Spring 1-1-2016

Transitioning the Traditional Business Model for Television: Personal Data Sharing by Streaming Video Mobile Apps

Matthew Thomas Guza
University of Colorado at Boulder, matthew.guza@colorado.edu

TRANSITIONING THE TRADITIONAL BUSINESS MODEL FOR TELEVISION: PERSONAL DATA SHARING BY STREAMING VIDEO MOBILE APPS

by

MATTHEW THOMAS GUZA

B.S. Information Networking and Telecommunications

Fort Hays State University, 2012

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado at Boulder in partial fulfillment

of the requirement for the degree of

Master of Science

Department of Interdisciplinary Telecommunications Program

2016

This thesis entitled:
Transitioning the Traditional Business Model for Television: Personal Data Sharing by Streaming Video Mobile Apps
written by Matthew Thomas Guza
has been approved for the Department of Interdisciplinary Telecommunications Program

Dr. David Reed (Thesis Chair)

Dr. Scott Savage (Thesis Committee)

Mr. Joe McManus (Thesis Committee)

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Guza, Matthew Thomas (M.S., Telecommunications [Department of Interdisciplinary Telecommunications Program])
Transitioning the Traditional Business Model for Television: Personal Data Sharing by Streaming Video Mobile Apps
Thesis directed by Dr. David Reed

Streaming video now represents more than half of all Internet bandwidth consumption and consumers are spending more of their time on mobile devices resulting in total video plays moving to mobile platforms. The rapid growth of streaming Internet video and the popularity of video on mobile platforms is leading to additional avenues of advertising. The new data collection vector from mobile ad providers creates new areas of concern for the user's privacy posture. This thesis investigates if streaming video apps implicate user privacy by sharing potentially sensitive information, how that information is shared, and the role of advertisers. Testing was conducted on 10 popular mobile video apps on both Android and iOS platforms using a Man-in-the-Middle proxy and specialized Wi-Fi access point to capture data flows originating from the apps. The captured data flows were analyzed using a list of keywords representing potentially sensitive information, resulting in a subset of data containing potentially sensitive data leaks. App vendor privacy policies were analyzed and compared to the captured app data leaks. The majority of the app privacy policies were complex, lacked transparency, and 6 apps were misaligned with their associated policy, sharing potentially sensitive information to third parties not conveyed to the user by the policy. Nearly all of the video apps tested in this thesis leaked potentially sensitive information about the user to third parties. The apps connected to 28 different ad networks and half of the apps shared data with these networks. Additionally, half of the apps shared potentially sensitive information with third parties in an insecure manner, creating the need for increased adoption of secure communications. Through interdisciplinary analysis, this thesis provides increased understanding of privacy implications from mobile video apps, and a description of how the advertising-based business model of television is transitioning to the online environment.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. David Reed, for all of his insight and support throughout the entire thesis process. Without a doubt, I am a stronger engineer and interdisciplinary researcher because of your guidance.

I would also like to thank my thesis committee members, Dr. Scott Savage and Joe McManus for their time and support of my endeavor.

I would also like to thank Elizabeth Golder for always being there to answer any of my ITP related questions. Additionally, I would like to thank all those who work to make ITP possible at a distance. Without your effort and flexibility in the unique situations introduced by distance education, people like myself would be unable to obtain this quality education while working in the industry.

Finally, I would like to thank my wife, Sarah. Without your unwavering support, I would have never embarked on any academic journey, let alone reached the height of graduate level achievement. Thank you for all you do.

CONTENTS

TABLES

FIGURES

FIGURE

GRAPHS

GLOSSARY OF TERMS

ARP: Address Resolution Protocol

CA: Certificate Authority

FTC: Federal Trade Commission

HTTP: Hypertext Transfer Protocol

IMEI: International Mobile Station Equipment Identity

MAC: Media Access Control

MITM: Man-in-the-Middle

PII: Personally Identifiable Information

SMOG: According to the author of the system (McLaughlin) the name is not a acronym, but rather a tribute to Gunning's Fog Index. Some argue it refers to “Simple Measure Of Gobbledygook”.

SMS: Short Message Service

SSID: Service Set Identifier

SSL: Secure Sockets Layer

Chapter 1

Introduction

1.0 Overview

Streaming video is now the dominant source of Internet bandwidth consumption in North America.[1] The rapid growth of television viewing on Internet devices, including smartphones, is driving increasing "cord cutting" from traditional cable and satellite television platforms and nearly 20% of young adults ages 18 to 29 are considered "cord cutters", having dropped their cable or satellite television platforms in favor of Internet streaming video options.[2]

Consumers now spend more time on mobile devices than any other platform with apps dominating the experience.[3] This pivot to mobile results in the majority of total video plays moving to mobile platforms, with smartphones becoming the device of choice over tablets for video.[4]

The increasing popularity of "cord cutting" and the move to mobile means advertisers focusing only on traditional television markets may be loosing out on a growing segment of customers. Not surprisingly, advertisers are beginning to redirect their ad distribution to the mobile space and mobile ad spending is on track to become the dominant digital ad spending segment.[5] This pivot to online provides a new, more interactive, platform of data and metric

[1] *Sandvine, "Global Internet Phenomena Report", 2015. Available: https://www.sandvine.com/trends/global-internet-phenomena/. [Accessed: 1- Apr- 2016].*
[2] *Pew Research Center, "Home Broadband 2015", 2015. Available: http://www.pewinternet.org/files/2015/12/Broadband-adoption-full.pdf. [Accessed: 1- Apr- 2016].*
[3] *Internet Society, "Global Internet Report 2015", 2015. Available: http://www.internetsociety.org/globalinternetreport/assets/download/IS_web.pdf. [Accessed: 1- Apr- 2016].*
[4] *Ooyala, "Global Video Index Q2 2015", 2015. Available: http://go.ooyala.com/wf-video-index-q2-2015. [Accessed: 1- Apr- 2016].*
[5] *"Mobile Ad Spend to Top $100 Billion Worldwide in 2016, 51% of Digital Market - eMarketer", Emarketer.com, 2016. Available: http://www.emarketer.com/Article/Mobile-Ad-Spend-Top-100-Billion-Worldwide-2016-51-of-Digital-Market/1012299. [Accessed: 12- Mar- 2016].*

collection for advertisers to better direct their advertisements over traditional cable or satellite television platforms, creating new areas of concern for user privacy posture.

1.1 Privacy

The first writing about privacy can be credited to Warren and Brandeis, who wrote about the "right to be left alone" in 1890.[6] This concept of privacy has progressed to include "the control over when and by whom the parts of us can be seen or heard, touched, smelled, or tasted by others".[7] But as technology and sharing change, the definition and very idea of privacy and its boundaries have struggled to evolve. However, many scholars argue that even with its complexities, privacy remains one of the most important concepts of our time.[8] The majority of mobile users agree that privacy, even as elusive as it may be, is extremely important and a recent study found 52% of mobile users will delete an app if concerned about privacy or security.[9]

The concept of privacy has crossed over into the digital space bringing with it additional unique and interdisciplinary challenges. As users integrate with Internet devices, including smartphones, the amount of private and sensitive information being stored and collected will continue to increase. The concept of privacy can be extended to this information, centering around "who" or "what" has access to the private and sensitive information stored on the device.

That being said, smartphone apps have a varying degree of access to private and sensitive information stored on the device. Just as different apps require access to different resources on

[6] *S. Warren and L. Brandeis, 'The Right to Privacy', Harvard Law Review, vol. 4, no. 5, p. 193, 1890.*
[7] *R. Parker, 'A Definition of Privacy', Rutgers Law Review, vol. 27, no. 2, p. 275, 1974.*
[8] *D. Solove, 'A Taxonomy of Privacy', University of Pennsylvania Law Review, vol. 154, no. 3, p. 477, 2006.*
[9] *On Device Research, "Global Consumer Trust Report 2016", 2016. Available: http://www.mobileecosystemforum.com/solutions/consumer-trust/global-consumer-trust-report-2016/. [Accessed: 1- Apr- 2016].*

the device for proper functionality, apps also differ in the way they handle user privacy and information collection, remote host connections, and data flows can vary from app to app.

In an effort to communicate privacy of the apps to some extent, most vendors publish privacy policies, attempting to outline "what" and "how" they collect, use, disclose, and manage user information. Popular app marketplaces, including Google's Play Store and Apple's App Store, encourage developers to include a privacy policy with each app. However, even when privacy polies are included, they can vary significantly from vendor to vendor.

1.2 Problem Statement

Do popular streaming video apps implicate user privacy by sharing potentially sensitive information with themselves and third parties? What and how is this information shared if any, and what role do advertisers play?

1.3 Scope

The scope of this thesis encompasses investigating popular smartphone streaming video apps for user privacy implications and the role of advertisers and ad networks. The method for measuring privacy concentrates on the concept of "data leaks", defined as information shared to third parties with or without permission of the associated privacy policy. In general, privacy implications are defined in this thesis as personal data leaks encompassing Personally Identifiable Information (PII), location details, unique device characteristics, and user behavior to a third party.[10]

The research for this thesis was carried out in Germany and as a result, only apps without

[10] *PII used in this thesis encompasses NIST Special Publication 800-122 and other community standards.*

geographic content licensing restrictions in Europe, which would limit testing in Germany, were selected and tested. Additionally, only those apps whose primary function was to provide streaming video, both free and subscription-based, were chosen. The list of apps that fit into this category was much larger than what was capable of being tested in a timely manner and as a result, 10 streaming video apps from the two most popular app marketplaces, Google's Play Store and Apple's App Store, were tested.

Testing of apps took place utilizing a Man-in-the-Middle (MITM) scenario that allowed Hypertext Transfer Protocol (HTTP) data flows to be intercepted and inspected for information leaks. In certain cases, apps could encrypt or obfuscate information before placement into the communications flow, making it unintelligible for the purposes of the testing carried out for this thesis. Only data flows that were human-readable after any decoding, such as Base64, were tested for information leaks.[11]

1.4 Motivation

As time spent watching video on smartphones increases to become the dominant video platform, third parties will strive to collect additional information on the user. Advertisers in particular will strive to deliver better "interest-based" advertisements to the user, requiring deeper insights into their behavior. From an advertiser perspective, there is a mounting need for more efficient and increasingly targeted advertising, creating new areas of privacy implication for the streaming video app user. Many mobile streaming video apps are free to the user, operating under an ad-based business model which often includes selling in-app ad space to the highest bidder with little to no vetting.

[11] *Base64 is a group of binary-to-text encoding schemes used to represent binary data in an ASCII format.*

Studies have found apps violating user privacy, misalignment with privacy policies, and noncompliance with marketplace terms, but none have focused exclusively upon this growing market segment and the role of advertisers. Additionally, these studies have traditionally focused on either technical or policy investigations, creating a need for an interdisciplinary analysis encompassing streaming video app personal information leaks, vendor privacy policies, and the evolution of the traditional business model for television and advertisers.

1.5 Structure of Thesis

This first chapter presented an overview of video and privacy, including information on current streaming video trends moving toward the mobile space and how the concept of privacy can be applied to smartphone data and apps. Additionally, the problem statement, scope, and motivation for this thesis and its associated research was provided.

The second chapter will present a literature review. This will include coverage of research related to this thesis in the areas of app privacy policies, data leaks, and advertising.

The third chapter will present comprehensive coverage of the research methodology. This will include details on the apps tested, the testing environment, how the apps were tested, and how vendor privacy policies and app data leaks were analyzed.

The fourth chapter will present the results and findings from the vendor privacy policy analysis and app data leak analysis. This will include findings on advertisers and ad networks.

The fifth chapter will conclude the thesis with an overview of observed streaming video app user privacy, an explanation on the role of advertisers and ad networks, limitations, recommendations, and suggestions for future work.

Chapter 2

Related Work

2.0 Overview

In order to provide enhanced insight into streaming video app user privacy posture and the role of advertisers, an interdisciplinary analysis of the technical and policy segments of privacy is necessary. Each aspect provides an important piece in framing what is conveyed to the user from a policy standpoint, creating a baseline of what the user might reasonably expect in terms of information handling and their privacy, and how the app actually handles user information and privacy under normal operations. The following sections provide coverage of related work necessary for understanding the privacy policy and data leak aspects of this thesis.

2.1 App Privacy Policies

Privacy policies provide a platform for app vendors to convey to users what data are collected and how that information is handled. The FTC urges app developers to include privacy policies, constructed with a focus on transparency, for all mobile apps.[12]

For the user, privacy policies should provide a clear explanation of the level of privacy they can expect from the app, providing insight into the app's data collection and sharing practices. However, a study of the top 600 mobile health (mHealth) apps found only 30.5% of the apps provided a privacy policy and of those apps that did, two-thirds did not specifically provide coverage of the app.[13] Specifically, most of the privacy policies surveyed provided

[12] *"Mobile Privacy Disclosures: Building Trust Through Transparency: A Federal Trade Commission Staff Report", Ftc.gov, 2013. Available: https://www.ftc.gov/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission. [Accessed: 15- Mar- 2016].*

[13] *A. Sunyaev, T. Dehling, P. Taylor and K. Mandl, "Availability and quality of mobile health app privacy policies", Journal of the American Medical Informatics Association, 2014.*

little useful information to the user as they did not focus on the app at all. Additionally, the examined policies were found to be complex and require a high level of literacy, making them inaccessible for some users. Although the study did not focus on streaming video apps, it provides insight into the state of app privacy policies that is useful during the policy analysis stage of this thesis.

Another study investigated the privacy policies of 211 Android diabetes apps finding only 19% of the apps had privacy policies.[14] Nearly half of the 41 apps that did have privacy policies shared information with partners or third parties. The study also found that 39% of the policies stated data collected might be used for advertisement purposes. These findings support the need to investigate third party information leaks and sharing, as well as connections with advertisers.

Expanding beyond health-related apps, a study investigated 110 Android and iOS app privacy policies and affirmed the incomplete adoption of privacy policy implementation for mobile apps.[15] The study found only 75% of Android apps and 67% of iOS apps tested had working privacy policy links. The study also analyzed the privacy policies for information on encryption, finding that 61% of privacy policies stated that data would be encrypted. This finding is important as it influences the research methodology of this thesis by making it necessary to implement a MITM proxy capable of intercepting encrypted data flows utilizing Secure Sockets Layer (SSL), to check for data leaks. Additionally, by reading a privacy policy, users need to know if their personal information and privacy is protected by proper encryption mechanisms.

[14] S. Blenner, M. Köllmer, A. Rouse, N. Daneshvar, C. Williams and L. Andrews, "Privacy Policies of Android Diabetes Apps and Sharing of Health Information", JAMA, vol. 315, no. 10, p. 1051, 2016.
[15] J. Graves, "An Exploratory Study of Mobile Application Privacy Policies", Technology Science, 2015. Available: http://techscience.org/a/2015103002. [Accessed: 1- Apr- 2016].

2.2 App Data Leaks

Privacy policies provide insight into how the vendor and apps are supposed to handle user information and privacy. However, testing apps under normal circumstances is necessary to investigate how they truly handle information, analyze if they perform in accordance with their associated privacy policy, and how they interact with advertisers.

A recent study conducted at the FTC, by researchers from Harvard University, Carnegie Mellon University, and the Massachusetts Institute of Technology, examined privacy information leaks of the top 110 free Android and iOS mobile apps, finding significant discrepancies in the actual flow of user data versus what the user has visibly approved to share.[16] The study found that 73% of the Android apps tested shared personal information with third parties and 47% of the iOS apps tested shared location data with third parties. Furthermore, the study concluded that many apps tested shared potentially sensitive user data without requesting visible permission from the user to access the data. App testing was conducted using a MITM proxy and did not concentrate on the streaming video app segment. The success of this study, however, supports the main testing methodology of this thesis (testing apps for potentially sensitive data leaks to third parties utilizing a MITM proxy).

A study by the Privacy Rights Clearinghouse analyzed 43 free and paid mobile health and fitness apps for privacy implications and found 52% of free apps and 40% of paid apps send data to third parties.[17] More importantly, the study found that only 15% of apps sending data to third

[16] *J. Zang, K. Dummit, J. Graves, P. Lisker and a. Sweeney, "Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps", Technology Science, 2015. Available: http://techscience.org/a/2015103001. [Accessed: 1- Apr- 2016].*

[17] *"Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications", Privacy Rights Clearinghouse, 2013. Available: https://www.privacyrights.org/mobile-medical-apps-privacy-technologist-research-report.pdf. [Accessed: 1- Apr- 2016].*

parties utilized SSL to encrypt the communications stream. Additionally, the study analyzed app connections and data leaks to advertisers concluding that apps with advertisements pose a significantly higher privacy risk. Like the research conducted at the FTC, this study did not focus on the streaming video app segment, however, a MITM proxy was utilized to test the mobile apps for data leaks.

2.3 Advertising

Free (and sometimes subscription) apps are often made possible by the inclusion of mobile ad networks. App developers can include ad network libraries in their app, enabling third party advertisers to bid for placement in the in-app advertising slots. When a user views an in-app ad, the ad network assigns a unique ID to the user and begins building a profile. As the user interacts with more apps containing the ad network libraries, more information can be gathered enriching user profiling to provide more targeted advertisements in the future.[18] When apps are noncompliant with marketplace privacy guidelines, their own privacy policies, or simply mishandle user private data, potentially sensitive and privacy-violating information may leak to ad networks. Since many free streaming video apps utilize an ad-based business model, privacy concerns may arise from advertising associated with the apps.

A study of 100,000 Android apps identified over 100 in-app ad libraries finding that most of the ad libraries collect private information from the device or user.[19] Some of the information

[18] *A. Seneviratne, K. Thilakarathna, S. Seneviratne, M. Kaafar and P. Mohapatra, "Reconciling bitter rivals: Towards privacy-aware and bandwidth efficient mobile Ads delivery networks", 2013 Fifth International Conference on Communication Systems and Networks (COMSNETS), 2013.*

[19] *M. Grace, W. Zhou, X. Jiang and A. Sadeghi, "Unsafe exposure analysis of mobile in-app advertisements", Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks - WISEC '12, 2012.*

probed by the ad networks utilizing the libraries included location data, web browser bookmarks, contacts, calendars, Short Message Service (SMS) messages, and phone information. The study discovered apps accessing data on the device that would provide no immediate benefit to advertisements, including SMS service details and call history. This questionable access to sensitive information could negatively impact users by allowing ad networks to uncover a users' true identity, resulting in privacy concerns stemming from explicit sensitive data leaks to ad networks.

Additionally, ad targeting could introduce privacy risks for the user when advertisement personalization provides a link to the users true identity. A recent study at the Georgia Institute of Technology surveyed more than 217 users and their associated ad viewings to try and understand privacy concerns stemming from in-app ad networks.[20] The study found the popular Google AdMob ad network delivered specialized and personalized ads based on the interests and demographic of the user with a high degree of accuracy providing a significant correlation between the users' profile and ads observed. Furthermore, the study found that through observing the personalized ads to the user, sensitive personal information such as gender can be determined with a 75% accuracy.

[20] *W. Meng, R. Ding, S. Chung, S. Han and W. Lee, "The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads", Georgia Institute of Technology, 2016. Available: http://www.cc.gatech.edu/~wmeng6/ndss16_mobile_ad.pdf. [Accessed: 15- Mar- 2016].*

Chapter 3

Research Methodology

3.0 Apps Tested

Ten apps (*Tables 1, 2*) focusing on streaming video for entertainment purposes, were tested February 3-17, 2016.

Table 1 - Android Apps Tested

App Name	Locale	App Type	Version	Dates Tested	Store Category
Dailymotion	EU	Free	4857	2/9/16 - 2/10/16	Entertainment
Maxdome	EU	Subscription	2.2.1	2/14/16 - 2/15/16	Entertainment
Netflix	US	Subscription	4.2.0	2/3/16 - 2/4/16	Entertainment
Popcornflix	US	Free	2.7	2/15/16 - 2/16/16	Entertainment
TubiTV	US	Free	2.5.2	2/11/16 - 2/12/16	Entertainment
Twitch	US	Free	4.6.5	2/4/16 - 2/5/16	Entertainment
Vevo	US	Free	2.2.17	2/10/16 - 2/11/16	Music & Audio
Viewster	Other	Free	5.2.1	2/12/16 - 2/13/16	Entertainment
Vimeo	US	Free	2.0.2	2/8/16 - 2/9/16	Entertainment
YouTube	US	Free	4.2	2/9/16	Media & Video

Table 2 - iPhone Apps Tested

App Name	Locale	App Type	Version	Dates Tested	Store Category
Dailymotion	EU	Free	5.2	2/6/16	Entertainment
Maxdome	EU	Subscription	3.2.3	2/16/16 - 2/17/16	Entertainment
Netflix	US	Subscription	7.2.5	2/3/16 - 2/4/16	Entertainment
Popcornflix	US	Free	3.3.1	2/15/16 - 2/16/16	Entertainment
Twitch	US	Free	3.6.3	2/4/16 - 2/5/16	Entertainment
Vevo	US	Free	5.0.5	2/10/16 - 2/11/16	Music
Viewster	Other	Free	5.2.1	2/12/16 - 2/13/16	Entertainment
YouTube	US	Free	11.03	2/8/16 - 2/9/16	Photo & Video

Testing occurred on both Android and iOS platforms, and all apps were obtained from Google's Play Store and Apple's App. Two of the apps were subscription-based and tested under a free trial period and the rest of the apps were free to the user. Eight of the ten apps allowed user registration, and new accounts from new email addresses were created, one for Android apps and one for iOS apps. The apps were from an international mix of vendors stemming from the United States, European Union, and Switzerland.

3.1 Testing Environment

The Android apps were tested using a HTC One Mini running Android version 4.4.2 and the iOS apps were tested using an iPhone 4S running iOS version 9.2.1. Each phone was factory reset prior to testing and all software updates were applied. Additionally, each phone was associated with a new email address as its primary account. No further customization was made to the phones.

3.2 Privacy Policies

The apps were checked for linked privacy policies prior to data testing. All of the apps included at least a privacy policy, of varying coverage and lengths, and two of the apps included a separate cookie policy (*Table 3*). Each policy was analyzed for length and vendor information collection practices and information use post-collection.

Table 3 - App Provider Policy Overview

App Name	App Type	Policy	Pages
Dailymotion	Free	Privacy, Cookie	1, 1
Maxdome	Subscription	Privacy	9
Netflix	Subscription	Privacy	6
Popcornflix	Free	Privacy	3
TubiTV	Free	Privacy	5
Twitch	Free	Privacy, Cookie	5, 3
Vevo	Free	Privacy	8
Viewster	Free	Privacy	8
Vimeo	Free	Privacy	3
YouTube	Free	Privacy	9

When each policy was analyzed for information collection, data types such as name, email address, or location were recorded if they were explicitly mentioned as collected by the policy. The collected information data types were categorized as information being provided by the user to the app provider, information collected by the provider in a automatic fashion,

information collected by third parties, and information obtained by the app vendor from a third party.

Additionally, each policy was analyzed to determine how the collected information was being used by searching for language dictating why the vendor collected the information. Most of the privacy policies included an explanation for data collection from the user. Based on the reasons provided, a summary of different types of information use categories were established to including service delivery, customer support, customer contact, market research, advertising, business management, contests, enforcement of terms and policies, and legal reasons. The vast majority of reasons for collecting information by each vendor were able to fit under one or more of these categories.

Finally, each policy was analyzed for language regarding the use of encryption. This analysis checked for the explicit mention of encryption of data before being pushed into a data flow, either by means of a true encryption algorithm or encoding such as Base64. Additionally, each policy was analyzed for language regarding secure communications utilizing SSL.

3.3 MITM Proxy Background

In order to test the apps for personal information leaks to third parties, data flows to and from the apps must be intercepted and inspected. Since many apps employ SSL to provide a secure communications channel for HTTP, a MITM web proxy was utilized to intercept both secure and insecure communications.

Most MITM attacks on SSL work either by exploiting the Certificate Authority (CA) trust model that devices utilize to make trust determinations, Address Resolution Protocol (ARP)

poisoning, or by utilizing social engineering techniques designed to lure users into trusting a self-signed or unexpected SSL certificate.[21,22]

The general process starts with the creation of a SSL certificate issued by the MITM attacker. When the client requests a secure communications session from a server, the MITM attack intercepts the request and responds with the MITM certificate instead of the proper server certificate. If the MITM attack certificate, masquerading as the legitimate server, is trusted by the user, the attacker will have the capability to intercept and decrypt information from the secure communications session. Once a SSL handshake is completed between the client and the MITM attacker, the attacker subsequently completes an SSL handshake with the legitimate server.[17] At this point, there is a secure communications session established between the client and the MITM attacker and a second secure communications session established between the MITM attacker and the legitimate server. Throughout on-going communication, the MITM attacker will intercept requests to the server on behalf of the client, decrypt, and forward them on to the legitimate server. The opposite process is completed when the server responds. The MITM proxy appears to the client as a single secure communications session with the legitimate server, allowing the MITM attacker to eavesdrop on SSL-protected communications.

3.4 MITM Proxy Method

App data flows were captured utilizing the open source Mitmproxy, available for most Linux distributions and bundled with the Kali 2.0 Linux distribution by default.[23] The proxy runs

[21] *I. Dacosta, M. Ahamad and P. Traynor, "Trust No One Else: Detecting MITM Attacks against SSL/TLS without Third-Parties", Computer Security – ESORICS 2012, pp. 199-216, 2012.*
[22] *P. Pateriya and S. Kumar, "Analysis on Man in the Middle Attack on SSL", International Journal of Computer Applications, vol. 45, no. 23, pp. 43-46, 2012.*
[23] *Docs.mitmproxy.org, "Introduction - Mitmproxy 0.15 documentation", 2016. Available: http://docs.mitmproxy.org/en/stable/. [Accessed: 10- Feb- 2016].*

as a laptop connected to an Internet-facing network running Mitmproxy and the smartphones are configured with the proxy's IP address and port as their Wi-Fi network HTTP proxy. The proxy utilizes a self-signed certificate that is untrusted by the device CA local trust stores. In order to circumvent this for the purposes of the testing methodology, the proxy's certificate is manually installed on the smartphones as a trusted certificate. When the proxy is running, all client HTTP data flows run through the proxy. The flows of interest include HTTP GET, used to request resources from a remote web server, HTTP POST, used to submit data to a remote web server, and HTTP PUT, used to update data on a remote web server.

Subsequently, the proxy performs masquerading of the legitimate server and interception and decryption of communications, while recording all traffic to a log file. Mitmproxy provides a console for real-time and differed analysis of the communications information (*Figure 1*) as well as a tool called Mitmdump, which allows for manipulation and searching of the log files.

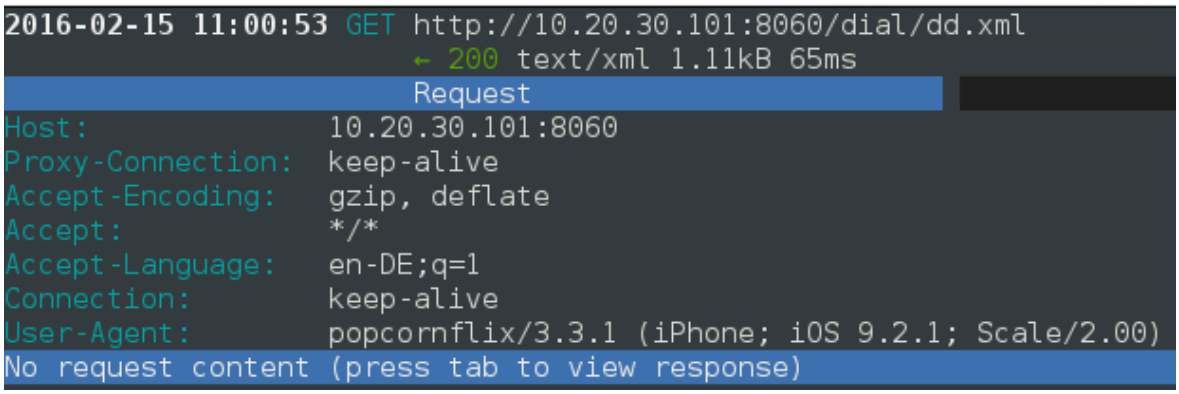


Figure 1 - Mitmproxy Real-Time Console

Both platforms provide interesting information from the the data flows, including the remote host the app is sending data to, whether or not the connection was utilizing SSL, decrypted payload information of the request, and decrypted payload information of the response. Utilizing the MITM proxy method, both insecure and secure HTTP communications flowing to and from the

smartphone streaming video app can be inspected for potentially sensitive data leaks that could have a negative impact on a user's privacy posture.

A list of 65 keywords representing potential sensitive information leaks was created and fed into a series of Linux bash scripts that searched the Mitmproxy data flow dump files for matches to the keywords (*Appendix A*). The keywords were categorized and given a designation based on the type of data they represent, including PII, behavior, location, and device. Whenever a match was found, the scripts created a separate Mitmproxy dump file containing the data flow of the match. Each one of these matched data flow files was inspected manually for potential sensitive information leaks and verified suspected leaks were recorded in a log file with with the keyword that triggered the match, type of data, data designation (e.g. PII, behavior), HTTP method, what stage of testing the leak was observed, remote host, if SSL was utilized, file ID of the associated Mitmproxy dump file, and comments from the manual check.

3.5 Wi-Fi Access Point Method

The MITM proxy method allows the capture of insecure and secure HTTP data flows to and from a video app. However, if the app is utilizing protocols outside of HTTP for transmitting and receiving data, the MITM proxy will be unable to capture these data flows. In order to verify if other protocols outside of HTTP are being utilized to leak potentially sensitive information, a specialized Wi-Fi access point was implemented, utilizing the open source Hostapd and Wireshark network tools to sniff data flows to and from the video apps.[24,25]

[24] *"hostapd: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/EAP/RADIUS Authenticator", W1.fi, 2016. Available: https://w1.fi/hostapd/. [Accessed: 16- Mar- 2016].*
[25] *"Wireshark User's Guide", Wireshark.org, 2016. Available: https://www.wireshark.org/docs/wsug_html_chunked/. [Accessed: 16- Mar- 2016].*

Running on the laptop utilized by the Mitmproxy, Hostapd was configured as an open Wi-Fi network with the network name (Service Set Identifier or SSID) of "NONO". The smartphones were associated with this network, in the same manner as any other Wi-Fi network. Alongside Hostapd, Wireshark was utilized to capture all data flowing from the Wi-Fi interface card to the wired Ethernet interface card, saving it to a log file. The log files can be analyzed with Wireshark to check for apps utilizing protocols other than HTTP (*Appendix B*).

3.6 App Testing Protocol

A protocol was developed for testing each app, integrating manual testing, the MITM proxy method, and the Wi-Fi access point method. The following summary represents how data was collected from the apps and utilized for analysis in Chapter 4:

Step 1 - ***Install App*** - Install the app on the phone utilizing the MITM proxy to record data flows. Record app permissions during the installation process if possible.

Step 2 - ***Check Permissions*** - Check app post-install permissions manually.

Step 3 - ***Test App 1*** - Play two short videos on the app utilizing the MITM proxy to record data flows. Each video should be around 5 minutes in length.

Step 4 - ***Test App 2*** - Connect app to Hostapd Wi-Fi network and play two short videos on the app utilizing Wireshark to record data flows. Each video should be around 5 minutes in length.

Step 5 - ***Menu Test*** - Navigate through the functionality of the app for 5 minutes utilizing the MITM proxy to record data flows.

Step 6 - ***Phone Test*** - Access the calendar, contacts, and web browser while the video app is running, utilizing the MITM proxy to record data flows.

Step 7 - ***Idle Run*** - Run the app on the phone in the background for roughly 12 hours utilizing the MITM proxy to record data flows.

Chapter 4

Results and Discussion

4.0 Privacy Policy Analysis

The privacy policy of each streaming app was analyzed (*Table 4*). The average page length was 5.8 pages, with the shortest policy belonging to Dailymotion consisting of 1 page and the longest of 9 pages belonging to YouTube (Google) and Maxdome. The average word count for the privacy policies was 3013.3 words. To assess policy accessibility for all app users, an online tool was utilized to score each policy using the standardized SMOG index, providing the number of years of education needed to understand the policy.[26] The average reading level of the policies was 11.28, with only one policy (Dailymotion) scoring below a 10, resulting in a reading level above the recommended target of 8 for general populous accessibility (based on National reading level statistics) (*Table 4*).

Table 4 - App Provider Policy Inventory

App Provider	Type	Policies	Pages	Words	SMOG
Dailymotion	Free	Privacy, Cookie	1	707	8.9
Maxdome	Subscription	Privacy	9	4336	10.1
Netflix	Subscription	Privacy	6	2905	12.6
Popcornflix	Free	Privacy	3	1279	11.8
TubiTV	Free	Privacy	5	4610	11.8
Twitch	Free	Privacy, Cookie	5	3278	12.3
Vevo	Free	Privacy	8	4336	11.5
Viewster	Free	Privacy	7	2150	11.8
Vimeo	Free	Privacy	5	3711	11.5
YouTube	Free	Privacy	9	2821	10.5

[26] *H. McLaughlin, "SMOG Grading - A New Readability Formula", Journal of Reading, vol. 12, no. 8, pp. 639-646, 1969.*

The analysis found 80% of the privacy policies contain language approving sharing of information with third parties (*Table 5*). Additionally, 3 out of 10 policies include language regarding the encryption or obfuscation of data before being placed into either an insecure or secure communications channel. While data payload encryption may not be necessary if a secure communications channel, such as SSL is implemented, only 4 out of 10 policies include language indicating they utilize SSL to protect data in transit. If the privacy policies are directly representative of the actual app configuration, a topic explored more in section 4.1, security and privacy issues may result from sending potentially sensitive user information over insecure (unencrypted) communications channels.

Table 5 - App Provider Policy Findings

App Provider	US / EU / Other	Third Party Sharing or Collection	Encryption	SSL
Dailymotion	EU	No	No	No
Maxdome	EU	No	Yes	No
Netflix	US	Yes	No	No
Popcornflix	US	Yes	No	No
TubiTV	US	Yes	Yes	Yes
Twitch	US	Yes	No	No
Vevo	US	Yes	No	No
Viewster	Other	Yes	No	Yes
Vimeo	US	Yes	Yes	Yes
YouTube	US	Yes	No	Yes

We also investigated each privacy policy for language indicating what information is collected from the user, resulting in a privacy policy information collection rubric (*Table 6*). The rubric frames what the policy explicitly states it will collect from the user, represented by a checkmark, allowing for a visual comparison between the different policies.

Table 6 - Privacy Policy Information Collection Rubric									
App Provider	**Name**	**Email**	**Demog**	**Interest**	**Behavior**	**GPS**	**Locale**	**IP Address**	**Device Info**
Dailymotion		✓	✓		✓			✓	
Maxdome	✓	✓	✓	✓	✓		✓	✓	
Netflix	✓	✓	✓	✓	✓		✓	✓	✓
Popcornflix					✓			✓	
TubiTV	✓	✓	✓	✓	✓	✓		✓	
Twitch	✓	✓	✓	✓			✓	✓	✓
Vevo	✓	✓	✓	✓	✓	✓		✓	✓
Viewster	✓	✓		✓	✓			✓	
Vimeo	✓	✓	✓	✓				✓	
YouTube	✓	✓		✓	✓	✓		✓	✓
	PII			**Behavior**		**Location**		**Device**	

**** Demog = Demographic information such as gender, age, or race***

The last row of the table identifies the broader information designation categories associated with the data flow keywords discussed in chapter 3, providing another way to classify and analyze data that is collected by a vendor or shared with a third party (*Appendix A*). The designation standard across information collection and sharing provides consistency when comparing app data collection flows against privacy policies. Additionally, the designation allows us to visualize, in the case of the privacy policies tested, that all apps except Popcornflix collect PII.

Every privacy policy contained language stating the vendor collected, often automatically, the IP address from the user. Standard server logs normally include IP address information, which explains the ubiquitous IP address collection. The majority of policies state the apps do not collect device-specific information such as such as serial number, International Mobile Station Equipment Identity (IMEI), or Media Access Control (MAC) address, which vendors, such as Apple, are blocking in favor of other privacy-centric tracking solutions

solutions.[27] The rubric also reveals that GPS information is not normally collected by the apps, from a privacy policy perspective. Recent studies have discovered apps collecting GPS information and there has been a renewed focus on the privacy implications of apps tracking the geographic location of users.[28]

4.1 App Data Leak Analysis

The apps initiated HTTP requests to a total of 515 unique hosts, with 349 unique host connections originating from Android apps and 403 unique host connections originating from iPhone apps. On both platforms, Dailymotion connected to the most remote hosts with 154 Android app connections and 166 iPhone app connections. The MITM proxy captured a total of 20,468 HTTP requests with 9,076 originating from Android apps and 11,392 originating from iPhone apps.

Only 2 apps, Netflix and YouTube, did not leak any data to third parties as detected by the MITM proxy method. Evaluating the apps that leaked data to third parties revealed 6 apps leaked PII on both platforms, 7 apps leaked behavior information (7 on Android, 5 on iPhone), and 5 apps leaked device information (5 on Android, 3 on iPhone) (*Tables 7, 8*). Full GPS coordinates were leaked to third parties by Popcornflix, TubiTV, and Viewster on the Android platform (*Table 7*). In total, Android apps leaked data to third parties across more information designation categories than iPhone apps.

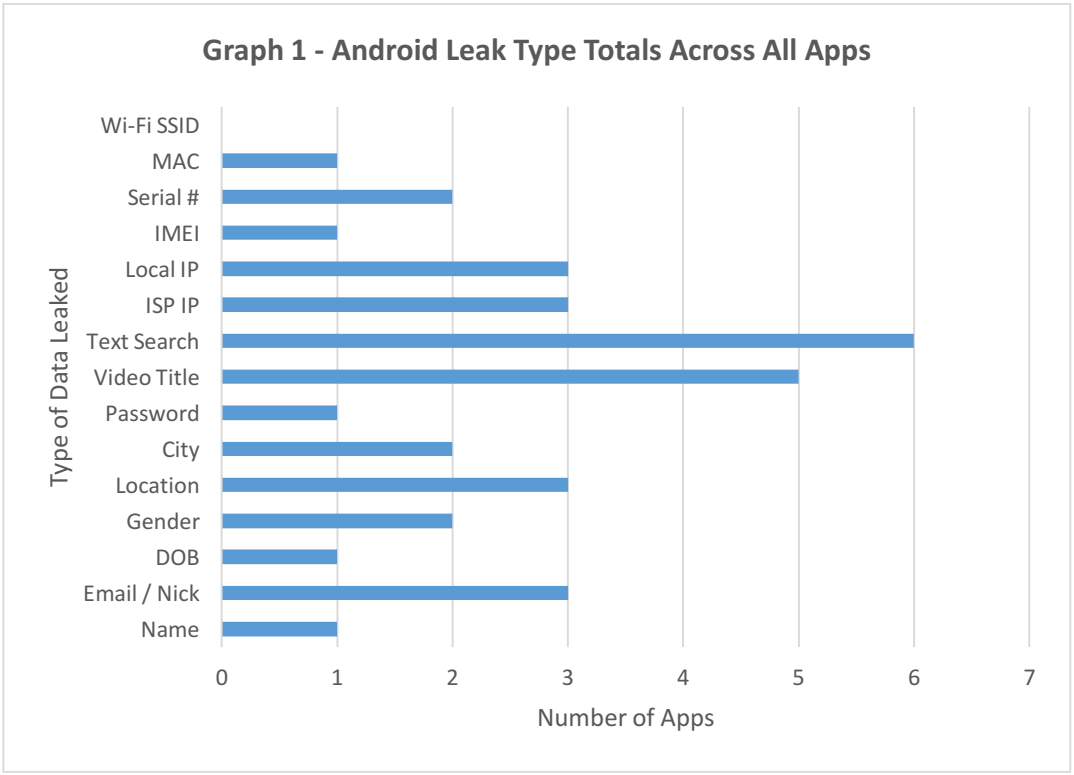
[27] *S. Perez, "iOS 7 Eliminates MAC Address As Tracking Option, Signaling Final Push Towards Apple's Own Ad Identifier Technology", TechCrunch, 2013. Available: http://techcrunch.com/2013/06/14/ios-7-eliminates-mac-address-as-tracking-option-signaling-final-push-towards-apples-own-ad-identifier-technology/. [Accessed: 05- Apr- 2016].*

[28] *E. Dwoskin, "Apps Track Users—Once Every 3 Minutes", WSJ, 2015. Available: http://www.wsj.com/articles/apps-track-usersonce-every-3-minutes-1427166955. [Accessed: 05- Apr- 2016].*

Table 7 - Android Data Leak to Third Party - Information Designation				
App Name	**PII**	**Location**	**Behavior**	**Device**
Dailymotion	✓		✓	✓
Maxdome	✓			
Netflix				
Popcornflix	✓	✓	✓	✓
TubiTV		✓	✓	✓
Twitch	✓		✓	
Vevo	✓		✓	✓
Viewster	✓	✓	✓	✓
Vimeo			✓	
YouTube				

Table 8 - iPhone Data Leak to Third Party - Information Designation				
App Name	**PII**	**Location**	**Behavior**	**Device**
Dailymotion	✓		✓	✓
Maxdome	✓			
Netflix				
Popcornflix	✓		✓	✓
Twitch	✓		✓	
Vevo	✓		✓	
Viewster	✓		✓	✓
YouTube				

Data leaks were spread across different categories with the largest number of apps on both Android and iPhone leaking user behavior information consisting of text search terms and video titles to third parties (*Graphs 1, 2*). On the Android platform, Vevo and Viewster leaked the most different types of data and Vevo, Viewster, and Popcornflix leaked the most different types of data on iPhone (*Tables 9, 10).* Android apps leaked more device-specific identifiers capable of tracking a user, such as serial number, IMEI, or MAC address, to third parties than iPhone apps. Additionally, Vevo was the only app that leaked both the full name of the user and their associated date of birth to third parties, which occurred on both platforms (*Tables 9, 10*).


Graph 1 - Android Leak Type Totals Across All Apps
Type of Data Leaked
Wi-Fi SSID
MAC
Serial #
IMEI
Local IP
ISP IP
Text Search
Video Title
Password
City
Location
Gender
DOB
Email / Nick
Name
0
1
2
3
4
5
6
7
Number of Apps

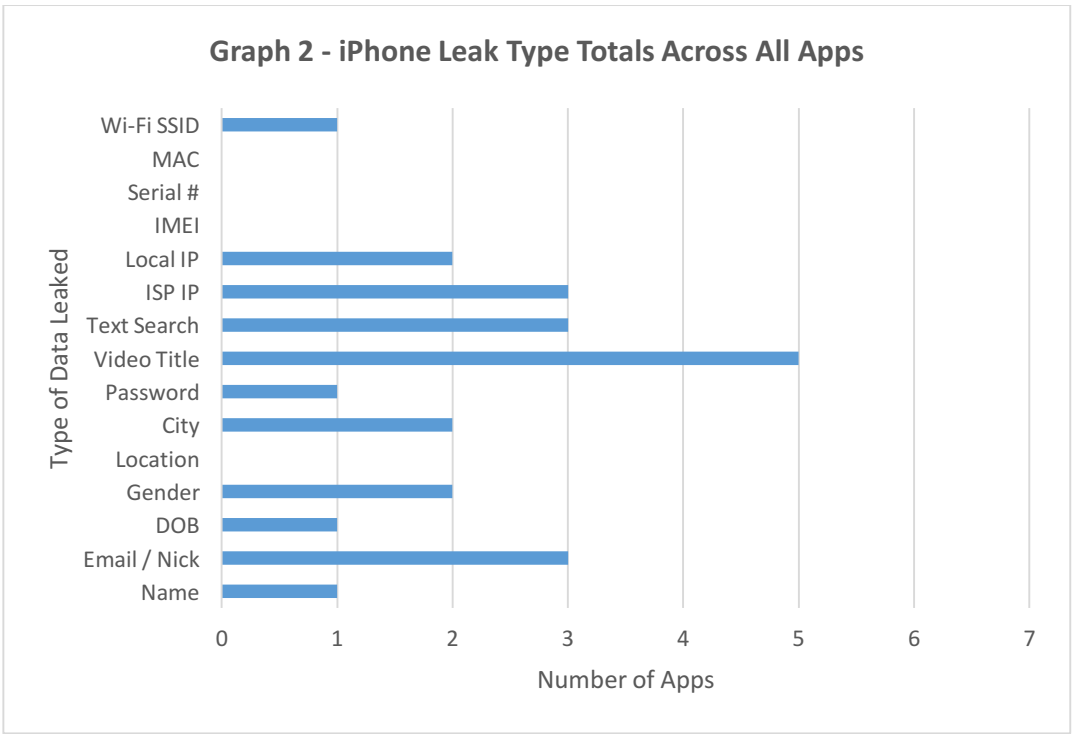

Graph 2 - iPhone Leak Type Totals Across All Apps
Type of Data Leaked
Wi-Fi SSID
MAC
Serial #
IMEI
Local IP
ISP IP
Text Search
Video Title
Password
City
Location
Gender
DOB
Email / Nick
Name
0
1
2
3
4
5
6
7
Number of Apps

Table 9 - Android Data Leak Details - Third Party											
	Dailymotion	Maxdome	Netflix	Popcornflix	TubiTV	Twitch	Vevo	Viewster	Vimeo	YouTube	
Name							✓				*PII*
Email / Nick						✓	✓	✓			*PII*
DOB							✓				*PII*
Gender		✓					✓				*PII*
City	✓			✓							*PII*
Password								✓			*PII*
GPS Location				✓	✓			✓			*Location*
Video Title	✓			✓		✓	✓	✓			*Behavior*
Text Search	✓			✓	✓		✓	✓	✓		*Behavior*
ISP IP	✓						✓	✓			*Device*
Local IP				✓			✓	✓			*Device*
IMEI								✓			*Device*
Serial #							✓	✓			*Device*
MAC					✓						*Device*
Wi-Fi SSID											*Device*

Table 10 - iPhone Data Leak Details - Third Party									
	Dailymotion	Maxdome	Netflix	Popcornflix	Twitch	Vevo	Viewster	YouTube	
Name						✓			*PII*
Email / Nick					✓	✓	✓		*PII*
DOB						✓			*PII*
Gender		✓				✓			*PII*
City	✓			✓					*PII*
Password							✓		*PII*
GPS Location									*Location*
Video Title	✓			✓	✓	✓	✓		*Behavior*
Text Search				✓		✓	✓		*Behavior*
ISP IP	✓			✓			✓		*Device*
Local IP				✓			✓		*Device*
IMEI									*Device*
Serial #				✓					*Device*
MAC									*Device*
Wi-Fi SSID				✓					*Device*

Half of the apps on both platforms leaked data to third parties over insecure, non-SSL HTTP communications (*Tables 11, 12*) and 4 apps on both platforms were found utilizing Base64 encoding to obfuscate information before placement in the communications channel (*Tables 13, 14*).

Table 11 - Android HTTP/S Leak Overview		
App Name	**HTTP**	**HTTPS**
Dailymotion	✓	✓
Maxdome	✓	
Netflix		
Popcornflix	✓	✓
TubiTV	✓	✓
Twitch		✓
Vevo	✓	✓
Viewster	✓	✓
Vimeo		✓
YouTube		

Table 12 - iPhone HTTP/S Leak Overview		
App Name	**HTTP**	**HTTPS**
Dailymotion	✓	✓
Maxdome	✓	
Netflix		
Popcornflix	✓	✓
Twitch		✓
Vevo	✓	✓
Viewster	✓	✓
YouTube		

Table 13 - Android Base64 Leak Overview	
App Name	**Base64 Found**
Dailymotion	✓
Maxdome	
Netflix	
Popcornflix	✓
TubiTV	
Twitch	✓
Vevo	
Viewster	✓
Vimeo	
YouTube	

Table 14 - iPhone Base64 Leak Overview	
App Name	**Base64 Found**
Dailymotion	✓
Maxdome	
Netflix	
Popcornflix	✓
Twitch	✓
Vevo	
Viewster	✓
YouTube	

A rubric was created to provide enhanced organization of the collected data flows that included third party sensitive information leaks (*Table 15*). The rubric provides the state of data leaks to third parties from each app and on each platform. More specifically, the rubric provides a visual representation of what type of data each app leaked to third parties and if the leaked data represents PII, location, behavior, or device information, providing an efficient way to compare the data leaks of multiple apps. The rubric can also be used to isolate leaks to specific types of remote hosts, such as ad networks and their associated ad hosts (Section 4.2).

Additionally, Wireshark logs from the Wi-Fi AP method were analyzed finding no substantial implementation of additional protocols outside of HTTP that were used to leak data from the apps.

Table 15 - App Data Leak Rubric

	Dailymotion	Maxdome	Netflix	Popcornflix	TubiTV	Twitch	Vevo	Viewster	Vimeo	YouTube	
Name							B				*PII*
Email / Nick						B	B	B			*PII*
DOB							B				*PII*
Gender		B					B				*PII*
City	B			B							*PII*
Password								B			*PII*
Location				A	A			A			*Location*
Video Title	B			B		B	B	B			*Behavior*
Text Search	A			B	A		B	B	A		*Behavior*
ISP IP	B			I			A	B			*Device*
Local IP				B			A	B			*Device*
IMEI								A			*Device*
Serial #							A	A			*Device*
MAC					A						*Device*
Wi-Fi SSID				I							*Device*

A = Third party leak on Android; I = Third party leak on iPhone; B = Third party leak, both platforms

Through an interdisciplinary analysis, captured data leaks to third parties, such as email address or name, were compared to explicitly stated data collection language in the privacy policies, underscoring instances where the app was detected leaking a data type not mentioned in the policy. The result of the analysis revealed 6 apps leak potentially sensitive data to third parties when the associated app vendor privacy policy did not explicitly state the same type of data was collected from the user (*Table 16*). Between the two platforms, Android apps leak information across more categories not included in vendor privacy policies than iPhone apps.

Table 16 - Detected Leaks of Data Types to Third Parties Not Identified in Privacy Policies

App Provider	Name	Email	Demog	Interest	Behavior	GPS	Locale	IP Address	Device Info
Dailymotion							A		
Maxdome									
Netflix									
Popcornflix						A	B		I
TubiTV									A
Twitch					B				
Vevo									
Viewster						A			A
Vimeo					A				
YouTube									
	PII			**Behavior**		**Location**		**Device**	

**** Demog = Demographic information such as gender, age, or race***
A = Third party leak on Android; I = Third party leak on iPhone; B = Third party leak, both platforms

The following is a summary of areas of concern observed between leaks and privacy policies:

1) Dailymotion's policy does not include information about search queries. User search terms were sent to: ad4.liverail.com, ads.nexage.com, and ssl.google-analytics.com

2) Popcornflix does not include location or GPS collection information in their policy. The Android app sent full GPS coordinates to: ads.mopub.com and mopub-east3-

bidder.manage.com

3) Popcornflix does not include information about search queries in their policy. User search queries were sent to ssl.google-analytics.com on both platforms.

4) TubiTV (only on Android) sends the device's MAC address to event.yozio.com which is not mentioned in the policy.

5) TubiTV does not include information about search queries in their policy. User search queries were sent to cms.adrise.tv.

6) Viewster does not include location or GPS collection information in their policy but sent full GPS coordinates to dev.appboy.com on Android.

7) Viewster sent the device serial number dev.appboy.com and the IMEI to track.appsflyer.com on Android which is not mentioned in the policy.

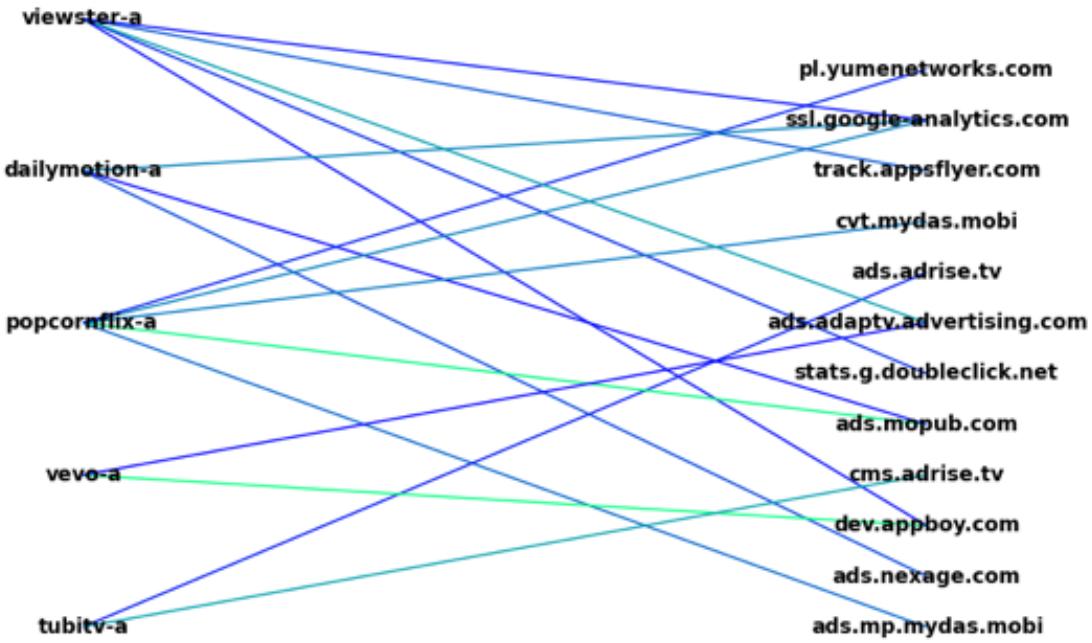
4.2 Ad Networks

Analyzing host entries from the MITM proxy method revealed the apps connected to 28 different ad networks (*Appendix C*). "Ad networks" are defined in this thesis as a third party advertising firm that runs a network of advertising across mobile platforms and apps. Each of the ad networks are constructed of one or more "ad hosts" representing the remote servers that apps communicate with (*Appendix D*). The following summarizes the analysis of data flows involving ad networks and their ad hosts:

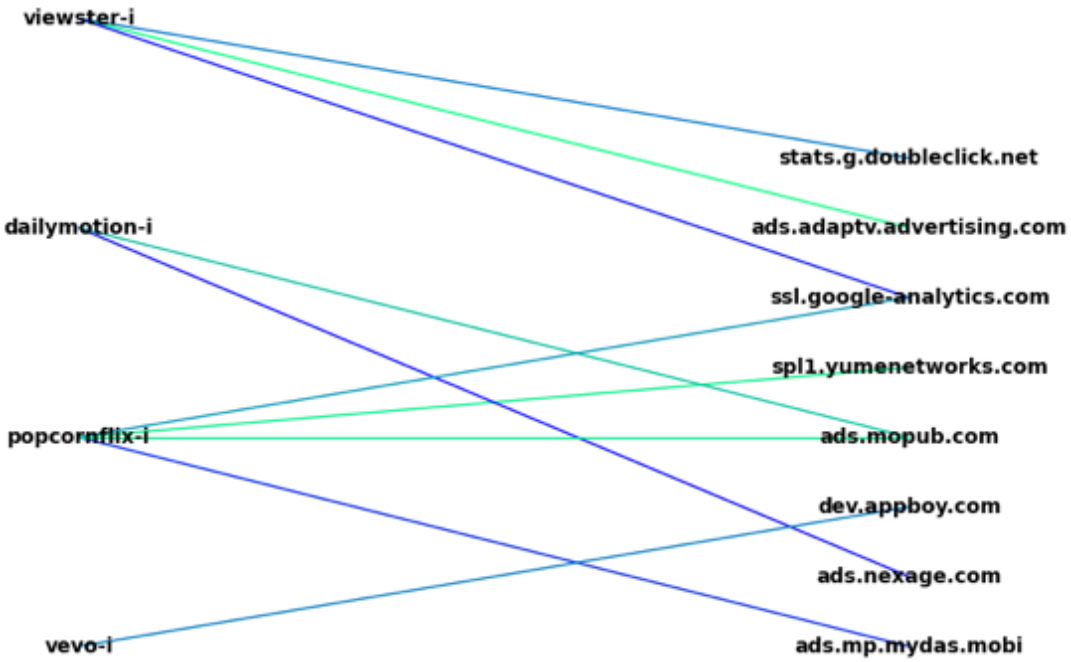
1) All apps except Netflix on Android and Maxdome on iPhone connected to at least 1 ad network and half of all apps tested shared data with third party ad networks (*Graphs 3, 4*).

2) Viewster topped the Android apps leaking to 5 different ad networks and Popcornflix topped the iPhone apps leaking to 4 different ad networks. (*Graphs 3, 4*).

Graph 3 - Android Data Leaks to Ad Networks by App



Graph 4 - iPhone Data Leaks to Ad Networks by App



3) On Android, Dailymotion topped all other apps by connecting to 17 ad networks, and on iPhone, Dailymotion and Viewster both connected to 14 ad networks. The amount of connections includes data flows not containing leaks and is a "total connection" statistic.

4) At a host level on Android, Dailymotion connected to 28 ad hosts and on iPhone, Popcornflix topped all other apps connecting to 38 ad hosts. The amount of connections includes data flows not containing leaks and is a "total connection" statistic.

5) When the app is offered on both platforms, Android apps leak to more ad networks than iPhone apps (*Table 17).*

Table 17 - App Data Leak Rubric Applied to Ad Networks

	Dailymotion	Maxdome	Netflix	Popcornflix	TubiTV	Twitch	Vevo	Viewster	Vimeo	YouTube	
Name							B				*PII*
Email / Nick							B	A			*PII*
DOB							B				*PII*
Gender							B				*PII*
City	B			B							*PII*
Password											*PII*
Location				A	A			A			*Location*
Video Title				B			A	B			*Behavior*
Text Search	A			B	A			B			*Behavior*
ISP IP	B						A	B			*Device*
Local IP				B							*Device*
IMEI								A			*Device*
Serial #							A	A			*Device*
MAC											*Device*
Wi-Fi SSID											*Device*

A = Third party leak on Android; I = Third party leak on iPhone; B = Third party leak, both platforms

4.3 Other Findings

The Wi-Fi network the test smartphones interacted with also hosted a Roku 2 streaming video appliance, capable of running many of the apps tested under this thesis. The Popcornflix app running on iPhone discovered the Roku on the Wi-Fi network and made numerous HTTP GET requests to XML pages running on port 8060 (on the Roku). The HTTP response from the

Roku to the Popcornflix app included device-specific information such as the full Roku model, a unique identifier, and the full Roku serial number printed on the bottom of the external device casing. The app combined the data collected from the Roku with information about the iPhone running the Popcornflix app, the Wi-Fi SSID, and the public IP assigned to the network by the ISP into a Base64 encoded stream. The encoded data was passed to a HTTP GET request, sending it insecurely (non-SSL) to a third party (*Figure 2*).

{"event":"DEVICE_DISCOVERED","properties":{"WIFI_SSID":"gnet","EXTERNAL_IP_ADDRESS":"37.201.225.30","distinct_id":"roku:12g34e026692","SCREEN_MODEL_NUMBER":"3050X","SCREEN_MODEL_NAME":"Roku 2 XD","WIFI_CONNECTED":true,"SCREEN_FRIENDLY_NAME":"Roku 2 XD","REMOTE_DEVICE_MAKE":"Apple","REMOTE_DEVICE_ID":"IPHONE:66B24918-B706-406E-8E36-088A6C0CA57C","REMOTE_DEVICE_MODEL":"iPhone 4S","SCREEN_MANUFACTURER":"Roku","SCREEN_ALLOWED_STATUS":"ALLOWED","SCREEN_SERIAL_NUMBER":"12G34E026692","REMOTE_DEVICE_TYPE":"iOS","token":"71eab1c776d8dfe7b4ebf881ae2bd780","APP_ID":"vzb2015616","SCREEN_DEVICE_VERSION":"UNKNOWN"}}

Figure 2 - Decoded Base64 Data Sent to Remote Host

As viewers transition video plays across multiple devices, advertisers are looking for ways to implement cross-device user tracking and cross-device advertising.[29] In the future, we may begin seeing apps becoming more aware of devices on the network, building a larger profile of the user and opening up new avenues for tracking and privacy concern.

During app testing, Netflix and Vimeo on the iPhone were unable to properly function, producing errors preventing video playback while connected to the MITM proxy. The app vendors may have opted to implement increased certificate validation using SSL pinning. SSL pinning provides a means for application developers to bundle hard-coded copies of trusted SSL certificates with the application before market release.[30] For example, if the app routinely needs

[29] *A. Schiff, "A Marketer's Guide To Cross-Device Identity | AdExchanger", AdExchanger, 2015. Available: http://adexchanger.com/data-exchanges/a-marketers-guide-to-cross-device-identity/. [Accessed: 29- Mar- 2016].*

[30] *V. Moonsamy and L. Batten, "Mitigating man-in-the-middle attacks on smartphones - A discussion of SSL pinning and DNSSec", in 12th Australian Information Security Management Conference, Edith Cowan University, Joondalup Campus, Perth, Western Australia, 2014.*

to connect to server A, the developer can bundle a copy of server A's SSL certificate with the application in a local trust store. When the application needs to connect to server A, a comparison between the received certificate from server A and the bundled certificate in the app trust store will reveal if the certificate is indeed the expected certificate. SSL pinning provides a way to mitigate many forms of MITM attacks, protecting the app and user from trusting a malicious, unexpected certificate placed in the device's local trust store. Apps that utilize SSL pinning take a positive step toward protecting a user's private information by making it harder for basic MITM attacks to take place.

4.4 Discussion

The analysis of collected HTTP data flows from the video apps reveal that nearly all of the apps sent potentially sensitive personal information to third parties, and many of the apps sent the data in an insecure manner. Additionally, an analysis of the verified data leaks against the app vendor privacy policies revealed more than half of the apps leaked certain types of data to third parties that was not mentioned as collected from the user in the policy. These findings confirm the apps can implicate a users' privacy posture by leaking private data to third parties and that a lack of complete data sharing transparency, or at the very least the inaccuracy of the provided privacy policies, introduces complexity for any user trying to evaluate how an app may affect their privacy prior to installation and use. If the lack of information and transparency troubles of the privacy policies are combined with unnecessary length, complex language, and lack of a standardized design, the app user will undoubtedly experience trouble trying to assess the apps baseline of privacy assurance. The resulting situation leaves the average user with an unclear view of how the vendor and app affect their privacy posture, while the apps leak

potentially sensitive data to third parties behind the scenes and unbeknownst to the user.

The collection of certain personal information, such as a user's name or email address, is often necessary for legitimate purposes such as new account registration. Sharing this information with third parties, however, can be harder to justify. For example, the collected HTTP data flows revealed some of the apps shared full GPS coordinates to third party ad networks such as mopub.com, adrise.tv, and appboy.com. Although advertisers may utilize GPS information to provide location-centric ad targeting, frequent collection of GPS coordinates, especially when combined with other unique identifiers, can easily transform into a user tracking mechanism negatively impacting a user's privacy.

As smartphones offer entirely new sources of data not previously available on traditional television platforms, advertisers will continue the pivot to mobile video platforms. Of the 28 ad networks found interacting with the tested apps, Adrise went a step further by releasing their own streaming video app, TubiTV, offering free movies and television episodes using an ad-based revenue model. If successful, Adrise may represent the future of ad-supported streaming video apps, joining them directly with the ad networks and removing any other layers of data collection oversight. On the other hand, streaming video providers utilizing a subscription-based revenue model, such as Netflix and Maxdome, shared significantly less information with third parties. Although these providers could share information they collect outside of the app, providing enhanced ad targeting through user data collection does not represent their primary revenue stream reported to shareholders.[31] While free ad-supported video apps cost the user less upfront, the underlying cost of user privacy from data leaks to third parties should be carefully evaluated.

[31] *"Q4 15 Letter to Shareholders", Netflix - Quarterly Earnings, 2016. Available: http://files.shareholder.com/downloads/NFLX/1793452058x0x870685/C6213FF9-5498-4084-A0FF-74363CEE35A1/Q4_15_Letter_to_Shareholders_-_COMBINED.pdf. [Accessed: 30- Mar- 2016].*

Chapter 5

Conclusion and Future Work

5.0 Video App User Privacy

Nearly all of the video apps tested in this study leaked potentially sensitive information about the user to third parties, often in an insecure manner, creating the potential for user privacy implications. Even with legitimate reasons for sharing such information with "partner" third parties, users should be aware that these streaming video apps collect information about their interests, their email address and location, and send it to other parties outside the app vendors control. Once this information leaves the app, there is no way to control how the third party will utilize or further share the user's information.

The streaming video app privacy policies evaluated by this thesis, often acting as the only insight into how the app may handle a user's privacy, are complex, unstandardized, and lack the transparency and clarity needed for a user to determine an expected baseline of privacy offered by the vendor and their app. Furthermore, many of the apps are misaligned with their associated privacy policies, sharing information with third parties not conveyed to the user as information collected by the vendor through the app.

Many of the streaming video apps collected device specific information that can easily be used as a unique identifier to track the user across other apps or platforms. A smaller subset of the apps also collected precise GPS coordinates, enabling tracking of the user by their geographic location. The associated app privacy policies did not include a clear explanation of why the location information is collected, how it is used, and who it is shared with leaving the user uncertain when it comes to their privacy.

5.1 Role of Advertisers

As viewers pivot their video playback time to mobile devices such as smartphones, advertisers will naturally expand from their traditional television markets in order to access this growing online and mobile consumer segment. Smartphones offer advertisers significantly more tracking options and profiling information than traditional cable and satellite platforms. Advertisers can gain deeper insights into user interests through other on-device app affiliations, obtain the user email address enabling tracking across multiple apps, and receive real-time GPS location information.

Non-provider ad networks are also motivated to collect more information from users to construct in depth user profiles, enabling enhanced ad targeting and a positive influence on the bottom line. As smartphone ads become more "useful" or "effective" for both the user and ad customer, the bid price in the ad auction may increase, resulting in additional revenue, supporting the advertising-based business model. From an advertiser perspective, expanded data collection may result in a method to achieve a deeper customer awareness and yield greater ad click-through ratios, resulting in increased revenue and a renewed motivation for data collection.

Apps such as Popcornflix, which can discover other video-capable connected devices such as a Roku, may allow advertisers to track user interests across multiple platforms leading to additional avenues of ad impressions and additional sources to collect user data. Although this may refine ad relevance and improve the overall ad experience, cross-device ad targeting can also allow for cross-device sensitive information leaks and tracking.

5.2 Limitations

The MITM proxy method allowed for decryption of secure HTTP (SSL) communications

and data encoded by Base64 was decoded and checked for data leaks. The data flow analysis did not check for encoding schemas other than Base64, resulting in the possibility that apps could have leaked data utilizing different encoding. However, the data flows analyzed revealed most apps, including those implementing Base64 encoding on select HTTP flows, leaked data to third parties across other data flows without utilizing encoding.

5.3 Recommendations

The thesis findings generate the following recommendations to industry and relevant policy makers:

1) App vendors should simplify and streamline privacy policies, removing any complex language, making them accessible to the average user. The policies evaluated in this thesis had SMOG scores significantly higher than the average recommendation.

2) The industry should create a standard template for privacy policies. All of the policies evaluated in this thesis "looked" different and utilized vastly different language without covering the same topics and concerns. Alternatively, Google and Apple, should insist on a standardized privacy policy template as part of the management of their mobile app platforms.

3) App vendors should include more language about advertisers in privacy policies, explicitly detailing to the user how potentially sensitive information is being shared with advertisers. Additionally, the policies should detail which ad networks are utilized by the app so that a user may conduct their own outside research into ad network privacy implications if the policy still falls short.

4) Apps should utilize SSL encryption on all HTTP data flows as a security mechanism to help protect user privacy.

5) Apps should implement SSL pinning to prevent MITM attacks on the app, which can adversely affect user privacy.

5.4 Future Work

Although this thesis produced meaningful insights into streaming video app privacy, further studies can expand the usefulness of this research by testing a larger group of apps. The Google Play App Store and Apple App Store offer additional video streaming apps that were not included in the pool of apps tested in this thesis.

Future studies might also investigate streaming video apps for encoding mechanisms other than standard Base64, used to obfuscate data sent out of the app. A recent study at the University of Toronto's Citizen Lab found a popular Chinese app, Baidu, leaking potentially sensitive data utilizing weak encryption and encoding mechanisms in conjunction with Base64.[32]

The apps covered in this study are not associated with traditional television broadcast networks, such as CBS, NBC, or ABC, due to content licensing restrictions in Europe. Study of the streaming video apps offered by these broadcast networks may provide further insight into the transition of traditional television advertising into online environments and how data collection and leaks differ from non-broadcast affiliated apps.

[32] *J. Knockel, S. McKune and A. Senft, "Baidu's and Don'ts: Privacy and Security Issues in Baidu Browser", The Citizen Lab - University of Toronto, 2016. Available: https://citizenlab.org/2016/02/privacy-security-issues-baidu-browser/. [Accessed: 31- Mar- 2016].*

BIBLIOGRAPHY

A. Seneviratne, K. Thilakarathna, S. Seneviratne, M. Kaafar and P. Mohapatra, "Reconciling bitter rivals: Towards privacy-aware and bandwidth efficient mobile Ads delivery networks", 2013 Fifth International Conference on Communication Systems and Networks (COMSNETS), 2013.

A. Schiff, "A Marketer's Guide To Cross-Device Identity | AdExchanger", AdExchanger, 2015. Available: http://adexchanger.com/data-exchanges/a-marketers-guide-to-cross-device-identity/. [Accessed: 29- Mar- 2016].

A. Sunyaev, T. Dehling, P. Taylor and K. Mandl, "Availability and quality of mobile health app privacy policies", Journal of the American Medical Informatics Association, 2014.

Docs.mitmproxy.org, "Introduction - Mitmproxy 0.15 documentation", 2016. Available: http://docs.mitmproxy.org/en/stable/. [Accessed: 10- Feb- 2016].

D. Solove, 'A Taxonomy of Privacy', University of Pennsylvania Law Review, vol. 154, no. 3, p. 477, 2006.

E. Dwoskin, "Apps Track Users—Once Every 3 Minutes", WSJ, 2015. Available: http://www.wsj.com/articles/apps-track-usersonce-every-3-minutes-1427166955. [Accessed: 05- Apr- 2016].

H. McLaughlin, "SMOG Grading - A New Readability Formula", Journal of Reading, vol. 12, no. 8, pp. 639-646, 1969.

"hostapd: IEEE 802.11 AP, IEEE 802.1X/WPA/WPA2/EAP/RADIUS Authenticator", W1.fi, 2016. Available: https://w1.fi/hostapd/. [Accessed: 16- Mar- 2016].

I. Dacosta, M. Ahamad and P. Traynor, "Trust No One Else: Detecting MITM Attacks against SSL/TLS without Third-Parties", Computer Security – ESORICS 2012, pp. 199-216, 2012.

Internet Society, "Global Internet Report 2015", 2015. Available: http://www.internetsociety.org/globalinternetreport/assets/download/IS_web.pdf. [Accessed: 1- Apr- 2016].

J. Graves, "An Exploratory Study of Mobile Application Privacy Policies", Technology Science, 2015. Available: http://techscience.org/a/2015103002. [Accessed: 1- Apr- 2016].

J. Knockel, S. McKune and A. Senft, "Baidu's and Don'ts: Privacy and Security Issues in Baidu Browser", The Citizen Lab - University of Toronto, 2016. Available: https://citizenlab.org/2016/02/privacy-security-issues-baidu-browser/. [Accessed: 31- Mar- 2016].

J. Zang, K. Dummit, J. Graves, P. Lisker and a. Sweeney, "Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps", Technology Science, 2015. Available: http://techscience.org/a/2015103001. [Accessed: 1- Apr- 2016].

M. Grace, W. Zhou, X. Jiang and A. Sadeghi, "Unsafe exposure analysis of mobile in-app advertisements", Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks - WISEC '12, 2012.

"Mobile Ad Spend to Top $100 Billion Worldwide in 2016, 51% of Digital Market - eMarketer", Emarketer.com, 2016. Available: http://www.emarketer.com/Article/Mobile-Ad-Spend-Top-100-Billion-Worldwide-2016-51-of-Digital-Market/1012299. [Accessed: 12- Mar- 2016].

"Mobile Privacy Disclosures: Building Trust Through Transparency: A Federal Trade Commission Staff Report", Ftc.gov, 2013. Available: https://www.ftc.gov/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission. [Accessed: 15- Mar- 2016].

On Device Research, "Global Consumer Trust Report 2016", 2016. Available: http://www.mobileecosystemforum.com/solutions/consumer-trust/global-consumer-trust-report-2016/. [Accessed: 1- Apr- 2016].

Ooyala, "Global Video Index Q2 2015", 2015. Available: http://go.ooyala.com/wf-video-index-q2-2015. [Accessed: 1- Apr- 2016].

Pew Research Center, "Home Broadband 2015", 2015. Available: http://www.pewinternet.org/files/2015/12/Broadband-adoption-full.pdf. [Accessed: 1- Apr- 2016].

P. Pateriya and S. Kumar, "Analysis on Man in the Middle Attack on SSL", International Journal of Computer Applications, vol. 45, no. 23, pp. 43-46, 2012.

“Q4 15 Letter to Shareholders", Netflix - Quarterly Earnings, 2016. Available: http://files.shareholder.com/downloads/NFLX/1793452058x0x870685/C6213FF9-5498-4084-A0FF-74363CEE35A1/Q4_15_Letter_to_Shareholders_-_COMBINED.pdf. [Accessed: 30- Mar- 2016].

R. Parker, 'A Definition of Privacy', Rutgers Law Review, vol. 27, no. 2, p. 275, 1974.

Sandvine, "Global Internet Phenomena Report", 2015. Available: https://www.sandvine.com/trends/global-internet-phenomena/. [Accessed: 1- Apr- 2016].

S. Blenner, M. Köllmer, A. Rouse, N. Daneshvar, C. Williams and L. Andrews, "Privacy Policies of Android Diabetes Apps and Sharing of Health Information", JAMA, vol. 315, no. 10, p. 1051, 2016.

S. Perez, "iOS 7 Eliminates MAC Address As Tracking Option, Signaling Final Push Towards Apple's Own Ad Identifier Technology", TechCrunch, 2013. Available: http://techcrunch.com/2013/06/14/ios-7-eliminates-mac-address-as-tracking-option-signaling-final-push-towards-apples-own-ad-identifier-technology/. [Accessed: 05- Apr- 2016].

S. Warren and L. Brandeis, 'The Right to Privacy', Harvard Law Review, vol. 4, no. 5, p. 193, 1890.

"Technical Analysis of the Data Practices and Privacy Risks of 43 Popular Mobile Health and Fitness Applications", Privacy Rights Clearinghouse, 2013. Available: https://www.privacyrights.org/mobile-medical-apps-privacy-technologist-research-report.pdf. [Accessed: 1- Apr- 2016].

V. Moonsamy and L. Batten, "Mitigating man-in-the-middle attacks on smartphones - A discussion of SSL pinning and DNSSec", in 12th Australian Information Security Management Conference, Edith Cowan University, Joondalup Campus, Perth, Western Australia, 2014.

"Wireshark User's Guide", Wireshark.org, 2016. Available: https://www.wireshark.org/docs/wsug_html_chunked/. [Accessed: 16- Mar- 2016].

W. Meng, R. Ding, S. Chung, S. Han and W. Lee, "The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads", Georgia Institute of Technology, 2016. Available: http://www.cc.gatech.edu/~wmeng6/ndss16_mobile_ad.pdf. [Accessed: 15- Mar- 2016].

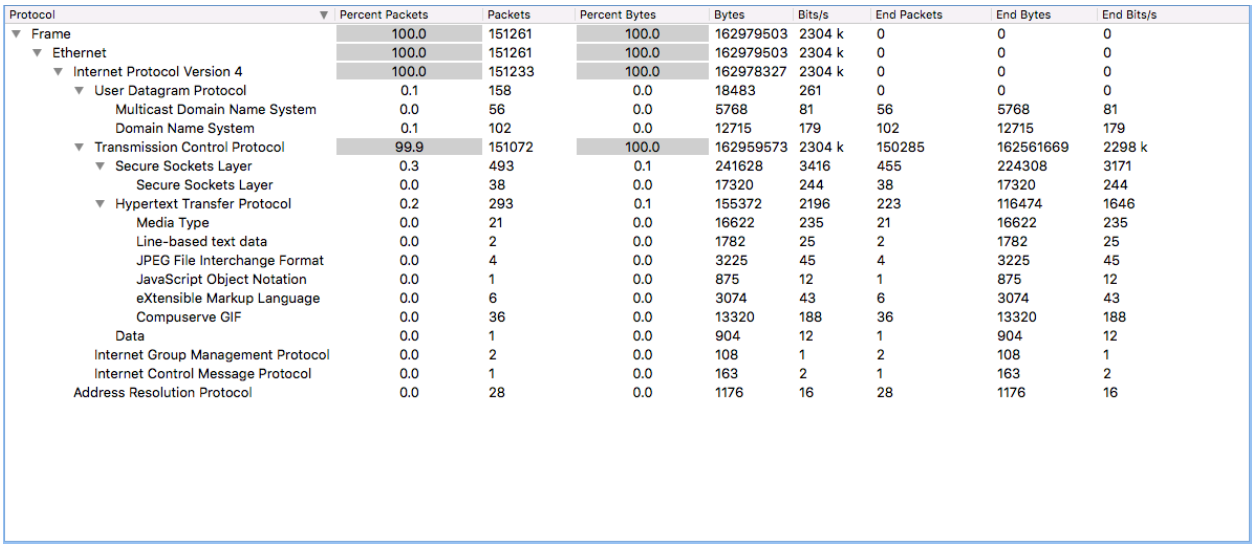
APPENDIX A - KEYWORD LIST

Keyword List for Android and iOS			
ID	Designation	Type of Data	Keyword
1	PII	Name	Guza
2	PII	Address	Pempelforter
3	PII	Address	Melrose
4	PII	Email Address / Nickname	mtgandroid01@gmail.com
5	PII	Email Address / Nickname	mtgiphone01@gmail.com
6	PII	Email Address / Nickname	mtgandroid01
7	PII	Email Address / Nickname	mtgiphone01
8	PII	Birthdate	01011980
9	PII	Birthdate	01-01-1980
10	PII	Birthdate	1-1-1980
11	PII	Birthdate	1-Jan-1980
12	PII	Birthdate	1 January 1980
13	PII	Birthdate	1 Jan 1980
14	PII	City	Dusseldorf
15	PII	City	Boardman
16	PII	State	Ohio
17	PII	Gender	Male
18	PII	Gender	male
19	PII	Gender	MALE
20	PII	Zipcode	40211
21	PII	Zipcode	44512
22	PII	Password	Pass1238
23	PII	Payment	PayPal
24	PII	Payment	paypal
25	Location	Testing Location	51.22
26	Location	Testing Location	6.79
27	Location	Testing Location	latitude
28	Location	Testing Location	longitude
29	Behavior	Text Search	fun
30	Behavior	Audio Search	test one two
31	Behavior	PIN	1234
32	Behavior	Interest	Big Bang Theory
33	Behavior	Interest	House of Cards
34	Behavior	Interest / Video	Family Guy
35	Behavior	Video	How I Met Your Mother
36	Behavior	Video	Call of Duty
37	Behavior	Video	Admiralbulldog
38	Behavior	Video	Tamah

39	Behavior	Video	Between Times
40	Behavior	Video	Deadpool
41	Behavior	Video	Spectre
42	Behavior	Video	Danny Daze
43	Behavior	Video	Jaden Smith
44	Behavior	Video	Little Mix
45	Behavior	Video	Olly Murs
46	Behavior	Video	The Comedian
47	Behavior	Video	Julien and Claire
48	Behavior	Video	Angel on my Shoulder
49	Behavior	Video	Virus
50	Behavior	Video	Zoo
51	Behavior	Video	Atlas
52	Behavior	Video	After the Rain
53	Behavior	Video	Future Report
54	Device	Wi-Fi SSID	gnet
55	Device	Wi-Fi SSID	NONO
56	Device	ISP IP	37.201
57	Device	Local IP	10.20.30
58	Device	MAC Address - Android	84:7
59	Device	MAC Address - iPhone	6:31
60	Device	IMEI - Android	356482050
61	Device	IMEI - iPhone	0129400020
62	Device	ICCID - iPhone	89492099147
63	Device	Serial Number - Android	HT387WA
64	Device	Serial Number - iPhone	C39GGLR
65	Device	OS Version - iPhone	13D15

Key	
Color	**Meaning**
	Identified as PII by NIST SP 800-122
	Designated as PII by other sources
	Behavior information
	Information that might be sensitive
Color	**Meaning**
	User-supplied during app registration (if applicable)
	Location data or Internet / network connectivity information
	OS / Hardware manufacturer information

APPENDIX B - WIRESHARK PROTOCOL OVERVIEW EXAMPLE



Protocol	Percent Packets	Packets	Percent Bytes	Bytes	Bits/s	End Packets	End Bytes	End Bits/s
Frame	100.0	151261	100.0	162979503	2304 k	0	0	0
Ethernet	100.0	151261	100.0	162979503	2304 k	0	0	0
Internet Protocol Version 4	100.0	151233	100.0	162978327	2304 k	0	0	0
User Datagram Protocol	0.1	158	0.0	18483	261	0	0	0
Multicast Domain Name System	0.0	56	0.0	5768	81	56	5768	81
Domain Name System	0.1	102	0.0	12715	179	102	12715	179
Transmission Control Protocol	99.9	151072	100.0	162959573	2304 k	150285	162561669	2298 k
Secure Sockets Layer	0.3	493	0.1	241628	3416	455	224308	3171
Secure Sockets Layer	0.0	38	0.0	17320	244	38	17320	244
Hypertext Transfer Protocol	0.2	293	0.1	155372	2196	223	116474	1646
Media Type	0.0	21	0.0	16622	235	21	16622	235
Line-based text data	0.0	2	0.0	1782	25	2	1782	25
JPEG File Interchange Format	0.0	4	0.0	3225	45	4	3225	45
JavaScript Object Notation	0.0	1	0.0	875	12	1	875	12
eXtensible Markup Language	0.0	6	0.0	3074	43	6	3074	43
Compuserve GIF	0.0	36	0.0	13320	188	36	13320	188
Data	0.0	1	0.0	904	12	1	904	12
Internet Group Management Protocol	0.0	2	0.0	108	1	2	108	1
Internet Control Message Protocol	0.0	1	0.0	163	2	1	163	2
Address Resolution Protocol	0.0	28	0.0	1176	16	28	1176	16

APPENDIX C - AD NETWORKS OBSERVED

71i.de
ad-score.com
adform.com
adform.net
adgoji.com
adition.com
adnxs.com
adrise.com
adrise.tv
adscale.de
advertising.com
appboy.com
appsflyer.com
doubleclick.net
eyereturn.com
eyeviewads.com
fiksu.com
google-analytics.com
googleadservices.com
liftoff.io
millennialmedia.com
mopub.com
mydas.mobi
nexage.com
pxlad.io
smartadserver.com
yume.com
yumenetworks.com

APPENDIX D - AD NETWORK AND AD HOST STATS

Android App Ad Net Conns	
App Name	**Number of Networks**
Dailymotion	17
Maxdome	1
Netflix	0
Popcornflix	6
TubiTV	4
Twitch	3
Vevo	5
Viewster	8
Vimeo	2
YouTube	3

iPhone App Ad Net Conns	
App Name	**Number of Networks**
Dailymotion	14
Maxdome	0
Netflix	2
Popcornflix	11
Twitch	4
Vevo	3
Viewster	14
YouTube	3

Android Ad Net Hosts	
App Name	**Number of Hosts**
Dailymotion	28
Maxdome	1
Netflix	0
Popcornflix	14
TubiTV	7
Twitch	4
Vevo	7
Viewster	13
Vimeo	4
YouTube	6

iPhone Ad Net Hosts	
App Name	**Number of Hosts**
Dailymotion	25
Maxdome	0
Netflix	5
Popcornflix	38
Twitch	5
Vevo	3
Viewster	20
YouTube	5

Android Ad Net Leaks	
App Name	**Number of Networks**
Dailymotion	3
Maxdome	0
Netflix	0
Popcornflix	4
TubiTV	1
Twitch	0
Vevo	2
Viewster	5
Vimeo	0
YouTube	0

iPhone Ad Net Leaks	
App Name	**Number of Networks**
Dailymotion	2
Maxdome	0
Netflix	0
Popcornflix	4
Twitch	0
Vevo	1
Viewster	3
YouTube	0