and offer rich feature sets. The PhysX engine, in particular, offers performance benefits because it can draw on hardware support in the form of a dedicated physics processing unit (PPU) or a parallel implementation on a graphics processing unit (GPU) [3]. This engine also supports our material model, especially with regard to friction effects and general collision detection for the various geometric shapes used to model scenarios.

However, even with hardware acceleration support, some of the sub-models described in this paper (especially those in Sections 2.1 and 2.3) cannot be represented by a naïve object model for individual particles. Real-time performance falls to unacceptable levels when more than 2,000 objects are used (this occurs when the frame rate of rendering nodes drops below 30 frames/second). We have, therefore, developed a dedicated "particle model" that can draw on the physics engine to employ a "sampling strategy" to capture the underlying object behavior in addition to that dictated by sub-models using small objects that do not require individual collision modeling but aggregate, explicit models and surface collision models. Figures 1 and 2 show an example of a simple (fire) particle model using the sampling approach in conjunction with surface interactions that include friction, deflection and heat exchange.

Although this approach provides adequate fidelity at relatively low computational complexity for small volumes and time-scales, it is still undesirable to model very small particles individually over larger time and spatial ranges; this is because other influences (e.g., sub-model effects) can dominate. Consequently, we have chosen to capture meso-scale effects using a "vector field

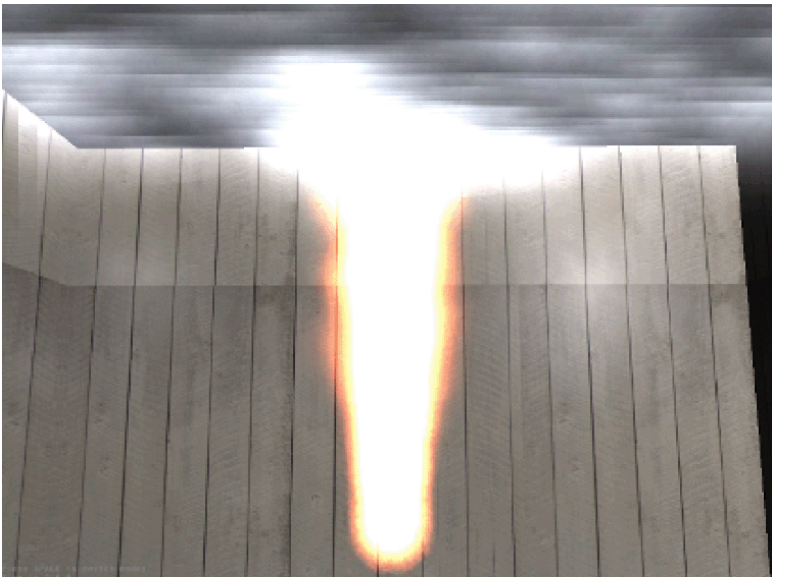
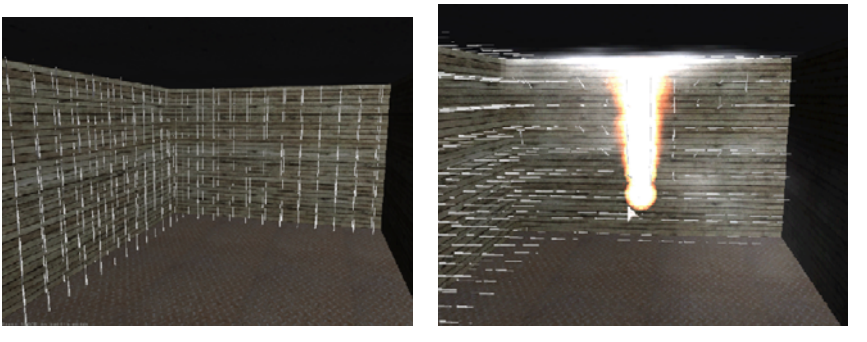


Figure 1. Interpolated model using sampling and surface interactions.



(a) Start of simulation. (b) Updated vector field.

Figure 2. Vector field meso-scale particle model simulation.

model layer" using gradient vector fields $\nabla f = \langle f_x, f_y, f_z \rangle(t)$ and a Hamiltonian dynamics model [1]. A number of more efficient approximations do exist, but we believe that this approach represents a straightforward "first principles" approximation (even though the reversible dynamics provided by the approximation are not used in our model). The choice of formalism restricts us to the use of smooth functions, but as we are mainly interested in modeling thermophysical effects, this is not an issue. However, for the vector field model, it is desirable to increase the update frequency and, hence, the temporal and spatial

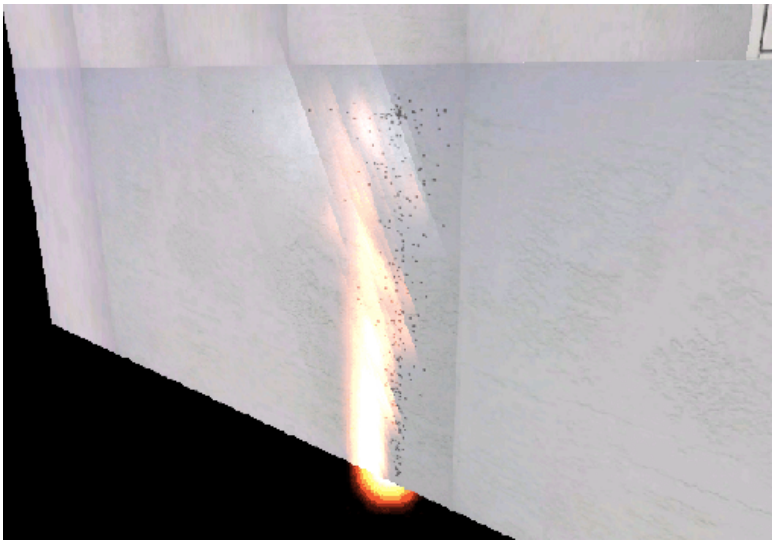


Figure 3. Soot particle aggregation and heat transfer interactions.

resolution to prevent interpolation errors from producing undesirable artifacts despite an otherwise smooth visual appearance in case of rapid dynamic effects. Note that the use of parallel models with discrete time steps can lead to artifacts, especially due to sampling effects.

2.1 Fire and Smoke Sub-Models

Fire and smoke are two demanding modeling and visualization domains. Several dedicated models are available in the literature (see, e.g., [11]). However, for the purposes of scenario development and visualization, a simple hydrodynamic model of heat transfer was chosen to obtain the requisite parameters for the particle model described above [20]; this resulted in a set of reference samples and a corresponding vector field for modeling heat transfer.

Of particular interest to critical information infrastructures are the adverse effects produced by soot particles that are affected by these flows. Consequently, soot particles and particle accumulations are modeled explicitly. Figure 3 shows soot particle aggregation and heat transfer interactions with a wall. The diagonal lines are superimposed artificially to enhance visibility of aggregations caused by heat transfer to the wall.

Note that our model is limited to simple scenarios and cannot accurately capture some effects found in more elaborate models. Also, the model is limited by the constraint on particle numbers to obtain support points for the vector field model. However, because we are primarily interested in observing the interactions of multiple sub-models, the limited fidelity of our model does not pose an insurmountable restriction.

Figure 4. Flooding effect simulation for unpressurized volumes.

2.2 Flooding Sub-Models

The scenarios considered in this paper do not engage detailed hydrodynamic models for water because dynamic effects are of limited interest in confined areas (unlike for a large-scale scenario such as a dam burst). Instead, we concentrate on spatially-constrained higher-pressure leaks and flooding effects. The former uses the aforementioned (spatially-constrained) particle model to capture the effects of leaking or ruptured pressurized pipes in a simplified fluid dynamics model [8] that considers a limited range of viscosity for the leaking material. Note that viscosity is assumed to be static in the model; this assumption may have to be adjusted for certain materials (e.g., liquefied natural gas) to consider temperature and heat transfer in the surrounding area.

The main emphasis is on flooding effects; the geo-referenced mesh and a coarser support vector mesh are used to establish whether fluids are retained in a contained volume as well as the flows between volumes and flow rates. Clearly, this is heavily dependent on an accurate mesh representation of the underlying model, especially when retrieving models from architectural drawings, which may suffer from gaps or format conversion errors and must, therefore, be validated manually. Nevertheless, the model keeps track of fluid levels in non-pressurized volumes to determine the pressure exerted on the sides of the container (see Figure 4). This, coupled with the material properties described earlier, allows leakage and rupture effects to be expressed. However, the model does not fully capture the dynamics of non-pressurized flooding and may, therefore, under-predict the dynamic effects compared with more complex

hydrological models [12]. Nevertheless, this can easily be compensated for by adjusting the parameters to obtain conservative upper bounds on the estimated effects.

2.3 Gas Spreading and Explosion Sub-Models

Models for gas and vapor cloud explosions include those by Alonso, *et al.* [5] for semi-open areas and by Cleaver, *et al.* [4] for confined spaces. We approximate explosions using the particle and vector field approach described above. However, as noted in Sections 2.1 and 2.2, a high-resolution sub-model is required to capture the speed and volatility of an explosion.

In the case of gas spreading, however, it is necessary to capture gas properties (e.g., relative density) in addition to flow models because aerosol formation increases the risk of an explosion. The model currently supports only a limited number of industrial gases. In the case of propane (C_3H_8), a common industrial gas, an aerosol is flammable when the propane concentration is between 2.1% and 9.5% at temperatures exceeding 723K assuming natural oxygen levels. However, if the aerosol is formed in the presence of pure oxygen, the upper bound increases to 48%. Note that the accuracy of our model is limited by the mesh resolution and concentrations must be interpolated between individual mesh points. Also, only binary reactions are supported; this captures common redox reactions.

2.4 Logical Dependencies

Having specified the individual sub-models, it necessary to cross-link the models. "Geospatial buffering" is used to identify affected entities as has previously been shown for meso-scale systems [18], albeit at a finer granularity than for our model.

After the volumes affected by events in a given simulation time step (more precisely, aggregate time step) are identified, the infrastructure components affected are identified by their geo-referenced coordinates in the scene graph of the overall model. This, in turn, can be used in a straightforward manner as a buffered multigraph in a conventional dependency analysis model [19]. Although it would be desirable to associate infrastructure components of different types with properties for physical simulations, we add these properties manually to the scenario components as discussed in Section 2.5.

2.5 Scenario Development

The development of simulation scenarios for the investigation of incidents or for evaluation and training requires considerable effort. The basis for the simulation is a faithful three-dimensional model, which must be of sufficient quality to ensure that all the structural elements are identified along with minor features such as ducting to ensure that the sub-models can be applied correctly. In the case of gas and smoke spreading or flooding simulations, even minor

imprecision due to volumes that are not fully convex may result in undesirable results. To minimize the effort required to generate models in the native representation of the visualization engine, the three-dimensional models may be imported in the AutoCAD Drawing Interchange Format (DXF). However, they must subsequently be analyzed to ensure correctness and for the presence of the desired properties.

Next, the model must be annotated with material and surface properties where needed. Although this is supported by AutoCAD DXF, the annotations may not reach the level of detail required for a given scenario and may, therefore, require additional manual refinement of the model or annotation. Typical examples include walls constructed of layers of different materials or walls containing ductwork and wiring. However, aggregate characteristics may be sufficient unless the properties of a given object or area are critical to scenario elaboration.

An additional step involves the placement of entities and objects that are not found in architectural drawings. In some cases, the required locations, dimensions and characteristics may be obtained from other databases (e.g., those storing cable and pipe management information), but manual placement is required for what-if scenarios. At this point, the logical dependencies also need to be specified, including geo-references associated with relevant entities and the physical representations of objects.

The development of scenarios can entail some effort, particularly when capturing details such as the placement of cable runs, pipes and ducting required to analyze complex, cascading faults and events. However, such baseline "scenes" are likely to be reused in multiple scenarios. Therefore, we concentrate on making scenario development instead of scene development interactive, especially since scene development typically requires manual validation in any case.

The final step in scenario development is the elaboration of events and activities. Here, the developer can use the interactive visualization environment to indicate where events (e.g., gas or water leakage, fire, etc.) occur and at which point in time. As these events are fed into the various interacting sub-models, the location and time of each event must be recorded to ensure that any developments from an event such as cascading failures are captured adequately. To facilitate the use of scenarios in what-if analyses and contingency training, events can be retained in a tree structure to permit interactive backtracking and the reuse of event chains in the simulation environment.

3. Visualization

The models described above capture event information for a scene graph. However, it is still necessary to transform this into a visual representation.

One of our objectives in developing the simulation and visualization environment was to provide maximum flexibility for deployment. Subject matter experts and participants in training exercises are often in different locations (hence the need for a distributed simulation and visualization environment) and substantial differences may exist in the computing resources available. To

maximize availability and bypass licensing and platform constraints, we selected jMonkeyEngine 2.0.1, a free, open source visualization engine. A Java-based system underlying the BSD open source license enables the engine to run satisfactorily on most platforms. The actual rendering mechanisms available depend on the platform with both LWJGL (Lightweight Java Game Library) and JOGL (Java OpenGL) bindings available. JOGL provides a more mature and higher-performance environment in cases where OpenGL rendering is supported by the underlying platform.

The Java implementation of the visualization system and models provides a largely platform-independent architecture for all the components. However, there is a performance impact because the visualization system and models can easily reach the limits of a particular platform. The fidelity of the underlying models can be enhanced by parallelized operation, but this is not the case for visualization. Therefore, a platform-specific rendering architecture may be necessary for scenarios where very high visual quality is required in real-time environments. We do not foresee a need for this at present; however, the sub-models described in Section 2 can be transformed to benefit from the localized parallelism afforded by general-purpose graphics processing unit (GPGPU) programming, especially for the master node(s) in order to scale up to larger numbers of visualization nodes.

3.1 Review Mechanism

As noted in the introduction, one of our goals is to permit the joint development and analysis of scenarios. For real-time training exercises, this may be considered a "serious game" in that it forces participants to react to complex events often with limited information and limited time to consider the most appropriate course of action or adverse effects resulting from a decision. Also, following a simulation, but especially after conducting training exercises, it may be important to review the events in a given scenario with multiple experts using a joint timeline.

To this end, simulation events are tagged with a timeline and the transactions required to reach a given configuration. When a sequence of events is to be reviewed, it is necessary to configure the time and location of the events and customize other aspects of the review such as the location and viewpoint of the reviewer and additional model information (e.g., material properties and logical dependencies). Where appropriate, visualization parameters can be adjusted to add fidelity or to reduce unnecessary aspects. For example, if the etiology of an electrical fire is to be investigated, it may be sufficient to understand the patterns in which soot deposits build up; fire and smoke simulations distract from understanding these patterns.

4. Evaluation

In addition to constructing individual scenarios for sub-models and minimal interacting sub-models and their visualization, we have constructed a larger

Figure 5. View of hospital area.

scenario in which all the sub-models developed are used. The scenario is based on a hospital environment. It does not represent an actual facility, but uses individual rooms and components in a composition.

The hospital scenario was chosen for several reasons. Hospitals are part of the critical infrastructure and their inability to provide services may have local as well as larger-scale effects. Also, they typically use hazardous materials and equipment (e.g., oxygen delivery systems and emergency generators) and are equipped with substantial information technology assets and capital equipment. The placement of these assets and the potential cascading effects of failures or accidents are not always fully understood. At the same time, responding to and mitigating faults requires cooperation among a large number of individuals ranging from information technology and facilities management specialists to fire rescue and emergency response units and hospital staff.

Figure 5 provides a composite illustration of the scene. A server room (lower left) is located one floor below an intensive care ward (upper left). The adjacent corridors (upper and lower right-hand side) are connected on both floors through HVAC ducting (visible in the upper right-hand quadrant). Additional cabling and pipes as well as hidden ducts that are part of the model are not visible in the illustration.

Based on the scene, we have developed several scenarios primarily to demonstrate the potential cascading effects. One scenario involved an HVAC steam pipe leak that caused localized water damage and led to an electrical fire in the server room. This resulted in the loss of the affected machine and cascading damage as the soot and smoke from the electrical fire affected adjacent computer systems. Ultimately, all information technology components in the server room were lost, requiring the use of a redundant system in another location. While circuit breakers and fire control systems kept the fire from spreading and causing further electrical outages, the hot smoke and soot escaped to the adjacent corridor and entered the intensive care ward by way of HVAC ducting, which could have had severe effects on patients.

5. Related Work

Visualization and physics engines have been used extensively in high-fidelity environments such as surgery simulations [9, 10]. According to Rohde and Toschlog [14], the fidelity of simulations provided by common gaming engines is comparable to that provided by military simulation environments. However, in the case of visualization for end users, it is the perceptual quality that is relevant in physics simulations. Yeh, *et al.* [21] have shown that these simulations can be adapted to avoid unnecessary computations. The gaming industry has considerable expertise in creating scalable multi-party environments; these can be leveraged in critical infrastructure simulations. Interested readers are referred to [17] for a detailed review of modeling and simulation environments.

Our sub-models were chosen for their simplicity and ability to capture sufficient detail for visualization and to understand the etiology of events rather than to establish ground truth. Large bodies of work exist in individual modeling areas; most of them are not based on first principles but are hybrids incorporating experimental results. One example of this approach is work focused on fire dynamics and concomitant effects (see, e.g., [6, 13]). A model in widespread use for simulating fire and smoke propagation in enclosed spaces is the Fire Dynamics Simulator [11] from the National Institute of Standards and Technology along with Smokeview [7], its visualization system. Similarly, explosions and blast damage (especially blast overpressure damage) have been studied experimentally and using various first-principle models (see, e.g., [4, 5]).

Polygon mesh models for analyzing flooding effects have been used in hydrological and hydrodynamic models, but typically at coarser levels of individual buildings rather than within individual structures [15]. At larger scales, the size of the model must be considered as in the case of small-scale models at higher resolutions. While fluid dynamics models are typically based on rasterization [2], more elaborate models typically rely on finite element approaches.

6. Conclusions

Capturing the cascading effects due to interdependencies between critical infrastructure assets requires the integration of multiple models. The distributed interactive visualization and analysis architecture presented in this paper is aimed towards constrained environments such as buildings and industrial facilities. Physical effects considered include flooding, fire, smoke and gas explosions, while the logical dependencies considered include telecommunications and electrical power. The visualization system allows first responders and subject matter experts to conduct scenario analyses and training exercises in an interactive and distributed manner.

Future work will focus on the development of a federated rendering mechanism, which is desirable from the information flow and information sharing perspectives. Also, future research will attempt to integrate explicit control-law-based physical plant models that would provide higher levels of automation for scenario-based analyses.

Note that the distributed interactive visualization and analysis system described in this paper is based entirely on free and open source software. The software is available upon request from the authors.

References

[1] R. Abraham and J. Marsden, *Foundations of Mechanics*, Addison-Wesley, Redwood City, California, 1987.

[2] P. Bates and A. De Roo, A simple raster-based model for flood inundation simulation, *Journal of Hydrology*, vol. 236(1-2), pp. 54–77, 2000.

[3] A. Boeing and T. Braun, Evaluation of real-time physics simulation systems, *Proceedings of the Fifth International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, pp. 281–288, 2007.

[4] R. Cleaver, C. Humphreys, J. Morgan and C. Robinson, Development of a model to predict the effects of explosions in compact congested regions, *Journal of Hazardous Materials*, vol. 53(1), pp. 35–55, 1997.

[5] F. Diaz Alonso, E. Gonzalez Ferradasa, J. Sanchez Perez, A. Minana Aznara, J. Ruiz Gimenoa and J. Martinez Alonso, Characteristic overpressure-impulse-distance curves for vapor cloud explosions using the TNO multi-energy model, *Journal of Hazardous Materials*, vol. 137(2), pp. 734–741, 2006.

[6] D. Drysdale, *An Introduction to Fire Dynamics*, John Wiley, Chichester, United Kingdom, 1999.

[7] G. Forney, Smokeview (Version 5) – A Tool for Visualizing Fire Dynamics Simulation Data, Volume II: Technical Reference Guide, NIST Special Publication 1017-2, National Institute of Standards and Technology, Gaithersburg, Maryland, 2010.

[8] P. Kundu and I. Cohen, *Fluid Mechanics*, Academic Press, Burlington, Massachusetts, 2008.

[9] A. Maciel, T. Halic, Z. Lu, L. Nedel and S. De, Using the PhysX engine for physics-based virtual surgery with force feedback, *International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 5(3), pp. 341–353, 2009.

[10] S. Marks, J. Windsor and B. Wunsche, Evaluation of game engines for simulated surgical training, *Proceedings of the Fifth International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, pp. 273–280, 2007.

[11] K. McGrattan, S. Hostika, J. Floyd, H. Baum, R. Rehm, W. Mell and R. McDermott, Fire Dynamics Simulator (Version 5), Technical Reference Guide, Volume I: Mathematical Model, NIST Special Publication 1018-5, National Institute of Standards and Technology, Gaithersburg, Maryland, 2009.

[12] J. Neal, T. Fewtrell, P. Bates and N. Wright, A comparison of three parallelization methods for 2D flood inundation models, *Environmental Modeling and Software*, vol. 25(4), pp. 398–411, 2010.

[13] S. Olenick and D. Carpenter, An updated international survey of computer models for fire and smoke, *Journal of Fire Protection Engineering*, vol. 13(2), pp. 87–110, 2003.

[14] M. Rohde and M. Toschlog, Toward the fusion of serious simulation and video games, *Proceedings of the Spring Simulation Multiconference: Military Modeling and Simulation Symposium*, article 71, 2009.

[15] J. Schubert, B. Sanders, M. Smith and N. Wright, Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding, *Advances in Water Resources*, vol. 31(12), pp. 1603–1621, 2008.

[16] R. Smith, Open Dynamics Engine (v0.5) User Guide, (www.ode.org/ode-latest-userguide.html), 2006.

[17] N. Svendsen, Interdependencies in Critical Infrastructures: A Qualitative Approach to Model Physical, Logical and Geographical Interdependencies, Ph.D. Thesis, Norwegian Information Security Laboratory, Department of Computer Science, Gjovik University College, Gjovik, Norway, 2008.

[18] N. Svendsen and S. Wolthusen, A framework for 3D geospatial buffering of events of interest in critical infrastructures, *Proceedings of the Second International Workshop on Critical Information Infrastructures Security*, pp. 37–48, 2007.

[19] N. Svendsen and S. Wolthusen, Multigraph dependency models for heterogeneous infrastructures, in *Critical Infrastructure Protection*, E. Goetz and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 337–350, 2007.

[20] B. Weigand, *Analytical Methods for Heat Transfer and Fluid Flow Problems*, Springer-Verlag, Heidelberg, Germany, 2004.

[21] T. Yeh, G. Reinman, S. Patel and P. Faloutsos, Fool me twice: Exploring and exploiting error tolerance in physics-based animation, *ACM Transactions on Graphics* vol. 29(1), pp. 5-1–5-11, 2009.

Chapter 13

ASSESSING THE ECONOMIC LOSS AND SOCIAL IMPACT OF INFORMATION SYSTEM BREAKDOWNS

Fabio Bisogni and Simona Cavallini

Abstract The pervasiveness of information systems raises security and business continuity issues related to their disruption. Policy makers involved in preventing and preparing for unexpected critical events need to understand the direct and indirect socio-economic impacts of potential information system disruptions. This paper presents a new methodology for assessing the sectors that are most vulnerable to critical information system breakdowns. The vulnerability of information systems (VIS) model, which is based on the well-known input-output paradigm, simulates information system disruptions and assesses their socio-economic impact in the dynamic framework of sector interdependencies and cascading failure effects. The VIS model represents an important step in research focused on the prevention, preparedness and impact assessment of unexpected information system breakdowns. It sheds light on how the effects of a disruption can spread and helps identify critical infrastructures from a socio-economic perspective.

Keywords: Information system breakdowns, socio-economic effects

1. Introduction

Information systems are an essential component of every critical infrastructure asset and, as such, are vital to the functioning of society. A major breakdown in information systems directly affects the individual infrastructures where they are used. However, the indirect effects of a breakdown in one infrastructure propagate and cascade throughout all the other infrastructures because of infrastructure interdependencies.

Information systems are considered to be the most strategic enabling factor for European socio-economic development. To this end, the European Commission [5] is implementing its i2010 strategy, which focuses on trustworthy,

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 185–198, 2010.

secure and reliable information and communication technologies. The European Commission [7] has emphasized the creation of plans for protecting Europe from large-scale cyber attacks and information system disruptions. It has issued a directive [6] that mandates the identification and designation of European critical infrastructures and the assessment of the need to improve their protection [6]. It has also created the European Programme for Critical Infrastructure Protection (EPCIP) [2], which seeks to counter threats and enhance preparedness, security and resilience.

Key components of these protection efforts are defining criteria for identifying critical infrastructures and developing a comprehensive knowledge base of disruptions and their effects on critical infrastructures. Equally important is the need to predict the economic effects (e.g., monetary loss and service degradation) and the social effects (e.g., fatalities and loss of public services) due to infrastructure disruptions.

Luiijf, *et al.* [15] have conducted an analysis of 1,749 serious infrastructure failures that have occurred in Europe since 2000. Their analysis underscores the importance of information systems as a component of national critical infrastructures. According to their study, telecommunications is the second ranked sector (after energy) that initiates cascading failures in other sectors. In particular, telecommunications caused 24.3% of the total identified sector outages – within the sector as well as in other sectors such as finance, government services, transportation, energy and health. A smaller number of critical events (3.6%) were caused by Internet failures, and these impacted government services, finance and telecommunications in addition to other sectors.

Although information system breakdowns are acknowledged to have severe effects, quantitative assessments of the impact at the sector, national and European levels are difficult to perform. The vulnerability of information systems (VIS) model, which is presented in this paper, helps address this issue. The model simulates information system disruptions and assesses their socio-economic impact on the affected sectors using several impact metrics. These metrics can be used to rank the sectors according to their vulnerability to an information system breakdown.

2. Related Work

Several research efforts have focused on investigating infrastructure interdependencies and the potential cascading effects due to critical infrastructure disruptions, but relatively few studies have examined the impact of critical infrastructure disruptions. Although it is generally agreed that an information system breakdown of short duration can lead to serious consequences that propagate throughout all the critical infrastructures, there is little understanding of the systemic damage and socio-economic effects of such breakdowns.

Due to the complexity of infrastructure interdependencies, critical infrastructures have mostly been studied as physical assets from an organizational perspective. These approaches provide useful knowledge to corporate decision

makers, but are of limited value to policy makers focused on incident prevention and preparedness, and infrastructure resilience.

Haimes and Jiang [11] were among the first researchers to consider the socio-economic perspective by expressing industrial relationships using input-output data pertaining to the production and consumption of each sector in an economic system. Rinaldi [18] subsequently engaged input-output models of economic flows described by Leontief [13] to critical infrastructures. According to Sarriegi, *et al.* [21], input-output models are one of the most effective modeling paradigms for dealing with critical infrastructure interdependencies. In particular, these models help identify the economic sectors that are most vulnerable to a critical infrastructure breakdown.

Haimes, *et al.* [12] have used input-output relationships as the foundation for risk analysis in interdependent infrastructures. Their inoperability input-output model (IIM) [11, 20] extends Leontief's static input-output model by expressing the effects of a shock to an infrastructure in terms of sector inoperability and economic impact. To overcome the time-invariant limitations of IIM, Lian and Haimes [14] proposed the dynamic input-output inoperability model (DIIM), which extends IIM by incorporating the time dimension to express sector recovery after a critical event. Despite its utility, DIIM does not completely capture the domino effects of a disruption that affects interdependent economic sectors.

Kujawski [16] has attempted to address the deficiencies in IIM and DIIM by proposing the multi-period model for disruptive events in interdependent systems (MPMDEIS). The premise of MPMDEIS is that a critical event affects an economic system according to a four-phase lifecycle, starting with a pre-event period and terminating in a post-recovery period. MPMDEIS models a critical event as a shock which, due to sector interdependencies, cascades from one sector to another and reduces their production capacities. MPMDEIS addresses the two main limitations of traditional Leontief input-output models [13]: the requirement that the technical coefficients of sector production remain constant during and after the shock, and the exogeneity (and the consequent perfect and immediate adjustment) of the demand side of the economy in the aftermath of a critical event.

3. VIS Model

The vulnerability of information systems (VIS) model [10] has two goals: (i) assess the socio-economic impact of unexpected critical breakdowns of information systems; and (ii) rank sectors in an economic system according to their vulnerability to information system disruptions.

The theoretical framework of the VIS model relies on input-output relationships and addresses the same limitations as MPMDEIS. The industrial sectors correspond to an economic representation of a national infrastructure. The structural input-output relationships among the various sectors and, in particular, data related to the intermediate consumption of technological goods and services represent the weights of information systems in the sectors.

The most detailed input-output tables provided by Eurostat [8] depict interdependencies among economic sectors according to two-digit NACE Rev. 1.1 classification levels [9]. Using the NACE classification of economic sectors (which yields 57 sectors in the VIS model) and input-output data for Year 2004 provided by Eurostat [8], the share of the Computer and Related Activities sector (NACE code 72) in the production function of each sector is considered to express the relevance of information systems to the specific sector. According to the NACE classification, Computer and Related Activities includes hardware consulting, software consulting and supply, data processing, database activities, maintenance and repair of office, accounting and computing machinery, and other computer related activities. Technically, the dependence of each sector on information systems is obtained by computing the share of information systems directly employed in the production function of the sector of interest and the shares that are indirectly employed in other sectors that produce intermediate goods for the sector of interest (i.e., information systems as a share of other productivity factors).

Interdependencies are modeled using the coefficients of the input-output tables. The economic system is represented in the VIS model in terms of a supply side and a demand side. The supply side is expressed using 57 simultaneous equations corresponding to the 57 sectors augmented with equations expressing labor and price effects. Each of the 57 equations represents a sector production function in which the output produced by the sector is defined by:

$$p_Y Y_j = f(p_1 X_{1,j},\ p_2 X_{2,j},\ \ldots,\ p_n X_{n,j},\ p_L L_j)$$

where $p_Y Y_j$ is the nominal price per quantity of the output of the sector j; $p_i X_{i,j}$ is the value of the production input i to sector j ($X_{i,j}$ is the amount of input i used by sector j and p_i is its price); and $p_L L_j$ is the value of labor in sector j.

The translog functional form of the sector production functions helps overcome the problem of fixed production coefficients and accounts for marginal substitution with other productivity factors and relative price effects in the event of a breakdown. In this way, the quantities of inputs employed during equilibrium conditions are the effective annual flows employed by each sector in the real world and are derived from the input-output tables provided by Eurostat for the geographical area of interest.

The supply side helps define a computational general equilibrium model along with the demand side, which expresses the consumption of each sector output based on monopolistic competition [1]. The demand D_j for the output of sector j is given by:

$$D_j = (\frac{p_j}{p})^{-\epsilon} D$$

where p_j is the price of the output of sector j in a monopolistic competition framework; p is the price index resulting from the Dixit-Stiglitz aggregator [4];

ϵ is the elasticity of substitution among differentiated products; and D is the aggregate demand.

The matching of supply and demand defines the equilibrium conditions that are perturbed by an information system breakdown. Upon solving the VIS model equations, the socio-economic effects of an information system breakdown are evaluated by negatively perturbing the production function of the Computer and Related Activities sector and, as a consequence, reducing the information systems input to all the other economic sectors, which, in turn, reduces their outputs and changes the prices accordingly. The VIS model recursively computes the impact at each time period (one day) taking into account the domino effects on the production of all the economic sectors and following a recovery path until the equilibrium conditions are established.

According to the hypothesis related to a specific information system breakdown scenario, featured by the extent of the shock and its persistence profile (e.g., duration and functional form of the recovery process), the socio-economic effects in each sector are measured in terms of deviations relative to the benchmark equilibrium scenario in which the shock does not occur. Damage at the sector level caused by an unexpected critical information system breakdown relative to the non-shock equilibrium situation is expressed using five impact variables:

- **Percentage Output Deviation (OD):** This variable (≤ 0) captures the reduction in the production output of each sector.
- **Monetary Loss (ML):** This variable (≤ 0) captures the absolute variation of the value of the sector production. It is computed as the output loss multiplied by the real-world price.
- **Percentage Labor Deviation (LD):** This variable ($-\infty$ to $+\infty$) captures the deviation in the labor employed in each sector. In some sectors, more labor is necessary to recover the "normal" production capacity; other sectors may lose human capital with respect to the equilibrium conditions.
- **Percentage Price Deviation (PD):** This variable (≥ 0) captures the deviation in the price necessary to restore equilibrium conditions (a reduction in the production output for a sector causes an increase in its price). The price variation is not the real-world value, but the deviation of the sector price index (Year 2000 = 100) consistent with equilibrium conditions.
- **Welfare Loss (WL):** This variable (≥ 0) synthesizes all the impacts of an information system breakdown and considers the damage from a societal perspective. The welfare analysis is performed by assuming a standard quadratic loss function that considers the output deviation, labor deviation and price deviation without any weights. The quadratic loss function ensures that negative argument values (e.g., percentage output deviations) contribute to an increase in the loss.

Table 1. Italian sectors ranked by dependence on information systems.

Rank	NACE	Economic Sector	Input
1	K72	Computer and Related Activities	25.02%
2	J65	Finance	17.54%
3	I64	Post and Telecommunications	15.05%
4	K73	Research and Development	12.97%
5	J67	Activities Auxiliary to Finance	8.82%
6	O91	Activities of Membership Organizations	8.05%
7	K74	Other Business Activities	7.90%
8	J66	Insurance and Pension Funds	7.15%
9	K70	Real Estate Activities	5.75%
10	I62	Air Transportation	4.93%

The final results are the rankings of the sectors according to the five socio-economic impact variables. The rankings enable policy makers to define the protection priorities for the sectors that are most affected by information system disruptions.

4. Italian Case Study

The vulnerability of an economic sector to an information system breakdown is estimated by computing the dependence of the sector on information systems in terms of the relevance of direct and indirect information systems input on the total productivity factor using the input-output table data (Year 2004).

Table 1 ranks the top ten Italian economic sectors according to their direct and indirect shares of information systems in their total inputs (based on the NACE two-digit codes). Note that the Computer and Related Activities sector is used as a proxy for information systems. The three most dependent Italian sectors are those that employ the largest share of information systems in their productivity factors: Computer and Related Activities (25.02%), Finance (17.54%) and Post and Telecommunications (15.05%). Note that the results in Table 1 and in all the subsequent tables are based on available Eurostat data.

By applying the VIS model, it is possible to obtain more refined evaluations of the vulnerabilities of the Italian economic sectors to information system breakdowns taking into account sector interdependencies and domino effects. The simulation scenario assumes that an unexpected breakdown lasting one day reduces the output of the Computer and Related Activities sector by 10% and that five days are required for the sector to recover 50% of the lost production capacity.

In the simulation, the recovery of production capacity of the Computer and Related Activities sector is assumed to follow an autoregressive first-order process (Figure 1). Feedback from case studies confirms that an unexpected, critical breakdown of information systems causes an immediate reduction in production followed by a gradual recovery that decreases over time [10].

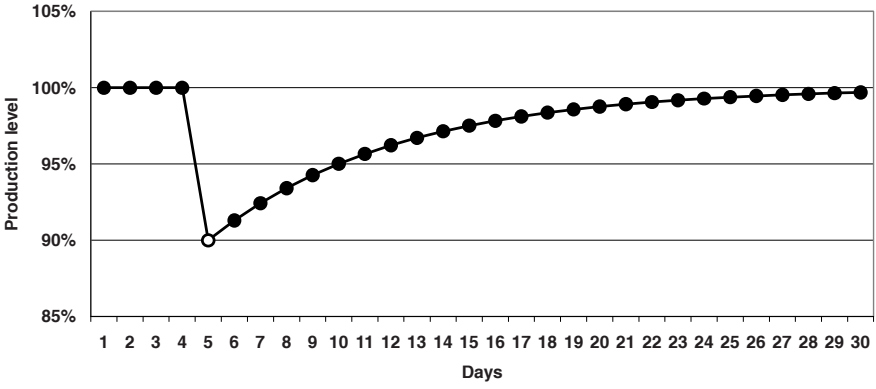


Figure 1. Recovery function for the Italian Computer and Related Activities sector.

Table 2. Sectors ranked by output deviation after an information system breakdown.

Rank	NACE	Economic Sector	1 Day	90 Days
1	K72	Computer and Related Activities	–11.01%	0.00%
2	K73	Research and Development	–1.87%	0.00%
3	I64	Post and Telecommunications	–1.38%	0.00%
4	J65	Finance	–1.25%	0.00%
5	J67	Activities Auxiliary to Finance	–1.20%	0.00%
6	I63	Supporting and Auxiliary Transportation Activities	–1.14%	0.00%
7	I62	Air Transportation	–1.09%	0.00%
8	J66	Insurance and Pension Funds	–1.08%	0.00%
9	K74	Other Business Activities	–1.07%	0.00%
10	K71	Renting of Machinery and Equipment	–1.05%	0.00%

Table 2 ranks the ten most vulnerable Italian economic sectors in terms of the output deviation (OD) percentage based on one-day data (two-digit NACE codes). The Computer and Related Activities sector has a deviation of –11.01% one day after the breakdown (instantaneous effect), followed by the Research and Development (–1.87%), Post and Telecommunications (–1.38%) and Finance (–1.25%) sectors.

The rankings in Tables 1 and 2 differ marginally for the top five sectors, all of which employ information systems to a significant degree. This shows the relationship between the intensity of use of information systems in a sector and its vulnerability to an information system breakdown. In the case of the top five sectors, 8% to 25% of the total production depends on the availability of information systems. The threshold of information systems usage relative to the other inputs can be assumed to represent the lower bound on the intensity of information systems adoption by a sector. Once this threshold is exceeded, any disruption in the Computer and Related Activities sector would cause major damage.

Table 3. Sectors ranked by output deviation after an information system breakdown.

Rank	NACE	Economic Sector	1 Day	90 Days
1	K72	Computer and Related Activities	–11.01%	–55.04%
2	J	Finance	–1.21%	–6.06%
3	I	Transportation, Storage and Communication	–1.12%	–5.60%
4	K	Real Estate and Business Activities	–1.04%	–5.19%
5	G	Wholesale and Retail Trade	–0.83%	–4.16%
6	E	Electricity, Gas and Water Supply	–0.74%	–3.72%
7	F	Construction	–0.70%	–3.51%
8	O	Other Services	–0.69%	–3.47%
9	C	Mining and Quarrying	–0.66%	–3.29%
10	A	Agriculture, Hunting and Forestry	–0.65%	–3.26%

NACE sectors can be aggregated at the one-digit level to provide a more comprehensive, albeit less detailed, view of the impact on the entire economic system. At the one-digit NACE level, the Italian economic system is represented in terms of sixteen macro-sectors.

Table 3 ranks the ten most vulnerable Italian economic sectors in terms of the output deviation (OD) percentage based on one-digit NACE codes. In particular, the table shows the cumulative effects after one day and after 90 days due to an unexpected 10% breakdown in information systems lasting one day with a 50% recovery after five days. Note that the ranking of the sectors is based on one-day data. Also, to simplify the analysis, the Real Estate and Business Activities sector (K) does not include the Computer and Related Activities sector (K72), which is shown separately.

The Computer and Related Activities sector (which is also affected by the initial 10% reduction), the Finance sector and the Transportation, Storage and Communication sector are impacted the most by the information system breakdown, with values of –11.01%, –1.21% and –1.12%, respectively. Note that the adverse effects on the economic sectors increase for a period of time. For example, the information system breakdown reduces the output of the Finance sector by 1.21% after one day and by 6.06% after 90 days; the adverse effects eventually converge to zero.

Table 4 ranks the ten most vulnerable Italian economic sectors in terms of the monetary loss (millions of euros) based on one-digit NACE codes. The table shows the cumulative effects after one day and after 90 days due to an unexpected 10% breakdown in information systems lasting one day with a 50% recovery after five days. Note that the ranking of the sectors is based on one-day data.

The Manufacturing sector suffers the most monetary loss compared with all the other sectors – 14.55 million euros after one day. In contrast, the output deviation for the Manufacturing sector is just –0.61% (Rank 11). The reason

Table 4. Sectors ranked by monetary loss after an information system breakdown.

Rank	NACE	Economic Sector	1 Day	90 Days
1	D	Manufacturing	–14.55	–72.74
2	K72	Computer and Related Activities	–13.53	–67.67
3	K	Real Estate and Business Activities	–10.68	–53.40
4	G	Wholesale and Retail Trade	–8.40	–42.03
5	I	Transportation, Storage and Communication	–6.94	–34.70
6	J	Finance	–3.57	–17.85
7	F	Construction	–3.51	–17.55
8	L	Public Administration and Defense	–1.91	–9.58
9	H	Hotels and Restaurants	–1.64	–8.20
10	N	Health and Social Work	–1.49	–7.49
Total (16 Sectors)			–70.89	–354.47

for the disparity is the significance of the Manufacturing sector in the Italian economy (i.e., proportion of the gross domestic product).

Monetary loss is undoubtedly the best measure for expressing the economic effects of a breakdown in information systems. After a 10% breakdown, the immediate monetary impact (one day) on the Italian economy exceeds 70 million euros. After 90 days, the monetary loss is more than 350 million euros. The Manufacturing sector, on the average, suffers 20% of the total monetary loss to the Italian economy.

Table 5. Rankings of the Italian economic sectors.

NACE	Economic Sector	OD	ML	LD	PD	WL
K72	Computer and Related Activities	**1**	**2**	**1**	**1**	**1**
J	Finance	**2**	6	5	4	**3**
I	Transportation, Storage and Communication	**3**	5	**2**	**2**	**2**
K	Real Estate and Business Activities	4	**3**	4	**3**	4
G	Wholesale and Retail Trade	5	4	7	6	6
E	Electricity, Gas and Water Supply	6	12	6	5	5
F	Construction	7	7	8	8	8
O	Other Services	8	11	10	9	9
H	Mining and Quarrying	9	15	**3**	7	7
N	Agriculture, Hunting and Forestry	10	13	11	11	10

Table 5 ranks the Italian sectors (one-digit NACE level) for each of the five impact variables. The top three ranks for each impact variable are highlighted using a bold font.

The sectors that are the most vulnerable to an information system breakdown (across the five impact variables) are Computer and Related Activities

and Transport, Storage and Communication. The other sectors differ considerably, especially when comparing the impact variables based on labor and price. However, when labor deviations, price increases and shortages of produced outputs are viewed as proxies for welfare reduction (WL), the Italian sectors that are the most affected after an information system breakdown are Transportation, Storage and Communication; Finance; Real Estate and Business Activities; and Electricity, Gas and Water Supply.

Rankings of the sectors based on the share of information systems as direct and indirect inputs in the sector production functions and using the five impact variables indicate that, regardless of the perspectives and objectives of public policy makers, a limited number of sectors can be considered to be highly vulnerable to information system breakdowns. These sectors are Transportation, Storage and Communication, which has tight interconnections with information systems; and Finance, which uses significant amounts of technology in its strategic business processes.

5. European Case Study

The impact of an unexpected information system breakdown on the Italian economic sectors strongly depends on the intensity of the use of information systems in the sectors of interest as well as in the national economy as a whole. The VIS model relies on data related to input-output relationships, prices and labor for the various sectors. Thus, the sector rankings for the Italian economy can be compared with the sector rankings for the entire European economy using input-output tables aggregated at the EU27 level.

The European information system breakdown scenario considered in this section is the same as that used in the Italian case study. The scenario involves a 10% disruption to information systems lasting for one day with a 50% recovery in capacity after five days.

Table 6 shows the top ten impacted sectors in the European and Italian economies based on the percentage output deviations after an information system breakdown (two-digit NACE codes). As before, the top three ranks and the corresponding deviations are highlighted using a bold font. Note that the percentage output deviations for the European economy are uniformly lower than those for the Italian economy. Moreover, the sectors in the European economy that are related to finance are the most vulnerable to an information system breakdown.

Table 7 shows the top ten sectors in the European economy for four impact variables (two-digit NACE codes). As expected, the sector rankings for the European economy exhibit less variation compared with the Italian sector rankings (Table 5). After, the Computer and Related Activities sector, the European sectors that are the most vulnerable to an information system breakdown are the Finance and the Real Estate and Business Activities sectors. The main difference in the top rankings for the European and Italian economies for the percentage output deviation variable is the reversed rankings of the Real Estate and Business Activities and the Transportation, Storage and Communication

Table 6. Rankings of the European and Italian economic sectors.

NACE	Economic Sector	EU27 Rank	EU27 OD	Italy Rank	Italy OD
K72	Computer and Related Activities	**1**	**–10.36%**	**1**	**–11.1%**
J67	Activities Auxiliary to Finance	**2**	**–0.63%**	5	–1.20%
J65	Finance	**3**	**–0.57%**	4	–1.25%
J66	Insurance and Pension Funds	4	–0.50%	8	–1.08%
I64	Post and Telecommunications	5	–0.48%	**3**	**–1.38%**
K71	Renting of Machinery and Equipment	6	–0.44%	10	–1.05%
K73	Research and Development	7	–0.43%	**2**	**–1.87%**
DL30	Manufacture of Office Machinery and Computers	8	–0.40%	27	–0.66%
K74	Other Business Activities	9	–0.40%	9	–1.07%
DA16	Manufacture of Tobacco Products	10	–0.36%	32	–0.62%

Table 7. Rankings of the European economic sectors.

NACE	Economic Sector	OD	LD	PD	WL
K72	Computer and Related Activities	**1**	**1**	**1**	**1**
J	Finance	**2**	**2**	**2**	**2**
K	Real Estate and Business Activities	**3**	**3**	**3**	**3**
I	Transportation, Storage and Communication	4	5	5	4
G	Wholesale and Retail Trade	5	7	7	6
E	Electricity, Gas and Water Supply	6	4	4	5
L	Public Administration and Defense	7	14	13	9
O	Other Services	8	10	10	8
A	Agriculture, Hunting and Forestry	9	6	6	7
C	Mining and Quarrying	10	9	9	10

sectors, which are ranked 3 and 4 in the European economy, but 4 and 3 in the Italian economy. Also, the Public Administration and Defense sector features in the European rankings, but not in the Italian rankings. The rankings of the European and Italian sectors based on the impact variables related to price and labor changes are similar. The main difference is the higher rankings of the Transportation, Storage and Communication, and the Financial sectors in the European economy compared with the Italian economy, A similar situation is seen for the welfare loss (WL) impact variable. The top six positions are held by the same sectors with slight changes – Transportation, Storage and Communication is ranked fourth in Europe, but second in Italy. Furthermore, as in the case of the percentage output deviation variable, the European sectors related to social services (Other Services, and Public Administration and Defense) are more vulnerable to an information system breakdown.

6. Conclusions

Due to the pervasiveness of information and communications technologies in the critical infrastructure, an information system breakdown can propagate throughout a nation's economic system, causing significant socio-economic damage. In general, the analysis indicates that the impact of an information system breakdown on a sector depends on the intensity of technology adoption.

The VIS model assists policy makers in understanding and preparing for unexpected critical events by ranking the economic sectors in terms of their vulnerability to an information system breakdown. Different impact variables incorporate perspectives ranging from economic loss to societal damage. In the Italian economy, the Computer and Related Activities, Finance, and Transportation, Storage and Communications sectors are most vulnerable to an information system breakdown when the impact is measured in terms of the percentage output deviation. In contrast, the Manufacturing sector is most vulnerable when monetary loss is used as a measure. For all the impact variables, the Finance sector is more critical at the European level than it is at the Italian level. The differences in the sector rankings are, nevertheless, useful because they provide valuable perspectives to policy makers in their decision making. For example, policy actions focused on labor can be more effective in a national economy in the long term than actions that merely minimize monetary losses.

Acknowledgements

This research was supported by the Prevention, Preparedness and Consequence Management of Terrorism and Other Security Related Risks Programme [3] of the European Commission's Directorate-General for Justice, Freedom and Security.

References

[1] O. Blanchard and N. Kiyotaki, Monopolistic competition and the effects of aggregate demand, *American Economic Review*, vol. 77(4), pp. 647–666, 1987.

[2] Directorates-General for Justice and Home Affairs, EPCIP – European Programme for Critical Infrastructure Protection, European Commission, Brussels, Belgium (ec.europa.eu/justice_home/funding/2004_2007/epcip/funding_epcip_en.htm).

[3] Directorates-General for Justice and Home Affairs, Prevention, Preparedness and Consequence Management of Terrorism and Other Security Related Risks, European Commission, Brussels, Belgium (ec.europa.eu/justice_home/funding/cips/funding_cips_en.htm).

[4] A. Dixit and J. Stiglitz, Monopolistic competition and optimum product diversity, *American Economic Review*, vol. 67(3), pp. 297–308, 1977.

[5] European Commission, i2010: A European Information Society for Growth and Employment, Commission Communication COM(2005)229 Final, Brussels, Belgium, 2005.

[6] European Commission, On the Identification and Designation of European Critical Infrastructures and the Assessment of the Need to Improve Their Protection, Council Directive 2008/114/EC, Brussels, Belgium, 2008.

[7] European Commission, Protecting Europe from Large-Scale Cyber Attacks and Disruptions: Enhancing Preparedness, Security and Resilience, Commission Communication COM(2009)149 Final, Brussels, Belgium, 2009.

[8] Eurostat, European Commission, Luxembourg (epp.eurostat.ec.europa.eu).

[9] FiFo Ost, Statistical classification of economic activities: Concordance tables NACE-ISIC, Munich, Germany (www.fifoost.org/database/nace/nace-en_2002c.php).

[10] FORMIT Foundation, VIS – The Vulnerability of Information Systems and its Inter-Sectoral, Economic and Social Impacts, Project Final Report, Rome, Italy, 2009.

[11] Y. Haimes and P. Jiang, Leontief-based model of risk in complex interconnected infrastructures, *Journal of Infrastructure Systems*, vol. 7(1), pp. 1–12, 2001.

[12] Y. Haimes, J. Santos, K. Crowther, M. Henry, C. Lian and Z. Yan, Risk analysis in interdependent infrastructures, in *Critical Infrastructure Protection*, E. Goetz and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 297–310, 2007.

[13] W. Leontief, *Input-Output Economics*, Oxford University Press, Oxford, United Kingdom, 1986.

[14] C. Lian and Y. Haimes, Managing the risk of terrorism to interdependent infrastructure systems through the dynamic inoperability input-output model, *Systems Engineering*, vol. 9(3), pp. 241–258, 2006.

[15] E. Luiijf, A. Nieuwenhuijs, M. Klaver, M. van Eeten and E. Cruz, Empirical findings on critical infrastructure dependencies in Europe, *Proceedings of the Third International Workshop on Critical Information Infrastructure Security*, pp. 302–310, 2009.

[16] E. Kujawski, Multi-period model for disruptive events in interdependent systems, *Systems Engineering*, vol. 9(4), pp. 281–295, 2006.

[17] P. Pederson, D. Dudenhoeffer, S. Hartley and M. Permann, Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research, Technical Report INL/EXT-06-11464, Idaho National Laboratory, Idaho Falls, Idaho, 2006.

[18] S. Rinaldi, Modeling and simulating critical infrastructures and their interdependencies, *Proceedings of the Thirty-Seventh Annual Hawaii International Conference on System Sciences*, 2004.

[19] S. Rinaldi, J. Peerenboom and T. Kelly, Identifying, understanding and analyzing critical infrastructure dependencies, *IEEE Control Systems*, vol. 21(6), pp. 11–25, 2001.

[20] J. Santos and Y. Haimes, Modeling the demand reduction input-output inoperability due to terrorism of interconnected infrastructures, *Risk Analysis*, vol. 24(6), pp. 1437–1451, 2004.

[21] J. Sarriegi, F. Sveen, J. Torres and J. Gonzalez, Adaptation of modeling paradigms to the critical infrastructure interdependencies problem, *Proceedings of the Third International Workshop on Critical Information Infrastructure Security*, pp. 295–301, 2009.

Chapter 14

MODELING INOPERABILITY PROPAGATION USING BAYESIAN NETWORKS

Zaw Zaw Aung and Kenji Watanabe

Abstract The modeling of critical infrastructure interdependencies is a challenging task. This paper discusses several interdependency modeling requirements and proposes a Bayesian network approach for modeling interdependencies and inoperability propagation. The approach is applied to a case study involving the Japanese critical infrastructure sectors. Survey data published by the National Institute of Land and Infrastructure Management and the Japanese National Information Security Center are used to generate conditional probability values for the Bayesian network. The approach has the flexibility to adapt to diverse critical infrastructure scenarios and interdependency structures.

Keywords: Inoperability propagation, Bayesian networks, risk assessment

1. Introduction

The modeling of critical infrastructure (CI) interdependencies is an important but challenging research problem. One of major requirements is adequate realistic data that can support the infrastructure modeling process [3]. However, data of sufficient detail, coverage and quality is not available for several critical infrastructure sectors. Due to the scarcity of data, many critical infrastructure modeling approaches are limited to certain domains, and most approaches are forced to engage scenario-based modeling.

This paper discusses the principal requirements for interdependency modeling and proposes an approach that uses a Bayesian network for interdependency modeling and inoperability propagation. The modeling approach is validated using a case study involving the Japanese critical infrastructure sectors.

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 199–212, 2010.

2. Related Work

The input-output inoperability model (IIM) developed by Haimes and coworkers (see, e.g., [4]) is based on the economic equilibrium model of Leontief [6]. Several extensions to IIM have been proposed (see, e.g., [1, 10, 11]).

The IIM formulation uses static economic data from "make" and "use" matrices provided by the Bureau of Economic Analysis. This formulation assumes that a direct correlation exists between national economic input-output data and economic sector operability/inoperability. However, such a correlation represents a crude approximation of reality. As discussed in [2], national input-output data can represent economic sector dependencies that are insignificant in some cases. In the case of Japan, almost all the defined critical infrastructures correspond to utility service sectors. These sectors have insignificant input-output table values, but they have high degrees of physical and functional interdependence.

Setola, *et al.* [12] have introduced an alternative IIM formulation. Instead of using national economic input-output data, their formulation derives the interdependency coefficients using expert interviews, where the expert data is expressed and processed using a fuzzy set methodology. Macaulay [7] has proposed a similar quantitative method for modeling interdependencies, with an emphasis on the financial sector. In particular, Macaulay develops tornado charts of economic dependencies from national input-output data, and derives data flow matrices based on a survey of experts in public and private critical infrastructure entities.

IIM yields useful estimates of sector inoperability and provides a simple method for translating these estimates into financial losses for each sector and for the economy as a whole. Nevertheless, adequate data for interdependency modeling is difficult to obtain. In Japan, for example, there have been a considerable number of service interruptions – Japan experienced six major earthquakes from 2007 to 2009 alone; and numerous post-disaster reports and case studies are available. However, the problem is that there is very little data specifically related to infrastructure interdependencies. For this reason, a thorough review of published reports is recommended as an alternative to acquiring hard data.

3. Modeling Interdependencies

Our infrastructure interdependency modeling approach is designed to address the requirements of flexibility, generality and reliability. While IIM is regarded as the most convenient way of estimating the economic impact of certain disruptions, our model uses a Bayesian network as a buffer between initial perturbations and IIM to allow flexible adjustment and risk management intervention (Figure 1). The propagated inoperability values obtained using the Bayesian network are input to the IIM for economic loss estimation and impact assessment.

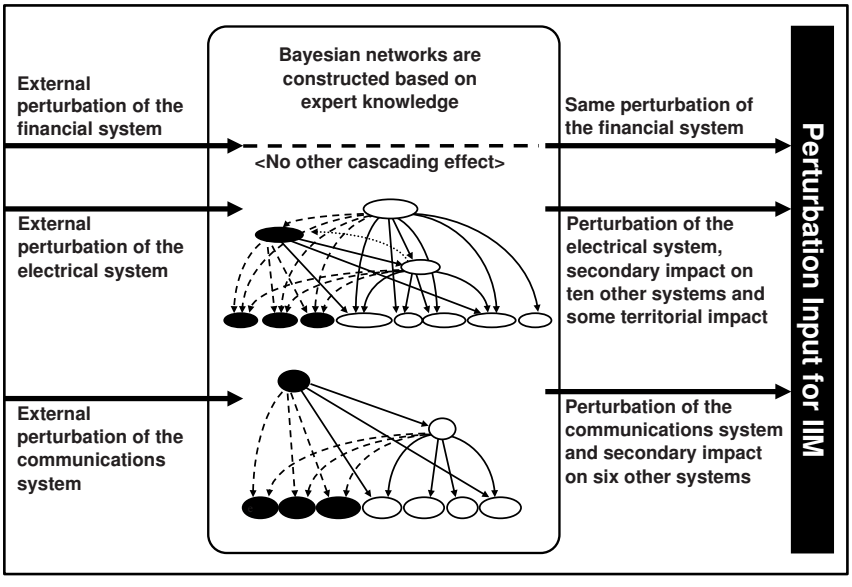


Figure 1. Bayesian network as a buffer between external perturbations and IIM.

We assume that the structure and strength of interdependencies change over time (daily, seasonal, etc.). The temporal changes may also occur during a disaster: outbreak period, emergency period and restoration period. The interdependencies during the outbreak period can be almost identical to those under normal conditions. However, the limited availability of resources during the emergency causes the interdependencies to be different from those during the outbreak period. Similarly, the interdependencies during the restoration period are different as a result of the recovery dynamics and resilience characteristics. Therefore, critical infrastructure interdependency modeling should address these situational dependencies and should adapt to the relevant disaster periods.

4. Data Sources

Our primary major data source for modeling critical infrastructure interdependencies was a 167-page technical report released in February 2009 by the National Institute of Land and Infrastructure Management (NILIM) of the Japanese Ministry of Land, Infrastructure, Transport and Tourism [5]. The report investigated the interdependencies between critical infrastructures in past disasters and presented the results in the form of tables and influence diagrams. The report has three major components: (i) data collection; (ii) two analytical models, one based on matrix equations and the other on system dynamics; and (iii) a simulation of earthquake damage spreading in the Tokyo metropolitan area. The data collection and matrix equations from the NILIM were used in our study.

Table 1. Dependencies during the Kobe and Niigata Chuetsu earthquakes.

Ref. No.	Influence Generated	Influence Received	Type	Details
1	Water	Health	Lifeline	8,850 m^3 of water had to be delivered by trucks for 47 days
16	Road	Water	Restoration	66.5% of employees could not reach work on the day of the disaster due to congestion
30	Water	Road	Alternative	Hospitals needed increased water delivery from 5-6 tons to 30 tons from locations as far as 7 km away
32	Elec.	Health	Lifeline	Failure of artificial respirators threatened sixteen lives; the respirators had to be operated manually
95	Water	Gas	Physical	Gas supply was halted to 12,463 locations due to water leakage into control systems
104	Comm.	Industry	Lifeline	One of two NTT Online Cable trunk lines between the computing center and the Kobe head office was cut
13	Road	Gas	Restoration	Gas supply system repairs in the Yamaguchi and Horikoe regions could not proceed for three days because of road damage
25	Gas	Waste	Functional	Sewage was used to cool pressurized gas
16	Comm.	Gas	Restoration	Vital SCADA data for the Nagaoka control center and Kawaguchi gas control unit was delayed by more than two hours

The data collection component of the NILIM report incorporates an exhaustive review of 65 reports on the Kobe earthquake and 52 reports on the Niigata Chuetsu earthquake. The unique CI-to-CI dependencies were extracted and categorized into the six groups listed below. Excerpts are listed in Table 1.

- **Physical Impact:** 18 cases
- **Functional Impact:** 33 cases
- **Restoration Delay:** 62 cases
- **Alternative Impact:** 43 cases
- **Common Failure:** 4 cases
- **Lifeline Impact:** 84 cases

Because it focuses on earthquake disaster management, the NILIM report does not cover all ten (officially-defined) Japanese critical infrastructures [9]. Nevertheless, it provides a good foundation for further interdependency analysis. Based on knowledge gained from literature surveys and government hearings, a survey questionnaire was created to assess the quantitative influence on the critical infrastructures.

Table 2. Survey results.

		Critical Infrastructure									Lifeline Services			
		Elec.	Gas	Water	Sewage	Comm.	Road	Rail	Harbors	Air	Transport	Finance	Health	Gov.
Critical Infrastructure	Elec.		1	3	8	4	9	16	12	8	12	8	12	16
	Gas	1		1	2	1	0	4	1	3	1	1	6	2
	Water	2	1		4	0	0	4	3	3	4	1	16	8
	Sewage	0	0	1		0	0	4	1	1	1	1	4	1
	Comm.	0	8	2	1		4	16	4	8	12	20	6	20
	Road	0	4	2	0	12		1	8	4	16	6	12	2
	Rail	0	1	0	0	0	6		3	4	4	2	0	2
	Harbors	0	3	0	0	0	0	0		0	1	0	0	12
	Air	0	0	0	0	0	0	0	0		1	0	0	0

The survey results are shown in Table 2. Each entry provides the influence of the row CI on the column CI (receiver). Note that the table shows the CI-to-CI influences as well as the influences on lifeline services.

Table 3. Influence matrix.

	Elec.	Gas	Water	Sewage	Comm.	Road	Rail	Harbors	Air
Elec.	0	0.016393	0.049180	0.131148	0.065574	0.147541	0.262295	0.196721	0.131148
Gas	0.016393	0	0.016393	0.032787	0.016393	0	0.065574	0.016393	0.049180
Water	0.032787	0.016393	0	0.065574	0	0	0.065574	0.049180	0.049180
Sewage	0	0	0.016393	0	0	0	0.065574	0.016393	0.016393
Comm.	0	0.131148	0.032787	0.016393	0	0.065574	0.262295	0.065574	0.131148
Road	0	0.065574	0.032787	0	0.196721	0	0.016393	0.131148	0.065574
Rail	0	0.016393	0	0	0	0.098361	0	0.049180	0.065574
Harbors	0	0.049180	0	0	0	0	0	0	0
Air	0	0	0	0	0	0	0	0	0

The CI-to-CI influence matrix (Table 3) was generated from Table 2 by normalizing the values based on the largest rowwise summation (= 61).

Table 4. Total dependency matrix.

	Elec.	Gas	Water	Sewage	Comm.	Road	Rail	Harbors	Air
Elec.	0.003027	0.060529	0.062064	0.139291	0.103214	0.185284	0.310372	0.249977	0.185895
Gas	0.017125	0.006658	0.018984	0.036822	0.019919	0.011660	0.079578	0.028158	0.061886
Water	0.033289	0.023033	0.003787	0.071028	0.005186	0.013347	0.082301	0.063595	0.062982
Sewage	0.000600	0.003098	0.016766	0.001303	0.001434	0.006833	0.067607	0.021723	0.022540
Comm.	0.003768	0.150483	0.039683	0.024776	0.021507	0.095532	0.284585	0.099075	0.169147
Road	0.003076	0.103339	0.042159	0.009882	0.202890	0.021722	0.080962	0.155817	0.106637
Rail	0.000625	0.029102	0.004504	0.001665	0.020331	0.100717	0.009461	0.065036	0.077227
Harbors	0.000842	0.049508	0.000943	0.001811	0.000980	0.000573	0.003914	0.001385	0.003044
Air	0	0	0	0	0	0	0	0	0

The DEMATEL method was used to obtain the total (direct + indirect) impact of the CI-to-CI influences. The resulting matrix is shown in Table 4.

Table 5. Total requirements of Japan's ten critical infrastructures.

	Elec.	Gas	Water	Finance	Rail	Logistics	Air	Comm.	Gov.	Health
Elec.	1.043578	0.025498	0.093584	0.008200	0.060693	0.011241	0.015587	0.015551	0.017824	0.024930
Gas	0.000534	1.012813	0.001717	0.001005	0.001095	0.000808	0.001316	0.001071	0.001191	0.003883
Water	0.001937	0.005211	1.105431	0.002248	0.006977	0.002976	0.003473	0.004135	0.004786	0.007713
Finance	0.059927	0.029559	0.034154	1.099556	0.232122	0.038274	0.071628	0.046012	0.020523	0.037563
Rail	0.002233	0.002142	0.002420	0.009354	1.003249	0.002407	0.002884	0.002962	0.006447	0.004207
Logistics	0.012923	0.020586	0.011587	0.008528	0.006226	1.006349	0.007863	0.015381	0.010985	0.013164
Air	0.000791	0.000630	0.000836	0.001372	0.000626	0.000505	1.005925	0.002804	0.001273	0.001525
Comm.	0.012735	0.016381	0.018865	0.032934	0.017441	0.015884	0.021257	1.154597	0.021695	0.017611
Gov.	0.001230	0.001310	0.001932	0.001498	0.000911	0.001205	0.002105	0.001109	1.000440	0.000994
Health	0.000007	0.000024	0.000054	0.000034	0.000030	0.000004	0.000006	0.000049	0.000014	1.023300

Table 5 shows the total industry-by-industry requirements for the Japanese critical infrastructures, which can be used to calculate the total industry requirements per dollar of industry output. This data, which was obtained from the input-output tables of Japan (Year 2000) published by the Statistics Bureau (Ministry of Internal Affairs and Communications), expresses the economic dependencies between critical infrastructures.

Figures 2 and 3 compare the operational dependencies obtained from Table 4 and the economic dependencies obtained from Table 5, respectively. Two indices are computed to enhance readability. The influence driving index (D) of a row infrastructure is the sum of the row entries. The influence receiving index (R) of a column infrastructure is the sum of the column entries. The indices D and R are plotted on the x-axis and y-axis, respectively.

The lifeline services (finance, health and government services) are omitted in the NILIM survey data (plotted diagram on the left); as a result, they appear to contradict each other. However, there are several interesting points to be discussed. Judging from its relative position, electricity is the high influence driving infrastructure in both figures. Communications in the right-hand-side diagrams (economic dependence) shows a high influence driving index similar to electricity. It underscores the similarity in the economic dependency patterns of electricity and communications while electricity has a much higher influence driving index than communications from the operational dependency point of

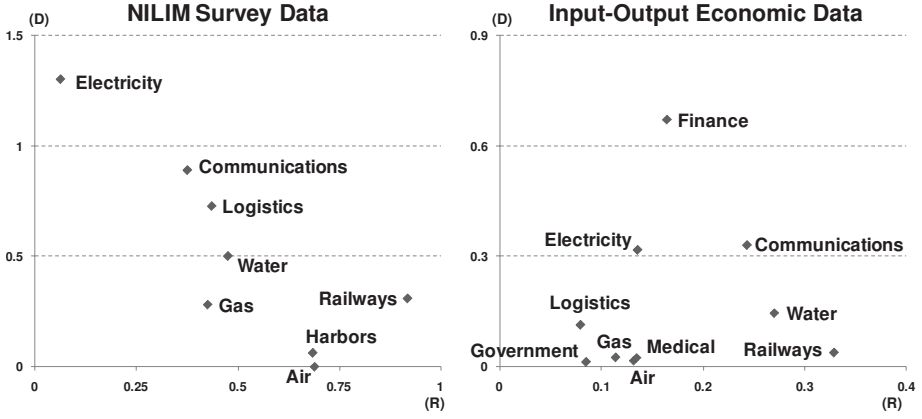


Figure 2. Influence driving (D) and receiving (R) comparison.

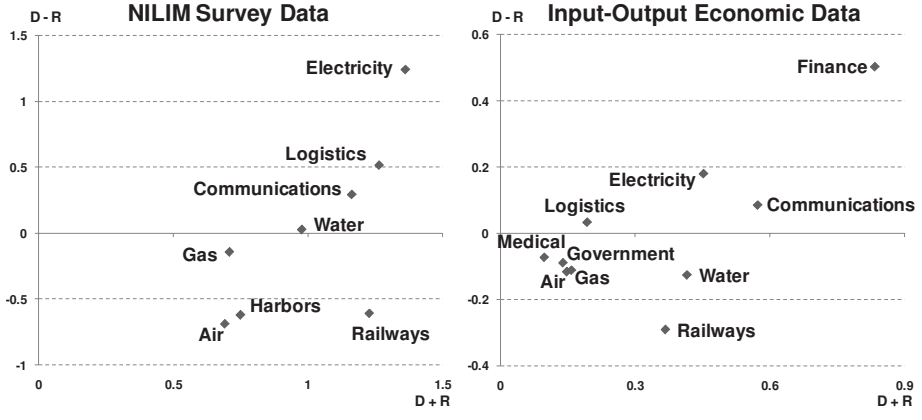


Figure 3. Net influence ($D - R$) and strength of relation ($D + R$).

view. Railways have the highest influence receiving indices from the economic and operational perspectives.

The net influence of a critical infrastructure is computed as $D-R$. If $D-R$ is positive, then the critical infrastructure has a net driving influence, otherwise it has a net receiving influence. The strength of relation of a critical infrastructure is computed as $D + R$. Note that $D - R$ and $D + R$ are used as the x-axis and y-axis, respectively, in Figure 4, which compares the operational and economic dependencies between critical infrastructures.

A net driving influence is observed for the electricity, communications and logistics infrastructures in Figures 2 and 3. The net influence of water is inconsistent because the NILIM survey considers operational and physical dependencies (water leakage into control systems is a serious threat after an earthquake). In the case of the $D+R$ metric, electricity and communications have anomalous results with respect to the economic and operational viewpoints. Communica-

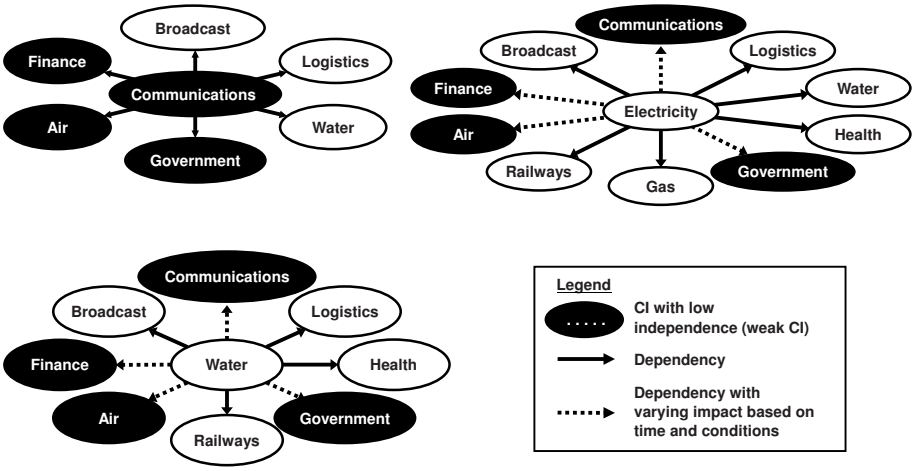


Figure 4. NISC interdependency analysis results.

tions in the right-hand-side diagrams (based on economic data) has a higher strength of relation than electricity, which reflects the higher investment in information technology by the communications sector. From an operational perspective, electricity is a fundamental requirement for every other critical infrastructure according to the NILIM survey. Logistics (road and transportation) has an insignificant strength of relation with respect to economic dependence. However, in the case of a disaster, road networks are vital for all the critical infrastructures, as demonstrated by the higher strength of relation in the NILIM survey.

Figure 4 shows the results of an interdependency analysis conducted by the Japanese National Information Security Center (NISC) [9]. The dark circles represent sectors with low dependencies (weak systems); the dotted arrows represent time-varying dependencies. Of the ten critical infrastructure sectors, broadcasting, railway, electricity, gas, medical services, water and logistics are termed as highly-independent (robust) systems. On the other hand, communications, finance, air transportation and government services are weak systems with low independence. Note that communications and broadcasting is defined as a single sector. However, they are treated separately because of their different dependency characteristics.

5. Causal Network

Figure 5 presents the causal network developed using the results of NISC's interdependency analysis. The diagram shows the first-level propagation of inoperability among the critical infrastructures. Quadrant I contains the major influence sectors – communications, electricity and water; any perturbation to one or more critical infrastructures in this quadrant will propagate to other critical infrastructures. Quadrant II contains communications and water to

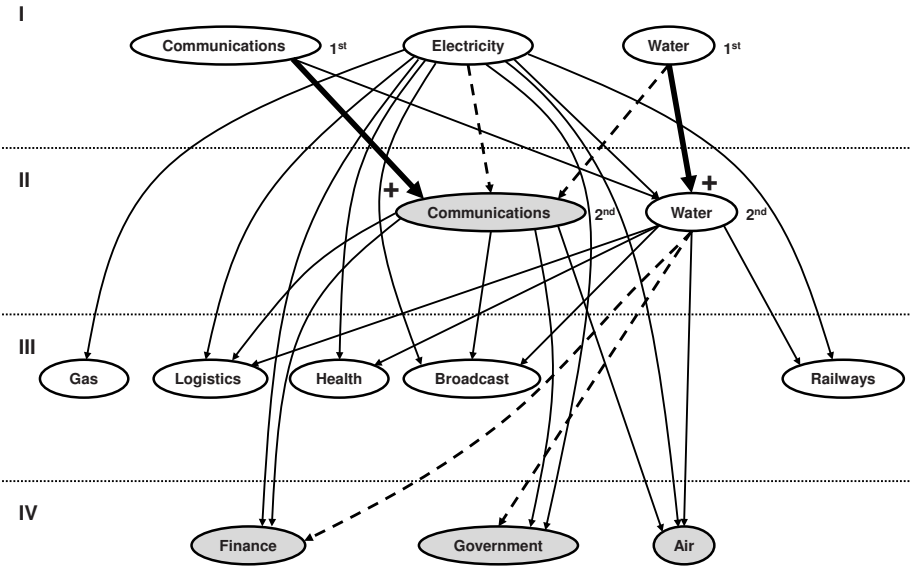


Figure 5. Causal network for Japan's critical infrastructures.

handle the interdependencies between the two infrastructures. The thick dark arrow between the first communications node in Quadrant I and the second communications node in Quadrant II expresses the fact that an inoperability perturbation in the first node (say 0.2) propagates to the second communications node as an identical value (0.2). If there are two external perturbations to communications and electricity of 0.2 and 0.2, respectively, then the propagated inoperability in the second communications node is the sum of 0.2 propagated from the first communications node and some portion of the inoperability influenced by electricity on communications. The nodes in Quadrants III and IV are infrastructures that have little or no influence on other infrastructures (that correspond to leaf nodes in the causal network). The critical infrastructures in Quadrant IV have low independence (i.e., they are weak systems) according to the NISC analysis.

Certain inconsistencies exist in the NILIM and NISC dependency results. The NILIM report focuses on earthquake damage spreading analysis and targets three types of dependency impact – physical, functional and restoration delay. On the other hand, the NISC study mainly focuses on the functional perspective. Our model focuses on the functional dependence and dependency structure of critical infrastructures during the disaster period.

6. Bayesian Networks

Bayesian networks provide a flexible formalism for expressing expert knowledge. Based on the causal network described above, we constructed a Bayesian network utilizing the influence matrices and qualitative assessments of CI-to-

CI dependencies presented in Section 4. Better results are obtained for a decision node with a larger number of states. However, it requires many more conditional probabilities and has a higher computational cost. For reasons of simplicity and for demonstration purposes, each node in the network is limited to having four states:

- **Normal:** The system is in a normal condition and is fully operational with an inoperability of 0.00.
- **Reduced:** The system is slightly perturbed and is 80% operational with an inoperability of 0.20.
- **Half:** The system is 50% operational with an inoperability of 0.50.
- **Down:** The system is completely out of service (0% operational) with an inoperability of 1.00.

We used Hugin Lite (version 6.8) to construct the Bayesian network for the first-order propagation of inoperability in the ten Japanese critical infrastructures. The network primarily targets functional dependencies and is modeled for a one-day period. The structure of the Bayesian network conforms to the NISC results and the influence levels (conditional probabilities) are based on the NILIM results and influence matrices. In addition, the qualitative assessments relied on the ratings and reasons provided by participants in the questionnaires, the functional impact obtained by mining data pertaining to previous disasters, interview notes, and NISC survey results such as the direct and time-varying impact and critical infrastructure interdependencies.

7. Future Tokyo Earthquake Case Study

Figure 6 shows the initial situation where all the critical infrastructures are in the normal operational state. In May 2006, the Tokyo Metropolitan Disaster Management Council [13] produced a damage estimate report for a predicted 7.3 magnitude earthquake occurring directly beneath Tokyo. This earthquake was assumed to occur together with a 6.9 magnitude quake beneath Tokyo Bay near the Shinagawa area [8].

Figure 7 shows the inoperability propagation due to a 6.9 magnitude earthquake. The results were obtained using estimated service disruptions of 20.5% to electricity and 18.2% to communications as the initial perturbations that were input to the Bayesian network.

Figure 8 shows the effects of a 7.3 magnitude earthquake in the same region. Estimated service disruptions of 48.6% to electricity and 38.4% to communications were used as the initial perturbations input to the Bayesian network. These external perturbations propagated into the other critical infrastructures creating varying levels of inoperability. The inoperability of communications increases from 48.6% to 58.3% due to its dependence on electricity and water supply. Of the other infrastructures, the financial system suffers the most with an inoperability of 9.5%.

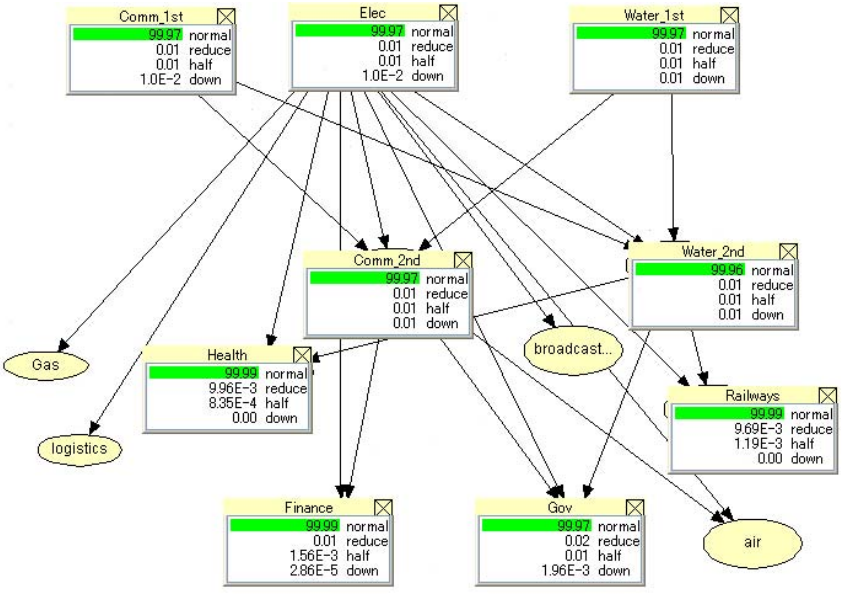


Figure 6. Initial situation in the critical infrastructure interdependency network.

8. Conclusions

The IIM is arguably the most popular method for estimating the economic impact of critical infrastructure disruptions. However, the Bayesian network described in this paper serves as a buffer between initial perturbations and the IIM, providing the flexibility to adapt to various scenarios and adjustments in interdependencies. The fidelity of the Bayesian network approach is, of course, dependent on the conditional probability assignments. The strength of the approach lies in its ability to combine expert judgment and objective data, and to refine the results as new data of higher quality becomes available.

References

[1] C. Anderson, J. Santos and Y. Haimes, A risk-based input-output methodology for measuring the effects of the August 2003 Northeast Blackout, *Economic Systems Research*, vol. 19(2), pp. 183–204, 2007.

[2] Z. Aung and K. Watanabe, A framework for modeling interdependencies in Japan's critical infrastructures, in *Critical Infrastructure Protection III*, C. Palmer and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 243–257, 2009.

[3] R. Bloomfield, N. Chozos and P. Nobles, Infrastructure Interdependency Analysis: Requirements, Capabilities and Strategy, Document No. D/418 .12101/3, Adelard, London, United Kingdom (www.csr.city.ac.uk/projects /cetifs/d418v13_public.pdf), 2009.

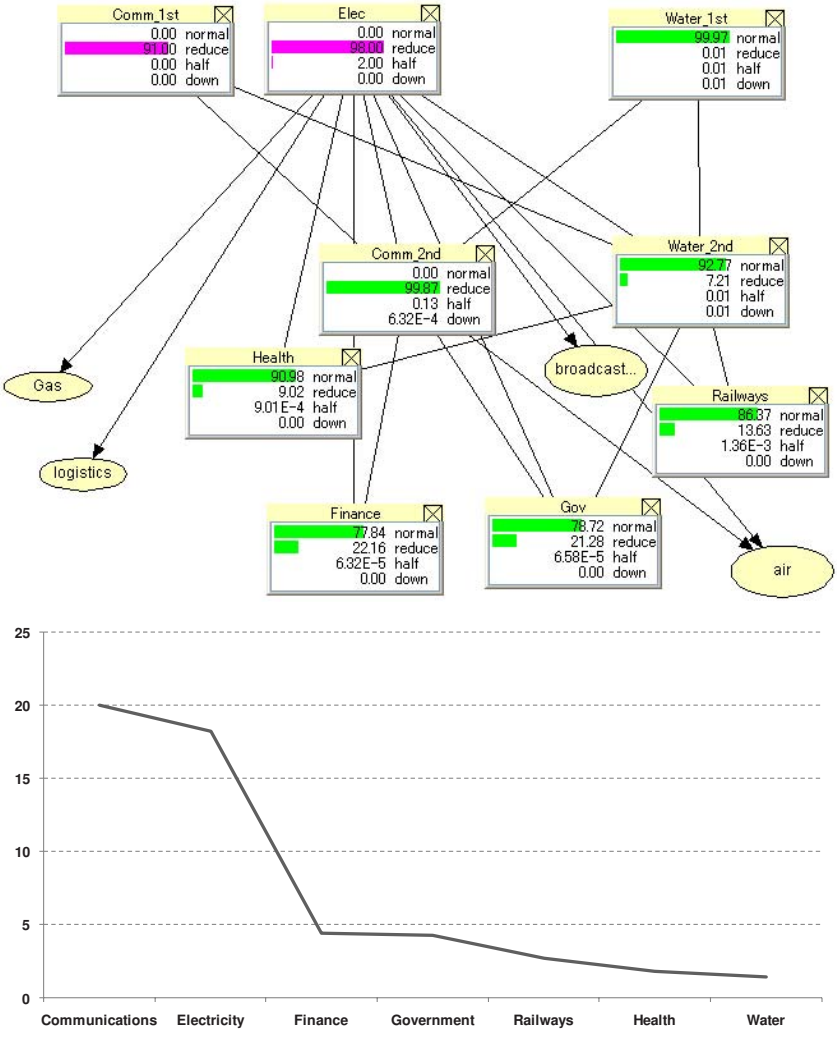


Figure 7. Inoperability propagation due to a 6.9 magnitude earthquake.

[4] Y. Haimes and P. Jiang, Leontief-based model of risk in complex interconnected infrastructures, *Journal of Infrastructure Systems*, vol. 7(1), pp. 1–12, 2001.

[5] S. Kataoka, M. Tsurata and Y. Shoji, Model Development of Interdependencies Among Critical Infrastructures and Simulation of Earthquake Damage Spreading, National Institute of Land and Infrastructure Management, Tokyo, Japan, 2009.

[6] W. Leontief, *Input-Output Economics*, Oxford University Press, Oxford, United Kingdom, 1966.

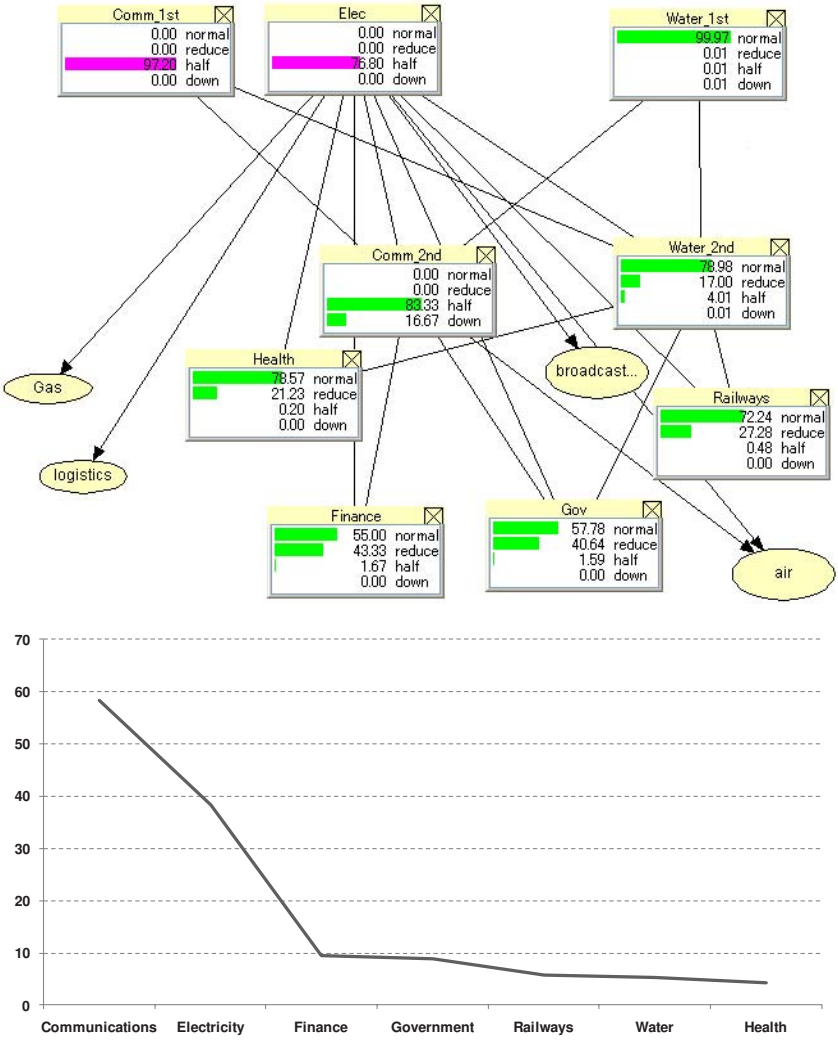


Figure 8. Inoperability propagation due to a 7.3 magnitude earthquake.

[7] T. Macaulay, Assessing operational risk in the financial sector using interdependency metrics, *International Journal of Critical Infrastructure Protection*, vol. 1, pp. 45–52, 2008.

[8] M. Nakamura, Estimated damage caused by future earthquakes occurring beneath Tokyo: From damage estimation research revised in 2006, *Journal of Geography*, vol. 116(3/4), pp. 504–510, 2007.

[9] National Information Security Center, 2007 Outputs of Interdependency Analysis, Document No. 5, Tokyo, Japan (www.nisc.go.jp/conference/seisaku/ciip/dai16/pdf/16siryou07.pdf), 2008.

[10] J. Santos, Inoperability input-output modeling of disruptions to interdependent economic systems, *System Engineering*, vol. 9(1), pp. 20–34, 2006.

[11] J. Santos and Y. Haimes, Modeling the demand reduction input-output inoperability due to terrorism of interconnected infrastructures, *Risk Analysis*, vol. 24(6), 1437–1451, 2004.

[12] R. Setola, S. De Porcellinis and M. Sforna, Critical infrastructure dependency assessment using the input-output inoperability model, *International Journal of Critical Infrastructure Protection*, vol. 2(4), pp. 170–178, 2009.

[13] Tokyo Metropolitan Disaster Management Council, Report on Damage Estimation for an Earthquake Directly Underneath Tokyo, Tokyo, Japan, 2006.

Chapter 15

RESILIENCE IN RISK ANALYSIS AND RISK ASSESSMENT

Stig Johnsen

Abstract Resilience is the ability of a system to react to and recover from disturbances with minimal effects on dynamic stability. Resilience is needed as systems and organizations become more complex and interrelated and the consequences of accidents and incidents increase. This paper analyzes the notion of resilience based on a literature survey and an exploration of incidents. In particular, resilience involves the ability of systems to undergo graceful and controlled degradation, the ability to rebound from degradation, the presence of redundancy, the ability to manage margins close to the performance boundaries, the establishment and exploration of common mental models, the presence of flexibility in systems and organizations, and the reduction of complexity and coupling. The paper describes how resilience can be included in system development and operations by considering organizations, technology and human factors. Also, it shows how past strengths and weaknesses can be considered in risk analysis to enhance safety, security and resilience.

Keywords: Safety, security, resilience, risk analysis

1. Introduction

Resilience engineering is an important aspect of safety and security due to the increased complexity and connectivity of systems and organizations. Safety is the "freedom from accidents or losses" while security is "the degree of protection against danger, loss and criminals" [10]. Resilience is "the ability of a system or organization to react to and recover from disturbances at an early stage, with minimal effect on the dynamic stability" [5]. Accidents and incidents are often due to a combination of vulnerabilities. The ability to foresee or rebound from accidents and incidents enhances both safety and security. Resilience involves the avoidance and reduction of the consequences of disturbances from a safety and security perspective.

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 215–227, 2010.

There are resilient systems that are not safe and safe systems that are not resilient; however, our goal is to ensure that systems are both safe and resilient. We focus on oil and gas installations in the North Sea, especially those that use integrated operations. Integrated operations leverage information and communication technology (ICT) to change work processes, improve decision making, enable remote operation of equipment and processes, and move functions and people onshore [11]. Integrated operations are complex and employ technologies so rapidly that learning from prior incidents is difficult because there is little, if any, experience regarding their use. Resilience in integrated operations is critical because the consequences of an accident or incident in an offshore facility can be catastrophic.

This paper attempts to define resilience in terms of a few key principles based on a review of the literature and of incidents in the oil and gas industry. The main questions are: (i) how can resilience be specified in more detail? and (ii) how can resilience be specified in order to enhance safety and security? To increase safety and security, we believe that resilience should be incorporated in a development lifecycle model [10] and in risk and hazard analysis.

2. Approach

Our approach involves the analysis of notions of resilience in the literature [4, 5, 10] along with accidents and incidents in the oil and gas industry related to integrated operations [7] in order to use resilience as a strategy to improve safety and security in complex systems. This is accomplished by specifying a few resilience (functional) principles (e.g., ability to manage margins) and using these principles to describe resilient operational techniques based on organizational, technological and human factors. A resilient organizational technique involving the management of margins could clarify organizational responsibility at the boundaries related to the overlaps between organizations and the interfaces between organizations.

In particular, our approach: (i) performs a literature review focusing on resilience as a strategy; (ii) identifies "tactical" resilience principles as goals, constraints and root causes to support resilience; and (iii) explores these principles in an operational setting in risk analysis. Our literature review attempts to explore previous incidents involving brittle practices and prior successes involving resilient practices. Based on the chain of events, we identify the conditions and the underlying constraints or root causes (e.g., management systems, culture and policies [10]), which we call "resilience principles."

2.1 Accident Models and Accident Avoidance

Accident models help identify resilient events, conditions and constraints [5]. Sequential models assume that accidents have simple linear dependencies and model accidents as malfunctions or failures using constructs such as fault trees. Epidemiological models assume that accidents have complex linear dependencies and model accidents as unsafe acts in combination with weak defenses.

Barrier models assume that accidents are caused by missing barriers or holes in barriers; in this context, resilience can be viewed as the improvement of barriers or better management of barriers using proactive indicators to signal the status of the barriers. Systemic models assume that accidents have non-linear dependencies and model accidents in terms of complex interactions, tight couplings and performance variability [12]. When interactions are complex and couplings are tight, the outcome is a normal accident. Resilience can be explored in this context as a mechanism to avoid normal accidents, i.e., to reduce complexity and/or reduce tight couplings.

Analysis of the positive aspects of safety and security can help avoid incidents and facilitate "bounce back." Consequently, we explore models and theories that have been used to describe positive characteristics of organizations and complex systems such as resilience, safety culture and high reliability organizations (HROs). We attempt to identify principles that would enable accidents to be foreseen and avoided, as well as to increase resilience in general, such as the ability to recover from an adverse situation or reduce the consequences. The notion of a safety culture can help explain accidents and avoid accidents. Indeed, the notion of a safety culture clarifies the differences between carriers in the airline industry [14] – the probability of occurrence of an airline accident varies by a factor of 42 across air carriers, regardless of the standardization of technology, organization and human competence in the airline industry [14]. Many alternative definitions of safety culture exist and there is disagreement about how the culture can be changed or improved; however, in this paper, we focus on the ability to improve safety using safety culture as an element of a change process. Finally, HRO has important positive properties [9, 15], which we explore in order to identify key resilience principles.

2.2 Improving Risk Analysis

A standard development lifecycle model [10] is a useful framework for positioning risk analysis. The steps in the lifecycle model are: conceptual development, design, implementation and operations. We attempt to integrate resilience in the lifecycle model in order to create a framework for improving safety [3]. There are several examples of how resilience can be used to increase safety throughout the lifecycle model. During the concept phase, the objectives and use of resilience can be identified. During the design phase, resilience and proactive indicators can be explored to remove or reduce hazards and incidents; scenario analysis and safety cases can help ensure that safety, security and resilience are integrated. During operations, hazards can be controlled using proactive indicators; the consequences of variability and incidents are reduced or contained by the focus on resilience.

3. Resilience Principles

Based on our exploration of the chains of events in accidents and accident recovery, and an analysis of the literature, we have identified several factors

that contribute to resilience and are applicable to integrated operations in the oil and gas sector.

Woods and Cook [19] describe an improvisation scenario involving manual system and organizational crosschecking to avoid medical administration errors. This is an example of graceful degradation that can be used as a resilience principle. Graceful degradation is a major challenge in information operations where the integration of ICT and process control systems can lead to unanticipated stoppages [7].

An HRO is alert and can foresee unwanted performance by efficiently handling local cues and local interactions. Such an organization has the ability to detect drifts towards boundaries or danger zones. On the other hand, brittle organizations do not read signals well and cannot foresee the occurrence of adverse incidents [18]. The management of margins is a good resilience principle that focuses on the boundaries of acceptable safety performance [13]. This is important in integrated operations, where the failure to manage critical operations is a major cause of incidents [7].

An HRO also has a strong focus on shared beliefs and values that facilitate collaboration, support organizational crosschecking and system insight. Also, common information and information flow across the organization enhances resilience [18]. Problems often arise in integrated operations due to the presence of multiple organizational silos with poor collaboration between ICT and process control personnel [7]. Consequently, engaging common mental models is a good resilience principle.

An HRO has the ability to handle deviations and unexpected chains of events using redundant solutions (organizations, personnel and technology) [15]. Jackson and Madni [6] stress the importance of handling incidents using alternate functions. Thus, redundancy is a key resilience principle. A major hazard in integrated operations is the loss of network communication, which can be mitigated by redundancy [7].

An HRO responds in a flexible manner to unexpected events [2]. Many accidents and incidents in the oil and gas industry can be prevented or mitigated by flexible responses [7]. Flexibility is, therefore, a key resilience principle.

Normal accidents often occur as a result of complexity and the tight coupling of systems [12]. Reducing complexity and tight couplings are key resilience principles. ICT and process control systems used in integrated operations are unduly complex and should be simplified [7].

Based on our discussion, we describe seven resilience principles:

- **Graceful and Controlled Degradation:** Proactive impact analysis must be performed and risky behavior should be identified and mitigated when system functions or barriers are failing. There should be an ability to perform a partial shutdown of functions; this should be designed in the system to ensure safety and security in the intermediate states during the shutdown process. The complementary principle is the ability of a degraded system to rebound or recover and return to normal conditions. The ability to recover is based on knowledge of the state of the system

and human intervention may be needed to aid in the recovery. Effective recovery is based on timely impact analysis and competent mobilization. Organizational competence and the appropriate technical systems can contribute to resilience. This abilities to achieve controlled degradation and rebound from adverse situations are key elements of resilience [16].

- **Management of Margins:** The ability to manage margins is a key aspect of resilience. The effective management of margins ensures that performance boundaries are not crossed; this is accomplished using proactive indicators. Another important aspect is to design for controllability. Extensive testing should be conducted to analyze the ability of a system to manage margins. In addition, testing should be based on worst case scenarios and scenarios involving human decision making in stressful environments. Sacrificial decisions, i.e., decisions that balance productivity versus safety or security, must also be a part of the scenarios. The management of margins should consider the slow erosion of margins and more dynamic sacrificial decisions that lead to the crossing of boundaries. When an optimum stress level is reached, it is necessary to identify the changes of states from positive to negative values using signals and indicators. Margins can be managed by examining trends (e.g., network traffic and network congestion) and reporting maintenance using proactive indicators. Decreases in error rates and increases in reliability can cause the risk of accidents to increase; it is important to measure and manage such drift. Awareness of risks can provide a better measure of accident potential than the actual evaluation of risk; this should also be explored when establishing proactive indicators.

- **Common Mental Models:** The use of common mental models ensures communication and collaboration across systems and organizations. Mental models play an important role in handling deviations and recovery; they also facilitate the understanding of the causes of accidents and learning from accidents [10]. Developing the appropriate mental models is important to improve resilience, but it needs careful analysis and reflection. Key stakeholders and management personnel should participate in the process; the involvement of personnel across organizational silos is key to creating a common understanding.

- **Redundancy:** Redundancy involves having alternate ways to perform a function. The function can be performed by different organizations, by different technical systems or by different procedures. Redundancy supports the ability of a system to degrade gracefully. Redundancy can be achieved via standby spares or through the concurrent use of multiple devices. However, redundancy can introduce complexity and increase the vulnerability to common cause failures. An alternative to redundancy is diversity, which is an aspect of flexibility. The use of redundancy should be assessed and improvements in safety and security should be evaluated

against the costs and unwanted side effects such as increased complexity and the risk of common failures.

- **Flexibility:** Flexibility involves diversity and having different ways of performing a function. Flexibility should incorporate error tolerance; errors should be immediately observable and reversible. Flexibility also involves improvisation (and "thinking outside the box") during stressful situations. Systems should be designed for improvisation and error tolerance.

- **Reduction in Complexity:** Complexity can be reduced by going from proximity to segregation, from common mode connections to dedicated connections, from interconnected systems to segregated systems, from limited substitution to easy substitution, from several feedback loops to few (or no) feedback loops, from multiple and interacting controls to single purpose and segregated controls, from indirect information to direct information, and from limited understanding to extensive understanding [12]. A reduction in the complexity of organizations can decrease the likelihood of accidents, especially those occurring as a result of inefficient organizational structures such as multilayered hierarchies with diffuse responsibilities and poor communication.

- **Reduction of Coupling:** Coupling can be reduced by enabling processing delays, flexibility in sequencing, flexibility in methods used, flexibility in resources, redundancies and availability of substitutes [12].

4. Resilience in Risk Analysis

The resilience principles should be incorporated in a standard development lifecycle model, which has four steps: (i) conceptual development; (ii) design; (iii) implementation; and (iv) operations. An accompanying hazard and resilience analysis must identify future risks as well as positive resilience attributes that can be engineered. Thus resilience, hazards and risks must be analyzed in terms of positive and negative factors.

Hazards, resilience and past successes (accident avoidance) should be identified using techniques such as preliminary hazard analysis (PHA), FMECA and HAZOP [10]. Accident avoidance should be explored in order to understand and support resilience. Resilience should be prioritized based on the impact on safety and cost as with other mitigating actions in regular hazard analysis. Hazards are deemed to be acceptable or not acceptable based on an assessment of hazard criticality. Unwanted side effects of resilience must be assessed and mitigated. The use of proactive indicators to signal safety levels should be discussed in all phases; also, there should be a focus on establishing common mental models. Stakeholders should participate in the entire process; they should reflect on the safety objectives, relevant hazards and resilience. Westrum [17] discusses such a process and emphasizes that organizations that focus on alignment, awareness and empowerment in the workforce are better at address-

ing underlying problems, which ultimately increases resilience. Key results from the phases must be discussed to ensure common understanding and that major hazards and resilience principles have been identified and applied properly. Operations usually involve collaboration across multiple organizations and organizational silos. The process should be performed during the conceptual development phase and should use a complete risk picture that involves perspectives from multiple organizations to ensure that safety and resilience are designed into the system.

4.1 Conceptual Development

Safety, control and resilience should be considered during the conceptual development phase. A list of key functions to be implemented in the system should be listed, and the hazards and relevant resilience principles corresponding to the functions should be identified. PHA may be used to identify hazards. The elimination of hazards and adjustments to achieve resilience should be evaluated by going through the seven resilience principles. Hazards related to boundary conditions should be described and high-level information needs related to boundary conditions should be identified together with proactive indicators.

The main results of the activities related to resilience during the conceptual development phase are:

- Specification of the safety, control and resilience objectives.
- Specification of the accountability (responsibility) of safety and resilience.
- List of functions with the appropriate hazards and resilience principles.
- List of the main boundary conditions to be controlled using proactive indicators.

4.2 Design

During the design phase, the functions to be performed are elaborated and hazard analysis is performed. Experiences from past accidents should be used to identify the hazards and risks; also, experiences from prior successes should be considered to ensure that resilience is propagated in future designs.

HAZOP analysis should be used to build in resilience during the design phase. HAZOP analysis, which is based on systems theory, assumes that accidents are caused by deviations. It has five main steps: (i) documenting and elaborating the design intentions; (ii) identifying the potential deviations from the design intentions; (iii) analyzing the reasons for the deviations from the design intentions; (iv) exploring the consequences of the deviations; and (v) exploring how the deviations and their consequences can be prevented, avoided or reduced.

The results of the HAZOP analysis include the deviations, the possible causes and consequences, and the mitigating actions that are devised with resilience in

mind. The management of margins is a key focus area in resilience engineering. Thus, the testing of boundary conditions and other resilience principles should be elaborated.

The main results of the activities related to resilience during the design phase are:

- List of major hazards in the system.
- Documentation of the critical margins, proactive indicators, and the information and reporting needs related to proactive indicators.
- Test plan focusing on the critical margins and the possibility of degraded operations and recovery.

4.3 Implementation

During the implementation phase, resilience should be integrated in the technical solution, in the organizational routines and in the knowledge and ability of the users of the system. The identified hazards and critical margins should be updated based on decisions made during the implementation phase.

Testing is a key issue related to safety and resilience; it ensures that deviations and degraded performance are handled properly. There should be at least one safe shutdown state and the transition to and from a fully operational state to each safe shutdown state should be defined and tested.

When the system has moved to a safe state, the ability to use the organization and manual procedures on the degraded system should be examined. Also, critical scenarios should be explored; these scenarios should be used in training to enhance the perception and understanding of risk.

The main results of the activities related to resilience during the implementation phase are:

- List of major hazards in the system.
- Documentation of critical margins and proactive indicators.
- Critical scenarios whose exploration increases safety and resilience, and creates the appropriate risk perceptions.

4.4 Operations

Safety and resilience should be managed during the operations phase. Hazards should be controlled and the consequences of variability or incidents should be reduced or contained. Key issues related to increasing resilience are the continuous monitoring of the system, and the tracking of indicators that identify boundary conditions and slow drift towards the boundaries.

An updated list of major hazards and indicators should be available to enhance risk perception and understanding. Dynamic indicators show the performance related to network load, stress levels of individuals in key positions,

levels of alarms and levels of gas emissions or small fires. Drift indicators, on the other hand, show long-range slow drift.

Technical and organizational drift can both impact safety. In many systems, minor daily modifications or small changes in operation can accumulate and create a risky environment. Organizational drift occurs as a result of complacency with regard to risk perceptions in the workplace, which can lead to serious incidents due to the erosion or ignorance of barriers. A safety climate questionnaire [1], which provides data about worker perceptions of safety, is a useful tool for evaluating drift.

It is important to measure and track the development of resilience in the organization as well as the system. Management plays a key role in prioritizing safety versus production. Scenario analyses [1] can be used to examine management prioritization in upward appraisals or managerial scripts. Safety cases should be used to explore emergency preparedness in the organization. Periodic audits and assessments of risk and resilience should be conducted based on unwanted incidents and successful recovery from incidents.

The main results of the activities related to resilience during the operations phase are:

- List of major hazards.
- Documentation of the critical margins and the relevant proactive indicators.
- Subjective assessment of risk.
- Audit of risks and resilience.

5. Discussion

This paper has attempted to answer two questions: (i) how can resilience be specified in more detail? and (ii) how can resilience be specified in order to enhance safety and security? With regard to the first question, based on a literature review, we have suggested a more detailed specification of resilience that describes root causes. The identification of resilience principles is based on accidents (brittle practices) and successful recovery (resilient practices). The three steps in identifying the resilience principles are:

- Identify a chain of events.
- Identify the conditions and lack of conditions.
- Identify the underlying constraints and root causes.

Different root causes are identified based on different perceptions. Thus, different approaches may engage different interpretations of resilience and identify different resilience principles. Clearly, there is no consensus on the list of resilience principles. It is, therefore, important that the principles be considered

as a set, not as individual standalone concepts. Two of the principles mentioned in this paper are also described by Rasmussen [13]: the ability to manage margins close to the performance boundaries, and the ability to achieve graceful and controlled degradation and rebound from adverse situations. These principles embody key issues related to resilience and their presence in the literature provides a degree of validation for our approach.

With regard to the second question of how resilience can be specified to enhance safety and security, we believe that the key is to consider the resilience principles during systems development and as a part of safety management. The resilience perspective improves the quality of a risk analysis. This is based on three arguments. First, the scope of past incidents explored in the risk analysis is increased; the understanding of how to avoid accidents and enhance recovery improves resilience and reduces the risk of future accidents. Second, considering current challenges in the analysis of future risk helps make the unexpected expected, leading to increased focus on graceful degradation and recovery. Note, however, that the ability to rebound and other resilience properties may increase system complexity, which can lead to accidents. Consequently, to avoid increased risk, resilience should be considered during risk analysis just like other mitigating actions. Third, there is increased focus on the management of margins and boundary conditions through the use of proactive indicators; this enhances the understanding of the key processes that influence safety.

Performing risk analyses with and without the consideration of resilience provides an opportunity to compare perspectives and mitigating actions and to identify differences. As suggested by Hale and Heijer [2], the results obtained should be measured in terms of the safety performance of the organization as well as productivity and quality gains.

6. Oil and Gas Production Systems

Safety and automation systems (SAS) are commonly used in integrated operations in the oil and gas sector. These systems comprise production control systems, process shutdown systems and safety instrumented systems. Undesirable ICT/SAS incidents typically involve general virus attacks or unanticipated network traffic. However, it is expected that directed attacks on the oil and gas infrastructure will be encountered in the future.

Key hazards impacting ICT/SAS used in integrated operations are the result of organizational, technical and human factors [8]. The hazards along with their mitigating resilience principles and associated resilience techniques are:

- **Common Failures:** Common failures impacting ICT/SAS are mitigated by graceful degradation. A technical assessment should be conducted to analyze the possibility of common failures due to the loss of power, communications and other common items. Graceful degradation should be achieved using redundant solutions. An organizational assessment should be conducted to identify structures that support graceful degradation.

- **High Network Traffic:** Large amounts of ICT network traffic that potentially impact SAS can be mitigated by graceful degradation. A technical assessment should be conducted to ensure that SAS can handle unanticipated ICT traffic.

- **Poor Collaboration:** Poor collaboration between ICT and SAS professionals can be mitigated by using a common mental model. An organizational assessment should be conducted with the goal of establishing design teams with cross-functional ICT and SAS competence. It is necessary to improve risk perceptions and awareness of the challenges when developing critical software that spans ICT and SAS. Also, it is necessary to conduct hazard analysis involving ICT and SAS personnel.

- **Virus Attacks:** Directed virus attacks that halt production can be mitigated by reducing complexity. A technical assessment should focus on hardening computers and reducing services that are connected to critical infrastructure components such as SAS. Virus attacks can also be mitigated by managing margins. A technical assessment should focus on establishing proactive indicators that identify hazard levels.

- **Communication Infrastructure Breakdown:** A communication infrastructure breakdown that causes the loss of connectivity to onshore facilities can be mitigated by graceful degradation. A technical assessment should be conducted to analyze the availability of an independent communication infrastructure. An organizational assessment should establish clear responsibility and routines; scenarios involving the loss of communications should also be tested. A communication infrastructure breakdown can also be mitigated by managing margins. In this case, a technical assessment should be conducted with a focus on indicators and reporting network traffic and loads.

An assessment of frequency and severity must be performed to prioritize the mitigating actions. Using the identified resilience principles tends to make the list of mitigating actions more complete and helps cover more of the relevant issues. This approach can also be used to improve hazard analysis when resilience is required.

7. Conclusions

Resilience is a highly desirable property for critical infrastructure assets. Resilient systems can react to and recover from disturbances with minimal effects on dynamic stability. Our strategy for incorporating resilience in system development and operations is accomplished by considering organizational, technological and human factors issues. The stragegy is also promising because it engages known strengths and weaknesses in risk analysis to enhance safety, security and resilience.

References

[1] R. Flin, Erosion of managerial resilience: From Vasa to NASA, in *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D. Woods and N. Leveson (Eds.), Ashgate, Aldershot, United Kingdom, pp. 223–233, 2006.

[2] A. Hale and T. Heijer, Defining resilience, in *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D. Woods and N. Leveson (Eds.), Ashgate, Aldershot, United Kingdom, pp. 35–40, 2006.

[3] Health and Safety Executive, Organizational Change and Major Accident Hazards, Chemical Information Sheet No. CHIS7, Caerphilly, United Kingdom (www.hse.gov.uk/pubns/chis7.pdf), 2003.

[4] E. Hollnagel, C. Nemeth and S. Dekker, *Resilience Engineering Perspectives – Remaining Sensitive to the Possibility of Failure*, Ashgate, Aldershot, United Kingdom, 2008.

[5] E. Hollnagel, D. Woods and N. Leveson (Eds.), *Resilience Engineering: Concepts and Precepts*, Ashgate, Aldershot, United Kingdom, 2006.

[6] S. Jackson and A. Madni, A practical framework for the architecting of resilient enterprises, *Proceedings of the Third Resilience Engineering Symposium*, pp. 125–132, 2008.

[7] S. Johnsen, Suggested proactive indicators to be used in the oil and gas industry based on a survey of accidents in the industry, presented at the *European Safety and Reliability Conference*, 2009.

[8] S. Johnsen, T. Skramstad and J. Hagen, Enhancing the safety, security and resilience of ICT and SCADA systems using action research, in *Critical Infrastructure Protection III*, C. Palmer and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 113–123, 2009.

[9] T. LaPorte and P. Consolini, Working in practice but not in theory: Theoretical challenges of "high-reliability organizations," *Journal of Public Administration Research and Theory*, vol. 1(1), pp. 19–48, 1991.

[10] N. Leveson, *Safeware: System Safety and Computers*, Reading, Massachusetts, 1995.

[11] Norwegian Ministry of Petroleum and Energy, Om Petroleumsvirksomheten, Stortingsmelding No. 38 (2003-2004), Oslo, Norway, 2004.

[12] C. Perrow, *Normal Accidents: Living with High Risk Technologies*, Princeton University Press, Princeton, New Jersey, 1999.

[13] J. Rasmussen, Risk management in a dynamic society: A modeling problem, *Safety Science*, vol. 27(2-3), pp. 183–213, 1997.

[14] J. Reason, *Managing the Risks of Organizational Accidents*, Ashgate, Aldershot, United Kingdom, 1997.

[15] K. Roberts, Some characteristics of one type of high reliability in organizations, *Organization Science*, vol. 1(2), pp. 160–176, 1990.

[16] G. Sundstrom and E. Hollnagel, Learning how to create resilience in business systems, in *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D. Woods and N. Leveson (Eds.), Ashgate, Aldershot, United Kingdom, pp. 235–252, 2006.

[17] R. Westrum, Removing latent pathogens, presented at the *Sixth International Australian Aviation Psychology Conference*, 2003.

[18] R. Westrum, A typology of resilience situations, in *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D. Woods and N. Leveson (Eds.), Ashgate, Aldershot, United Kingdom, pp. 55–65, 2006.

[19] D. Woods and R. Cook, Incidents – Markers of resilience or brittleness? in *Resilience Engineering: Concepts and Precepts*, E. Hollnagel, D. Woods and N. Leveson (Eds.), Ashgate, Aldershot, United Kingdom, pp. 69–76, 2006.

Chapter 16

A MANUFACTURER-SPECIFIC SECURITY ASSESSMENT METHODOLOGY FOR CRITICAL INFRASTRUCTURE COMPONENTS

Thomas Brandstetter, Konstantin Knorr and Ute Rosenbaum

Abstract Protecting critical infrastructure assets such as telecommunications networks and energy generation and distribution facilities from cyber attacks is a major challenge. However, because security is a complex and multi-layered topic, a foundation for manufacturers to assess the security of products used in critical infrastructures is often missing. This paper describes a structured security assessment methodology that is specifically designed for use by manufacturers during product development. The methodology, which incorporates risk analysis, theoretical assessment and practical assessment, anticipates operational security challenges before products are deployed in critical infrastructures.

Keywords: Critical infrastructure components, security assessment, risk analysis

1. Introduction

Security assessments of critical infrastructure components (CICs) differ from those of classical IT systems in that availability and integrity of the components trump confidentiality [18]. Also, it is often impossible to perform regular patching for these components; consequently, the patching cycles typically follow planned maintenance schedules.

Manufacturers of CICs such as control systems for energy generation are facing increasing security demands for their products from customers and regulatory bodies. The central question to be answered is: what security problems related to the products should be remediated? This paper describes a three-step security assessment methodology to help answer this question. The steps are: (i) evaluate the individual security risks associated with the design and architecture of the product, and identify the risks that cannot be tolerated and

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 229–244, 2010.

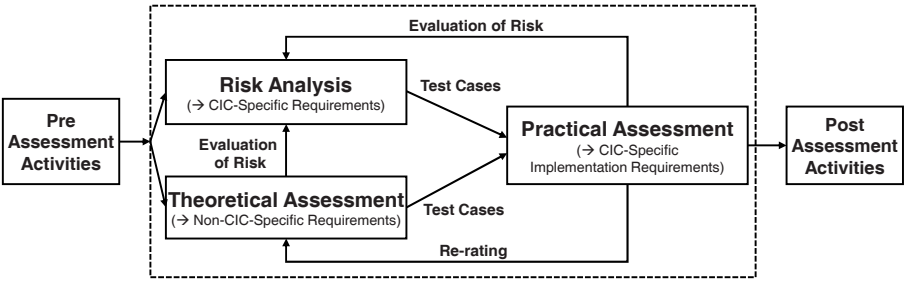


Figure 1. Security assessment methodology phases.

must be mitigated, accepted or transferred; (ii) determine how the product ranks with regard to security requirements published by potential customers and regulatory bodies; and (iii) perform practical tests of the CIC in operational environments to uncover implementation and configuration flaws.

The security assessment methodology described in this paper is intended to address the needs of manufacturers during the development of CICs. The methodology, which is pragmatic, cost-effective, generic, flexible and built on CIC industry standards, has been successfully applied to several CICs.

2. Security Assessment Methodology

Figure 1 presents a high-level overview of the security assessment methodology. The methodology starts with the pre-assessment phase, which involves the preparation and signing of the project agreement, and includes a definition of the assessment scope (CIC version and release), milestones, location, timeline, costs, staffing, liability, etc. The subsequent risk analysis phase determines the individual information security risk levels arising from the technical design and architecture of the CIC by performing a risk analysis and deriving specific security measures for the CIC. The theoretical assessment phase examines how far security measures mandated by standards, regulatory requirements and generic customer requirements are implemented in the CIC. This typically includes technical, organizational and process aspects. Security measures specific to the standard, but not specific to the component under test, are checked. The practical assessment phase involves the application of manual procedures and automated tools in a suitable testing environment to determine the potential for exploiting vulnerabilities. The final post-assessment activities involve presenting a final report to the product manager, issuing a security assessment methodology confirmation and suggesting solutions for the security flaws.

In recent years, many manufacturers have begun to tie security activities to the product development process. Our security assessment methodology follows this approach. The various phases can be performed during different development milestones of a CIC. Risk analysis and theoretical assessment should be completed as early as possible (e.g., during product planning or design). In the case of a practical assessment, the product must be in a "testable" state,

i.e., a suitable test environment must be available. Note that it is possible to perform only selected parts of the methodology, e.g., risk analysis and practical assessment, or theoretical assessment and practical assessment. However, partial assessment is not recommended because important synergies are lost.

2.1 Pre-Assessment

The pre-assessment phase is the preparatory phase of a security assessment. During this phase, the various participants agree on the project details. After an initial discussion and using a predefined questionnaire, the proposed target of evaluation (TOE) is briefly analyzed. This analysis identifies the security goals and the scope and depth of the assessment, which must be agreed upon by all the involved parties. The detailed specifications of all the subsequent assessment phases are determined in this initial analysis. Depending on the security goals of the TOE and its market placement, it is necessary to decide on the standards to be included in the theoretical assessment and the tests to be included in the practical assessment. This helps determine the overall effort for conducting the assessment and a realistic cost estimate and timeline, all of which assist in planning the subsequent phases.

2.2 Risk Analysis

ISO/IEC 27005:2008(E) defines information security risk as the potential that a given threat will exploit vulnerabilities in an asset or group of assets and thereby cause harm to the organization. It is measured in terms of a combination of the likelihood of an event and its consequence. Risk analysis is the practice of determining the threats to which an organization or system is exposed and the potential harm. The risk analysis approach in our security assessment methodology is based on established risk analysis techniques [9, 17], but is adapted to the specific needs of CIC manufacturers by defining a risk management framework that is designed to be cost-effective by using a workshop to conduct the analysis.

Risk Analysis Steps The risk analysis steps follow the ISO and NIST standards [9, 17]. First, the CIC assets are identified. Next, threats that exploit asset vulnerabilities are determined and the probability of a successful attack is estimated. Finally, the impact of an attack is described and classified. The associated risk is computed by combining the probability of a successful attack and its impact.

During these steps, the CIC risk must be seen from two points of view. First, what risks does the CIC pose to the manufacturer's business model? Second, what risks arise during CIC operation due to its technical architecture?

Both views have to be considered during risk analysis. Depending on the point of view, the assets are quite different. To the manufacturer, the assets may represent intellectual property and licensing schemes; often, considerable threats and associated risks can be identified for these assets. To the operator,

Table 1. List of potential attackers.

Attacker	Comment
Third-party consulting vendor	Attacks against licensing scheme, e.g., by selling high-end products to the customer but ordering and paying for low-end products with less functionality.
Competitor	Competitor seeks proprietary information about a product, e.g., to better position his products, or to copy the product or some of its functionality.
Hacker (organized)	Hacker attempts to control a CIC on behalf of a third party.
Hacker (curious)	Hacker accidentally breaks into a system and tries to gather information.
Malware	Malware infects a CIC network accidentally or intentionally.
Employee (manufacturer)	Employee with access to confidential development data steals or destroys the data.
Cyberterrorist	Cyberterrorist disrupts CIC service to cause panic or to extort money.

the assets are quite different; they may, for example, correspond to personal data that has to be kept confidential or systems whose availability must be maintained. Because the risk analysis is performed by the manufacturer and not by the system operator, several deployment scenarios may have to be analyzed to determine the possible impact to an operator.

Practical Risk Analysis As with all the phases of the security assessment methodology, security efforts are balanced with economic aspects. Therefore, the risk analysis is conducted in the form of a group workshop, which is typically one to three days long, depending on the complexity of the CIC. The workshop is conducted by an experienced assessor who is a security expert and can serve effectively as a moderator. The workshop participants represent all the various phases of the product lifecycle, e.g., product development, system testing, service, sales and marketing, and product management. Ideal participants would have comprehensive knowledge and significant experience in product design and architecture (product development); use cases and deployment in customer environments (service); and competitors and sales channels (sales and marketing) necessary to understand the risks related to intellectual property and license fraud related risks.

Predefined lists of potential attackers, targets, threats and impacts are used to provide examples, raise discussion and check for completeness. The list in Table 1 is derived from generic lists [9, 11] that are adapted to CIC needs.

Risk Matrix

Impact	Probability	Not Likely	Possible	Likely	Very Likely
	Negligible	1	1		4
	Moderate	1	2	6	
	Critical		6	5	1
	Disastrous	4	5	3	1

Risk Matrix After Measures

Impact	Probability	Not Likely	Possible	Likely	Very Likely
	Negligible	1	2	3	2
	Moderate	1	5	2	
	Critical	6	6		
	Disastrous	10	2		

Figure 2. Sample risk analysis results.

Experience has shown that this approach yields an efficient and useful risk analysis in a relatively short amount of time.

Risk Management In order to reduce the effort involved in risk analysis, a predefined risk management model is used following a qualitative rating. Four categories are defined for rating the probability of a successful attack and the impact; these categories take into account CIC-specific aspects such as availability. If necessary, the descriptions of the categories may be clarified and amended by product-specific aspects during the workshop. The results of the risk analysis are presented as a 4×4 risk matrix. The ratings of the risks, i.e., the definitions of the risks that are considered to be acceptable and those that need to be mitigated, are also predefined.

Figure 2 shows the results of a threat analysis of an energy management system. Initially, several non-acceptable risks were identified, one is classified as "Probability: Very Likely" and "Harm: Disastrous." However, the risk may be reduced to an acceptable level after selecting and implementing countermeasures. The initial analysis can be completed in a two-day workshop.

Four categories for the probability of a successful attack and the resulting impact are offered. Using an even number of values ensures that no midpoint value can be chosen, which eliminates indecision in arriving at an assessment.

Risk Analysis Results For the workshop participants, the immediate results provide a better understanding of the threats to which the CIC is exposed, because the participants themselves "discover" the threats to the CIC. They gain understanding of the need for security measures and learn to act accordingly. The workshop gives them a forum to discuss security aspects. Also, the workshop provides training and awareness opportunities for non-expert participants. The overall effect of the workshop is far superior to that of a risk assessment conducted by an external consultant, which is based entirely on technical input from the development team.

The risk analysis provides the project manager with a list of the identified risks along with their ratings, identifying the risks that must be mitigated according to their priority. The risk analysis also provides valuable inputs to the other phases of the security assessment (e.g., the list of critical assets and identified risks that form the basis of the practical assessment). Note, however,

that it is sometimes the case that the threats identified by the theoretical assessment and the practical assessment have to be added to the risk analysis.

2.3 Theoretical Assessment

This section discusses the theoretical assessment approach, which is designed to assess the security level of CICs with regard to generic security standards. Note that the term "standard" does not accurately fit the documents (guidelines, recommendations, regulatory documents and laws) referred to in this work. However, for the sake of simplicity we use this term throughout the paper.

In general, customers who operate critical infrastructure assets have published their own generic security requirements; sometimes based on regulatory requirements for operation, sometimes directly referring to existing standards. Fulfilling customer requirements is a prerequisite for selling products. Consequently, it is important for a manufacturer to know how well its products satisfy the requirements. The first step in the theoretical assessment approach is to decide which standard is relevant to the CIC. Next, a questionnaire is created based on the selected security standard if one is not yet available. Finally, the approach uses the results of the interviews of CIC experts based on the questionnaire to arrive at a security assessment.

Selecting Standards Numerous CIC-related security standards have been published (see, e.g., [5, 21] for a list of more than 50 important standards). For example, the NERC CIP standard [10] is published by an industry regulatory body and focuses on the operation of CICs. A U.S. information sharing center has published a "procurement language" [8] that focuses on the development of CICs. From the point of view of product management, the diversity of security documents presents a challenge and an opportunity. On the one hand, it complicates the task of selecting the standards used in an assessment. On the other hand, many of the documents contain agreed-upon security requirements that are seen in many tenders.

Developing Questionnaires A theoretical assessment uses one questionnaire per standard. A questionnaire has a generic structure that is independent of the standard, but its content is structured according to the pattern of the underlying standard. The content reflects the requirements of the standard being assessed.

Relevant requirements have to be derived when a standard does not specifically address a manufacturer's product. For example, NERC CIP [10], a standard for operators of bulk electric systems, requires that operators maintain logs of security events for 90 calendar days and that these logs be reviewed regularly. Merely checking if a product supports logging is insufficient because most systems already support logging. The intent of the standard is for products to incorporate state-of-the-art logging technology.

In contrast with the NERC CIP standard, the U.S. Cyber Security Procurement Language for Control Systems [8] summarizes security principles that should be considered when designing and procuring control system products. Therefore, it is well-suited as direct input for a questionnaire. Because the scope of the document is broad, some requirements will not be applicable to a given product and have to be marked as not applicable during the assessment. For each requirement of the selected standard, one or more corresponding questions are derived so that they can be answered with "Yes," "No" or "Not Applicable." Predefined intermediate answers such as "Dependent on Contract" are also permitted. It expresses the fact that a requirement is not fulfilled by the default product offering but, depending on the contract, can be offered as an additional feature. Without this option, different answers are possible: the answer "Yes" because the requirement can be fulfilled, and "No" because the standard offering does not fulfill the requirement. In the case of automatic evaluation, the predefined answers are mapped to a value in a predefined range and are used to calculate the "average compliance." Additionally, a comments field is provided for each question to enable respondents to clarify their answers.

Conducting Interviews The theoretical assessment is conducted in a workshop environment where experienced security assessors (who are not involved product development) conduct interviews of product experts and guide them through the questionnaire. Depending on the goal, different assessment depths are possible: (i) merely documenting the oral statements of the interviewees; (ii) checking and analyzing the available documentation; or (iii) deriving tests for the subsequent practical assessment phase. The assessment depth can be varied on a per-requirement or per-section basis.

In practice, we perform spot tests for some topics and also use some of the theoretical assessment topics to derive topics for the practical security assessment. This combination ensures that all the intended security mechanisms exist and are implemented securely, thereby raising the level of confidence of the assessment.

Analyzing and Reporting Results The results of the theoretical assessment include the level of compliance with the requirements and the identified deviations. The results support a detailed analysis of the shortcomings of the product, and are suitable for presentation to management. The degree of deviation is apparent without delving into the technical details; also, security becomes measurable.

Figure 3 shows the results for a NERC CIP benchmark of two versions of a CIC. The initial version of the product incorporates backup functionality without a documented recovery concept. The new version incorporates additional security functionality and documentation, with the documentation, in particular, improving the CIP 009 rating.

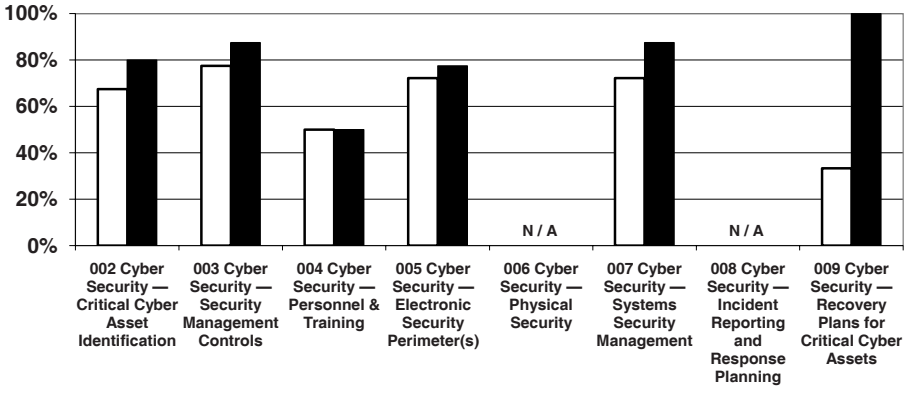


Figure 3. Sample NERC CIP compliance theoretical assessment results.

2.4 Practical Assessment

Practical assessment, the next phase of the security assessment methodology, evaluates the resilience of a CIC to hacking attacks. This phase is introduced to detect exploitable vulnerabilities and potential security flaws in a CIC taking into account state-of-the-art hacking techniques and tools. The results of the risk analysis and theoretical assessment phases are used as input when generating attack patterns and testing tasks. Practical assessment complements these phases by verifying the actual implementation of the security features.

We begin by discussing a sample test task to explain how a practical assessment works. NERC CIP 005-1 R4 requires a review of controls for default accounts, passwords and network management community strings. The first task is to identify the credentials in a target system; this is typically performed using an automated tool (e.g., Nessus Security Scanner [19]) or by manually reviewing the credential store on the system. Next, the credentials are reviewed for known default values or easily guessed credentials. Insecure credential combinations are then documented. Finally, the identified username/password combinations are tested to determine if they permit access to the system.

A suitable test system is necessary to conduct a practical assessment. A test system at a manufacturer is suitable for conducting in-house tests. Factory and site acceptance tests are typically performed during the handover of a CIC from the manufacturer to the customer; these tests also provide an excellent environment for a practical assessment. Alternatively, a practical assessment can be performed at the customer site. In this case, special care must be taken to define the testing tasks as they must not affect normal operations. As with the entire security assessment methodology, the practical assessment follows a structured process, which is presented in Figure 4.

Planning and Preparation The practical assessment test tasks are initially collected and categorized based on input from the preceding risk analysis and theoretical assessment phases, and on an agreement between the project

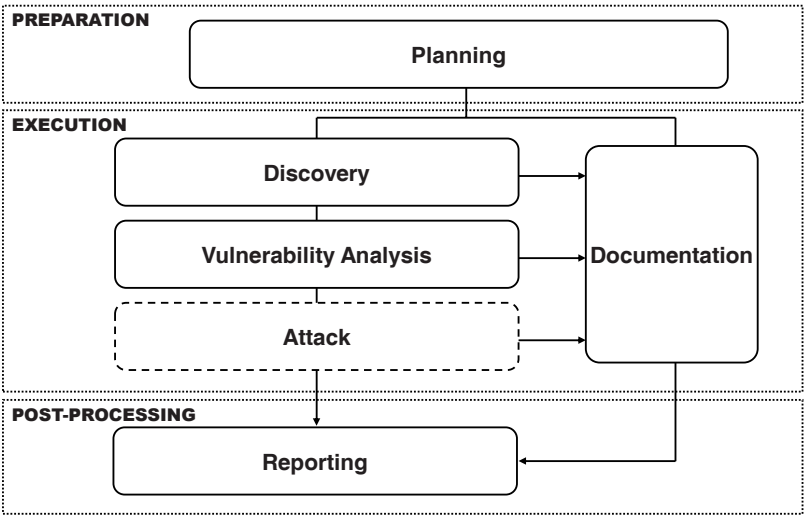


Figure 4. Practical assessment steps (with optional attack step).

manager and the assessment team. In this step, the assessor uses a structured assessment plan (Table 2) to evaluate the scope of the subsequent tasks, enabling the depth and intrusiveness of the assessment to be controlled.

Assessment The practical assessment process closely follows the steps used by real hackers. A hacker initially tries to gather information about the target via discovery and reconnaissance activities. This information is reviewed and analyzed for potential vulnerabilities during the vulnerability analysis phase. Finally, in the attack step, the hacker attempts to exploit certain vulnerabilities and launch real attacks against the system.

A practical assessment begins with the discovery step, where information gathering tasks are carried out to collect information about the target using active or passive tools. A vulnerability analysis is conducted using the collected information; this is done by manually reviewing the information for indications of potential flaws. For each flaw, the assessor attempts to estimate the potential of successful exploitation and the criticality. This must be done because hackers typically work their way from "low hanging fruit" to more complex attacks. The review helps identify the most promising entry points for further attacks.

In the practical attack step, the assessor attempts to exploit an identified vulnerability and document the extent to which the intrusion attempt is successful. The attack step involves both active and passive testing. Active testing uses invasive tools and techniques to gain access to the target or to crash a certain service. Passive testing mainly involves a configuration review (invasive tools may not be used because they can impact system availability). Strong dependencies exist between all the steps as new findings are fed back into succeeding test activities. The practical attack step is optional because it may be

Table 2. Sample practical assessment plan (N: network; P: platform).

ID	Sect.	Module	Task	Tool
101	N	Network survey	System enumeration	ipconfig/ifconfig
102	N	Network survey	System identification	nmap
103	N	Network survey	Information leaks	Wireshark
104	N	Port scan	Service enumeration	Nessus
105	N	Port scan	Service identification	Nessus
106	N	Port scan	Error checking	hping
107	N	Port scan	Protocol response verification	nmap
108	N	Port scan	Packet response verification	nmap
109	N	Port scan	Distributed TCP/IP analysis	Unicornscan
110	N	Perimeter review	Security analysis (Level 1)	cisecurity (rat)
111	N	Perimeter review	Network security review	Checklist
113	N	Perimeter review	Switch security configuration	Checklist
114	N	Perimeter review	Router hardening test	Cisco Torch
115	N	Perimeter review	Router security configuration	Checklist
116	N	Perimeter review	Firewall hardening test	ccsat
117	N	Perimeter review	Firewall security configuration	Checklist
118	N	Perimeter review	IDS security analysis	Manual checking
119	N	Perimeter review	Trusted sys. security analysis	Manual checking
121	N	DoS test	DoS vulnerability analysis	Manual checking
122	N	DoS test	DoS testing	datapool 3.3
123	N	DoS test	DoS testing	netcat
124	N	DoS test	DoS risk analysis	Manual checking
201	P	Windows/all	Baseline security analysis	MBSA
202	P	Windows/all	Security analysis (Level 1)	cisecurity (win)
203	P	Windows/all	Security testing (Level 1)	Manual testing
204	P	Windows/all	Security analysis (Level 2)	GFI Languard
205	P	Windows/Svr2003	Security testing (Level 2)	MS SCW
208	P	Unix/all	Security analysis (Level 1)	cisecurity (Unix)
209	P	Unix/all	Security testing (Level 1)	Manual testing
210	P	Unix/all	Security analysis (Level 2)	COPS
211	P	Unix/all	Security testing (Level 2)	Bastille
214	P	All	Login credential verification	John the Ripper

sufficient to gather information about the target and review it for indications of vulnerabilities rather than executing an attack.

Finally, note that each test task has two possible outcomes: the intrusion attempt either succeeds or it fails. Both outcomes must be noted to comprehensively document the test; this also gives product developers a better view of the security aspects of the product that have been addressed properly.

Reporting The findings (i.e., discovered security flaws) are documented in a report using a predefined structure. Table 3 presents example findings from the energy management system assessment described above. The example focuses on a test of the ability to log security-relevant information such as brute-force attacks on accounts at the operating system level. During the discovery phase, a port scan revealed typical server message block (SMB) ports in the TCP range of 135–139 and 445. During the subsequent vulnerability analysis phase, the ports were tested for the null-session feature, which enables an attacker

Table 3. Sample practical assessment finding.

Headline	Account login auditing disabled on application server.
Criticality	HIGH
Vulnerability Location	Windows OS auditing policy of application server with hostname `appserver.localdomain`.
Description	Logging and auditing settings at the OS level were reviewed to check for proper audit trail generation. During the review it was noted that login attempts at the OS level were not audited, regardless of whether they were successful or not. This enables an attacker to conduct a brute-force attack on an account without being detected. If security-critical information is not recorded, there is no trail for forensic analysis. Discovering the cause of problem or the source of the attack may become more difficult or impossible.
Prerequisites	For an actual attack (e.g., brute-forcing an account), the attacker would need network access to the system.
Standards Violated	NERC CIP 007-1 R 5.1.2; NERC CIP 007-1 R 6.3
CWE	778
Countermeasure	The logging level must be set appropriately for security-relevant items like account login activity. Enable account logging at the OS level.

to gather information about user account names and other account details at the operating system level. With this information, a brute-force attack for determining the passwords of existing user accounts was started, upon which the log entries were reviewed for appropriate tracking details. In the example, the logging subsystem failed to document the existence of the attack because of an inappropriate configuration.

The report is an important tool for quality control because it verifiably demonstrates that all the sections chosen in the planning step have been covered during the practical assessment. Also, it proves that the entire scope of the practical assessment phase has been completed.

2.5 Post-Assessment

The post-assessment activities of the security assessment methodology include, but are not limited to, the final report, the communication of the findings, the issuance of a security assessment methodology confirmation, and support for addressing the security flaws identified in the assessment. The security assessment methodology results are documented in a final report comprising the three sub-reports from the risk analysis, theoretical assessment and practical assessment phases along with their relationships. The content of the final report is typically confidential and is, therefore, delivered only to the project manager,

who then becomes the owner of the report. If required, a confirmation about the assessment is generated for the project manager that states the detailed CIC version, size and date of the assessment, and confirms that the CIC was assessed and describes the security issues addressed in the product.

The next step for the project manager is to decide how to proceed with the results of the security assessment, especially the identified risks, the shortcomings related to the standards, and the implementation and configuration flaws. Entries in the error tracking database corresponding to the product have been successfully used for emergency (short-term) mitigation projects. Other findings can be addressed via change requests and subsequently by new security requirements for the product. Support for these activities is not part of the security assessment methodology, but they are, nevertheless, very important to enhance product security.

3. Discussion

This section discusses the applicability of the security assessment methodology to CICs, compares the methodology with related work in the field and identifies future areas of research.

3.1 CIC Applicability

The risk analysis phase of the security assessment methodology uses generic security standards (e.g., [9, 17]). This has been done on purpose because the generic method described in these standards is well-suited to CICs. The general principle followed in designing the security assessment methodology was to reuse as much as possible of existing methodologies and adapt them to CIC needs where necessary. The adaptations for the three phases of the security assessment methodology are:

- **Risk Analysis:** While the risk analysis phase is based on [9, 17], the workshop and, in particular, its inclusion of participants with experience in CIC development and management are CIC-specific.
- **Theoretical Assessment:** The use of CIC-related standards [4, 8, 10] in the questionnaires makes this phase CIC-specific. In most critical infrastructure domains, the main standardization bodies have decided not to use generic security standards such as the ISO 2700x series, but to adapt these standards to reflect specific domain requirements.
- **Practical Assessment:** The tools and test cases are, by necessity, CIC-specific. For example, CIC-specific protocol fuzzers have to be used because CICs engage proprietary protocols.

3.2 Related Work

To our knowledge, this is the first security assessment methodology that combines the three phases, risk analysis, theoretical assessment and practical

assessment, in a pragmatic and cost-effective manner for use by CIC manufacturers.

Risk analysis is a fairly mature area (see, e.g., [9, 17]). The main focus of risk analysis as used in our work is to identify risks stemming from the design and architecture of a product. This is in contrast to other published methodologies, such as OCTAVE [1] and CRAMM [16], which deal with the risks faced by organizations that operate IT systems. Some risk assessment approaches created for operators of critical infrastructures (e.g., [12]) share the basic aspects of our risk analysis approach, but they cannot be directly applied by CIC manufacturers.

With regard to the theoretical assessment phase, certain parallels exist with the recommended use of the Control System Cyber Security Self-Assessment Tool (CS2SAT) [20], which includes a self-assessment step based on a questionnaire using recommended standards. Note, however, that CS2SAT is designed primarily for use by operators, and cannot be directly applied by CIC manufacturers.

Considerable work has been performed in the area of practical assessment. Several approaches either compete with or overlap with our practical assessment approach. Interested readers are referred to [2] for a detailed discussion of practical assessment approaches.

3.3 Future Work

Future work related to the security assessment methodology will focus on enhancing the risk analysis, theoretical assessment and practical assessment phases. The current risk analysis approach is relatively stable, but the opportunity exists to streamline and optimize the underlying process. Our future work related to theoretical assessment will address the identification and inclusion of new standards, and corresponding updates to the questionnaire; another research thrust is to devise approaches for leveraging the synergies existing in overlapping standards. Refinements to the practical assessment phase will concentrate on extending the assessment plan with new attacks and tools, and improving the test task descriptions.

Deficiencies identified by the security assessment methodology create new security requirements for the CIC, which should be implemented according to a requirements engineering process. Our future work will attempt to align the security assessment methodology findings with those obtained using common requirements engineering methods.

The security assessment methodology has been developed based on our experience with security assessments of CICs and CIC security needs. The methodology has been applied successfully to several CIC products. Because the methodology is generic, it can, in principle, be applied to other systems (e.g., corporate IT systems). However, it will be necessary to identify relevant security standards for these systems before the security assessment methodology can be applied.

4. Conclusions

The security assessment methodology presented in this paper has been applied to more than fifty products, including control systems, substation automation devices, and field devices in industrial and energy environments. The results indicate that the methodology is flexible and well-suited to assessing the security levels of CICs within a matter of man-days as opposed to man-weeks.

A key advantage of the methodology is that the security level of a CIC can be measured and quantified. This is accomplished by constructing a risk matrix. Changes to the risk matrix caused by implementing countermeasures as a result of risk analysis quantifies the resilience against the documented risks. Also, security capabilities are measured in the form of a benchmark against the requirements derived from relevant industry standards. This provides excellent input for subsequent security decisions.

The security assessment methodology is generic and can be adjusted to different CICs by using the relevant CIC security standards as the basis and applying the methodology to the CIC specifics. Finally, the security assessment methodology is lightweight and cost-effective in comparison with evaluation methods such as the Common Criteria [6]. In most cases, one to three assessors require a few weeks to complete a security assessment of a large CIC.

References

[1] C. Alberts, A. Dorofee, J. Stevens and C. Woody, Introduction to the OCTAVE Approach, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania (www.cert.org/octave/approach_intro.pdf), 2003.

[2] M. Braun, Process Optimization of Practical Security Assessments, Master's Thesis, University of Applied Sciences, Augsburg, Germany, 2008.

[3] Bundesamt fuer Sicherheit in der Informationstechnik, Durchfuehrung fuer Penetrationstests, Bonn, Germany (www.bsi.bund.de/cae/servlet/contentblob/487300/publicationFile/30684/penetrationstest_pdf.pdf), 2003.

[4] Bundesverband der Energie und Wasserwirtschaft, White Paper: Requirements for Secure Control and Telecommunication Systems, Version 1.0, Berlin, Germany (branchenkommunikation-energie.bdew.de/bdew.nsf/id/52929DBC7CEEED1EC125766C000588AD/$file/Whitepaper_Secure_Systems_Vedis_1.0final.pdf), 2008.

[5] R. Carlson, J. Dagle, S. Shamsuddin and R. Evans, A Summary of Control System Security Standards Activities in the Energy Sector, National SCADA Test Bed, U.S. Department of Energy, Washington, DC, 2005.

[6] Common Criteria Recognition Agreement Members, Common Criteria v3.1. Release 3, National Information Assurance Partnership, U.S. Department of Defense, Fort George Meade, Maryland (www.commoncriteriaportal.org/thecc.html), 2009.

[7] P. Herzog, OSSTMM – Open Source Software Testing Methodology, Institute for Security and Open Methodologies, New York (www.isecom.org/osstmm).

[8] Idaho National Laboratory, Cyber Security Procurement Language for Control Systems, Version 1.8, Technical Report INL/EXT-06-11516, Revision 3, Idaho Falls, Idaho (www.msisac.org/scada/documents/4march08scadaprocure.pdf), 2008.

[9] International Organization for Standardization, ISO/IEC 27005:2008(E), Information Technology – Security Techniques – Information Security Risk Management, Geneva, Switzerland, 2008.

[10] North American Electric Reliability Corporation, Critical Infrastructure Protection Program, Princeton, New Jersey (www.nerc.com/page.php?cid = 6—69).

[11] North American Electric Reliability Corporation, Security Guideline for the Electricity Sector: Identifying Critical Assets, Version 1.0, Princeton, New Jersey (www.nerc.com/docs/cip/sgwg/Critcal%20Asset_ID_Final_Clean.pdf), 2009.

[12] Office of Energy Assurance, Energy Infrastructure Risk Management Checklists for Small and Medium Sized Energy Facilities, U.S. Department of Energy, Washington, DC (www.esisac.com/publicdocs/assessment_methods/Risk_Management_Checklist_Small_Facilities.pdf), 2002.

[13] Open Information Systems Security Group, Information Systems Security Assessment Framework (ISSAF), Draft 0.2.1B, Colorado Springs, Colorado (www.oissg.org/downloads/issaf-0.2/information-systems-security-assessment-framework-issaf-draft-0.2.1b/download.html), 2006.

[14] OWASP Foundation, Open Web Application Security Project Testing Guide, Version 3.0, Columbia, Maryland (www.owasp.org/images/5/56/OWASP_Testing_Guide_v3.pdf), 2008.

[15] K. Scarfone, M. Souppaya, A. Cody and A. Orebaugh, Technical Guide to Information Security Testing and Assessment, NIST Special Publication 800-115, National Institute of Standards and Technology, Gaithersburg, Maryland (csrc.nist.gov/publications/nistpubs/800-115/SP800-115.pdf), 2008.

[16] Siemens Enterprise Communications, CCTA Risk Analysis and Management Method (CRAMM), Milton Keynes, United Kingdom (www.cramm.com/overview/howitworks.htm), 2009.

[17] G. Stoneburner, A. Goguen and A. Feringa, Risk Management Guide for Information Technology Systems, NIST Special Publication 800-30, National Institute of Standards and Technology, Gaithersburg, Maryland (csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf), 2002.

[18] K. Stouffer, J. Falco and K. Scarfone, Guide to Industrial Control Systems (ICS) Security, NIST Special Publication 800-82, National Institute of Standards and Technology, Gaithersburg, Maryland (csrc.nist.gov/publications/drafts/800-82/draft_sp800-82-fpd.pdf), 2008.

[19] Tenable Network Security, Nessus – The network vulnerability scanner, Columbia, Maryland (www.nessus.org/nessus).

[20] US-CERT, Cyber Security Self-Assessment Tool, Control System Security Program, U.S. Department of Homeland Security, Washington, DC (www.us-cert.gov/control_systems/satool.html).

[21] US-CERT, Standards and References, Control System Security Program, U.S. Department of Homeland Security, Washington, DC (www.us-cert.gov/control_systems/csstandards.html).

Chapter 17

AN ADVANCED DECISION-SUPPORT TOOL FOR ELECTRICITY INFRASTRUCTURE OPERATIONS

Yousu Chen, Zhenyu Huang, Pak-Chung Wong, Patrick Mackey, Craig Allwardt, Jian Ma and Frank Greitzer

Abstract A major failure in the electricity infrastructure would almost certainly lead to significant societal disruption and massive economic losses. The reliable operation of the electricity infrastructure is an extremely challenging task because human operators have to consider thousands of possible configurations in near real time to choose the best option. Nevertheless, the operation of the electricity infrastructure is largely based on operator experience with limited real-time decision support. This makes it difficult for operators to anticipate, recognize and respond to anomalies caused by human error, natural disasters or cyber attacks.

This paper proposes an advanced decision-support tool for electricity infrastructure operations. The tool converts large amounts of data into actionable information to help operators monitor the power grid status in real time. It performs trend analysis at the regional or system level to enable operators to foresee and discern emergencies; it performs cluster analysis to help operators identify the relationships between system configurations and affected assets; and it interactively assesses candidate actions to assist operators in making effective and timely decisions.

Keywords: Electricity infrastructure, decision support, visual analytics

1. Introduction

The U.S. electricity infrastructure has been called the most complex machine on earth [1]. However, much of the infrastructure was designed more than 50 years ago. A failure of the electricity infrastructure can lead to significant disruptions of people's lives and industrial and commercial activities, causing massive economic losses. Incidents such as the Western North America blackout of 1996 [4] and the Northeast blackout of 2003 [7] underscore the need to have

T. Moore and S. Shenoi (Eds.): Critical Infrastructure Protection IV, IFIP AICT 342, pp. 245–260, 2010.

a reliable power grid. The prediction, prevention and mitigation of blackouts have become the primary focus of the DOE Office of Electricity Delivery and Energy Reliability (DOE-OE) [8] as well as the central topic of power systems research.

The operation of the electricity infrastructure is an extremely challenging task due to its complex structure, geographical coverage, complex database and information technology systems, and highly dynamic and nonlinear behavior. The operation is also affected by a number of external factors, including physical attacks, cyber threats, human error and natural disasters. Because of the complex nature of the electricity infrastructure, large amounts of data and information have to be processed to gain adequate situational awareness and to adapt to emergency situations. Managing this complexity is a critical issue in electricity infrastructure operations.

Electricity infrastructure operations are largely driven by human operator experience and occur with little real-time decision support. The lack of effective systems that can manage the complexity of operations translates to an inability on the part of human operators to anticipate, recognize and respond to adverse and unexpected situations.

This paper presents an advanced decision-support tool that is designed to meet the immediate needs of electricity infrastructure operators. In particular, the tool improves situational awareness, enabling operators to recognize current and potential failures; it helps predict the consequences of potential failures; and assists in evaluating the effect of candidate actions. The tool has been successfully applied to real-world power grid models and data.

2. Enhanced Operational Framework

Electricity infrastructure operations involve highly complex computational processes and power grid models. Figure 1 presents a functional view of real-time power grid operations [2].

This paper focuses on two key functions: state estimation and contingency analysis. State estimation computes the various system parameters that are input to other operational functions, including contingency analysis. Contingency analysis studies "what-if" conditions in anticipation of potential power grid failures. It identifies operational violations that occur when certain operational limits, such as transmission line load capacity and substation voltage thresholds, are exceeded. The violations are presented to operators for review and to determine the appropriate candidate actions.

The North American Electric Reliability Corporation (NERC) operating standards [6] require that the loss of any single element in the power grid should not cause system instability. Networks that meet this standard are rated as "N-1 Secure." If the loss of one or more elements does not result in any limit violations, the system is deemed to be secure for the corresponding contingency. The contingencies that result in violations of operating limits are flagged and placed in a list for the operators to inspect. When contingency

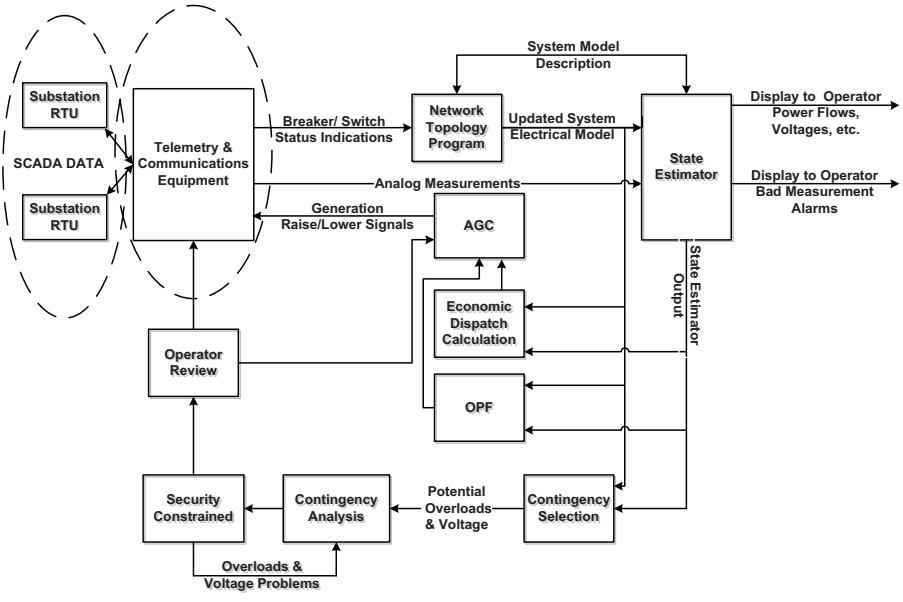


Figure 1. Electricity infrastructure operations.

violations occur, NERC mandates that operators take actions to mitigate the situation in a timely manner.

Due to the size and complexity of the modern power grid, the number of contingencies to be analyzed can be very large – it is not uncommon for several hundred contingencies or more to be examined. Conveying the contingency outputs to system operators in a meaningful and easy-to-understand manner is a real challenge. State-of-the-art commercial tools use a tabular form to display contingency violations (Figure 2). Each violation corresponds to a row in the form; note that no information is provided about the geographical context and relative severity of the violation. The tabular display may be adequate when few contingency violations are present. However, when the system is heavily stressed and there are many contingency violations, operators can be overwhelmed by the information presented in the tabular display. In such cases, it is almost impossible for operators to sift through large amounts of violation data and understand the system situation in minutes, let alone seconds. Of course, it is during these situations that operators need the information the most.

To address these challenges, we have developed an advanced decision-support tool that is intended to assist three important aspects of power grid operations:

- Improved situational awareness by visualizing and analyzing the change in risk levels as a result of violations.
- Prediction of the consequences of potential problems by analyzing the pattern of impact.

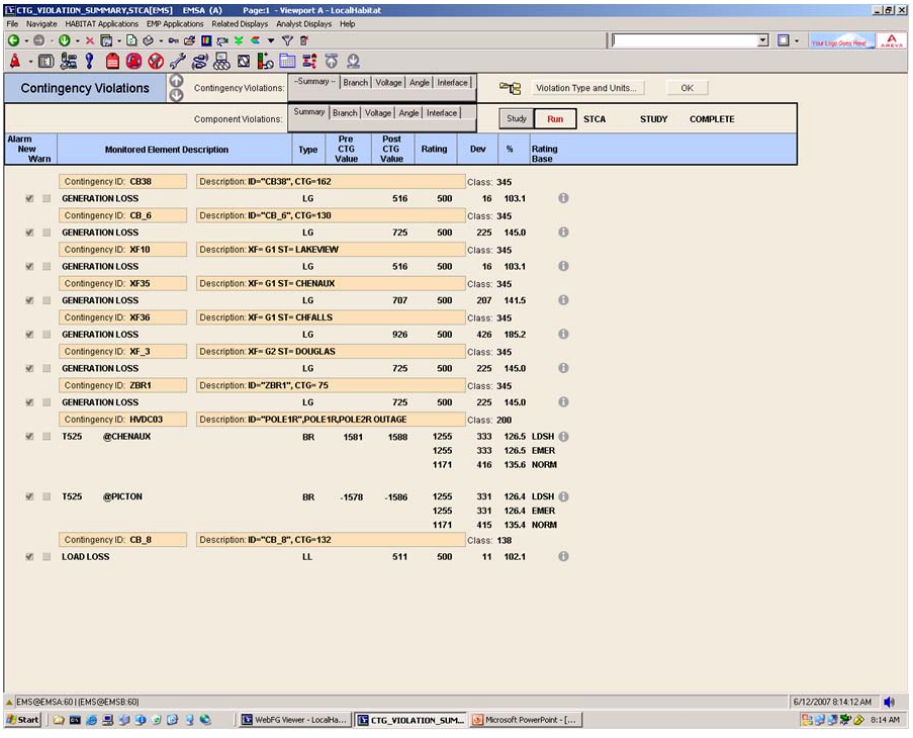


Figure 2. Tabular representation of violation data.

- Assessment of the effects of candidate actions by interactively analyzing the collective severity level.

The advanced decision-support tool is designed to provide information about the status of the electricity infrastructure, analyze historical data, generate system-trending information for prediction, identify relationships between system configurations and affected assets, and interactively assess candidate actions to determine the best solution. The tool presents operators with actionable information about the current system status and trends, enabling them to comprehend the situation and identify the best candidate actions in a timely manner.

3. Improving Situational Awareness

Operator situational awareness is enhanced using visual analytic and graphical trending techniques. The visual analytic technique converts large amounts of raw operational data to actionable information. The graphical trending technique provides operators with information about system trends at multiple levels of abstraction, enabling them to foresee and discern emergencies. Interested readers are referred to [3] for details about these two techniques.

3.1 Visual Analytic Technique

The visual analytic technique involves two steps: (i) defining and computing the risk level of contingency violations; and (ii) converting the risk level to a contoured map by adapting a novel visual analytic technique developed by the National Visualization and Analytics Center [5].

Instead of using the tabular form in Figure 2 to convey the status of the power system, contingency data is converted into a quantitative risk level that is presented to operators. The risk level of contingency violations is given by:

$$R_{ik} \in \begin{cases} [0, R_T) & \text{Safe} \\ [R_T, 100) & \text{Alert} \\ [100, \infty) & \text{Violation} \end{cases} \tag{1}$$

where R_T is the pre-specified alert risk level (expressed as a percentage) for each transmission line and substation. Note that $R_T = 97.5\%$ in our study.

The risk levels (expressed as percentages) for transmission lines (Equation (2)) and substations (Equation (3)) are defined in terms of the capacities of their power loading and voltage level parameters, respectively:

$$R_{ik} = \frac{P_{ik}}{P_{imax}} \times 100 \tag{2}$$

$$R_{ik} = \left| \frac{(V_{ik} - V_{imin}) - (V_{imax} - V_{imin})/2}{(V_{imax} - V_{imin})/2} \right| \times 100 \tag{3}$$

where ik denotes the i^{th} line or i^{th} substation for the k^{th} contingency; P_{ik} is the loading in the i^{th} line for the k^{th} contingency; P_{imax} is the loading limit of in the i^{th} line; V_{ik} is the voltage of the i^{th} substation for the k^{th} contingency; V_{imin} is the lower voltage limit of the i^{th} substation; and V_{imax} is the upper voltage limit of the i^{th} substation.

Note that the risk level definitions apply to the tabular violation data shown in Figure 2. Also, they specify how close the operational parameters are to the corresponding limits, even when no violations exist.

Because each contingency generates a set of contingency risk levels as defined by Equations (1–3), there will be k sets of risk levels for k contingency cases. Across all the contingencies, the risk level of the i^{th} element is defined as the maximum value over the set of risk levels:

$$R_i = max(R_{ik}); \quad k = 1, 2, \ldots, K \tag{4}$$

Note that the tool permits the use of other functions (e.g., mean or sum) instead of maximum. Also, the risk R_{ik} can be multiplied by the probabilities of the contingencies to obtain the risks in combination with the likelihood of the occurrence of failures. For simplicity, this paper assumes a unit probability for all contingencies.

The next step is to convert the risk levels as defined in Equation (4) to a contoured map with colors indicating different risk levels. The system status

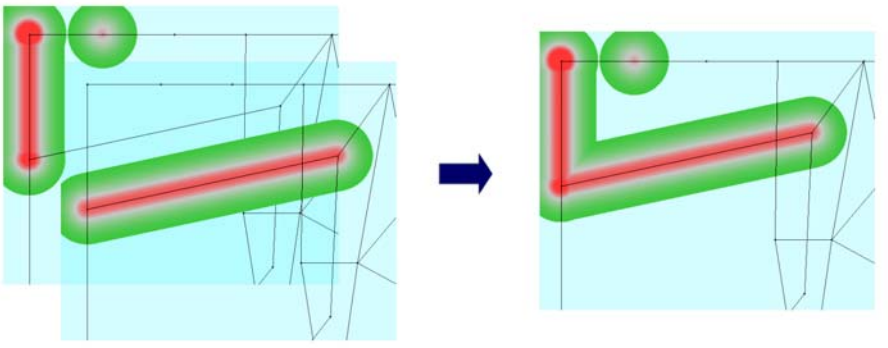


Figure 3. Collective risk area with superposition of individual risk areas.

is visualized as fading colors from the center as shown in Figure 3. The impact area of a substation has a circular shape, while that of a line has an elliptical shape. Individual risk areas are superimposed to form the collective risk area. The risk maps are overlaid to represent the collective risk of multiple possible configurations.

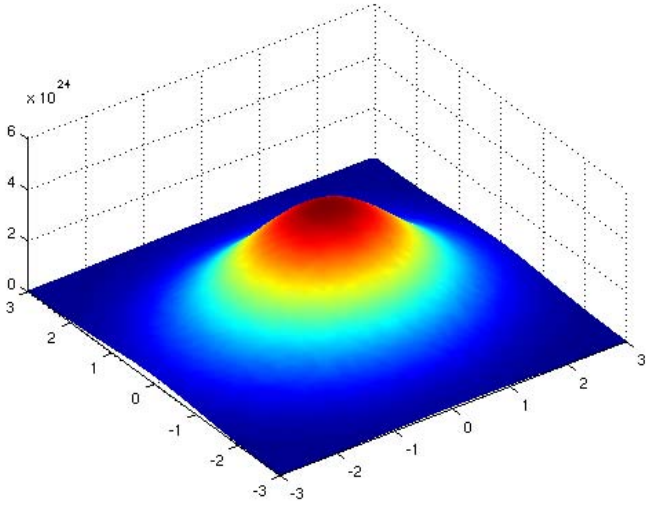


Figure 4. Gaussian color filter and color map.

The implementation uses a hash table to store all the pixels of the substations and lines, and a Gaussian color filter to display the collective risk. In the hash table, each pixel has a value determined by the contingency risk level of the substation or line. Only the largest value is retained in the table in order to represent the highest risk. The Gaussian color filter is circular with values conforming with a Gaussian curve (Figure 4). The output of the filter is the

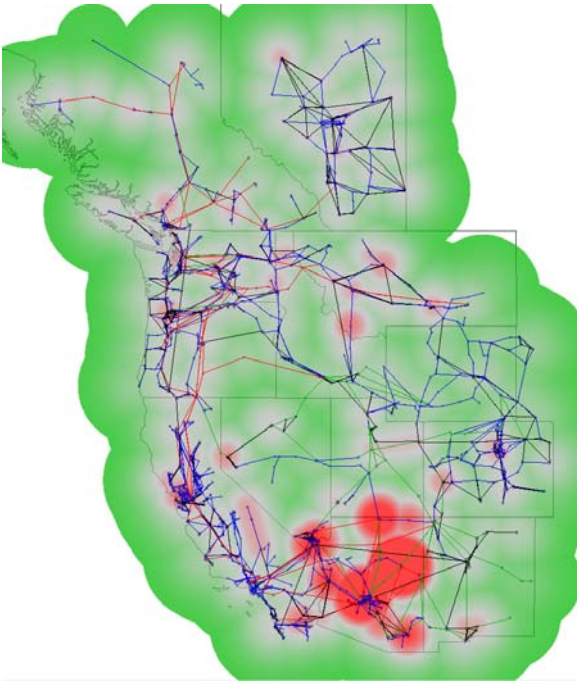


Figure 5. Risk map of the Western North American power grid.

output matrix M, which associates each point in the map with a floating point number (color value); each value is assigned to a color map to obtain the final contour. The colors, green, gray and red, in the color map correspond to the risk categories, safe, alert and violation, respectively.

The final visual representation uses HaveGreen [10] as the application framework, which provides an interface for navigating and zooming over the power grid. Figure 5 is created using the model and data from the 2005 HS2A Approved Operating Case of the Western Electricity Coordinating Council (WECC) [9]. The figure shows the Western North American power grid with 50 sets of contingency results overlaid on a single risk map to visualize the collective risk of the contingencies. Unlike the tabular representation in Figure 2, the color-contoured map enables operators to quickly identify the vulnerable portions of the power grid (represented in red).

3.2 Graphical Trending Technique

Trending analysis involves the observation and examination of the change in risk over time, and the prediction of whether or not the network is becoming more vulnerable, more compromised or more robust. Increases in the risk indices, which are expressed by the size of the region and color intensity in the risk map, indicate developing problems.

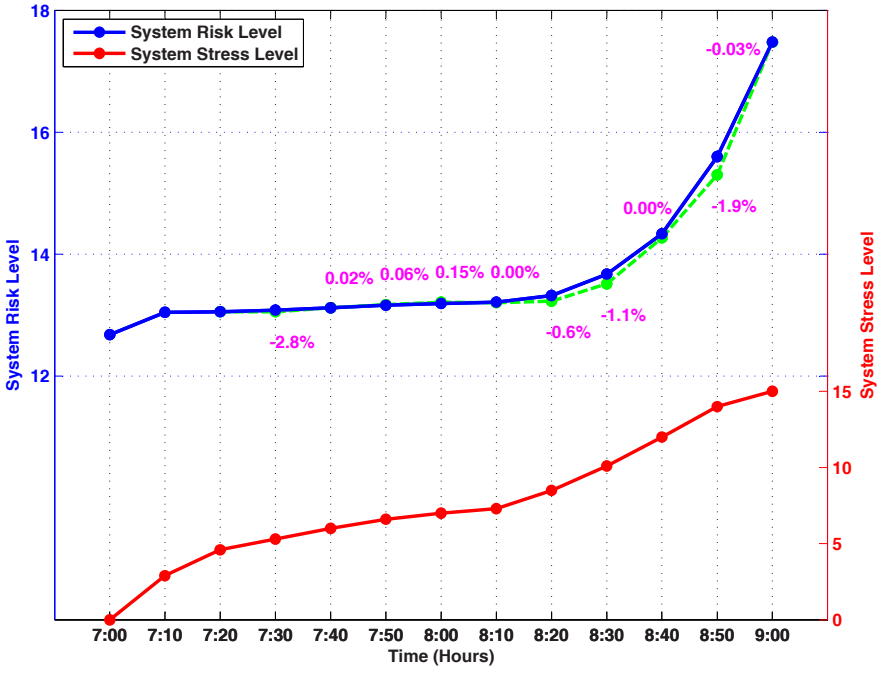


Figure 6. System risk level and stress level over time.

A trend analysis chart uses one line to represent the overall size and intensity of the critical regions in the power grid visualization; and multiple lines to represent the size and intensity of individual critical regions. A critical region is defined as a contiguous set of pixels where each pixel has a color value no less than a particular threshold. Each value is taken from the corresponding element in the output matrix M generated by the Gaussian color filter.

The first step in trending analysis is to find the calculated risk level for each individual region using a non-recursive breadth-first search. If the value of an element meets or exceeds the threshold, then all its neighboring elements are examined and marked if they meet the threshold. All the values in this contiguous region are added to produce the current regional risk value. After the breadth-first search is complete, a normalized regional risk value is obtained by dividing the regional risk value by the number of pixels in the image. Once all of the regions are found, the total risk value is calculated by summing the risk values of the individual regions.

Figure 6 presents the risk levels of the Western North American power grid during the morning load pick-up period. When the total power consumption is low at the beginning of the period, increasing the load does not increase the risk levels as much as when the total load is high (towards the end of the period). This is consistent with operational experience. To further validate the system risk levels, the contingency risk levels R_{ik} that are higher than 100% are

Table 1. System risk level and summation of contingency risk levels.

Time	Risk Level	ΣR_{ik}	Time	Risk Level	ΣR_{ik}
7:00	12.68	169.47	8:10	13.21	179.65
7:10	13.05	178.60	8:20	13.32	182.82
7:20	13.06	178.49	8:20	13.32	182.82
7:30	13.08	179.48	8:30	13.67	212.92
7:40	13.12	179.53	8:40	14.33	251.49
7:50	13.16	179.58	8:50	15.60	353.09
8:00	13.19	179.62	9:00	17.48	488.23

summed (Table 1). The results show that the system risk levels are consistent with the results of the contingency analysis.

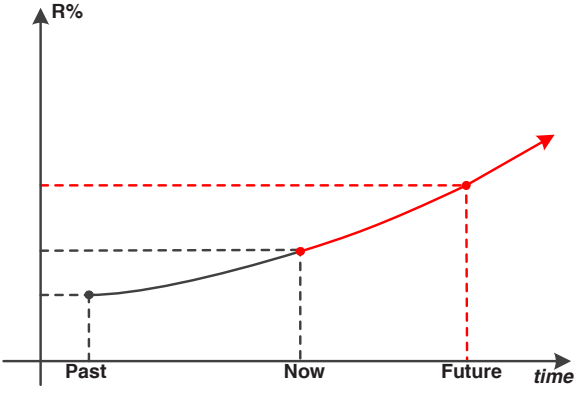


Figure 7. Illustration of visual trending analysis.

The next step is to conduct a visual trending analysis based on the system and regional risk levels. The trend is obtained by fitting a curve to the historical risk levels of the network or regions, and extrapolating them to predict the future system situation (Figure 7).

The green dashed line in Figure 6 is the predicted system risk level, where each point is computed based on the three preceding risk levels. Note that the prediction is reasonably close to the actual system risk level (i.e., within a 2.8% error range).

Complex evolving patterns may exist in the power grid network. As the risk values of different regions are computed, they are tracked to see how they relate to the previous risk regions. This helps determine if two new regions come from a previous region (defined as whether or not a region overlaps with a previous region). A region can originate from multiple regions, and a region can spawn multiple regions. This feature is depicted in the trend analysis chart as a line splitting into multiple new lines or combining multiple lines into one new line.

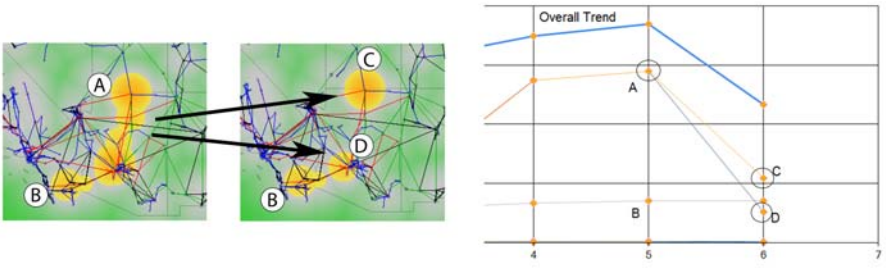


Figure 8. Complex evolving patterns of network risk impact areas.

Figure 8 shows an example of one area splitting into two areas (Region A splits to Regions C and D at Time 5). Note that the y-axis represents the risk level.

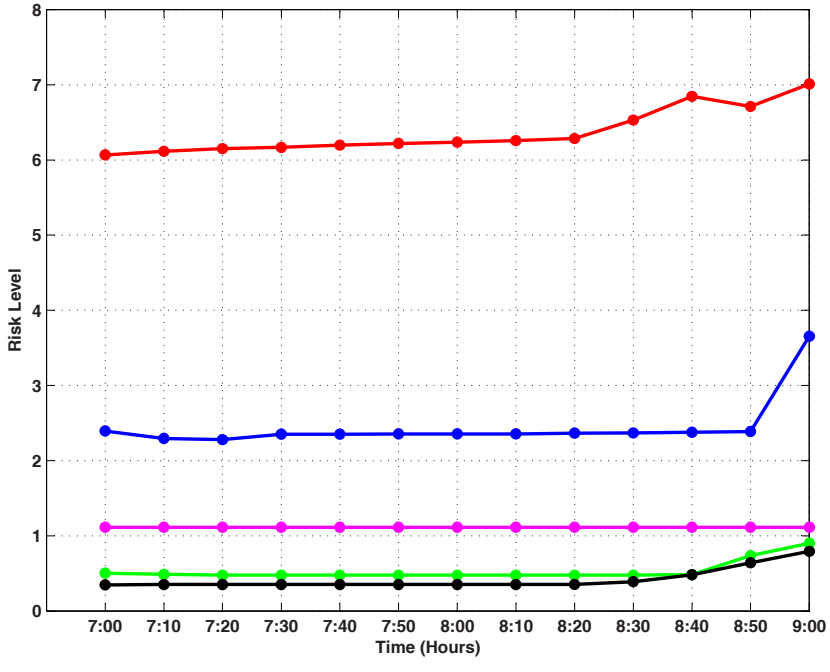


Figure 9. Regional risk trends in the Western North American power grid.

Figure 9 shows the trends for the five most critical regions for the same system conditions as in Figure 6. The overall system trend gives an overview of the system status. However, the system trend can be relatively flat because changes in different regions may cancel each other's impact. Therefore, it is important to observe regional trends to identify critical regions that demand immediate operator attention.

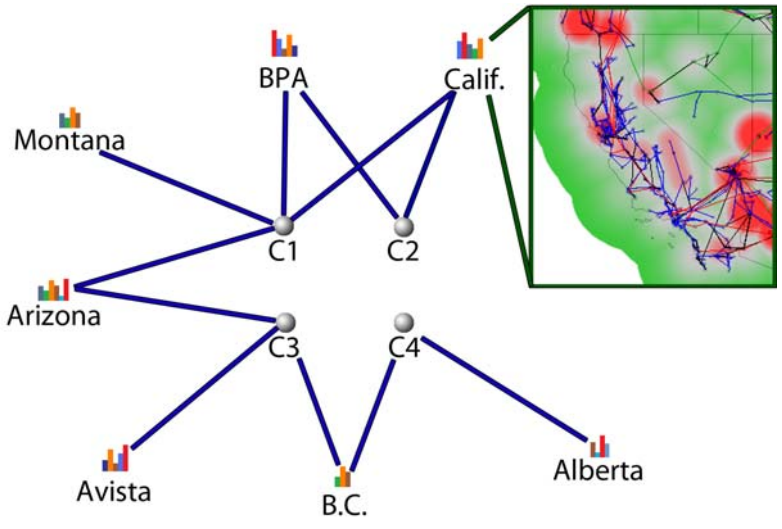


Figure 10. Clustering analysis.

4. Recognizing Failure Patterns

Power grids can have numerous configurations that result in stresses on substations and transmission lines. With the help of visual analytic techniques and visual trend analysis, operators can quickly gain wide-area situational awareness. To enable operators to focus on important information during network emergencies, clustering analysis is needed to identify system patterns and present them in association with the risk contour map. Clustering analysis can help operators identify the relationships between system configurations and affected assets. The criteria for configuration clustering are based on geographical characteristics, configuration types and impact types. Clustering analysis is combined with the contoured map to provide operators with a quick overview of the grid status while enabling them to drill down to the details if needed.

Figure 10 shows how clustering analysis can help reveal the relationships between network configurations and affected network assets. C1–C4 are critical contingency cases that cause violations (shown in the bar charts) in different locations. If an operator wishes to see more information in a specific area, he can go to a deeper level to investigate the contingency impact within the area. By applying the same method to multiple levels, a hierarchy of related contingencies from the area level all the way down to the individual configuration can be constructed. For example, in Figure 10, C2 causes violations in the BPA and California areas. An operator wishing to see more detail in the California area may use the hierarchy structure. At the same time, he has the option to study multiple configurations concurrently and to compare scenarios.

Figure 11 shows an example of clustering analysis. The yellow highlighted lines in the figure indicate the locations of the contingencies, which correspond

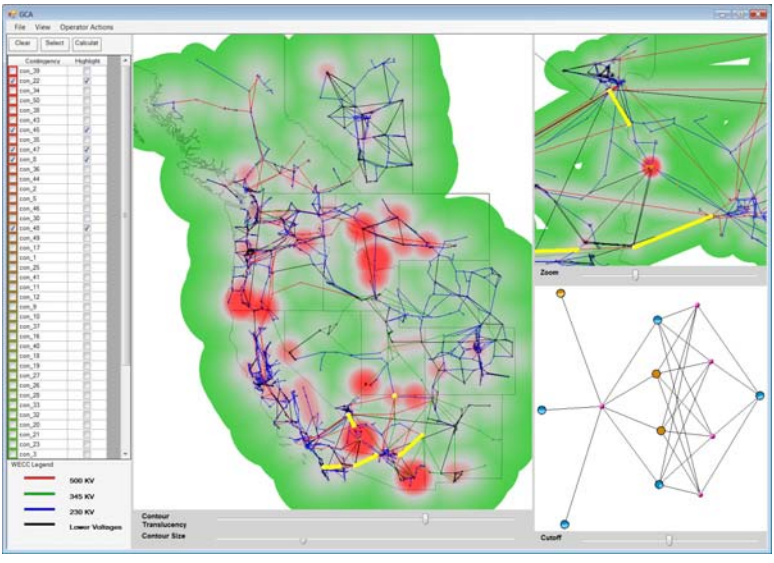


Figure 11. Clustering analysis function in the decision-support tool.

to the purple dots in the "spider web" plot in the lower-right corner. The location of a contingency is found by clicking the purple dot in the spider web. The blue circles in the spider web are the substation voltage violations, while the orange circles are the line loading violations. The shaded areas in the blue circles and orange circles indicate the severity of violations: the larger the shaded area, the more severe the violation. By clicking on different elements in the spider web, operators can easily identify the system patterns and focus on important information. The cutoff severity slider below the spider web can be used to change the minimum severity level to be shown, enabling only the more severe violations to be presented in a less crowded spider web.

The simple spider web example clearly shows how a contingency affects grid assets and how an asset is affected by various contingencies. A purple dot with many links to blue or orange circles indicates a critical contingency because it affects many power grid assets. These contingencies, if they occur, would have significant consequences; therefore, security enhancements or mitigation plans should be in place to ensure that the consequences are contained to acceptable levels. On the other hand, a blue or orange circle with many links to purple dots indicates a vulnerable asset because it can be affected by many contingencies. These assets should be protected by reliable backups or should be reinforced through new network development.

5. Assessing Candidate Actions

The interactive assessment of candidate actions provides additional decision support for power grid operators. With the help of visual analytic techniques,

graphical trend analysis and clustering analysis, problems can be recognized and their consequences identified. Normally, there are multiple options for responding to a specific problem and choosing the best action is a challenging task for operators. Operators often make their decisions based on their experience because there is little decision support to enable them to identify the best option. Consequently, there is no guarantee that the action will be successful; in many cases, an action worsens the situation or causes new problems.

The interactive assessment of candidate actions helps determine the effect of operator actions. Operator actions may include power grid reconfiguration, generator re-dispatch, load shedding, etc. Before the operator chooses a specific action to implement, the candidate options may be tested in a model simulation, and the new grid status visualized in the color-contoured map. The collective severity level (CSL) is used to quantify the effect of the candidate actions and rank them in a prioritized list. The CSL is defined as:

$$CSL = \sum_{i=1}^{N} \left(\frac{max(P_{ik})}{P_{imax}} \right)^2 \text{ where } max(P_{ik}) > P_{imax} \tag{5}$$

Note that i denotes the i^{th} transmission line; k denotes the k^{th} contingency case; N is the number of transmission lines; P_{imax} is the capacity of the i^{th} transmission line; and P_{ik} is the power carried on the i^{th} transmission line for the k^{th} contingency case.

Figure 12 presents an example of interactive assessment. The figure shows a Western North American power grid risk map with an overlay of the results of 50 contingencies. For simplicity, only the line loading violations are shown. The power grid is clearly stressed because many violations exist (indicated by red regions). Note that only simple load shedding actions are considered as operator actions to illustrate the functionality of the tool. More realistic actions such as generator re-dispatch, reactive compensation and network reconfiguration will be investigated in future work.

Consider a situation where an operator has five candidate actions labeled A through E, which represent five different load reductions: –8.4%, –7.7%, –4.9%, –3.0% and –1.0%. The tool provides an interactive function as a menu item. An operator can select the menu item and simulate the candidate actions and update the contoured map. Figure 13 displays the results of the interactive assessment of the five candidate actions with line loading violations only. These actions are sorted based on their effectiveness from best to worst.

Table 2 lists the actions, their collective severity levels and rankings. By referencing Figure 13, an operator can easily identify A as the best action because it removes almost all the red color from the map and it has the lowest CSL (0.00). This is expected because a load reduction will better alleviate the system stress level.

The same methodology can be applied to other operator actions. Indeed, Figure 13 helps operators identify the violations that remain along with their locations, enabling them to judge if an option is satisfactory without relying solely on the CSL metric (Table 2). An operator can also choose to fine-tune the

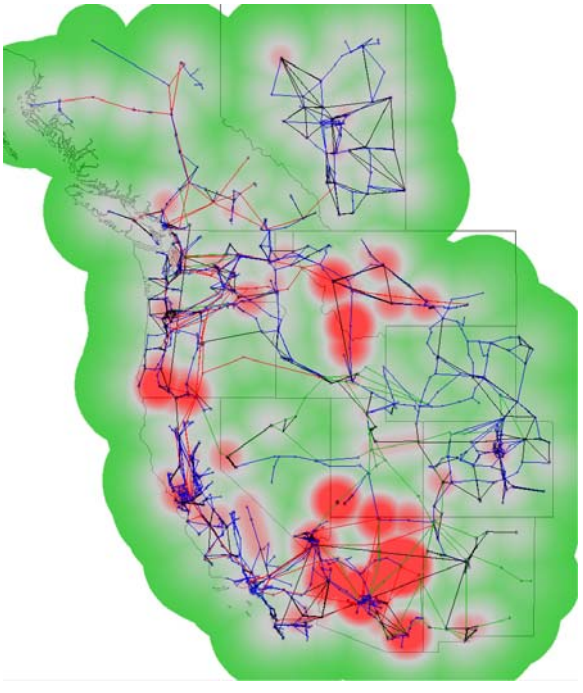


Figure 12. Western North American power grid risk map.

Table 2. Candidate actions and assessment results.

Option	Description	CSL	Ranking
A	8.4% load reduction off the current condition	0.00	1
B	7.7% load reduction off the current condition	53.30	2
C	4.9% load reduction off the current condition	74.03	3
D	3.0% load reduction off the current condition	90.28	4
E	1.0% load reduction off the current condition	117.23	5

options if none are deemed adequate. By adjusting the actions, the operator can reevaluate the option using the interactive function until a satisfactory option is determined.

6. Conclusions

Visual analytics techniques as implemented in the decision-support tool can significantly enhance power grid operations by converting large amounts of operational data into actionable information, translating the operational data into risk levels and presenting the risk levels in a color-contoured map. These features enable operators to quickly gain situational awareness of the power grid without sifting through large amounts of raw data. A predictive capability

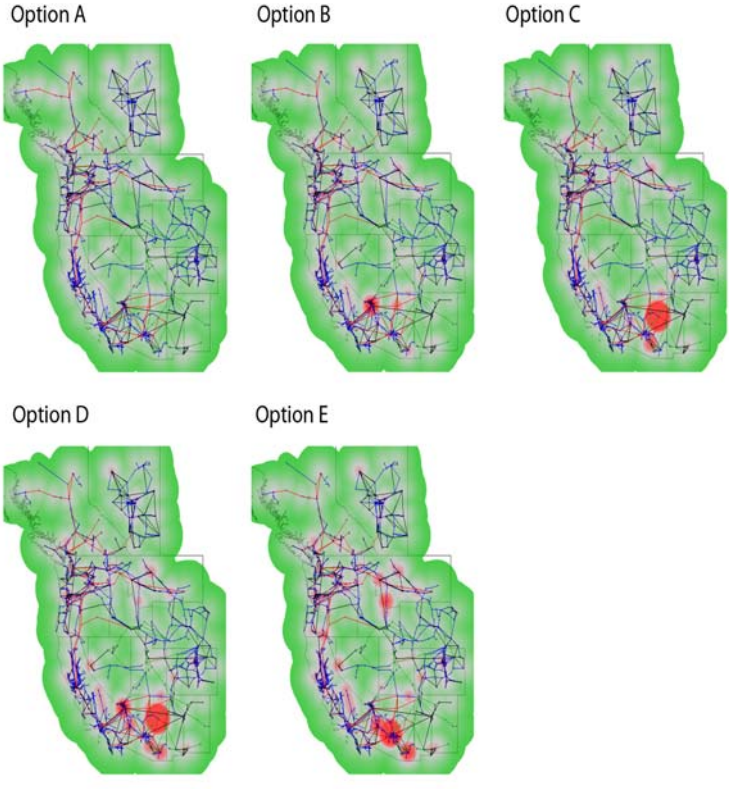


Figure 13. Sorted interactive assessment of candidate actions (from best to worst).

is established by analyzing network risk level trends using an approach that combines structural and statistical analyses; this assists operators in identifying system trends and foreseeing and discerning emergencies. The decision-support tool also performs clustering analysis to help operators identify the relationships between system configurations and affected assets. Additionally, operators can interactively evaluate candidate actions to identify the best action in a given situation.

The tool has received favorable reviews from power grid operators. It is currently being evaluated in collaboration with the WECC, which oversees the Western North American power grid. The results of the evaluation will be used to drive the refinement of the tool prior to its use in power grid control centers.

The decision-support tool engages a generic framework. Thus, it can be applied to other applications such as system planning and sensor data quality assessment. The tool can also be extended for use in other complex networks, including gas pipeline systems, telecommunications systems and aviation networks. Our future work will implement the hierarchical organization chart for clustering analysis, apply more realistic actions for interactive assessment, con-

duct usability studies to validate the utility of the tool, and integrate the tool with current commercial tools.

Acknowledgements

This work was supported by the Information and Infrastructure Integrity Initiative of the Pacific Northwest National Laboratory. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830. The authors also wish to thank Kevin Schneider, Ning Zhou, Jeff Dagle, Jim Thomas and Mark Hadley for their insightful comments and support of this work.

References

[1] R. Bush and G. Wolf, The bulk power grid seeks intelligent operation, *Transmission & Distribution World*, January 2, 2007.

[2] Z. Huang, R. Guttromson, J. Nieplocha and R. Pratt, Transforming power grid operations via high-performance computing, *Scientific Computing*, vol. 24(5), pp. 22–27, 2007.

[3] Z. Huang, P. Wong, P. Mackey, Y. Chen, J. Ma, K. Schneider and F. Greitzer, Managing complex network operation with predictive analytics, *Proceedings of the AAAI Spring Symposium on Technosocial Predictive Analytics*, pp. 59–65, 2009.

[4] D. Kosterev, C. Taylor and W. Mittelstadt, Model validation for the August 10, 1996 WSCC system outage, *IEEE Transactions on Power Systems*, vol. 14(3), pp. 967–979, 1999.

[5] National Visualization and Analytics Center, Pacific Northwest National Laboratory, Richland, Washington (nvac.pnl.gov).

[6] North American Electric Reliability Corporation, Transmission Operations, Standard TOP-004-2, Princeton, New Jersey (www.nerc.com/files/TOP-004-2.pdf), 2007.

[7] U.S.-Canada Power System Outage Task Force, Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, Department of Energy, Washington, DC (reports.energy.gov/BlackoutFinal-Web.pdf), 2004.

[8] U.S. Government, Energy Policy Act of 2005, Public Law 109–58, *United States Statutes at Large*, vol. 119, pp. 594–1143, 2005.

[9] Western Electricity Coordinating Council, 2005 HS2A Approved Operating Case, Salt Lake City, Utah (www.wecc.biz/committees/StandingCommittees/PCC/TSS/BaseCases/Pages/default.aspx), 2005.

[10] P. Wong, G. Chin, H. Foote, P. Mackey and J. Thomas, Have green – A visual analytics framework for large semantic graphs, *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pp. 67–74, 2006.