Lecture 23: Network surveillance and censorship

Anirudh Sivaraman

2019/04/08

In today's lecture, we'll discuss two threats to the Internet that have emerged over the last few years: network surveillance and network censorship. Network surveillance is the act of monitoring a user's network activity (e.g., who the user is talking to, what the user is talking about, and when the user is talking). Network surveillance is typically done in secret so that the user being monitored does not alter their behavior because they know someone is watching them. Network censorship is the act of preventing network access to certain kinds of content (e.g., specific web sites, specific Google searches, or specific Wikipedia pages) to users.

Network surveillance is passive and doesn't involve modifications to the user's network experience.[1] On the other hand, network censorship is active, and the censor modifies the user's experience in some way. For example, when accessing specific web pages, the user is redirected to another web page that informs them that they are not permitted to access the original web page. Or the censor could drop the user's packets, causing a decrease in TCP's congestion window, and hence degraded performance.

The reason we are studying these two together is that the adversary in both cases is most commonly an entire country that wants to monitor its citizens' activity or censor their Internet access. We should note here that this lecture uses the term adversary in the computer security sense. The term adversary denotes an opponent that is seeking to monitor/censor a user's traffic despite the user's best efforts.

Notably, this lecture does not use the term adversary in a moral or ethical sense and does not take a position on the morality or legality of the issue. There might be legitimate reasons for both surveillance and censorship such as identifying national security threats or preventing access to objectionable content. On the other hand, surveillance and censorship may violate a user's civil liberties, e.g., unauthorized monitoring of a user's activity or preventing access to content on the Web that espouses idealogies that run counter to the idealogies of the ruling party.

An example of country-wide surveillance is the surveillance conducted by the United States' National Security Agency (NSA), as revealed by the Edward Snowden leaks in 2013. The Snowden leaks also revealed the participation of intelligence agencies from other nations such as Australia, Germany, and the United Kingdom in these surveillance programs. An example of country-wide censorship is the Great Firewall of China (GFW), which blocks access to many different web sites from China. Similar censorship programs are operated by several countries in the middle east.

When an entire country is the adversary, network security properties are much harder to achieve. In this case, the relevant security properties are the ability to circumvent censorship and the ability to access the Internet unmonitored. In fact, there is very little anyone can do in this situation because the adversary (the country) is much more *resource-rich* (in terms of money, computation power, legal provisions such as court subpoenas, etc.) relative to the average user of the Internet in any of these countries. In the reminder of this lecture, we'll go over a few ways in which access to almost infinite resources allows such an adversary to monitor or censor as much as they wish.

1 Network surveillance

Because network surveillance is done passively and in secret, a good deal of recent information pertaining to network surveillance is based on indirect inferences from publicly available documents from the Snowden leaks.

[1] Indeed, in many cases, the goal is for the user to be unaware that surveillance is even happening.

Network surveillance at the scale of an entire country can make many of the security vulnerabilities that we briefly discussed last lecture much more pronounced.

For instance, secret court orders allow the government to retrieve the phone records of all customers of a phone company [3]. A similar mechanism, called the PRISM surveillance program [3] allows the NSA to collect user-specific information from Internet companies such as Google. These are both examples of a resource-rich adversary whose resources include the extremely uncommon ability to issue court orders that compel a company to reveal private customer data in the interest of national security.

If data is unencrypted (e.g., it does not use HTTPS/TLS), the job of the adversary becomes much easier. This was the case for data that was being transmitted unencrypted within Google's datacenters [11]. There are also reports of tampering with routers to enable surveillance of traffic going through these routers [4]. If the adversary can't get access to the router, the adversary could still tap into the physical cables carrying packets. This isn't science fiction: the Government Communications Headquarters, a British spy agency, is known to tap into fibre optic cables all over the world. Further, through a program codenamed Tempora, this British data is also shared with the NSA [3]. This tap probably costs a fortune, but a country's surveillance program has a gigantic budget, e.g., the Intelligence budget for the U.S.A is about 50B USD and about 10B of this is for the NSA [13].

Encrypting traffic does help, but has its limitations. First of all, TLS only encrypts the TCP payload; the TCP port, the IP addresses and the MAC layer addresses are all still in clear text and can be snooped by a modestly determined adversary. In many cases, information on who is communicating with whom is as valuable as what they are talking about. TLS also relies on a chain of trust where the root certificate authorities are expected to be trusted entities. If one of the certificate authorities along this chain is compromised (e.g., DigiNotar [2]), it could issue fake certificates. It is possible that the DigiNotar compromise was exploited by the NSA [8]; however, it is also possible that it was exploited by a different intelligence agency [9].

Another reasonable assumption that TLS makes is that the private keys are truly private. If a server gets hacked into, it's quite easy to steal its private keys. The Snowden leaks revealed that the NSA has a specific program called Tailored Access Operations to infiltrate servers [14]. Finally, the cryptography underlying TLS is only as good as long as the mathematical problem (e.g., prime factorization) truly remains hard in a practical sense—despite advances in computing technology and access to large computing clusters. The Logjam attack demonstrated what is called a *brute force attack*, i.e., that it was possible to solve some of these mathematical problems at the core of modern cryptography through brute force with a budget of a few million dollars, well within the capabilities of an organization such as the NSA [16].

All these issues are symptoms of the resource-rich nature of the adversary where the adversary has far more resources than an average user: (1) the ability to compel companies into handing over user data, (2) the ability to solve hard mathematical problems through brute force on a large array of computers, (3) the ability to spend money on physical tap infrastructure, (4) the ability to spend money on hacking into machines,[2] and so on.

2 Network censorship

Network censorship, by contrast to network surveillance, is more active. The adversary, again usually an entire country, actively modifies the user's experience by redirecting the user to a different web page, dropping the user's packets, delaying them, or sending them to a blackhole. Network censorship can span all layers of the Internet stack.

Censorship can start at the human, organizational, and government layers, colloquially referred to as layers 8, 9, and 10 [6]. Unlike the 5 layers of the Internet we have discussed in this course, these layers aren't part of any standard and are just a reference to concerns that logically rest on top of the application layer.[3]

At these higher-than-application layers, the Government can compel a company to not display certain kinds of content that are deemed objectionable or inappopriate. For instance, web searches for the word "Tiananmen"

[2]For instance, the NSA is believed to have exploited security vulnerabilities without publicly revealing these vulnerabilities if it is convinced that no one except the NSA is aware of them [10].

[3]These layers start at 8 and not 6 because there was a competing layered architecture called the OSI architecture that had 7 layers unlike the Internet's 5-layer stack.

(a reference to the violent military crackdown on the Tiananmen Square protestors in 1989 [15]) on the Chinese web site Weibo return a result that says "According to the relevant laws and regulations, search results for [this phrase] cannot be displayed." [5]. Similarly, a search for "June 4th" (the date of the 1989 protests) returns filtered search results containing posts about birthdays and wedding anniversaries [5].

At the application layer, a country-wide censorship apparatus can respond to DNS queries for a domain name with an IP address that does not correspond to a server that hosts content for that domain name. Such censorship is used to prevent access to sites such as youtube.com, google.com [18], or github.com (which might store archived copies of sites within GitHub repositories). Censorship can also be accomplished by inspecting the search query in web searches, which is visible as part of the HTTP headers if packets are not encrypted [1].

At the transport layer, specific ports could be blocked. For instance, SSH traffic that runs on port 22 can be used to connect to a remote host and circumvent censorship under the guise of remote terminal access. In an SSH-based circumvention strategy, the SSH connection between a user and a remote host is used as a means to carry both (1) HTTP requests that the remote host forwards to the web server from the user and (2) HTTP responses that the remote host forwards from the web server to the user. Because of SSH's role in censorship circumvention, SSH traffic is itself blocked in parts of China [7].

At the routing/network layer, the Internet's routing protocol, BGP, can be used to advertise a blackhole route for particular IP address ranges that simply drops packets destined to these IP address ranges. This was the technique used by Pakistan Telecom to prevent access to YouTube in 2008 [12]. Specific IP addresses can also be blocked. For instance, one censorship circumvention technique is to use the Tor network, which relays an Internet connection over several intermediate relays for anonymity. However, in initial versions of Tor, the list of IP addresses for these relays was publicly known, allowing censors to easily block all IP address corresponding to Tor relay nodes [17].

Encryption might circumvent censorship that looks at the HTTP headers, but it is only effective if the censor permits encrypted traffic to flow through it. For instance, TLS traffic can be easily detected because it typically uses TCP port 443.[4].

3 Closing thoughts

As is hopefully clear, security is an arms race. In the case of surveillance, the adversary develops increasingly powerful methods over time to snoop on traffic (e.g., brute force attacks), while the users try and develop increasingly powerful methods to not be snooped on (e.g., using more powerful encryption). Similarly, as users develop techniques to circumvent censorship, the censors develop methods to block these circumvention techniques. It's not very different from playing a high-stakes game of Whac-A-Mole :)

Network surveillance and censorship are big threats to the open and free culture of the original Internet where anyone could access any content privately. That said, there are legitimate reasons for surveillance and censorship as well, and there is an important policy question here as to what should and should not be allowed. The Internet is no longer just about technology. Today, these policy questions are just as important to the future of the Internet as technical questions about how to make the Internet faster or more reliable.

4 Further reading

If you're interested in learning more about surveillance, Jennifer Granick's book is an excellent start: `https://www.amazon.com/American-Spies-Modern-Surveillance-Should/dp/1107501857`.

[4] In practice, dealing with encrypted traffic is a bit more complicated and can't be as blunt as blocking all traffic on port 443. The censorship apparatus might want to allow some uses of TLS for companies within a country to communicate securely with their offices in other countries. At the same time, it may want to disallow circumventing uses of TLS. To handle these use cases, censorship engines use increasingly clever methods to detect and block circumventing TLS traffic, while allowing other non-circumventing TLS traffic [17]

References

[1] Censorship of Google Searches in China — GreatFire Analyzer. https://en.greatfire.org/search/google-searches.

[2] DigiNotar - Wikipedia. https://en.wikipedia.org/wiki/DigiNotar.

[3] Edward Snowden: The 10 Most Important Revelations from his Leaks. http://mashable.com/2014/06/05/edward-snowden-revelations/#ty98pJTNQPqx.

[4] Glenn Greenwald: how the NSA tampers with US-made internet routers. https://www.theguardian.com/books/2014/may/12/glenn-greenwald-nsa-tampers-us-internet-routers-snowden.

[5] How China has censored words relating to the Tiananmen Square anniversary — Public Radio International. https://www.pri.org/stories/2016-06-03/how-china-has-censored-words-relating-tiananmen-square-anniversary.

[6] Layer 8 - Wikipedia. https://en.wikipedia.org/wiki/Layer_8.

[7] My Experience With the Great Firewall of China. http://blog.zorinaq.com/my-experience-with-the-great-firewall-of-china/.

[8] New NSA Leak Shows MITM Attacks Against Major Internet Services - Schneier on Security. https://www.schneier.com/blog/archives/2013/09/new_nsa_leak_sh.html.

[9] No, the NSA was not behind the DigiNotar hack - Koen Rouwhorst. https://koen.io/2013/09/14/no-the-nsa-was-not-behind-the-diginotar-hack/.

[10] NOBUS - Wikipedia. https://en.wikipedia.org/wiki/NOBUS.

[11] NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say. https://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html?noredirect=on&utm_term=.4ccc00857a30.

[12] Pakistan Cuts Access to YouTube Worldwide - The New York Times. http://www.nytimes.com/2008/02/26/technology/26tube.html.

[13] The Black Budget. http://www.washingtonpost.com/wp-srv/special/national/black-budget/.

[14] The NSA Uses Powerful Toolbox in Efffort to Spy on Global Networks - SPIEGEL ONLINE. http://www.spiegel.de/international/world/the-nsa-uses-powerful-toolbox-in-effort-to-spy-on-global-networks-a-940969.html.

[15] Tiananmen Square protests of 1989. https://en.wikipedia.org/wiki/Tiananmen_Square_protests_of_1989.

[16] David Adrian, Karthikeyan Bhargavan, Zakir Durumeric, Pierrick Gaudry, Matthew Green, J. Alex Halderman, Nadia Heninger, Drew Springall, Emmanuel Thomé, Luke Valenta, Benjamin VanderSloot, Eric Wustrow, Santiago Zanella-Béguelin, and Paul Zimmermann. Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice. In *CCS*, 2015.

[17] Roya Ensafi, David Fifield, Philipp Winter, Nick Feamster, Nicholas Weaver, and Vern Paxson. Examining How the Great Firewall Discovers Hidden Circumvention Servers. In *IMC*, 2015.

[18] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. Global Measurement of DNS Manipulation. In *USENIX Security*, 2017.