

Do We Care About Surveillance?

Edward Snowden's Impact and Policy Implications



Explorations in Cyber
International Relations
Massachusetts Institute of Technology Harvard University

Workshop on
Cyber Security & the Governance Gap
MIT, January 6 and 7, 2014



Evan Marshall, 4th Year Undergraduate and M'Eng Candidate

Start: September 2013

Researched for Cyber Politics in International Relations (17.447)

Introduction

Public outcry was especially strong in June of 2013, after the Guardian and the Washington Post simultaneously released materials leaked by NSA contractor Edward Snowden that describe unprecedented mass surveillance programs. A day after the release, President Barack Obama addressed the issue during a visit to Silicon Valley, saying, "it's important to recognize that you can't have 100% security and also then have 100% privacy and zero inconvenience. You know, we're going to have to make some choices as a society."

In the months since, the American people have responded in various ways. Some have taken to social media sites to discuss the Snowden revelations. Others have gone a step further, not only discussing the topics of the Snowden leaks, but also the underlying concepts of network security, encryption, metadata. Using quantitative social media and internet traffic data, my research identifies three stages of this online response: Discovery and Dissemination, Information Aggregation, and Behavioral Shifts. Knowing how these cohorts behave is a powerful step to making the policy choices about privacy and security.

Methods

With the goal of identifying quantitative effects in social networks and various privacy protection related applications, I decided to collect numerical data to quantify social media responses and the traffic data associated with the appropriate applications. Analyzing social media reveals how information spreads and collects, such as informing users about the Edward Snowden leaks, or identifying the spread of related privacy applications and tools like Tor. I then used the traffic data to represent the individual behavior changes as individuals learning about and potentially adopt new tools to protect their privacy. I ended up collecting ~15,000 tweets and ~50,000 Reddit posts as well as yearly traffic data for 12 different security and privacy related software.

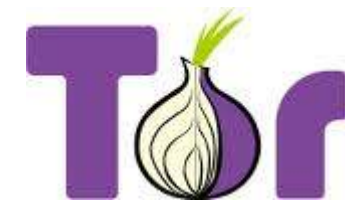
Data Sources



Twitter is a micro-blogging platform that is used to share short messages (140 characters or less) across potentially vast networks of users.



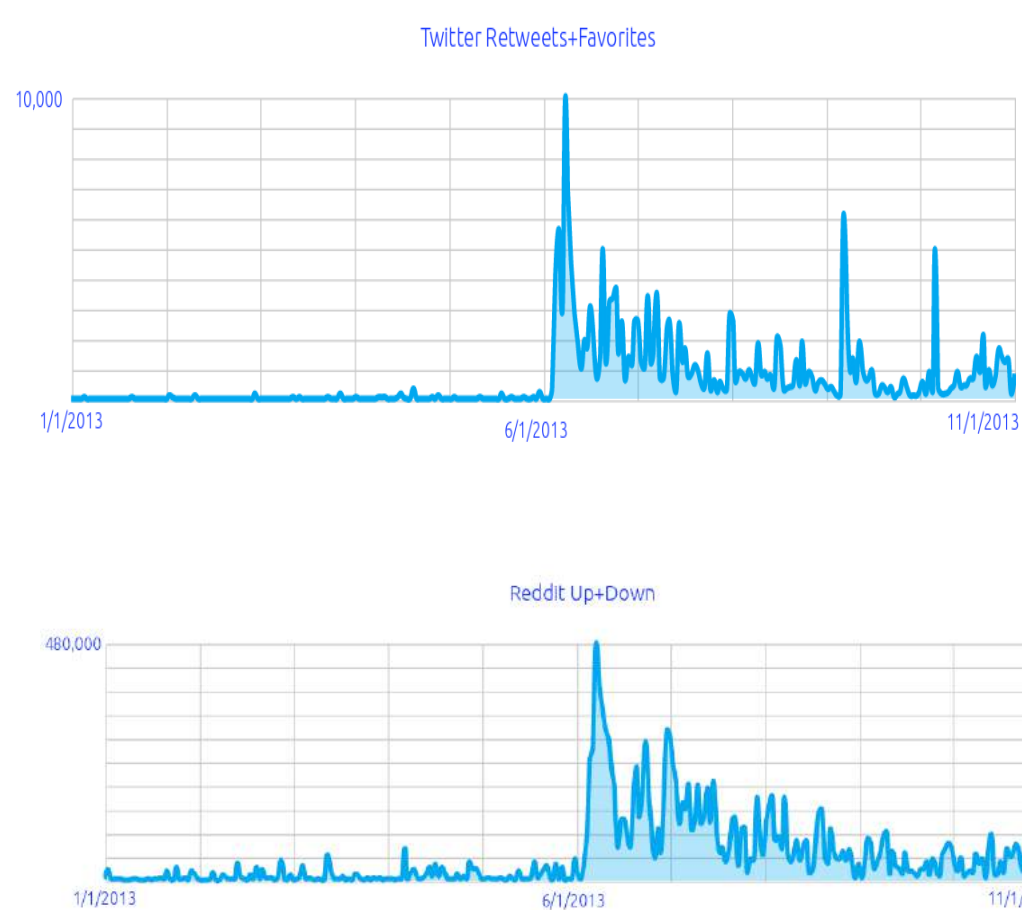
Reddit is a social news and entertainment website where users submit either link or text based posts. About 6% of Americans use Reddit



TOR utilizing onion routing to provide anonymous internet access to its users. Whereas some tools only secure data, TOR secures full anonymity.

The Stages of the Online Response

Stage 1: Discovery and Dissemination

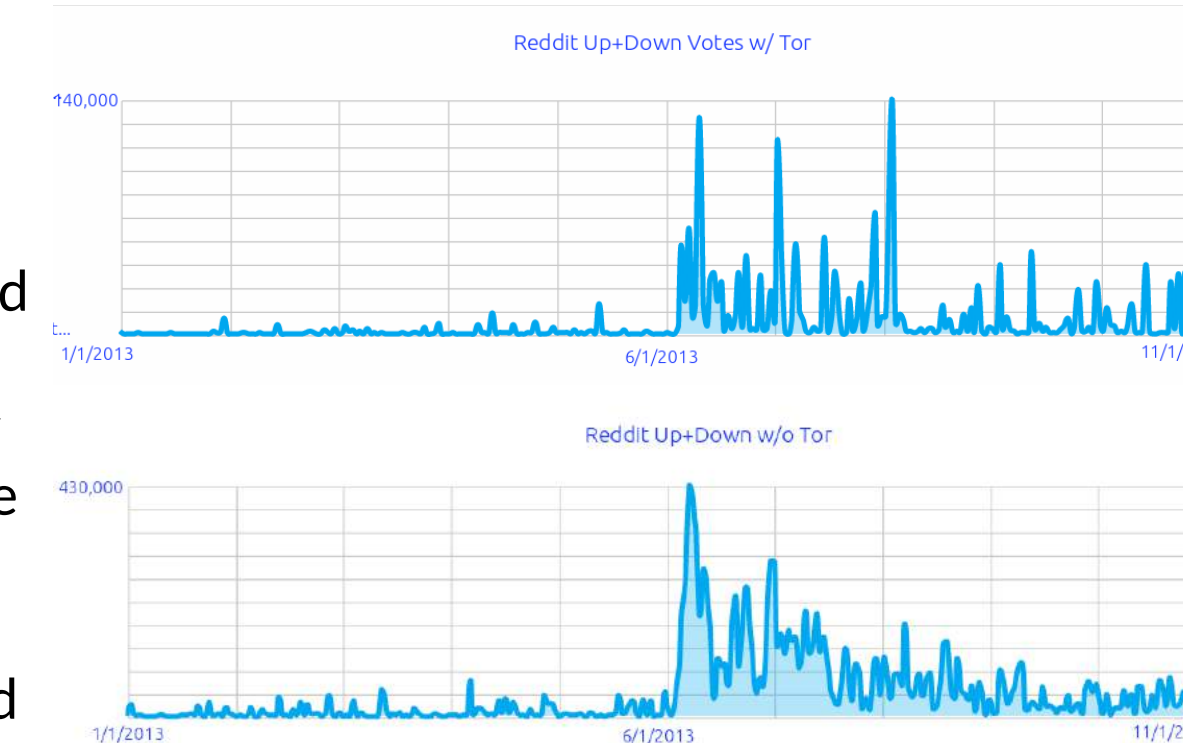


The first stage is discovery and simple discussion of these topics. For example, a person may read an article about NSA surveillance and write a tweet saying, "Edward #Snowden on the run in Hong Kong." This kind of activity indicates the reach and impact of a news headline, press release, or other catalyst for discussion.

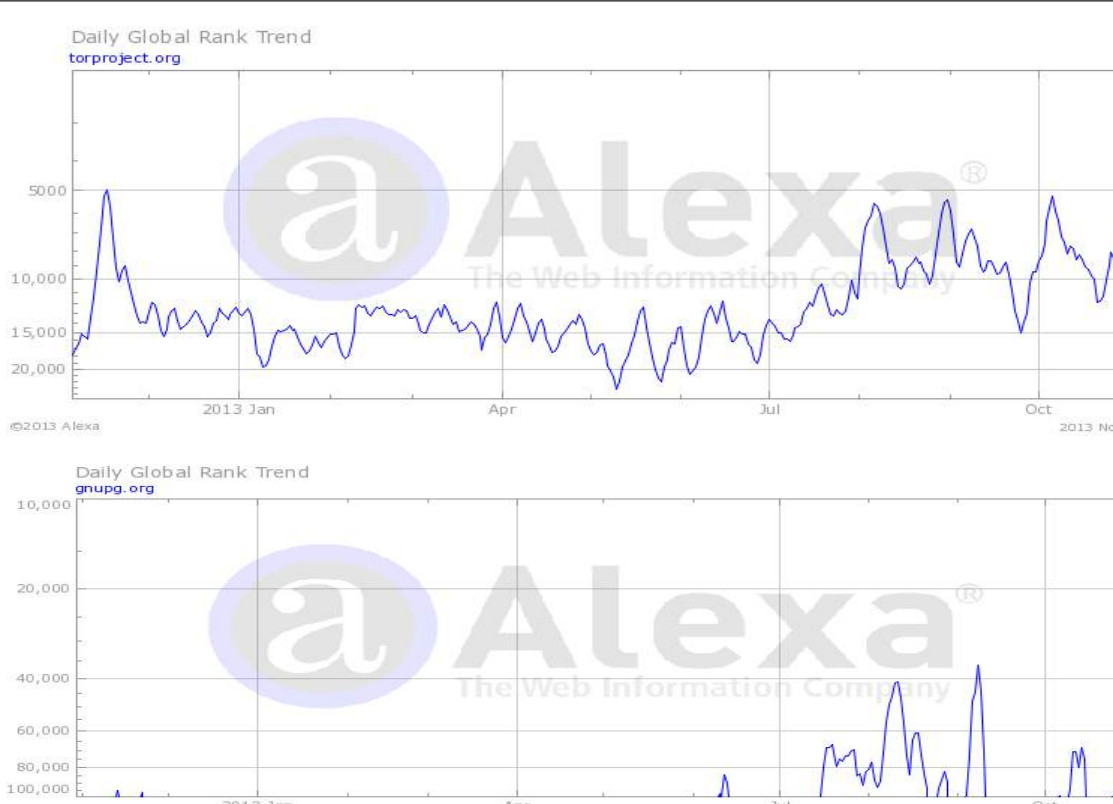
There is an extremely significant double-spike in Twitter activity across all keywords in the first few days of June 2013. Twitter activity carries on with a decay that appears bi-exponential in form, generally attenuating through June and July, during which time Snowden traveled to Hong Kong and Moscow, and began filing asylum applications to a number of countries. The Reddit data appears to closely correlate with the Twitter response.

Stage 2: Information Aggregation

The second stage is aggregation of information and discussion about the underlying concepts, indicating more than a passing interest in a news headline. In the following two graphs, we separate Tor from the other keywords. The first graph shows Reddit activity for the Tor keyword, and the second graph shows Reddit activity for all keywords except Tor. It is clear that discussion for Tor saw a sustained increase in activity after the Snowden release. This suggests that some people were curious enough about encryption to go beyond the content of news headlines, searching online and asking questions about encryption mechanisms. Examining the popular posts confirmed the aggregation of a body of information about Tor on Reddit.



Stage 3: Behavioral Shifts



The third stage is action based on the goal of securing data and anonymity. For example, a person may query Google with the search terms "how to use Tor to be anonymous online," and navigate from there to the project website where Tor is available for free download. This hit to the Tor project website, which we can observe via Alexa, reflects something deeper than discussion and aggregation of information. It suggests that the user may want to install Tor, or at the very least learn how it works. The graphs to the left reflect the traffic ranking of two sites: Torproject.org and gnupg.org. Both experience unusual traffic spikes starting soon after the Snowden leaks. 10 other sites with related tools experience similar traffic spikes.

Discussion

If the increase in encryption site traffic from July onward is due in part to an increase in awareness and curiosity resulting from the Snowden leaks, it means two things. First, it means that the third stage of the three-stage model, in which internet users take action to improve the privacy of their data, is strong. Second, it means that there is a lead-lag effect of about one month between the dissemination of information (first stage in our model) about NSA programs, and the actions that internet users took to realign their use habits in an equilibrium where they increasingly protect themselves from government intrusion by using encryption. This indicates that there is a time in which internet users researched and learned about encryption programs like Tor. Indeed, the dataset shows that the aggregation of information about Tor on Reddit was highest in June and July of 2013. The increase in pagerank for Tor and other encryption sites began in the middle of that time period.

Further Analysis and Next Steps

In addition to the research I have done, many related inquiries may be addressed with a few follow up studies. Three such issues include: the specific types of events for which these steps and variants occur, the quantification of results, and the ability to predict events in real time. By answering some of the related questions, real policy and other impacts may be realized. The subject remains young in its study and requires further inquiry.

By determining the existence of separate phases in online response to the Edward Snowden leaks, it becomes important to understand the types of events for which these phases occur and any other variants. Many specific questions may rise independently such as did Bradley Manning provoke a similar internet effect? If it becomes realized that many internet users adopt or leave online applications in a similar process, the value of quantification rises significantly.

After understanding the various subject matters that we can study through this 3-stage model, it quickly becomes important to quantify the effects. Comprehending the numerical implications of real world events may help researchers develop political and legal models for understanding the interface between social media and the real world. The key to these models relies in large data collection as well as statistical and computational techniques. Utilizing big data algorithms may assist future study with greater depth and accuracy.

Thank You!

I give thanks to Professor Choucri who provided guidance and feedback that substantially helped me progress through this research project.

Also, I extend my deepest appreciation to the many members of the Internet community that have contributed their time building the foundational tools which enable this research and future research.