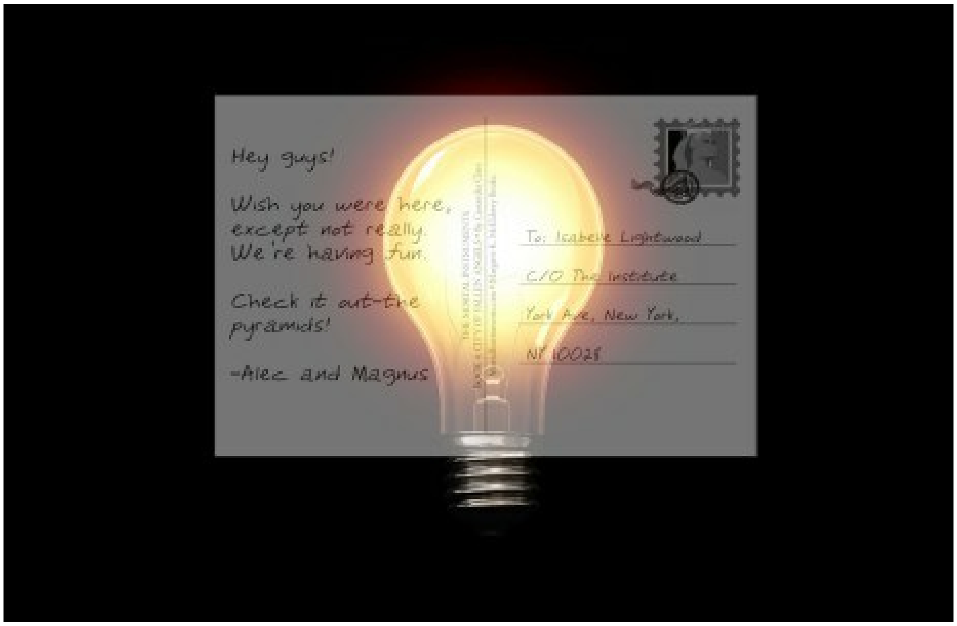
Data-mining the Haystack: Should You Attempt to Overload the NSA's Servers with "Suspicious" Email Keywords?

www.thelastbastille.com/2016/02/18/data-mining-the-haystack-should-you-attempt-to-overload-the-nsas-servers-with-suspicious-email-keywords/

*"The problem was not that we weren't collecting information, it wasn't that we didn't have enough dots, it wasn't that we didn't have a haystack – it was that we did not understand the haystack that we have. The problem with mass surveillance is that we're piling more hay on a haystack we already don't understand. And this is the haystack of the human lives of every American citizen in our country. If these programs aren't keeping us safe, they making us miss connections, **vital** connections, on information we already have, if we're taking resources away from traditional methods of investigation, from law enforcement operations that we know work, if we're missing things, like the Boston Marathon bombings, where all of these mass surveillance systems, where every domestic dragnet in the world didn't reveal guys that the Russian intelligence service told us about **by name**, is that really the best way to protect our country? Or are we trying to throw money at a magic 'solution' that's not just costing us our safety, but our rights and our way of life?"*

– Edward Snowden



Misinformation abounds when it comes to mass data-mining. Typically, the federal government will either deflect attention by stressing they are collecting **only** foreign intelligence (*not* domestic intelligence) or that they are **only** collecting metadata (*not* the content of your communications). Whether it be by way of the USA PATRIOT Act's § 215 or FISA's § 702, the fact of the matter is that any hypothetical "limits" imposed by these statutory laws don't matter, and the constitutionality of them even less so, considering the history of the unconstitutional National Security Agency (NSA).

All of these excuses for **strictly** foreign and/or metadata intelligence gathering by the NSA is really quite spurious for two major reasons – ECHELON and the Utah Data Center. Back in July of 2001, the European Union Parliament issued a report that admitted to the existence of ECHELON as a signals intelligence (SIGINT) network of the Five

Eyes intelligence alliance, which was established by the UK-USA Agreement. The Utah Data Center was exposed a few years ago as being part of both STELLAR WIND as well as the Department of Defense's Global Information Grid, which handles data to the magnitude of yottabyes (that's equivalent to 1 trillion terabytes). Between ECHELON and the Utah Data Center, the NSA has no *practical* or *technical* reason to "limit" its own power, given that the NSA collects **domestic** intelligence (including **content**) through their foreign allies who spy on Americans, and then they store all that information in approximately $2 billion worth of cloud computing.

Despite all this, there is still a naïve belief in the virtues of reformism. It matters not that the FISA Amendments Act ("the reform of the reform," as it were) *expanded* the scope and authority of § 702, but rather the fact that surveillance programs like PRISM, MUSCULAR, and XKeyscore were only possible because each one of them were an *expansion* upon the Five Eyes' ECHELON system, which itself exploits the Internet's backbone. What I inferred from a few declassified FISC rulings from over two years ago was that this whole "bundling" claim is disingenuous too, because if the NSA knew from a purely technical standpoint that indiscriminate data-mining was unavoidable, then they shouldn't be doing it *at all*, simply due to the fact that such bulk data collection is *at least* a search, if not also a seizure, of one's "papers and effects" through the Internet, and is therefore a blatant violation of the Fourth Amendment. No amount of "reforming" the NSA is even applicable to GCHQ, CSE, ASD, or GCSB in the first place, especially in terms of curbing their espionage against the communications of the American citizenry.

Taking this as a given, the NSA has got to put all that SIGINT they got from their foreign allies, *somewhere*, right? William Binney, a NSA whistleblower, warned that the Utah Data Center engaged in **deep packet inspection** of email, specifically. This distinction made by government apologists between foreign versus domestic spying is largely irrelevant, not only because of the Five Eyes network, but also the fact that many email servers are themselves located offshore, and therefore susceptible to SIGINT espionage by the NSA. The reason the Utah Data Center is geared towards storing yottabytes worth of information is simply because that very intelligence is not just metadata, but content as well; otherwise, they wouldn't need that kind of storage space now, would they?

Remember, the federal government still insists on exercising its "states secrets privilege" as a matter of **public policy**, so you have to ask yourselves – do you honestly think that *any* high-level government official would be willing to brazenly lie to the public if by doing so he would be preserving classified "national security" procedures? Congresscritter Hank Johnson questioned then-NSA chief General Keith Alexander about the NSA's technical capability in intercepting SIGINT during a congressional subcommittee hearing on March 20th of 2012 (*before* the Snowden leaks); the general flat out denied in his testimony that the NSA collected bulk data of Americans' communications, including content. Consider also these statements made by Barack Obama back on July 7th of 2013 (post-Snowden), where he did admit that the NSA collected metadata on American citizens:

> *"Nobody is listening to your telephone calls."*
>
> *"Now, with respect to the Internet and emails, it does not apply to U.S. citizens and it does not apply to people living in the United States…"*
>
> *"They do not involve listening to people's phone calls, do not involve reading the emails of U.S. citizens or U.S. residents."*

Now, if all of this email and phone surveillance is unquestionably and completely legal, then would *el presidente* please be kind enough to explain to everyone why there are as many credible NSA whistleblowers as there are? Is the Tyrant-in-Chief implying that they all just exaggerated a bit too much, and therefore *deserved* all of the harassment and sometimes outright persecution they faced? Or is there serious foul play afoot?

Decades ago, William Knowles suggested that it might be worth "overloading" the NSA's servers with "suspicious" keywords. Fast-forward to 2013 when ***Vice*** attempted to do just that with its **Hello, NSA** random keyword laced message generator; yet, this was more of an attempt to hamper the ability of the NSA to data-mine "social media" websites for open-source intelligence, as evidenced by DHS' redacted 2011 *Analyst's Desktop Binder*. Concurrently, there was a similar attempt to obfuscate the value of any bulk data collection, as well as call the President's bluff that the NSA does not listen to phone calls or read emails, by **Operation Troll the NSA,** which encouraged the public to call and/or email a script containing these "suspicious" keywords on the same day. Naturally, as is typical with all "critical mass" protests, there was no measurement metric taken by which to gauge the effectiveness of such a (presumably) widespread action.

Let's talk shop, shall we? The average email size is roughly 75 kilobytes, without attachments. If we make an assumption that the Utah Data Center has a maximum storage capacity of a single yottabyte and divide that by the average email size, then it would take 40 quintillion emails to fill it (given that one yottabyte equals one sextillion kilobytes). Since the total American population is estimated to be 323 million people, and assuming that each one of them sent out 100 emails a day without fail, then it would take them 80 billion years to reach the Utah Data Center's presumed maximum capacity.

I think you can tell by the results of my calculations that the average email is the smallest conceivable file size, whereas the Utah Data Center is the largest conceivable set of cloud computing servers; therefore, this presumption that nearly innumerable swarms of keyword laden emails could ever make the slightest of dents into overwhelming the NSA's servers is pure foolishness. If you were to attempt such a dent, it would've been better to use the largest conceivable file size that has the highest possibility of successful file transferal without interruption or glitches, such as those videos that feature 10 hours of elevator music. Obviously, this would be something akin to the black fax method, which attempts to run the black ink within printers dry through sheer waste.

It is more than fair to say that the efficacy of email keywords "overloading" the NSA's servers has not been proven; what evidence and arguments have been presented here are a critique that is offered in good faith. Advocates of this method bear the burden of proof for demonstrating its efficacy, not me or any other critics. Suffice it to say, I don't think those proponents have met their burden of proof, and what evidence is available suggests that their purported method is unfeasible, especially considering the obvious failure that was Operation Troll the NSA. What is the solution here, then, if not "overloading" the NSA's servers with "suspicious" email keywords?

The solution is, therefore, *not* using keywords to either "trip" or "overload" anything, but rather encryption. In this case, since the concern is about emails being data-mined and read by intelligence analysts, the specific encryption type would be Pretty Good Privacy (PGP). Nearly two years ago, I wrote two different tutorials for configuring PGP with your operating system; one for a Windows OS, and another for a Mac OS. Within the EU Parliament's 2001 report about ECHELON, they recommended the use of encrypted email:

> *"In contrast [to telephones], e-mails can and should be encrypted by everyone. The oft-repeated claim that a person has no secrets and thus has no need to encrypt messages must be countered by pointing out that written messages are not normally sent on postcards. However, an unencrypted e-mail is nothing other than a letter without an envelope. The encryption of e-mails is secure and relatively straightforward and user-friendly systems, such as PGP/GnuPG, are already available, even free of charge, to private individuals on the Internet. Unfortunately, they are not yet sufficiently widely distributed. The public authorities should set a good example and themselves employ encryption as a standard practice in order to demystify the process."*

Now, GNuPG is sufficiently widely distributed, and considering the prevalence of crypto-currencies like Bitcoin, any excuses of computer illiteracy simply get tossed out of the window for the rubbish that it is. As long as you have electricity and Internet access, you can use PGP encrypted emails, once it's been configured.

Since the NSA is indiscriminately data-mining nearly everyone, I'm sure they've got a copy of Hillary Clinton's missing Benghazi emails, am I right? Either way, the more serious issue is that too many American dissidents of various ideologies are, oddly enough, stubbornly consistent in their practice of making all manner of excuses in order to *not* practice good security culture. If they took one-tenth of that effort in *avoiding* the practice of security culture techniques and redirected it towards learning these methods, then they'd already be using them well enough for it to be beneficial to them. Why there is this tendency to needlessly increase one's opportunity costs is a mystery I intend to solve, for it would explain their aversion towards what does work, as well as their attraction for what does not work, in terms of helping them secure their own personal liberty.