LOVELY HORSE

From GCWiki
Jump to: navigation, search

LOVELY HORSE.JPG

LOVELY HORSE2.JPG

LOVELY HORSE is a TCP Task Order 144 initiative as part of CDO (formerly NDIST) and the Cyber Theme towards developing Open Source capability. So far, we have worked towards making structured datasets available on the high side for analysts to use - this data is available within HAPPY TRIGGER. We are now looking towards making use of more unstructured information (blogs, forums, Twitter). LOVELY HORSE seeks to experiment with provision of an indexed repository of unstructured information that can be used to push content of interest to individual analysts via a variety of mechanisms.

The initial LOVELY HORSE prototype can be accessed from here. See below for more details.

*See also — **BIRD SEED***

Contents

[edit] Problem statement

Analysts are potentially missing out on valuable open source information relating to cyber defence because of an inability to easily keep up to date with specific blogs and Twitter sources. Accessing these resources involves using specific JEDI terminals, or reading up at home. We need to make this information available to analysts on the high side at their normal terminals.

However, there is a balance to be found - analysts don't have the time to spend hours and hours reading through loads of blogs. In addition, we don't want this repository to be yet another tool that analysts have to access - this information needs to be incorporated into existing workflows.

We need to find a way to index this information so that analysts **only get a relevant subset** of this information **pushed to them**.

[edit] Initial prototype

We are working with JTRIG to make use of the existing BIRDSTRIKE architecture for capturing tweets from Twitter. We are also working with CISA around techniques they are developing to capture blog content. Both of these obviously take time, and are slower burn objectives.

In the meantime, we are running an initial prototype, where Twitter and (and subject to legal/security approval) blog content is manually scraped and uploaded to GCDesk. This content is accessible by way of personalised RSS feeds. Individual users can choose their preferences, in terms of which Twitter accounts and blogs they want to follow, and a personalised RSS feed is generated automatically for them to which they can subscribe.

This can be used by anyone, and can be accessed from here. Your personal RSS is linked from the LOVELY HORSE website.

As stated previously, this is a manual update at the moment, and will initially be **maintained on a best endeavours basis** (hopefully roughly daily). Once the BIRDSTRIKE architecture comes on line, this will be updated in real time.

For any requests for new Twitter feeds you wish to be able to subscribe to, please get in touch.

[edit] Current sources

Currently, we're bringing in the following list of Twitter accounts. To request new ones, please submit your requests, with a brief justification, via the suggestion box on LOVELY HORSE

- 0xcharlie
- alexsotirov
- anonops
- anonymousirc
- anon_central
- anon_operations
- bradarkin
- CeRTFi
- danchodanchev
- daveaitel
- dinodaizovi
- diocyde
- egyp7
- GoVCeRT_NL
- halvarflake
- hdmoore
- hernano
- JaNeTCSiRT
- kevinmitnick
- lennyzeltser
- lulzsec
- mdowd
- mikko
- msftsecresponse
- operationleaks
- owasp
- pusscat
- Shadowserver
- snowfl0w
- taosecurity
- taviso
- teamcymru
- thegrugq
- TheHackersNews
- tinman2k
- VuPeN
- WTFuzz

[edit] Future development concept

The rest of this page is constituted from ideas that we currently have about LOVELY HORSE.

[edit] Team

It will be delivered by TCP's TO144 team.

[edit] Sources

Initially we need to identify a series of sources. We currently have a list of around 60 blog and Twitter sources that have been identified by CDO analysts and cyber defence experts from Detica, and most of these have been approved for collection by MP-LEG.

Information will arrive in unstructured 'information articles'. In the context of a blog, an article would be a post; on Twitter, an article would be a tweet.

Blog sources:

These sources have currently been approved by MP-LEG (see approvals spreadsheet in DISCOVER)

- http://www.secureworks.com/research/blog/
- http://www.secureworks.com/media/blog/
- xs-sniper.com/blog
- bugix-security.blogspot.com/feeds/posts/default
- carnalOwnage.attackresearch.com/rss.xml
- intrepidusgroup/insight/feed
- www.offensivecomputing.net/?q=node/feed
- rdist.root.org/feed/
- www.darknet.org.uk
- imthezuk.blogspot.com
- dogber1.blogspot.com/
- www.ragestorm.net/blogs/
- blog.mandiant.com
- www.openrce.org
- feeds.trendmicro.com/Anti-MalwareBlog
- blogs.technet.com/b/msrc/rss.aspx
- blogs.adobe.com/psirt/feed

These sources are currently not approved by MP-LEG

- www.f-secure.com/weblog/weblog.rtf
- www.f-secure.com/exclude/vdesc-xml/latest_50.rss
- feeds.feedburner.com/GoogleOnlineSecurityBlog
- securityvulns.com
- feeds.feedburner.com/infosecResources
- targetedemailattacks.tublr.com

Twitter sources:

The advice from MP-LEG on this issue is that "provided the accounts you are selecting for acquisition meet the criteria as agreed in the approvals spreadsheet, i.e. those of "academics specialising in the identification and investigation of vulnerabilities and malware", there is no need to seek authorisation for each individual Twitter account." Our selection of Twitter sources is currently as listed above, but will undoubtedly increase over time.

Further potential sources of interest are found at Computer_security_news_and_views

[edit] Processing

Initially, these articles get processed into three components:

[edit] Content

The content will be the full textual content of the article. This will be stored as some sort of CLOB in a database.

[edit] Metadata

We would strip metadata from the article such as

- Author/Source
- Datetime of submission

and used this to update a Source Directory - information about the individual sources. For example:

- Author
- Number of articles in LOVELY HORSE
- Average usefulness rating - see feedback mechanism
- Tags of subject matter linked with this source? - see indexing

[edit] Index

This is the important bit. The aim is to index the unstructured information so that it can be linked back to

- An analyst's particular interest
- As enrichment to an existing investigation

The proposed idea is to make use of tagging (defining 'indexing' as 'identifying keywords'). Each article would be tagged with information that had been extracted from it. These tags could be IP addresses, domains, or any text string from within the content of the article. Effectively these tags are the output of entity extraction, and this list of tags would then be associated with that article.

Similarly, lists of tags are associated with individual analysts, to define their specific interest set.

[edit] How to generate a pot of tags?

We would need a pot of tags that becomes our entity set which we're extracting from new articles coming in. How to generate this pot of tags?

- Simple idea would be to regex for IP addresses and domains to start off with.
- Could index every capitalised word in a blog title.
- Could get analysts to provide a list of keywords they are specifically interested in.
- Could we extract keywords from existing analyst toolsets - for instance, do analysts tag investigations within Palantir?
- Analysts should be encouraged to tag articles they read

There is potential to link this entity extraction initiative in with corporate entity extraction tools that may provide more sophisticated matching.

- Could try and analytically identify tags. Whole articles could be tokenized and a word count generated. If a particular term appeared, say, 4 or 5 times in the current week, but not last week, then maybe that's a new trend? In which case we should add this term to the pot of tags.

[edit] Feedback mechanism

Important to allow analysts easy ability to appraise usefulness of information. Analysts should be able to 'like' content from whichever interface they're accessing the content. If an analyst likes a particular article, tags from that article are automatically added to their personal tag list.

Articles can have a usefulness rating assigned to them - generate some metric on the lines of (number of 'likes'/number of views). Articles that have a usefulness rating over a specific threshold could be pushed to all analysts. An average of the usefulness ratings across all articles from one source can be used to appraise different sources - almost becomes a crude 'confidence factor' in the information - should I trust/act upon this information?

[edit] Visualisation and Access

Need to be different ways analysts access and view this content.

- Palantir - as enrichment to existing investigations. Similarly to the current enrichment helper, any articles that had tags which are entities within the investigation are flagged up. The content should then be viewable in a human readable format within Palantir.
- Alerts - analysts should be alerted when a new article is tagged with a tag from their interest set. How should this alerting happen? Email? RSS feed?
- General search, there should be LOVELY HORSE front end that can be used for analysts to search across the whole repository. Would want to investigate tools that can provide Google-like searching (need to investigate MERA PEAK, NSA's LEXHOUND).
- May need to be a timeframe element in the enrichment, content that is 2 years old may not be relevant.

POC: [REDACTED]< email)

[edit] Your Thoughts

If you've got any thoughts on this initiative, please get in touch either directly to [REDACTED], or feel free to edit this section and add them below:

-
-
-

Retrieved from "[REDACTED]"

Open Source for Cyber Defence/Progress

From GCWiki
< Open Source for Cyber Defence
Jump to: navigation, search

Many structured datasets are now available in the HAPPY TRIGGER database. Unstructured datasets are being worked on and will go to LOVELY HORSE. Other integration with TWO FACE and ZooL is in place, and more will come to XKEYSCORE.

Contents

[edit] Data currently gathered

Data source	Nature of the data	OPP-LEG Status	In HAPPY TRIGGER?	In LOVELY HORSE?	In ZooL?	In TWO FACE?	Update frequency
alexa.com	Top domains list; has previously been used to find popular social networking sites in foreign countries to help with analyst investigations.	Approved	☐		☐	☐	Automatic updates on daily basis
user-agents.org	User agent strings, useful for finding spoofed or malicious entries	Approved	☐		☐		Manual update
www.nsrl.nist.gov	Access to hashes of known COTS files	Approved (for free scrape)	☐				Manual update every three months
www.maxmind.com (ASN list)	Used to help map out IP ranges of networks being monitored.	Approved (for free scrape)	☐		☐		Manual update on best endeavours basis
ZeusTracker.abuse.ch	Zeus specific malware tracking including IPs, binaries and domains to be used by the e-crime team.	Approved	☐		☐	☐	Automatic updates on hourly basis
SpyEyeTracker.abuse.ch	SpyEye specific malware tracking including IPs, binaries and domains to be used by the e-crime team.	Approved	☐		☐	☐	Automatic updates on hourly basis
amada.abuse.ch	Useful for declassifying information about known malicious IPs and domains.	Approved	☐		☐	☐	Automatic updates on hourly basis
http://torstatus.blutmagie.de/	TOR consensus document, useful for identifying whether a target was using TOR and the status of the individual nodes.	Approved	☐		☐	☐	Automatic updates on hourly basis
EmergingThreats.net	Snort rules used for network monitoring purposes	Approved (for Free data)	☐		☐	☐	Manual updates on best endeavours basis
PremiumDrops.com	Daily newly registered domains to alert analysts to suspicious domains worth investigating for malicious activity	Approved					Currently unavailable, need to find covert access method for paid content
verisign.com	Monthly updates of newly registered domains to alert analysts to suspicious domains worth investigating for malicious activity	Approved					
MalwareDomainList.com	General malware tracking resource	Approved	☐				Currently one-off sample
twitter.com	Real-time alerting to new security issues reported by known security professionals, or planned activity by hacking groups e.g. Anonymous. For more information about the sources currently being brought into the building see source list on the LOVELY HORSE wiki	Approved		☐			Prototype currently running. For more information see LOVELY HORSE
ContagioMiniDump.com	Most recommended blog by CDO analysts. Highly regarded for malware analysis relevant to APT investigations. Can be useful to declassify information for reporting purposes	Approved					
metasploit.com	Access to new zero-day exploits for the malware team to analyse	Approved (for free data)					
exploit-db.com	Access to an archive of exploits and vulnerable software. Exploits from submittals and mailing lists collected into one database.	Approved					
ics.sans.edu (Internet Storm Center)	Already used by GovCertUK on a daily basis for timely and relevant security news and incident reporting.	Approved	☐		☐		Currently updated on best endeavours basis
POSITIVE PONY	IP address to company and sector mapping. See the POSITIVE PONY wiki page for more details.	Approved Further approvals pending	☐ (dev)				Currently a static data set
NETPLATE	Multiple data types - details will be included on this page when releasable						

(POSITIVE PONY screenshots.)

[edit] Future ones to work on

Knowledge required	Available from	Update frequency	Filtering	Volumetrics	Comments
whois records	From the Passive Sigint system, or buy from RIRs (Regional Internet Registeries)? Or can we find another way of getting all updates copied to us? What about NSA's FOXTRAIL? Or our own GeoFusion? And there's now REFRIED CHICKEN from [REDACTED] ("It's a database of passively intercepted domain WHOIS records, searchable by any word in the record. Since Feb 2011. There are legal and policy constraints which mean you cannot search domains, or terms within records, that may be sensitive on grounds of location or nationality without appropriate authorisation. If you would like an account please let me know. Access to the data relies on having a Global Surge Account.")	every few days	don't know	don't know: ask the NAC?	NSA's FOXTRAIL is in this space, and needs more checks to see whether it isn't suitable. And GeoFusion (poc: [REDACTED]).
recent domain registrations	maybe an analytic run against the main DNS records to find the new domains -- or is there a more definitive source? Companies like Cyveillance are able to obtain feeds of new domain registrations (for 'brand monitoring', so I imagine we'd be able to get hold of something similar... [REDACTED]@gchq 09:51, 7 September 2011 (BST)	ready for morning and afternoon 'shifts'?	none?	very small (MB)	NSA's FOXTRAIL is in this space, and needs more checks to see whether it isn't suitable

Site	Type of data	Legal status
Pastebin	An increasing number of tip-offs are coming from the Pastebin website, as this is where many hackers anonymously advertise and promote their exploits, by publishing stolen information. An automated, regular search (say, weekly) across Pastebin for certain keywords such as .gov.uk or GSI or HMG etc. would be very valuable to ensure that GovCertUK is always notified if any information that they need to be concerned about appears in open source. "30-11-2011 GovCertUK briefed about an attack on a UN server. This tip came from open source and specifically from Pastebin where the stolen emails and passwords had been posted online."	NOT APPROVED: This nature of this site means that it would be very difficult to demonstrate the proportionality of scraping the whole site to identify the small proportion of information that would be of value to CDO and therefore approval cannot be given for scraping of the site.
OVAL List	for NDR to feed into HIDDEN SPOTLIGHT vulnerability database	APPROVED
Afraid.org	[REDACTED]: This lists domains which are publically available for anyone to add a sub-domain to. CDO analysts have suggested that this should be another resource they check alongside whois and robtex when investigating a domain.	
Joe Stewart's blog for Dell Secure Works	[REDACTED]: this regularly includes SNORT rules and other information that can be signatured.	APPROVED
scadasec mailing list	[REDACTED] request	APPROVED

[edit] Vulnerability Intelligence

Knowledge required	Available from	Update frequency	Filtering	Volumetrics	Comments
twitter traffic for vulnerabilities	use twitter API in standard way	hourly?	by twitter names of known malware/vulnerability researchers	very small (MB)	Current work is BIRD SEED. JTRIG's BIRDSTRIKE provides the scraping already, but only for handfuls of IDs, and doesn't repeat. The tweets requires data mining. Experiment run by CDT for NDR using Cyber Cloud, and has OPP-LEG approval already.
certain blogs and CERT web sites for vulnerabilities	direct web scrape (if allowed). MHS OSINT pages have examples?	hourly?	by list of specific sites/pages	small (GB)	TR-CISA have previously run several contracts looking at this problem, with a view to delivery to CNE. Final wrap up work is scheduled to automate the derivation of SEM rules (see TR-FSP) from open source information such that machines matching those rule (vulnerabilities) can be found in passive. Wanted by NDR (ref MARBLE POLLS) and GovCERT. See Open source vulnerability sources.
certain CERT IRC chatrooms for vulnerabilities	direct IRC access (if allowed)	hourly?	by list of specific IRCs	v.small (MB)	NB: Assume will include some encrypted IRCs. Wanted by GovCERT. Maybe a MARBLE POLLS source.
certain CERT email lists for vulnerabilities	direct reception	hourly?	by list of specific mailing lists	v.small (MB)	NB: Assume will include some encrypted email (including PGP). Wanted by GovCERT. Maybe a MARBLE POLLS source.
Commits to open source code repositories and security patch check-ins	GitHub etc.	daily?	by specific code projects, presumably	small (GB)	Requested by NDR [REDACTED].
Emerging Threats 'Open'	Scraped via SHORTFALL framework	Daily?	By updated Snort rules	???	Approval granted from OP-LEG to scrape info.

[edit] Bulk Infrastructure Data

Knowledge required	Available from	Update frequency	Filtering	Volumetrics	Comments
known malware/bot/spam servers/orbs/relays	eg, SpamHaus block lists, DNS block lists (dnsbl.abuse.ch), DNS blackholing lists (malwaredomainlist.com), Drive-by downloads (blade-defender.org) etc.	several times a day	none	small (GB)	SpamHaus import is already an exploit-level service from ITServices. TR-CISA have just completed an initial study of open sources of this sort of information, with an initial delivery of sample data to CDO. Longer term, we can set up an automated service to fetch this regularly from the Internet, although initially we will use JTRIG infrastructure. Some directly requested by CDO via [REDACTED].
known good lists	eg, Clean MX (support.clean-mx.de), and perhaps Google's Safe Browsing API could be used (see blog entry?	several times a day	none	small (GB)	Directly requested by CDO via [REDACTED]
known ORB servers	from sources eg, GhostNet	daily	none	very small (MB)	idea from CDO

[edit] Miscellaneous

Knowledge required	Available from	Update frequency	Filtering	Volumetrics	Comments
UK address to protect	need to find out how we get them at the moment.	weekly?	none	small (GB)	[REDACTED] apparently got complete list of .gov.uk domains via JANET in June 2011. [REDACTED] trawled KED (and therefore probably Akamai whois data) to find some List X network info.
USER_AGENT strings, sources, and expected frequency	?	weekly?	none	small (GB)	see User Agent prototype by [REDACTED]. Of wider interest.
Malware development and hacking techniques being discussed in forums	requires covert monitoring of forums	weekly?	?	?	CKX currently working with E-crime to identify and evaluate forums of potential interest. This project may extend to active monitoring of and reporting on discussions in selected forums. CKX Ops Manager is [REDACTED].

POC: [REDACTED]
POC: [REACTED]
POC: [REDACTED]

- Discussion
- Edit
- History
- Delete
- Move
- Watch
- Additional Statistics

Personal tools

- [REDACTED]
- My talk
- My preferences
- My watchlist
- My contributions

Navigation

- Main Page
- Help Pages
- Wikipedia Mirror
- Ask Me About...
- Random page
- Recent changes
- Report a Problem
- Contacts
- GCWeb

Search

Go Search

Toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

- SIDtoday Article
- Letter to the Editor
- SIGINT-y Social Media Page

(TS//SI//REL) Who Else Is Targeting Your Target? Collecting Data Stolen by Hackers

FROM: (U//FOUO) Menwith Hill Station (F77)

Run Date: 05/06/2010

(TS//SI//REL) Hackers are stealing the emails of some of our targets... by collecting the hackers' "take," we 1) get access to the emails themselves and 2) get insights into who's being hacked.

(TS//SI//REL) People who open attachments from unknown senders (gasp) or respond to "Nigerian" money laundering emails aren't the only individuals on the internet being hacked. Some of ***our*** targets are also being targeted by outside forces, both by state-sponsored and freelance hackers. Could your target's communications be the target of other countries or groups?

(TS//SI//REL) Recently, Communications Security Establishment Canada (CSEC) and Menwith Hill Station (MHS) discovered and began exploiting a target-rich data set being stolen by hackers. The hackers' sophisticated email-stealing intrusion set is known as INTOLERANT. Of the traffic observed, nearly half contains category hits because the attackers are targeting email accounts of interest to the Intelligence Community. Although a relatively new data source, TOPIs have already written multiple reports based on INTOLERANT collect.

(U) **Technique**

(TS//SI//REL) To the analyst using SIGINT databases, collected INTOLERANT data looks like Simple Mail Transfer Protocol (SMTP) mail. In this case, though, the traffic fairy has been hard at work... To hide the traffic, the hackers' programs split a victim's email into pieces. Each piece is then obfuscated, given a different, spoofed, source IP address and sent to a different destination IP address. Having different destination IP addresses serves to route the pieces across separate channels[1] of a satellite signal. The channels being used carry large amounts of traffic, allowing INTOLERANT data to hide as background noise. Much collaboration between CSE, MHS, GCHQ and NSAW has brought about the transformation of INTOLERANT data we collect into "readable" SMTP mail.

(U//FOUO) **Victim Set**

(TS//SI//REL) INTOLERANT traffic is very organized. Each event is labeled to identify and categorize victims. Cyber attacks commonly apply descriptors to each victim - it helps herd victims and track which attacks succeed and which fail. Victim categories make INTOLERANT interesting:

A = Indian Diplomatic & Indian Navy
B = Central Asian diplomatic
C = Chinese Human Rights Defenders
D = Tibetan Pro-Democracy Personalities
E = Uighur Activists
F = European Special Rep to Afghanistan and Indian photo-journalism
G = Tibetan Government in Exile

(TS//SI//REL) New victims appear to flood out their entire inbox, going back months or, even, years. Then only new mail is transmitted. Hundreds of emails are seen on an average day.

(U) **Attribution**

(TS//SI//REL) Within the world of cyber exploitation, attribution is always difficult and INTOLERANT is no exception. Initial analysis points toward a likely state sponsor based on the level of sophistication and the victim set. Determining which state is sponsoring the activity has yet to be done. Since the traffic is traveling over satellite, the culprit must be within the satellite beam's footprint to receive the stolen emails. There was hope the footprint would point to which state was responsible, but that hope was not realized as shown in the image.

(TS//SI//REL) Attribution of INTOLERANT data is difficult, since the satellite beam footprint is so large. Eventually, the virtual team working this effort would like to know who is hacking whom.

(U) **Way Forward**

(TS//SI//REL) Analysis continues with the goal of learning more about the attacks as well as improving attribution. Efforts are also being made to inform relevant parties, including NTOC, due to the obvious operations security (OPSEC) concerns where US and UK authorities have contact with Indian diplomats or the European Special Representative, for instance.

(TS//SI//REL) ***So the next time you scan your target's email, pay special attention to the case notation.*** If it contains 4PXFIL[2] (E9BDJ4PXFIL.targetNumber in the case of INTOLERANT), then the email is likely available because somebody else has hacked your target. For additional details, send an email to mhsindex@nsa.ic.gov.

(U//FOUO) POCs: [REDACTED]; INDEX team (MHS)

(U) Notes:

1. (U//FOUO) Packet Identifiers, PIDs are used in satellite hub signals to designate sub-channels.

2. (TS//SI//REL) 4PXFIL stands for "fourth party exfil" or "out-sourcing SIGINT." These terms are used within the SIGINT community to refer to the practice of collecting data as it transits the Internet going from the victim's computer to the attacker's.

(U//FOUO) SIDtoday editor's note: This article is reprinted from MHS's Horizon newsletter, March edition.

Comments/Suggestions about this article?

Information Owner: [REDACTED] Page Publisher: [REDACTED]
Last Modified: 11/10/2012 / Last Reviewed: 11/10/2012