CipherCloud DMCA notice

CipherCloud just filed a DMCA notice with stack exchange to take down the question How is CipherCloud doing homomorphic encryption? (now deleted by stackexchange).

Since I obviously can't post the full question+answers here, a short summary of what it contained:

- The question itself asks if/how CipherCloud uses homomorphic encryption.
- The answers use public information (from their website, papers, promotional videos etc.) to make educated guesses about how their encryption might work.
- Answers guess quite a few technical details (from that public information)
- CipherCloud doesn't give any clear description of how their crypto works. So the answers might not be entirely correct, but seem reasonable.
- Some answers are pretty critical of the security
- Some answers use a few screenshots as evidence for their speculation (three total)

I'd like to get it online again, minus the infringing material if there is any. Some thoughts:

- It's not really clear to me how this question or any answer violate CipherCloud's rights. The texts look like they're written by the posters, and not taken from the CipherCloud website.
- The only copyrighted material should be the screenshots.

  So perhaps we could edit them out them out to get it online again? Do a few screenshots of a piece of software really count as copyright infringement? Or does that fall under fair use?

- Or is the takedown not about copyright, but rather slander or something similar? I'm not too familiar with the US legal system and don't know if DMCA can be used for that.

  While critical, the answers look like reasonable speculation based on the few available facts to me.

- Or is their crypto itself so secret that analyzing it in public already infringes their rights? Seems very unlikely to me that IP rights extend that far.

Is there any information in the notice beyond the url and the sender? The email from stackexchange didn't contain anything else specific to this case.

Update: SE sent a copy of the DMCA notice to involved users. I'll leave publication to SE, but here is a summary:

The notice consists of two parts. The first is a DMCA notice for copyright infringement. In particular they claim that using the three images infringes their copyright.

The second part is about SE ToS and "false and misleading statement"s. They go through the answers disputing the accuracy of technical statements. In particular they claim that the system is not deterministic, defeats frequency analysis and that CipherCloud does not incorporate 1:1 mapping.

There is some statement that I read as a claim that CipherCloud's product offering could different from the public demo sid observed. But the wording is a bit ambiguous.

The first part of the notice has some merit: The images were taken from CipherCloud material. I still think using them falls under fair-use. But since that's a tricky area of law, I probably won't challenge that.

The second part is interesting. Some of the statements they challenged were dubious, in particular the statements about ECB mode and xor are probably not correct.

But at least the determinism claim has some clear evidence in the screenshots. For example in one screenshot (From their 5 minute product tour video around 2:53) the words `meet` , `to` and `want` occur multiple times with matching ciphertexts and `new` occurs twice with differing case and slightly different (but obviously related) tokens. So at least in that screenshot some form of deterministic encryption with 1:1 mappings between words and ciphertext tokens was used.

discussion deleted-questions

edited 1 hour ago asked 18 hours ago

CodesInChaos
5,538 1 3 11

1 The screenshots can be defended as fair use. The "false and misleading" part is specious; it's not even a valid part of a DMCA takedown notice. They are, of course, completely within their rights to post a rebuttal answer on its own merit, but apparently they chose to abuse the DMCA Takedown system instead. – Robert Harvey 1 hour ago

6 Answers

http://webcache.googleusercontent.com/search?q=cache:FYBbAFUycYQJ:crypto.stackexchange.com/questions/3645/how-is-ciphercloud-doing-homomorphic-encryption+&cd=1&hl=en&ct=clnk&gl=us

answered 16 hours

ago
Josh
101 1 3

13 Saved it to my server: lajm.eu/emil/dump/ciphercloud-security.html – Emil Vikström 9 hours ago

5 Getting something off the internet is like getting pee out of the swimming pool... – TheHippo 6 hours ago

@TheHippo, Dilution is the solution to pollution. – mikeazo ♦ 5 hours ago

I have a copy as well incase anyone needs it. I guess not, given the kind of attention it gets now, but just in case – Luc 5 hours ago

It would be great if Stack Exchange's legal department would post a copy of the DMCA notice. DMCA notices have to contain several pieces of information, one of which is the work being infringed. If they don't provide that (and I'd be curious to see what that work is), then it's not a valid takedown notice per the DMCA. If it does provide that, then it should be clear which part of the answer(s) needs to be removed and that part can be removed. The answers themselves probably aren't copyrighted, but if they are, the notice will say that.

Finally, Stack Exchange's legal department should contact any user who posted an answer and ask them if they're aware that their content was subject to a copyright takedown notice. Per several legal cases, SE needs to have a repeat infringer policy, and notification of the end-user is a necessary part of that policy (how can the user be expected to stop infringing if you don't notify him that he's been infringing?). If you send notice to the users, then you should also let them know that they can file a counter-notice stating that the material is *not* copyrighted by any other party and that the DMCA notice was issued in error. You can restore content if the user provides this counter-notice.

As always, I am not a lawyer and you should not take what I have written above as legal advice. Consult a lawyer for clarification on these options or any additional ones.

edited 10 hours ago
D.W.
6,694 1 9

answered 15 hours ago
hrunting
101 2

I don't have access to the original notice. SE sent me (and probably everybody else who posted on that question) an email. That email did't contain any case apecific information beyond url and sender. – CodesInChaos 11 hours ago

Community moderators do not have any information at all, not even the reason why the question was deleted. We moderate the site in our spare time, just like you contribute to it. I am confident that staff will handle the situation in a constructive way. And now something completely different: Wow, that is a nice photo of a nice house. – Hendrik Brummermann ♦ 11 hours ago

3 See related question: Request: Please make takedown notice public, and upload it to Chilling Effects, where I request that Stack Exchange corporate share the DMCA takedown notice from CipherCloud. – D.W. 10 hours ago

hrunting, Stack Exchange *did* contact affected users. I had one of the answers, and they sent me an email with details. I think you can delete that paragraph. – D.W. 10 hours ago

2 In that case, you can file a counter-notice. I see below that someone has posted the copy of the notice sent to users with instructions for filing a counter-notice. File the counter-notice. It sounds like everyone involved has a "good-faith belief" that the material was removed mistakenly or because of misidentification. – hrunting 6 hours ago

Another point is that SE should be displaying a 451 instead of a 404 on the OP's page.

answered 16 hours ago
hafichuk
101 3

1 +1. Here's a link to the specification describing 451: datatracker.ietf.org/doc/… – L0j1k 15 hours ago

I just brought this feature request up on meta: meta.stackoverflow.com/questions/177264/… – L0j1k 15 hours ago

1 As mentioned in @L0j1k's feature request, `451` isn't a valid status yet so this "answer" is invalid. – Marcel 13 hours ago

2 @Marcel: 451 is a valid status code. Read RFC 2616 section 6.1.1, especially the part where it says that "HTTP status codes are extensible." It says clearly that clients that don't know 451 MUST interpret it as a 400 error. – Rhymoid 9 hours ago

2 "The 451 status code is optional; clients cannot rely upon its use. It is possible that certain legal authorities may wish to avoid transparency, and not only demand the restriction of access to certain resources, but also avoid disclosing that the demand was made." – Bobby Jack 8 hours ago

For those interested, here is the full text of the email that the Stack Exchange legal folks sent to me (and, I assume, to all other users who had posted answers). It describes how to send a counter-notification. I encourage everyone who posted an answer to read carefully and decide whether you want to send a counter-notification.

> Dear Member:

> This is to notify you that we have removed or disabled access to the following material as a result of a third-party notification by CipherCloud claiming that this material is infringing:
>
> How is CipherCloud doing homomorphic encryption?: http://crypto.stackexchange.com/questions/3645/how-is-ciphercloud-doing-homomorphic-encryption
>
> Please Note: In order to avoid future strikes against your account, please delete any text to which you do not own the rights, and refrain from uploading additional text that infringe on the copyrights of others.
>
> For more information about Stack Exchange's copyright policy, please read http://stackexchange.com/legal
>
> If you elect to send us a counter notice, to be effective it must be a written communication provided to our designated agent that includes substantially the following (please consult your legal counsel or see 17 U.S.C. Section 512(g)(3) to confirm these requirements):
>
> (A) A physical or electronic signature of the subscriber.
>
> (B) Identification of the material that has been removed or to which access has been disabled and the location at which the material appeared before it was removed or access to it was disabled.
>
> (C) A statement under penalty of perjury that the subscriber has a good faith belief that the material was removed or disabled as a result of mistake or misidentification of the material to be removed or disabled.
>
> (D) The subscriber's name, address, and telephone number, and a statement that the subscriber consents to the jurisdiction of Federal District Court for the judicial district in which the address is located, or if the subscriber's address is outside of the United States, for any

> judicial district in which the service provider may be found, and that the subscriber will accept service of process from the person who provided notification under subsection (c)(1)(C) or an agent of such person.
>
> Such written notice should be sent to our designated agent as follows:
>
> DMCA Complaints Stack Exchange, Inc. One Exchange Plaza 26th Floor New York, NY 10006 Email: dmca@stackexchange.com
>
> Please note that under Section 512(f) of the Copyright Act, any person who knowingly materially misrepresents that material or activity was removed or disabled by mistake or misidentification may be subject to liability.
>
> Sincerely,
>
> Stack Exchange Inc.New

answered 10 hours ago
D.W.
6,694 1 9

This DMCA is not specific enough, it should never have been processed by Stack Exchange Inc. You can safely issue a counter notice on those grounds. – seanieb 4 hours ago

3 @seanieb, that is the email from Stack Exchange staff to contributors. This is not the copyright violation complain sent to Stack Exchange by the company. – Hendrik Brummermann ♦ 3 hours ago

> Or is the takedown not about copyright, but rather slander or something similar? I'm not too familiar with the US legal system and

> don't know if DMCA can be used for that.

No. In a takedown, they have to declare that they are making a good faith complaint against something that they believe they own the copyright to. DMCA is about copyright.

answered 16 hours ago
rox0r
101 2

3 There has been a lot of news coverage in Germany about copyright being used to silence critical opinions in the USA. The EFF has a small collection of incidence that got press coverage in the USA itself: eff.org/takedowns – Hendrik Brummermann ♦ 11 hours ago

The DMCA is a very large bill and besides the safe harbor provision, which contains takedown notices and the like, it also has a bit which, IIRC, makes it illegal to describe DRM circumvention methods in some situations. I don't believe that the regular DMCA takedown notices have anything to do with this, but you could certainly send someone a letter threatening action under this part of the act. That's all from my shaky memory, but here's some real info:

http://chillingeffects.org/anticircumvention/faq.cgi

answered 13 hours ago
aptwebapps
101 1

The provision in the DMCA prohibits reverse engineering to discover, utilize and publish circumventions to protection mechanisms intended to protect copyrighted material. (I'm paraphrasing, and I'm not a lawyer). It's hard to see how CipherCloud could make such a claim from some people on the Internet merely making some deductions about their software through casual observation. – Robert Harvey 53 mins ago

How is CipherCloud doing homomorphic encryption?

Much of the literature and latest papers suggest that homomorphic encryption is still not practical yet.

How is CipherCloud able to achieve this? Does anyone have an idea? Their website does not provide much information about how their system works.

homomorphic-encryption database

edited 5 hours ago
D.W.
5,913 6 19

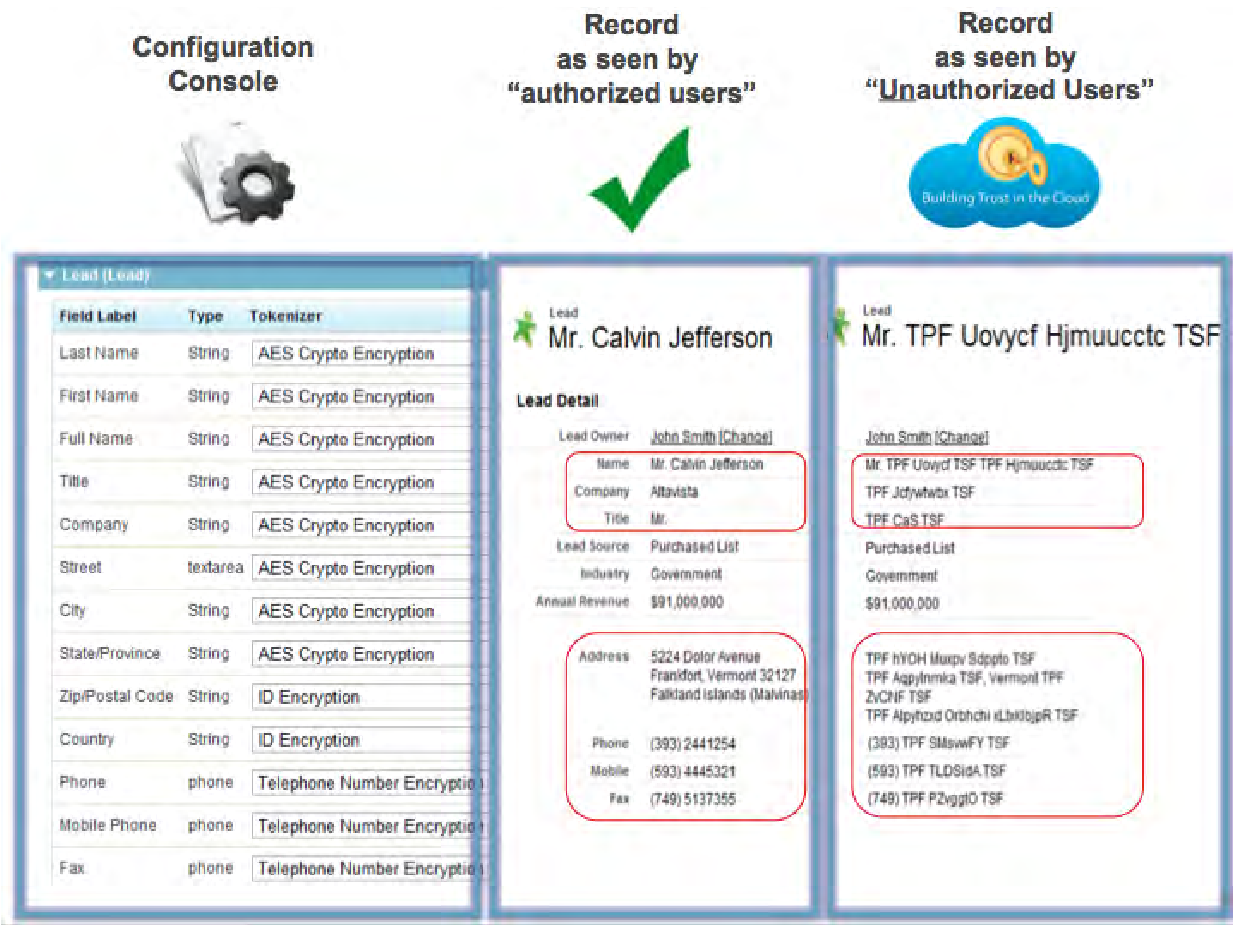
asked Aug 25 '12 at 15:09
sashank
449 1 11

6 Answers

My impression from their whitepaper is that they don't use homomorphic encryption. They seem to encrypt fields individually with symmetric encryption(mentions AES). The scheme is deterministic and format preserving.

When a field consists of several parts(say `firstname` and `lastname`) they split it into those parts and encrypt them individually.

Some important fields, such as "annual revenue" aren't encrypted at all, probably because they need to do use it in calculations.


Configuration Console
Record as seen by "authorized users"
Record as seen by "Unauthorized Users"
Building Trust in the Cloud

▾ Lead (Lead)

Field Label	Type	Tokenizer
Last Name	String	AES Crypto Encryption
First Name	String	AES Crypto Encryption
Full Name	String	AES Crypto Encryption
Title	String	AES Crypto Encryption
Company	String	AES Crypto Encryption
Street	textarea	AES Crypto Encryption
City	String	AES Crypto Encryption
State/Province	String	AES Crypto Encryption
Zip/Postal Code	String	ID Encryption
Country	String	ID Encryption
Phone	phone	Telephone Number Encryptio
Mobile Phone	phone	Telephone Number Encryptio
Fax	phone	Telephone Number Encryptio

Lead
Mr. Calvin Jefferson

Lead Detail

Lead Owner	John Smith [Change]
Name	Mr. Calvin Jefferson
Company	Altavista
Title	Mr.
Lead Source	Purchased List
Industry	Government
Annual Revenue	$91,000,000
Address	5224 Dolor Avenue Frankfort, Vermont 32127 Falkland Islands (Malvinas)
Phone	(393) 2441254
Mobile	(593) 4445321
Fax	(749) 5137355

Lead
Mr. TPF Uovycf Hjmuucctc TSF

John Smith [Change]
Mr. TPF Uovycf TSF TPF Hjmuucctc TSF
TPF Jcfywfwbx TSF
TPF CaS TSF
Purchased List
Government
$91,000,000

TPF hYOH Muxpv Sdppto TSF
TPF Aqpylnmka TSF, Vermont TPF
ZvCNF TSF
TPF Alpyhzxd Orbhchi xLbxlbjpR TSF
(393) TPF SMswFY TSF
(593) TPF TLDSidA TSF
(749) TPF PZvggtO TSF

An inherent weakness in such a scheme is that it **doesn't offer semantic security**. If the same string gets encrypted in different places, an attacker can see that the same string was used in both places.
If he can figure out one of them, he automatically known the other too. He could also employ some kind of frequency analysis on tokens. For example the most common firstname might be "John", allowing him to find all "John"s.

An interesting question is how they achieved their format-preservation. There are relatively secure methods, but it's easy to get wrong. I couldn't find any specification of what they're doing, so I wouldn't trust them on getting it right.

edited Aug 25 '12 at 15:50

answered Aug 25 '12 at 15:31

CodesInChaos

5,245 1 6 25

Yeah , security through obscurity is long dead , but they seem to be having soaring sales. – sashank Aug 25 '12 at 15:53

I don't think they have implemented homomorphic encryption at all. They have just implemented regular AES encryption (they have a FIPS 197 certificate for their AES), but in what appears to be a very insecure way. Why would they choose to do that? Because they had no choice. Here's what I mean:

The challenge for cloud encryption providers like CipherCloud that have a lightweight architecture (no required database, small storage requirements), is that they need the back-end SaaS application (Salesforce, GMail, etc.) to be able to perform all the search and reporting functions on the data as if it were in clear text. To make this possible, you must ensure that the same string gets encrypted the same way every time. As CodesInChaos suggested in an earlier answer, this makes the solution extremely vulnerable to frequency-analysis attacks.

But SaaS encryption implementation has a much larger problem. Searching for exact matches is an easy problem to solve - just send the encrypted value for the match to the SaaS application an you're all set. But that is not the only kind of search you need to support. What happens if a user does a search for all names that begin with John? (e.g. "John*") There are (at least) two options: The first is to store a mapping to every instance of every string that begins with John* in the encryption appliance, then send all the instances of the encrypted text for every mapped string that matches John* to the SaaS application so it can perform the search. That becomes problematic if there are a lot of strings that begin with "John*" - you have to send all those matching strings to the SaaS application in order to make the search work. But imagine a search for John* + Jon* + Jame* + Smith*. You could run out of query parameters pretty easily. It's even worse when running reports.

You also have to have a mapping infrastructure (a database would be the enterprise-grade way to do this) on the encryption appliance to make this work, but CipherCloud do not appear to require a database, making this approach unlikely in their case. And CipherCloud does not seem to use this approach, as it appears from their publicly available documents.

But they may have implemented a worse one, because the second way to address searches for "John*" is what they appear to have actually done. This method preserves string order within the ciphertext, such that John becomes XXyzzz, Johnathan is XXyzzzAAddBBaaBB, and Johnson is XXyzzzDdffsss (this is not their algorithm just a representation of the net effect.) That way, a search for John* means I only have to send "XXyzzz*" to the SaaS application in order to properly fulfill the search. But this approach greatly weakens the security of the data. This is because once I deduce that John is XXyzzz, anytime I see a string beginning with those characters I know it is some form of the name "John*" and I can really start attacking the data. CipherCloud claims to use AES, which should not have this problem, so how can they preserve this string order using AES? Well, the first thing to do is not use padding, or to use the same padding everywhere. Yikes! The second thing is to use the same IV (initialization vector, aka nonce) for all strings. Yikes again! Without padding and IV diversity, AES becomes a glorified version of XOR. Who would bet the security of their data on that? (This probably explains why they do not have, and are not even in process to obtain, FIPS 140-2 validation, which pertains to the proper implementation of an approved algorithm.)

More recent demonstrations of the CipherCloud solution appear to use multi-byte characters in the ciphertext, which makes the patterns harder to see by the naked eye (ok, to eyes used to parsing western character sets) but certainly no harder for a computer to crack.

I'm not sure if they are still there, but there used to be some good videos of their solutions on Vimeo and Youtube, so you can look at those and see for yourself what I'm talking about. I'm sure you can also download whitepapers from their site. I'll leave it to someone else to really dig into the available data and figure out exactly how they are doing what, but it's worth mentioning to any would-be investigators that CipherCloud also appears to be preserving certain punctuation in clear text. (I saw an instance of " I'm " encrypted in a way that preserved the apostrophe!)

As always, but doubly so when it comes to security products, Caveat Emptor! If you are looking at CipherCloud, or any SaaS encryption solution, you'd do well to ask a lot of specific questions and make sure the answers are clear and unambiguous.

answered Aug 26 '12 at 15:25

AdrenaLion
81 1

My impression is that they don't use order preserving encryption everywhere, and in the case of names they simply split it in parts, which wouldn't allow prefix-matches. But the available information is so sparse that it's hard to say anything concrete. – CodesInChaos Aug 26 '12 at 15:33

I haven't posted in a while, so long in fact that the email tied to my Stack Exchange account is no more, I forgot my StackEx password, and I had to create a new account. (I'll leave it to the reader to decide if this is the real me.)

But I did want to just to follow up here, because there were some unanswered questions from my last post and the follow-up posts from others. Since I wrote the above post, I had been wondering myself how this searchable encryption could actually work without being incredibly weak from a security standpoint. As it happened, I was at the RSA Security 2013 conference this week where Ciphercloud was exhibiting. In between sessions I had time to visit their booth to learn more.

They do claim to do "military grade encryption", and it does appear that they can use third-party FIPS 140-2 encryption modules. However, in the demonstration I was given, where they were encrypting data in a SalesForce setup, the encryption was definitely NOT using FIPS 140-2 or anything close. In fact, I could see on their large demo screen the exact issues I had expected to see with their encryption algorithm, plus some things that just made me shake my head.

For example, it turns out that they are indeed preserving clear-text patterns in their ciphertext. Searching for "John" is easy if it is encrypted the same way (eg "XXyyZ123") everywhere. But they also appear to individually encrypt each word within a string, such as you would see in an Account Name field. I know this because they showed their demo of a side-by side comparison of clear text and encrypted Accounts. There were two Accounts with "United Oil & Gas" in the name. Both the encrypted names were the same. That means they are using the same key, nonce (IV), and padding for the Account Names. Since the whole point of encryption is to promote randomness in the ciphertext, this is a pretty weak, non-random implementation. Would you entrust your data to what amounts to XOR? I sure wouldn't.

But here is the part that had be shaking my head, mainly because it takes almost zero crypto cracking skills to determine the true value of the data: They appear to have issues, for reasons I cannot completely fathom, encrypting the punctuation characters in the string. In the example they showed me, "United Oil & Gas" was encrypted as something like "fgt^e3s3 SD72d & 3edf" (Note: they also have prefixes and suffixes that wrap their encrypted strings, but I have not included them because they appeared to be pretty consistent and may be there to identify the strings as encrypted, but would do nothing to protect the data.)

So, if you are looking for a customer that named "United Oil & Gas", you have a pretty simple way to narrow down which records that could be - just search for the "&" character, and narrow those results to the one where it appears between the second and third words. Then, in that list, look at the word lengths in the name, and the strings with the short second and third strings are your best bet. This is in part because in "United Casualty & Life", the word "Casualty" would have longer ciphertext than the word "Oil". (Remember they are using the same padding to make this all searchable.) The bottom line is that encryption hasn't really protected the data here. Cost with no benefit.

But it gets worse: Once you knew you had the ciphertext for each of the words “United”, “Oil”, and “Gas”, you could just search for matches for those ciphertext patterns, and you would know all the Account Names (and perhaps all the other fields as well) that had those words in them, as well as the placement of the words in the multi-word strings stored in those fields.

But then, it may be even worse: You may even be able to derive new words based on the words you have already derived. This is because for those three clear text words, you now know the ciphertext patterns for any words that begin with those strings. (Full disclosure: Here is where I am speculating a bit because the guy showing me the demo couldn't tell me the AES modes that they use. I am assuming they use something like CBC, and that they still process in 16 bit blocks - two characters - at a time.) With CBC, the same key, nonce, and padding will preserve the patterns of the strings at the beginning of words. So "United", "Un" would share two common character patterns to start their ciphertext, and "United" and "Unit" would share the first four. So if you derive "United" you could find any word that also began with "Unit", "Un" etc.

Using the example I saw at Ciphercloud’s booth at RSA, you could use that pattern preservation to find out any other account with the word "United", or "Oil", or "Gas", as well as any words that began with the same character strings as "United", "Oil", or "Gas."

Now, I know this was a demo, and the guy showing it was probably a marketing guy with no concept of security. But this was the 2013 RSA Security show. You are going to be viewed by people like me who know a thing or two about encryption, and poke holes in the shoddy stuff. I will also say, in their defense, these shows are coordinated by their marketing department and may not have the most up to date demonstration materials. So, perhaps they could have shown a better (newer?) implementation of their product that would have satisfied me or any security professional.

But the fact remains that they did not. And at one of the largest, most influential security shows in the US, if not the world , you shouldn't put up for demonstration something so easily defeated.

Caveat Emptor!

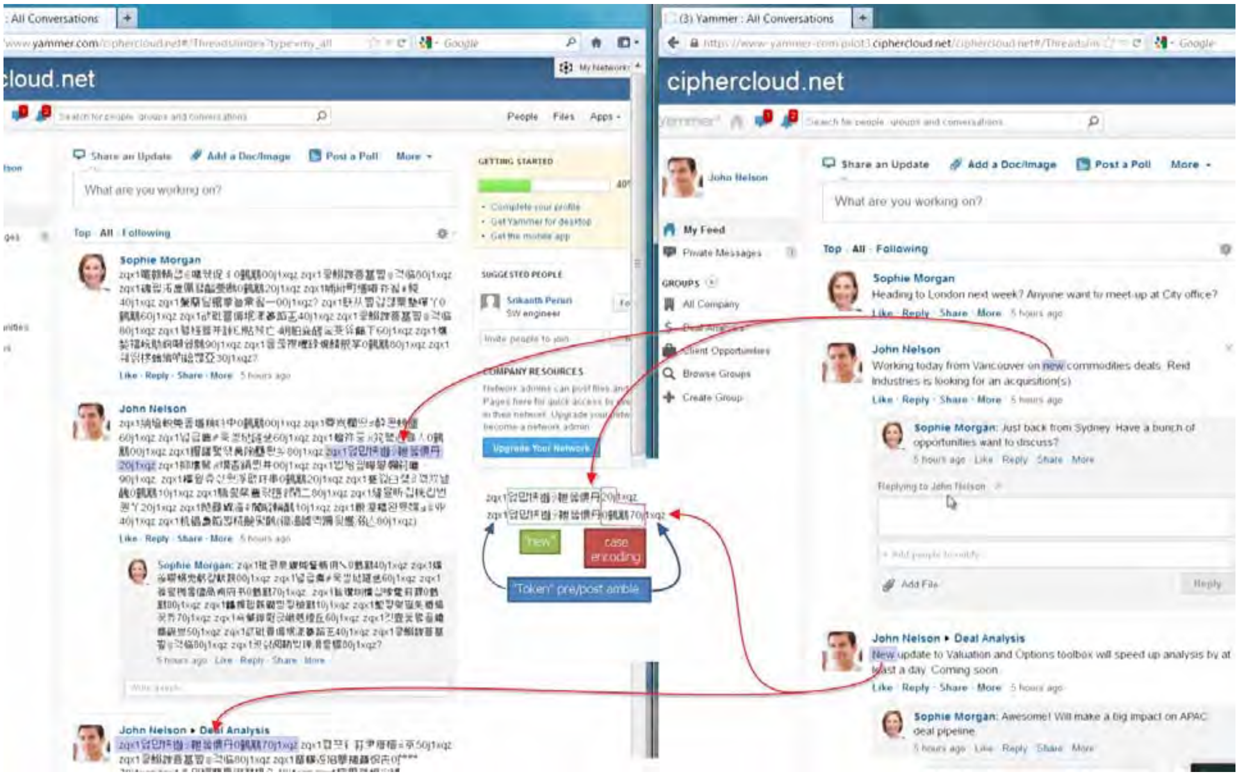
edited Mar 2 at 1:52
adrenalion
23 2

answered Mar 1 at 20:35
adrenalion
51 1 1

1 @adrenalion Welcome back! Note that while moderators can't merge your accounts, it either is possible for you to do so yourself (for the two new accounts used on this post) or request help from Stack Exchange to merge the original account. Use the merge user profiles link on the help page. – Paŭlo Ebermann ♦ Mar 3 at 21:29

They are not using any exotic encryption. In fact I don't even think they are doing any encryption, just 1:1 mapping (tokenization) after lowering the case on plain text data. For details, I did some basic crypt-analysis on a still from their publicly visible demo video.

Click this image to see the weakness! (**Full resolution image here**)


ciphercloud.net

John Nelson

Share an Update · Add a Doc/Image · Post a Poll · More

What are you working on?

Top · All · Following

Sophie Morgan
Heading to London next week? Anyone want to meet up at City office?
Like · Reply · Share · More · 5 hours ago

John Nelson
Working today from Vancouver on new commodities deals. Reid Industries is looking for an acquisition(s)
Like · Reply · Share · More · 5 hours ago

Sophie Morgan: Just back from Sydney. Have a bunch of opportunities want to discuss?
5 hours ago · Like · Reply · Share · More

Replying to John Nelson

John Nelson ▸ Deal Analysis
New update to Valuation and Options toolbox will speed up analysis by at least a day. Coming soon
Like · Reply · Share · More · 5 hours ago

Sophie Morgan: Awesome! Will make a big impact on APAC deal pipeline
5 hours ago · Like · Reply · Share · More

Basically they end up with a 1:1 mapping of lower case words. No matter if they circle the galaxy, suck all the energy of a star or perform AES256. At the end of the day it's just 1:1, at lower case word level! So you can run the entire "encrypted" conversation into a statistical analyzer and based on the frequency of regular English words uncover that 1:1 mapping. If you add the logic that word level patterns exist ("The the" is extraordinarily rare vs "extra extra") i.e. Markov chains modelling - then you need even fewer copies of "encrypted data" to peel off the security.

There is NO way I would trust my Aamzon S3 or Azure Blob storage to be encrypted by these guys.

Overall I would say this is borderline commercial *snake oil* because

- There is no security. AES is misleading/wrong (ECB + 0/constant padding + 0/constant IV => simple 1:1 mapping => why don't you just rotate the bits and call it a day!)
- Framing it as secure means IT managers might make assumptions

and relax policies they would otherwise not allow => introduces new holes on corporate security

- Additional overhead of buying their gateway boxes/maintenance with NO security => waste of corporate money/time
- Vendor lockin - I don't know if they supply migration tools because if you one day decide to upgrade to real security from their "secure" offering, you can't. Because their box/gateway is the only one that would know these 1:1 mappings. Unless you want to formalize hacking yourself as the official upgrade and data extraction path.

edited Mar 13 at 16:24

answered Mar 7 at 23:12
Sid
156 5

I don't know how CipherCloud works. However, a related question is: How could you encrypt data in a database, in a way that allows you to achieve these goals? What are the best cryptographic techniques currently known, for that goal?

As it happens, that question has a good answer. Take a look at CryptDB, a system built by MIT researchers to encrypt all the data in your database while still allowing your application to manipulate the data. In their system, the application can execute SQL queries on the encrypted data (even though it is encrypted!) and do some limited computation on the data.

CryptDB uses a combination of techniques that have been developed by cryptographers over the past decade or two, to achieve these goals. They show that the result is practical, with good performance and ability to use it with existing systems (like phpBB). It's a brilliant system, and a significant advance for the field. Read their research paper for more on how they do it:

- CryptDB: Protecting Confidentiality with Encrypted Query

Processing. Raluca Ada Popa, Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan. SOSP 2011.

In summary, the techniques in CryptDB are what Ciphercloud *ought* to be doing. I have no clue whether Ciphercloud is *actually* doing that (you'd have to ask Ciphercloud that), but CryptDB represents about the state of the art in this area right now.

edited 5 hours ago answered Aug 25 '12 at 18:04

D.W.

5,913 6 19

I also watched the video (thanks Sid, for the link) and after looking at it, it reveals some of the other methods that Ciphercloud appears to be using to preserve search. Nothing appears to be an implementation of any sort of homomorphic encryption.

I snapped a copy of one screen after the response from John is entered and encrypted, and have attached an image below (apologies for the crude highlighting). Look at the word "meet" in John's post and then "meet-up" in the first post from Sophie. The pattern of ciphertext for the string "meet" is the same in both, which would be required if you were to perform searches by encrypting the user input of "meet" and sending the ciphertext to the cloud to actually perform the search.

I have not had time to fully explore this, but note that in "meet-up" the hyphen is preserved in the clear within the ciphertext. I suspect that this is because there is a requirement to enable search for the word "up", which basically requires setting the IV back to one of its static values like the one used when "meet" was encrypted or perhaps the one (assuming that there are any other IVs) used to encrypt other instances of "up". This is the only way to guarantee that the suffixed "up" will match the singular instance of "up".

I didn't highlight it, but you can also see that terminating punctuation such as question marks are preserved in the clear. Again, if you want to perform an exact match search for "meet", you need to strip the extra character because the ciphertext for "meet?" would be different than from "meet" so the search would not return results that a human would expect.

http://i.stack.imgur.com/oBXZJ.jpg

But, the implication here from a security perspective is that if I am able to plainly see punctuation such as hyphens, and preservation of patterns in the ciphertext is so critical that I have to strip (and then reveal!) trailing punctuation, then an attacker is provided a head start in breaking down the encryption. If you are not promoting randomness in your ciphertext you are not encrypting. What Ciphercloud appears to be doing is not random, therefore it is not truly encryption, and certainly not homomorphic encryption.

So, the answer to the original question is that Ciphercloud is NOT doing homomorphic encryption.

As always, Caveat Emptor! adrenalion

answered 12 hours ago
adrenalion
23 2