Are You Watching TV or Is TV Watching You?

Data Collection and Privacy Issues in Modern Smart TVs

Jacob Ehrlich
COMP116
Tufts University
December 12th, 2018

I. Abstract

In a competitive market, media companies vie for success by providing the most personalized experience to their users and the most profitable model for the company. In order to provide this authentic user experience and sell targeted advertisements, these companies need user data—and lots of it. By understanding these services' motivations, we can start to better understand why many methods of data collection harvest your private information with little warning or notice. This paper will explore not only the apparent necessity of data collection, but also its legality by examining case studies such as the Vizio privacy lawsuit. In order to better understand the vulnerabilities propagated by data collection, we will examine and analyze the Amazon FireStick, an internet connected device you can connect to your TV. How much data does one of these Smart TVs collect? What kind of data is it? In addition to our study of data, we will evaluate the urgency of security risks that result from the collection process by utilizing networking tools such as arp-scan, nmap, and adb. We will begin our study of the Smart TV by examining its place in the larger context of "internet of things" (IoT) devices. Can we logically group Smart TVs like the FireStick with other IoT devices? How can the FireStick possibly be different?

II. Introduction

As media consumption continues to shift from network television and live programming to online streaming, so does the way we consume it. Subscription based streaming services such as Netflix, Hulu, and Amazon, commonly used via personal laptops or home desktops have begun integrating their streaming platforms into home televisions known as 'Smart TVs.' These TVs may seem like a great new way to access the programming we love, yet their integration comes at a cost. Many of these systems take advantage an internet connection by collecting a constant stream of user data behind the scenes. While Smart TV companies generally collect this data through legal means, these same companies increasingly blur the line between full and partial disclosure of privacy terms, which leaves users unaware of both the type of data transmitted and the insecurity of many of those transmission protocols. As this data collection increases, so does the risk of private data leakage; however, data collection allows companies for direct feedback from their users and thus potentially the most rapid route to technological progress. Should we even care about these insecurities if the data we provide gives us more targeted programming recommendations and a better user experience? How bad could they really be?

II. To The Community

I cannot fairly delve into the many issues data collection presents before I have made clear its many benefits. As a filmmaker and a film student at Tufts, I am deeply fascinated with media production and content creation. What better way to keep the media production industry booming than to make the stuff your audience will pay you to see? And what better way to know what an audience wants to see than monitoring and harvesting their viewing data? With my background in basic computer security, I believe I can do this topic justice by fairly considering both the great benefits and atrocious vulnerabilities of data collection.

Apart from my interest and biases, why is this topic important in general? The Internet of Things (IoT) refers to a group of objects, sensors and smart devices that have a network connection and thus the capability of intercommunication without intervention.[1] A Denial of Service (DoS) attack occurs when a botnet—a group of infected devices, brings down a server with an overflow of network packets. Botnets are commonly made up of IoT devices because smart dishwashers, refrigerators, and light switches usually lack proper authentication protocols and can be easily bypassed by an attacker to amplify a signal.[2] Problem one—lack of proper credentials. As an IoT device, Smart TVs like the Amazon FireStick suffer from this vulnerability, but what makes Smart TVs different and worth our examination? In addition to easy access, Smart TVs open themselves up to man-in-the-middle attacks as they collect and send data, unencrypted, over a network. With multiple points of entry, Smart TVs may be the ultimate exploitable IoT device.

III. Legality of Data Collection

Before assessing the process of data collection, we must first answer the question—is collecting data even legal? On October 4th, 2018 after years of court debate, lawyers settled a lawsuit against Smart TV company Vizio, which cost the company 17 billion dollars for hidden data collection.[3] Vizio's Smart TVs analyze watching habits using software that first determines the programming being watched then sends data like IP address and MAC address across a network back to their servers.[4] They then sell this information to advertisement companies. Vizio claims that they only allow third parties to view "non-personal identifiable" information; however, this information is far from non-personal—sex, age, income, marital status and education.[5] Vizio sends an IP associated with this data to ad companies with no promise of encryption. These Smart TVs do a great deal of data juggling, sending it from one place to another, increasing the risk of a man-in-the-middle attack, potentially exposing private information. Interestingly, the lawsuit has no mention of encryption concerns.

While the network vulnerabilities of data collection fall to the background, lawyers obsess over Vizio's deception of their customer base. Vizio marketed their data collection as a *feature* of their product, not an *option*. Users were unwillingly forced into this agreement where Vizio sold their data to third parties over insecure network protocols. So perhaps data collection is legal as long as users know about it. I am afraid, however, that the methods of acquiring and sending this data were completely overlooked as a result of this case. There were no clauses forcing Vizio to change *how* they collect data, only the warning messages they must issue to

---

[1]Conti, Mauro, et al. "Internet of Things Security and Forensics: Challenges and Opportunities." *Future Generation Computer Systems*, vol. 78, no. 2, Jan. 2018, pp. 544–546.

[2]Lu, Yang, and Li Da Xu. "Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics." *IEEE*, 2018, doi: 10.1109.

[3]Gardner, Eriq. "Spying TVs: Legal Settlement Provides a Few Bucks for Vizio Owners; Millions for Lawyers." *The Hollywood Reporter*, The Hollywood Reporter, 5 Dec. 2018, www.hollywoodreporter.com/thr-esq/spying-tvs-legal-settlement-provides-a-few-bucks-vizio-owners-millions-lawyers-1149262.

[4]Ibid.

[5]Angwin, Julia. "Own a Vizio Smart TV? It's Watching You." *ProPublica*, ProPublica, 9 Nov. 2015, www.propublica.org/article/own-a-vizio-smart-tv-its-watching-you.

their users in the future. So how exactly is this data collected and sent to third parties? What does it mean to send something unencrypted?

IV. How Your Data Gets Around

Via an analysis of the following Amazon FireStick network traffic, we will illuminate both the types of data being sent and their alarming vulnerabilities. A Pi-Hole, known as "a black hole for advertisements," intercepts network traffic associated with a given IP address and stores the data in easy to read tables.[6] The following data was acquired via a Pi-Hole and an Amazon FireStick over the course of 24 hours (Fig 1.1/2).[7] How can we begin to make meaning from this table? What is a query? When two computers want to communicate via a network, they use the smallest unit of communication known as a network packet—think a sentence for human interaction. Computer A will send a request to establish a connection to computer B, computer B will respond by sending back an acknowledgement packet, computer A will respond with one last acknowledgement packet and a connection is established. This process is known as the TCP/IP 3-way handshake. Using this vocabulary, we can understand the Pi-Hole data more clearly—the nearly 6,000 queries are really 6,000 requests to establish a connection with another computer. So should we be alarmed? Well not entirely. Many queries are harmless and well protected by secure TCP protocols like HTTPS—this includes queries for fetching the program currently watched or a search on YouTube. What we are interested in are the queries blocked by the Pi-Hole—the queries to advertisement services. Are these queries, a.k.a. the data collected behind the scenes, sent using secure protocols?

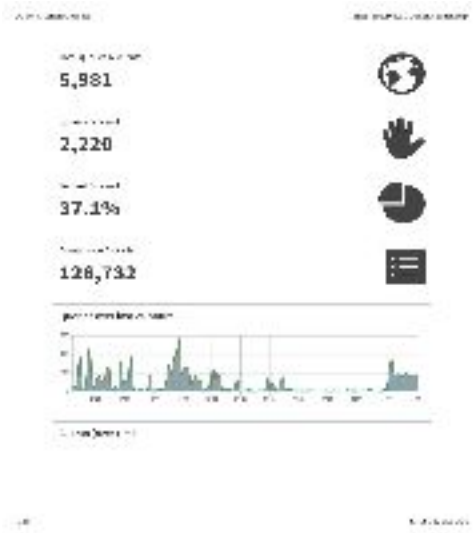

Figure 1.1

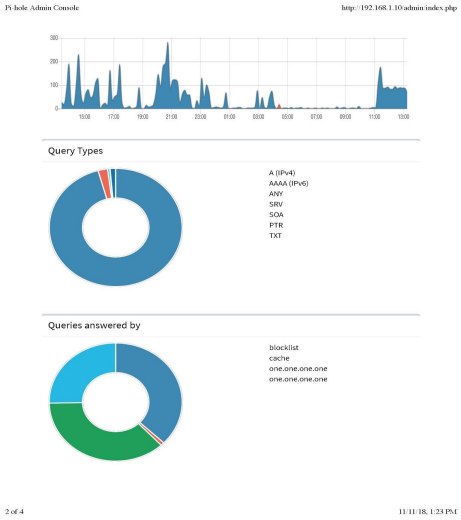

Figure 1.2

[6]"Pi-Hole®: A Black Hole for Internet Advertisements." *Pi-Hole®: A Black Hole for Internet Advertisements*, 2018, pi-hole.net/.

[7]Chow, Ming. Figure 1.1, 1.2. "PiHole Admin Console Data" 11 Nov. 2018.

Top Blocked Domains

Domain	Hits	Frequency
cws.conviva.com (queries.php?domain=cws.conviva.com)	368	
api.mixpanel.com (queries.php?domain=api.mixpanel.com)	260	
nbcu.demdex.net (queries.php?domain=nbcu.demdex.net)	127	
nbcume.sc.omtrdc.net (queries.php?domain=nbcume.sc.omtrdc.net)	124	
reports.crashlytics.com (queries.php?domain=reports.crashlytics.com)	99	

Figure 2.1

With nearly 40% of the total queries blocked, we would hope that if our devices are sending sensitive collected data like in the Vizio case, it is at least sent using secure protocols; however, upon closer inspection, the results are rather bleak. Above are the top blocked domains for the same Amazon FireStick data previously observed (Fig. 2.1).[8] Each domain is merely a receiver of a request. A webscan of the top domain, cws.conviva.com, reveals misconfigurations and weaknesses with the protocols used to send data to Amazon.com. The overall grade by htbridge.com was an F due to the multitude of HTTP packets sent with vulnerabilities such as version disclosure, cross-site-scripting, and lack of content security policy (CSP) (Fig 2.2).[9] When two computers communicate via the TCP/IP 3-way handshake, sensitive data should always be transferred with encryption via HTTPS or FTPS (secure protocols). HTTP, on the other hand, sends data 'in the clear,' unencrypted for an attacker to harness with basic networking tools. Furthermore, the domain does not implement CSP, which attempts to mitigate attacks across networks.[10] Because many of the blocked queries (requests for collected data) use HTTP (and a misconfigured HTTP header that discloses version at that), your data sent via a Smart TV like the Amazon FireStick does not protect your viewing data with encryption.

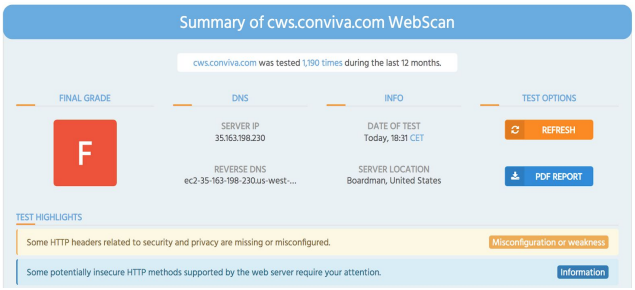

Figure 2.2

So the Amazon FireStick does not guarantee to protect collected data with network security features, but what about other Smart TVs? Unfortunately data collection poses a security issue in many other cases. At the start of Smart TV production, large companies LG and Samsung had glaring security issues data sending. LG sent both user viewing data and user uploads or downloads completely unencrypted over a network.[11] Similarly, security editor Dan Goodin at Ars Technica was able to remotely connect to early Samsung Smart TVs without any

[8] Chow, Ming. Figure 2.1. "PiHole Admin Console Data" 11 Nov. 2018.
[9] "ImmuniWeb® WebScan | High-Tech Bridge." *High-Tech Bridge - Web and Mobile Application Security*, 10 Dec. 2018, 18:31, www.htbridge.com/websec/. Figure 2.2.
[10] "Content Security Policy (CSP)." *MDN Web Docs*, 24 Aug. 2018, developer.mozilla.org/en-US/docs/Web/HTTP/CSP.
[11] Goodin, Dan, et al. "LG Smart TV Snooping Extends to Home Networks, Second Blogger Says." *Ars Technica*, Ars Technica, 21 Nov. 2013, arstechnica.com/information-technology/2013/11/lg-smart-tv-snooping-extends-to-home-networks-second-blogger-says/.

credentials.[12] The two security problems I mentioned at the start, lack of credentials and unencrypted network traffic, persist across these platforms. Since these cases occurred many years ago and have since been resolved on those platforms, how come recent technology like the Amazon FireStick suffers from insecure network protocols, i.e. HTTP headers containing plaintext user data? Have we actually gotten any better? As a demonstration, I wrote a python script that remotely connects to an Amazon FireStick, controlling it remotely, just like in the Samsung example years ago. Using an android developer feature known as ADB debugging, users can remotely connect to a FireStick without any credentials.

V. How to Protect Your Data

While most Smart TVs suffer from security vulnerabilities, many of them can be mitigated with targeted strategy. To disallow access to your Amazon FireStick's remote controls, do not enable ADB debugging. With ADB off, your FireStick closes port 5555 leaving no way for an attacker to connect from another device. In this same way, check requests to connect to your TV—do not blindly permit warnings from your device, read them carefully. As far as man-in-the-middle attacks go on your data, there are fewer options. If Smart TV companies send your data via HTTP, there is not much you can do. When researching, prioritize devices that use protocols like HTTPS, which uses encryption to protect your data. However, an ad blocking device, like a Pi-Hole, would avoid any insecure data leakage. Like many issues in cyber security, this mitigation is a double-edge sword: at the risk of allowing your data to be sent out in the open, you restrict data collection entirely, and thus disable useful features like 'recommended programming.'

VI. Conclusion

As an IoT device, Smart TVs suffer from lack of proper credential authentication. As a data collecting device, Smart TVs send thousands of requests each day, many of which unprotected by encrypted protocols like HTTPS. In this way, Smart TVs may be the ultimate exploitable IoT device, with multiple points of entry for attackers. Although mitigating these vulnerabilities can protect data in many cases, we must ask ourselves if we care to protect our data at all. Do you consider your viewing data private? Yes, marital status, age, sex, and education are personal pieces of information, but if the only association to those pieces of information is how many hours you spent watching Netflix programming, do you really care? If we can prevent attackers from remotely connecting to our Smart TV devices by simply disabling features, perhaps the HTTP plaintext data collection going on behind the scenes cannot harm us. Of course users take a risk sending information like IP and MAC address unencrypted across a network. Sending a plaintext IP address is like posting your home address on social media, opening yourself up to attackers who now know where to find you. Personalized user experiences come at the cost of information exposure; computer scientists and cyber security researchers have yet to solve this issue.

---

[12] Goodin, Dan, et al. "LG Smart TV Snooping Extends to Home Networks, Second Blogger Says." *Ars Technica*, Ars Technica, 21 Nov. 2013, arstechnica.com/information-technology/2013/11/lg-smart-tv-snooping-extends-to-home-networks-second-blogger-says/.

References

Angwin, Julia. “Own a Vizio Smart TV? It's Watching You.” *ProPublica*, ProPublica, 9 Nov. 2015, www.propublica.org/article/own-a-vizio-smart-tv-its-watching-you.

Chow, Ming. “PiHole Admin Console Data” 11 Nov. 2018.

“Content Security Policy (CSP).” *MDN Web Docs*, 24 Aug. 2018, developer.mozilla.org/en-US/docs/Web/HTTP/CSP.

Conti, Mauro, et al. “Internet of Things Security and Forensics: Challenges and Opportunities.” *Future Generation Computer Systems*, vol. 78, no. 2, Jan. 2018, pp. 544–546.

Gardner, Eriq. “Spying TVs: Legal Settlement Provides a Few Bucks for Vizio Owners; Millions for Lawyers.” *The Hollywood Reporter*, The Hollywood Reporter, 5 Dec. 2018, www.hollywoodreporter.com/thr-esq/spying-tvs-legal-settlement-provides-a-few-bucks-vizio-owners-millions-lawyers-1149262.

Goodin, Dan, et al. “LG Smart TV Snooping Extends to Home Networks, Second Blogger Says.” *Ars Technica*, Ars Technica, 21 Nov. 2013, arstechnica.com/information-technology/2013/11/lg-smart-tv-snooping-extends-to-home-networks-second-blogger-says/.

Goodin, Dan. “How an Internet-Connected Samsung TV Can Spill Your Deepest Secrets.” *Ars Technica*, Ars Technica, 12 Dec. 2012, arstechnica.com/information-technology/2012/12/how-an-internet-connected-samsung-tv-can-spill-your-deepest-secrets/.

“ImmuniWeb® WebScan | High-Tech Bridge.” *High-Tech Bridge - Web and Mobile Application Security*, 10 Dec. 2018, 18:31, www.htbridge.com/websec/.

Lu, Yang, and Li Da Xu. “Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics.” *IEEE*, 2018, doi:DOI 10.1109.

“Pi-Hole®: A Black Hole for Internet Advertisements.” *Pi-Hole®: A Black Hole for Internet Advertisements*, 2018, pi-hole.net/.