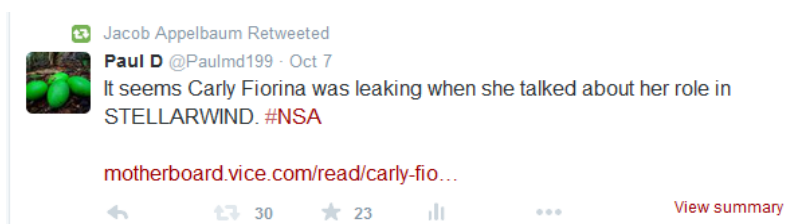


Oct 18, 2015 (Updated Oct 23, 2015)
Paul Dietrich (@paulmd199)

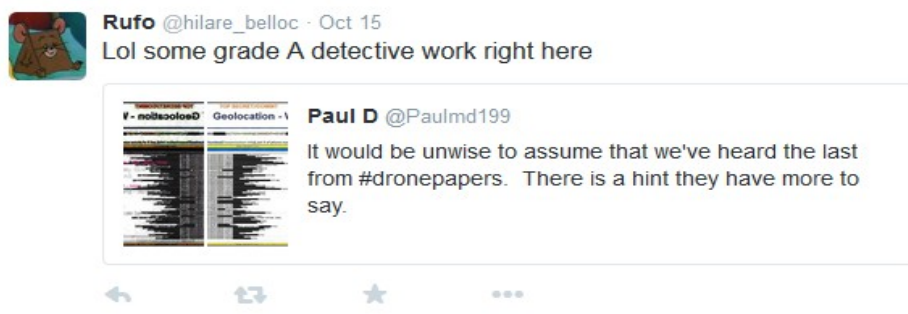
Adventures in Twitter Censorship

It started simply enough, I woke up this morning to an unusually high number of notifications mentioning several posts I had made. Usually this means that an account with a large following retweeted something I said. I was curious to find out who.

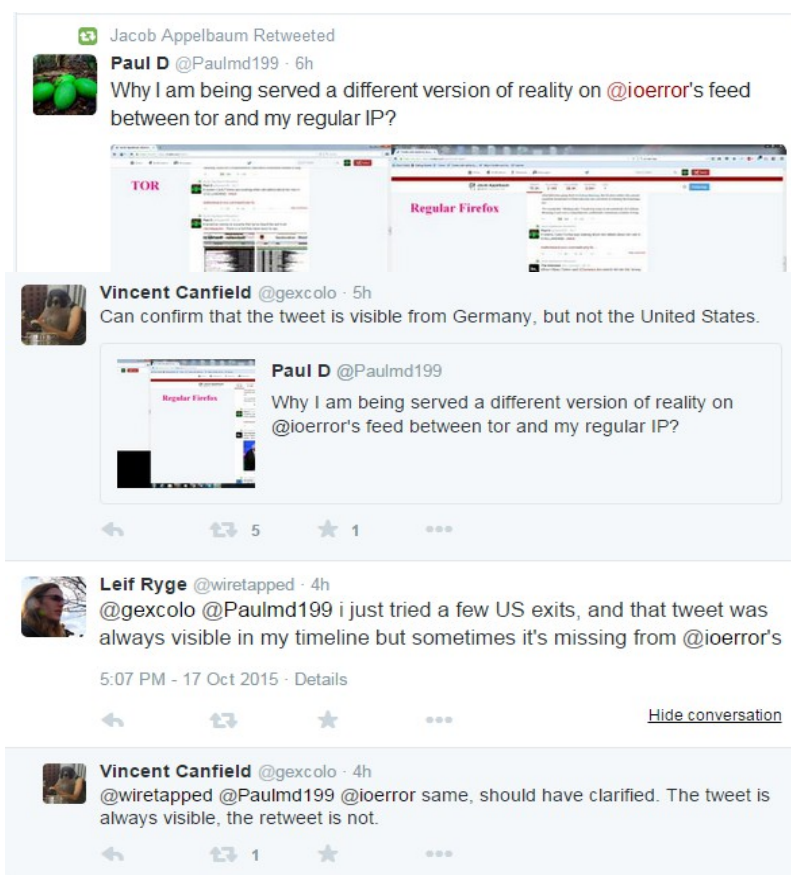
I determined that Jacob Appelbaum (@ioerror) had retweeted a [news article](#) I was involved with. But a second [post](#) I made also suddenly had a large number of retweets. I saw a notification that he had favorited it, but no indication that he retweeted. Favorites usually don't have a snowball effect. I thought it might be the case that he perhaps had undid the retweet.



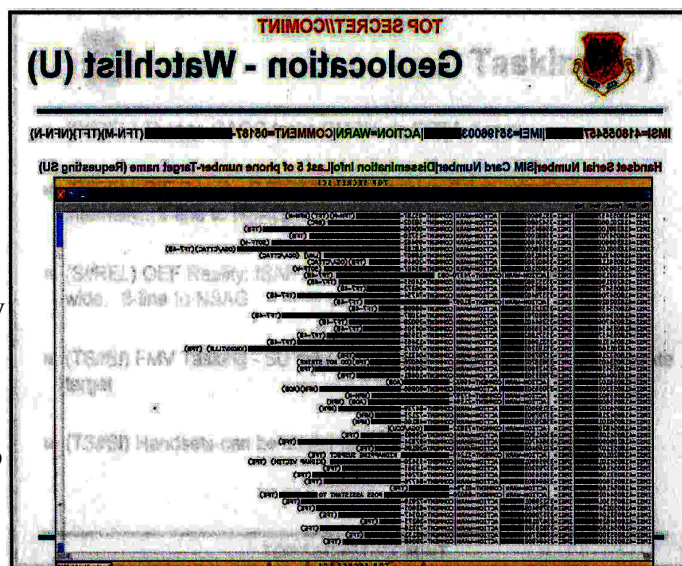
And that would be the end of it, but for a memory of a similar incident the day before. A keyword search on “drone papers” returned a [quote-tweet](#) of the same post. But when I checked the user's profile, it had vanished. Hmm, perhaps he deleted it. But his tweet was still appearing in search results, and when I clicked on the link to the tweet itself, not the user's profile, it was still there. Then I checked his profile with the tor browser. And sure enough, there it was. “That was odd,” I thought, but dismissed it as Twitter weirdness. And when it showed up in my regular browser, the next day, it was forgotten.



With this memory, I decided to check with tor browser. And there it was all along, in Appelbaum's twitter feed. I thought this was worthy of a [post](#), which prompted another retweet from Appelbaum, and several helpful responses, among them were posts from two [other users](#) (one of whom has worked with the Snowden Archive and been [published](#) in *Der Spiegel*) confirming that they too were experiencing the same issue. The tweet had disappeared from view from US, in Appelbaum's feed. From these posts, and my own work, it appears the issue was not limited to only one ISP (it affected both CenturyLink and Verizon), to one browser (Firefox, Chrome, IE and Opera were tried), one operating system (it affected both Windows and Android), or one login (at least 3 different users affected, I was also unaffected whilst logged in, but still using tor). It was also not a case of tor vs non-tor (US tor exits affected, as were my regular home IP addresses). Apparent location of the user's IP address seemed to be the main factor. Germans could see the post, Americans couldn't.



And there was something else, the post contained fragments of a top secret document relating to drones, that were not published anywhere else. It happens that the “[drone papers](#)” were printed double-sided and then scanned, and the verso image shows through. In most cases, that image was published anyway more plainly. But in one case, it shows fragments of an unpublished page. The takeaway is that *The Intercept* hasn't yet published the entire cache. Which I said: “It would be unwise to assume that we've heard the last from [#dronepapers](#). There is a hint they have more to say.” I thought Top Secret documents would be high on the list of things that might be deliberately censored.



Writing on the reverse side shows through.

Here is a short analysis of the document fragment itself.

- Tasking, NSA-speak for the process of collecting information from specific sources to be forwarded to analysts.
- “OEF....: ISAF... wide... line to NSAG”
 - OEF, refers to [Operation Enduring Freedom](#)
 - ISAF, International Security Assistance Force
 - NSAG, refers to the NSA facility in Georgia
- “FMV tasking,” FMV means Full Motion Video. In other words, this refers to the collection and analysis of drone video footage. Perhaps the word “SOCOM” appears, which would refer to U.S. Special Operations Command. Additionally the phrase “...ate target” appears. Perhaps this is “eliminate target”
- “Handsets can be...”, handsets are primarily what drones target, they can be geolocated, listened to, and hit with hellfire missiles.

It became apparent that Twitter itself was the cause. Moreover, that it was likely not a bug, but a deliberate feature of the system, one that has been [documented](#).

The social media company also says, “**we have begun to test a product feature to help us identify suspected abusive Tweets and limit their reach.**” This feature takes into account a wide range of signals and context that frequently correlates with abuse including the age of the account itself, and the similarity of a Tweet to other content that our safety team has in the past independently determined to be abusive. It will not affect your ability to see content that you've explicitly sought out, such as Tweets from accounts you follow, but instead is designed to help us limit the potential harm of abusive content. This feature does not take into account whether the content posted or followed by a user is controversial or unpopular”.

If this tool is in use, it has several interesting capabilities. To suppress posts from individual feeds, for certain viewers, depending on where they are, and for a specific time. Most views and retweets usually occur within 24 hours. If a post can be hidden for a time, its reach will be severely limited. And if the original post is a retweet, the person controlling the feed is unlikely to notice (especially if it shows up in his own view). If notifications that someone has retweeted are suppressed, the poster himself is less

likely to notice. On the odd chance someone notices, and goes back the next day to check, the post will be right where it was supposed to be, and that person says “oh well, guess it was nothing,” and moves on. But worse than that, even if a person is dead sure, the effects soon become non-reproducible, which reduces the chances that the person can prove it. Censorship that doesn't look like censorship. It deliberately reduces the spread of information that might otherwise go viral. Vicious. “You can say what you like but no one will hear you. And also, you'll think no one cared, so you'll give up trying.” Subtle, deniable, and quite ruthless.

Who controls this software? Does Twitter itself set all the rules? Do governments set policies for Twitter to administer? Or can they drive it directly? Is it all automatic, censorship by algorithm? Some mutant hybrid? (The tool's existence is not in question, as Twitter has acknowledged it, the only question is whether it was being used here, and if so, what it implies for the capabilities and policies governing the tool.)

This time, I was lucky enough to notice, to have a post that credibly would be censored, and have my claims investigated and confirmed by others within the window. It is also a case where the censorship is not legitimate. Twitter definitely has some explaining to do.

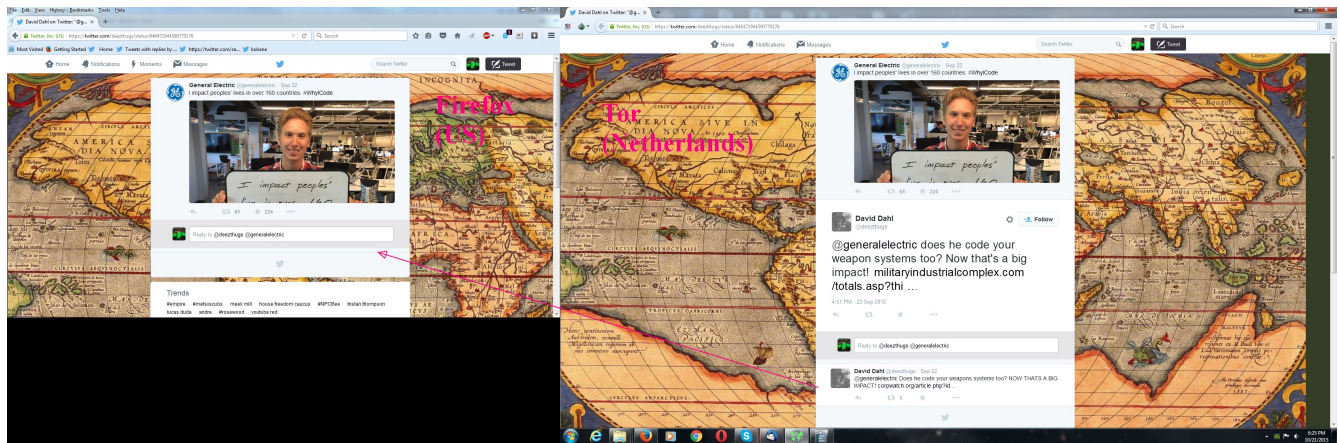
Update (Oct 21, 2015): Twitter has issued a terse [reply](#): “Earlier this week, an issue caused some Tweets to be delivered inconsistently across browsers and geographies. We've since resolved the issue though affected Tweets may take additional time to correct.”

I have reason to doubt the timeline provided by Twitter. One [post](#) that was brought to my attention was a criticism of GE, by David Dahl that disappeared September 22, and remained hidden when I went to verify it on October 18. Secondly, a tweet criticizing

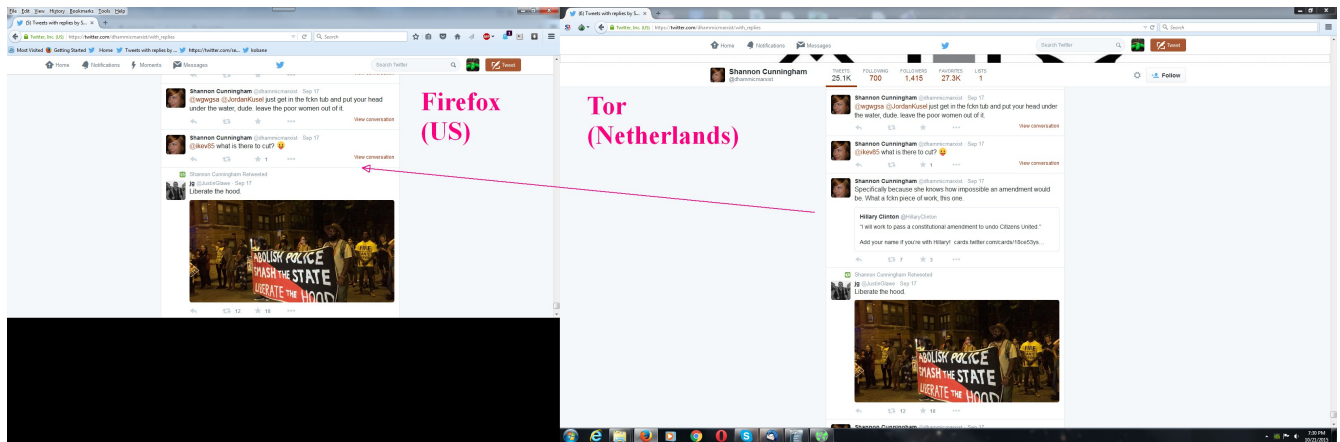
Hillary Clinton vanished September 17, it likewise still remains hidden from US view. These tweets didn't just disappear last week, both parties noticed at the time. Shannon Cunningham [wrote](#): “What the fuck happened to my tweet criticizing [@HillaryClinton](#)?” and David Dahl [wrote](#) in reply to the user who had [pointed out](#) the disappearance: “[@helveticade](#) It is apparently gone. WTF. [@twitter](#)”



Screncap of tweet criticizing GE sent September 22



Updated screencap of Dahl's Tweets criticizing GE, taken October 21

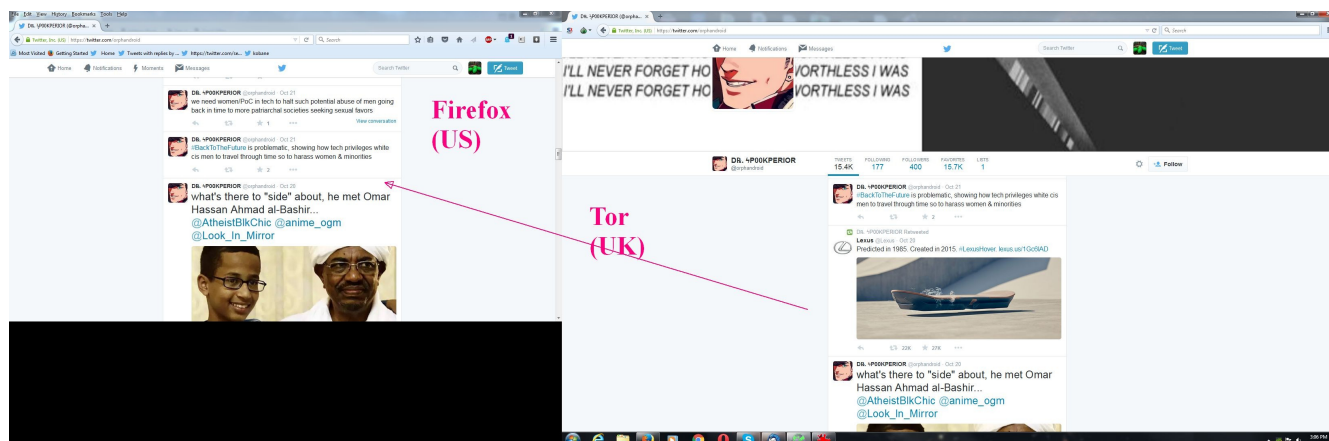


Screencap of tweet Cunningham's tweet criticizing Hillary Clinton, taken October 21

Twitter's reply is still not factually correct, and lacks a genuine explanation of the cause of this issue. I am still interested in perusing this further. If anyone can provide additional data, I would encourage them to contact me. I am particularly curious about posts dated after October 21 (which would disprove twitter fixed the issue), or before September 17 (which would bring the start of the issue even further back), especially if they have political overtones of any kind.

Update 2 (Oct 23, 2015)

I found a new one yesterday. This one is interesting: Lexus' tweet about its hoverboard gone missing from a certain [timeline](#). Which doesn't sound like much, but it is. When you read the rest of the fellow's timeline, it looks pretty much like he's a misogynist, neo-Nazi, and all-around Bad Person (or plays one on Twitter). It's as if Lexus might not to be associated with people like him, and who could blame them?



Screncap taken October 22

I found him not quite by accident. He is tangentially connected to a Breitbart [article](#) about censorship. (click on the link that reads "[non-abusive messages](#)" and he is among the [replies](#).) I thought his handle "DR. 4P00KPERIOR" sort of looked like the SS logo, and his timeline would be a good place to hunt.

Then I've made another connection: why had Appelbaum's innocuous tweet about a Chromebook gone missing? Perhaps because he's Jacob Appelbaum: Assange friend, Snowdenista, activist, journalist, tor developer and all-around troublemaker, and Google might feel embarrassed to be associated. Only a censor-bot could be so dumb, and a censor-bot is what we're dealing with.

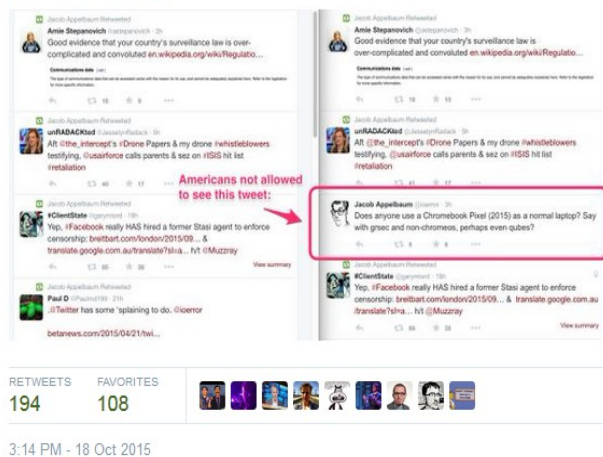
In other words: advertising interest. It's of a piece with other missing tweets: Dahl's criticism of GE, and Cunningham's criticism of Hillary Clinton. In Clinton's case, it's less politics and more money. Clinton, like other candidates, pays Twitter to promote their brand. For them, she's just another advertiser.

This was what Yancy Faith mentioned too -- criticism of advertisers sorta disappears. The technology to "limit the reach" of abuse instead serves the bottom line: providing revenue for Twitter by making advertisers happy.

That being the case, we can look for the effects dating back to April or May. I'll keep digging.



Fun game: look at [@ioerror](#)'s timeline through [@torproject](#) and compare to what you see. 😊



YancyFaith @YancyFaith · Oct 18

We noticed similar phenomena re: tweets criticizing a major ad buyer, others criticizing an activist [#Jack](#) supports

Cryptome @Cryptomeorg

Adventures in Twitter Censorship
cryptome.org/2015/10/Advent...
w/ @Paulmd199