2016-10-01

The State of Man-in-the-Middle TLS Proxies: Prevalence and User Attitudes

Mark Thomas ONeill
Brigham Young University

The State of Man-in-the-Middle TLS Proxies:

Prevalence and User Attitudes

Mark Thomas O'Neill

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Daniel Zappala, Chair
Kent Seamons
Christophe Giraud-Carrier

Department of Computer Science

Brigham Young University

ABSTRACT

The State of Man-in-the-Middle TLS Proxies:
Prevalence and User Attitudes

Mark Thomas O'Neill
Department of Computer Science, BYU
Master of Science

We measure the prevalence and uses of Man-in-the-Middle TLS proxies using a Flash tool deployed with a Google AdWords campaign. We generate 15.2 million certificate tests across two large-scale measurement studies and find that 1 in 250 TLS connections are intercepted by proxies. The majority of these proxies appear to be benevolent, however we identify over 3,600 cases where eight malware products are using this technology nefariously. We also find thousands of instances of negligent, duplicitous, and suspicious behavior, some of which degrade security for users without their knowledge. Distinguishing these types of practices is challenging in practice, indicating a need for transparency and user awareness.

We also report the results of a survey of 1,976 individuals regarding their opinions of TLS proxies. Responses indicate that participants hold nuanced opinions on security and privacy trade-offs, with most recognizing legitimate uses for the practice, but also concerned about threats from hackers or government surveillance. There is strong support for notification and consent when a system is intercepting their encrypted traffic, although this support varies depending on the situation. A significant concern about malicious uses of TLS inspection is identity theft, and many would react negatively and some would change their behavior if they discovered inspection occurring without their knowledge. We also find that a small but significant number of participants are jaded by the current state of affairs and have lost any expectation of privacy.

Keywords: SSL, TLS, Proxy, MITM, man in the middle, measurement, survey, AdWords, security, malware, firewall, censorship

ACKNOWLEDGMENTS

Thanks to my wife, Leah, for taking care of me during the long, sleepless nights and offering encouragement in rough times. Thanks to my son, Veritas, for deleting lines of code and injecting characters into my papers while I pursued my degree. Thanks to Dr. Zappala for giving me the freedom to explore what I wanted. Thanks to Dr. Seamons for guiding my technical learning. Thanks to Dr. Giraud-Carrier, for forcing me to finish this.

Thanks to Scott Ruoti for his collaborative work in designing and assessing the results of our two surveys. Thanks to Rich Shay for providing feedback on the wording of questions in our first survey. Thanks to Alexander Lemon, JJ Lowe, Brent Roberts, and Justin Wu for help with coding the survey data.

This work was supported by a 2014 Google Faculty Research Award, Sandia National Laboratories, and the National Science Foundation under Grant No. CNS-1528022. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction

Transport Layer Security (TLS) [60], is the most popular security protocol used on the Internet today.[1] Many network applications leverage TLS, including email clients, VPN clients, instant messaging services, and all web browsers. TLS, when used correctly, provides a variety of security guarantees to a connection between a client and server, including confidentiality, integrity, and authentication. Throughout the last decade, these pieces have proven to be relatively resilient, requiring only minor patches to address weaknesses. However, the authentication portion of TLS has serious flaws that subject the whole protocol to one of the very attacks it was meant to prevent: Man-in-the-Middle (MITM) attacks.

1.1 Background

Normally TLS authentication is performed by only one of the parties in the communication, although mutual authentication is supported by the protocol. A typical scenario is shown

[1]Prior versions of TLS were known as Secure Socket Layer (SSL). Unless otherwise specified, TLS will be used in this thesis to mean both TLS and SSL.

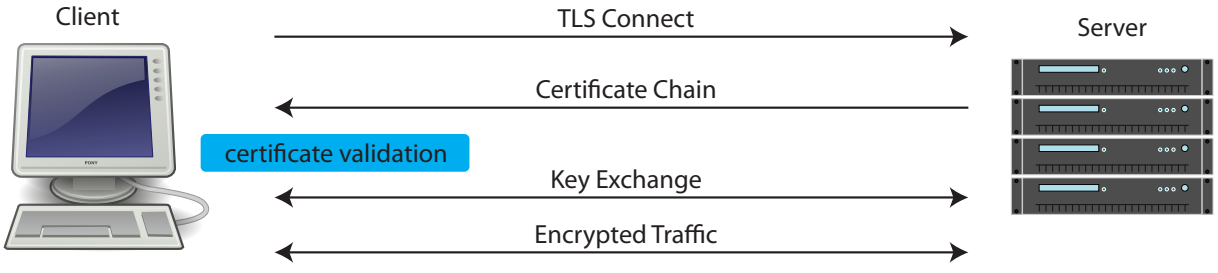

Figure 1.1: High-level example of server authentication under TLS

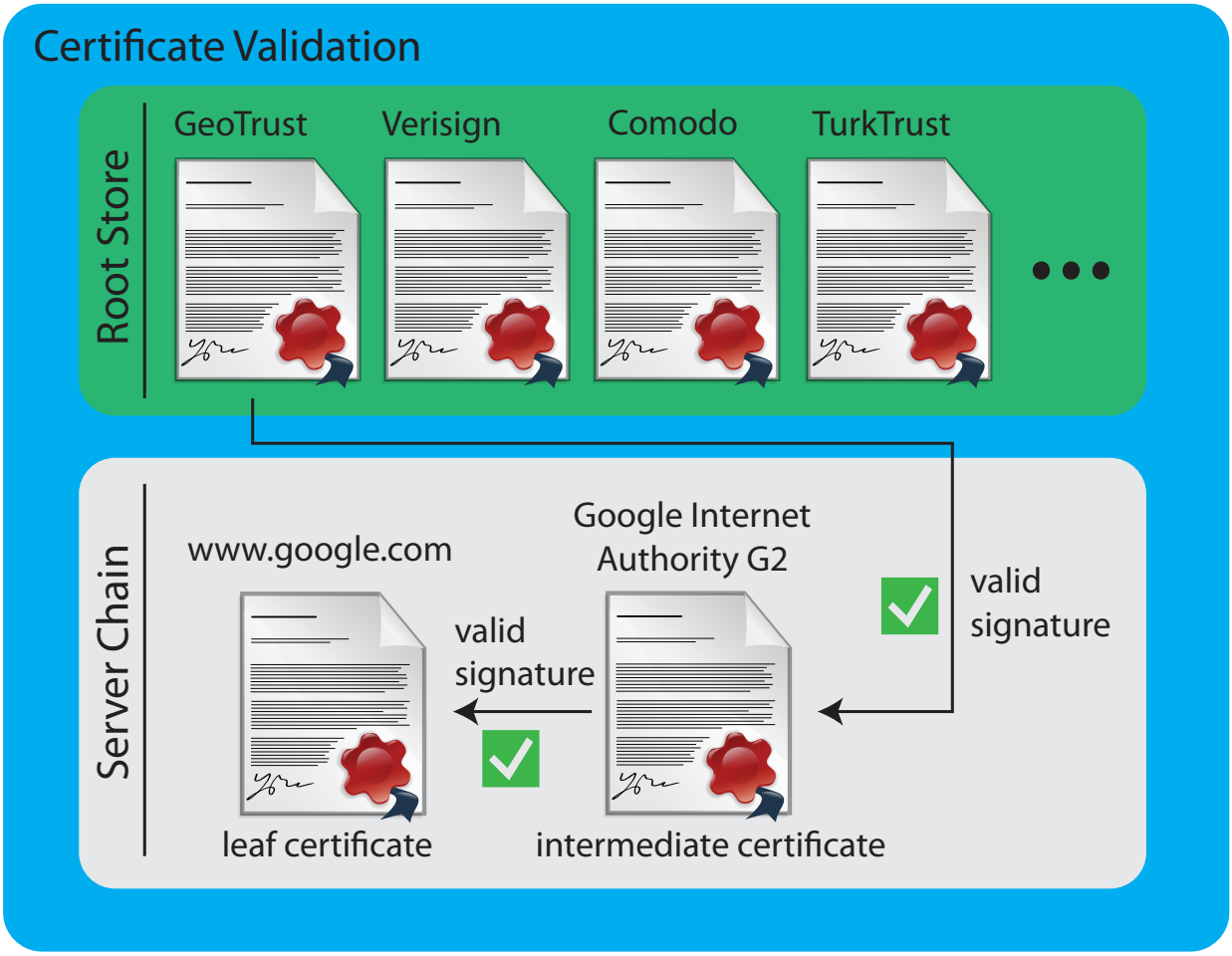

Figure 1.2: Example of signature chain validation for TLS authentication

in Figure 1.1. First, a client (e.g., browser) connects to an intended remote host (e.g., `www.google.com`). Before any data is sent from the server it will send a chain of digital certificates to the client, which vouches for the authenticity of the server. After validation and acceptance of the certificate chain, the client is assured that the host public key found within the certificate chain, is in fact the proper public key for intended host. The client then encrypts some preliminary data used to bootstrap the encryption tunnel using that public key. The server proves that it owns the corresponding private key by successful decryption of that data, and subsequent response. All data sent between the two endpoints is then encrypted.

To validate a received certificate, the client must perform a series of checks. Among these is ensuring that the certificate has a valid status (e.g., is not expired, is not revoked,

etc.), ensuring the certificate is issued to the hostname to which the client expected to connect (e.g., making sure the certificate bears the name amazon.com), and validating that the certificate has a proper signature from a Certificate Authority (CA).

This last step—validating the signature in the certificate—is shown in Figure 1.2. Each certificate in the chain must be digitally signed by the succeeding certificate in the chain, and it is up to the client to validate these signatures. Often, intermediate certificates are presented by the server in addition to the one it had issued to it from a signing authority. The client ensures that each certificate in the chain is properly signed by private key of the next certificate in the chain, and, finally, that the last certificate in the chain is signed by the private key of a root CA. For this last signature check, the client utilizes a locally-installed root store of trusted CAs. In this example, the leaf certificate for `www.google.com` is properly signed by the intermediate certificate owned by Google Internet Authority G2, which is in turn signed by the GeoTrust CA certificate, which resides in the root store of the browser. The root store is usually shipped with a browser or operating system. If the leaf certificate is to be trusted, its chain of signatures must link back to a CA from the root store. This constitutes the trust that the server is the website it claims to be. If this validation fails, the connection is aborted, as the absence of a valid chain indicates the possible presence of a MITM attacker. We call this authentication scheme the CA system.

Though the CA system may seem sufficient, in use it exhibits three alarming traits:

1. *Clients trust too many (untrustworthy) CAs:* A recent study found 1,832 browser-trusted signing certificates in use on a single day, controlled by 683 organizations[20]. Each of these CAs must follow best practices and be worthy of the implicit trust given to them. However, the greatness of their numbers make it prohibitive for concerned citizens to audit their behavior and many CAs have been compromised by hackers or found accidentally providing their private keys to the public [46]. In addition, there are reasons why some CAs should not be trusted. For example, some governments, such as

Russia and China, have their own CAs that are trusted by most user devices by default, presenting a conflict of interest between the user and CA in many situations.

2. *CAs can sign for any hostname:* All CAs are authorized to vouch for the authenticity of (sign a certificate for) any hostname (e.g., domain name, IP address) [19]. This includes domains for which certificates are already issued. Hosts do not have control over which entities are signing certificates for their name, nor are they notified when it takes place. Thus at any time there may be many different valid certificates signed by many different CAs for a single domain. A security breach of a single CA can result in the signing of forged certificates for any domain. Thus TLS authentication is only as strong as the weakest CA.

3. *Flawed implementations are proliferating:* The application logic to properly perform authentication using certificates is not trivial. The number of applications and libraries implementing secure connectivity via TLS has exploded in recent years, especially in the mobile space. Studies show that there are a number of implementation flaws that leave users vulnerable to man-in-the-middle attacks, e.g., [11, 24, 27]. In many cases these applications are failing to utilize TLS library calls appropriately, as there are many important misunderstood nuances regarding validation functions and a severe lack of sufficient documentation. In other cases application developers fail to even attempt certificate validation, or turn it off during development and unintentionally leave it off in production code [25].

These traits make it possible for attackers to perform a TLS MITM attack. This scenario is illustrated in Figure 1.3. In this case an attacker positions himself in between the client and server and poses as the legitimate server to the client. The client connects to the attacker instead of intended host, accepts a forged certificate chain generated by the attacker, and establishes an encrypted connection with the attacker. Optionally, the attacker also establishes a TLS connection with the original server as the client would have normally. In this fashion the attacker is free to read and modify any encrypted traffic being sent between

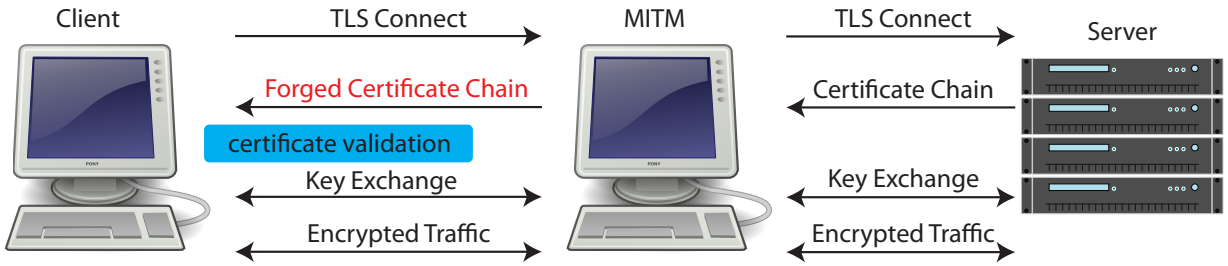

Figure 1.3: Example TLS Man-in-the-Middle

the client and server, decrypting and reincrypting traffic as it marshals data between those endpoints. Note that this behavior is not detectable by the client (or server), as the client has accepted the forged certificate chain.

Acceptance of the forged chain can occur for a variety of reasons, two of which are shown in Figure 1.4. In this case, the attacker delivers a substitute certificate for the intended host, `www.google.com`, which has been signed by a CA that he has hacked or by his own CA that he somehow injected into the root store of the victim machine. Since the certificate is otherwise valid and has a valid signature path to an entity in the root store, it is accepted, despite the fact that the certificate does not come from Google and that Google actually uses GeoTrust to sign its certificates.

These attacks are not just theoretical. For example, in 2011 when DigiNotar's servers were hacked and more than 500 certificates were fabricated by the intruder, including a certificate for Gmail that allowed the intruder to access stored email for 300,000 Iranians [39]. This happened despite the fact that Gmail does not use DigiNotar to sign its certificates. Other CAs have also been hacked, such as Comodo in 2011 and StartCom in 2016 [46, 64], which left dozens of trusted certificates in the hands of attackers. Note, however, that this is not needed if the attacker owns his own CA (e.g., a government), has otherwise managed to infect the root store of the victim machine, or if the targeted application does not perform correct certificate validation. Furthermore the CA system is further disrupted by the use of enterprise TLS inspection products, which add a CA to user root stores and MITM

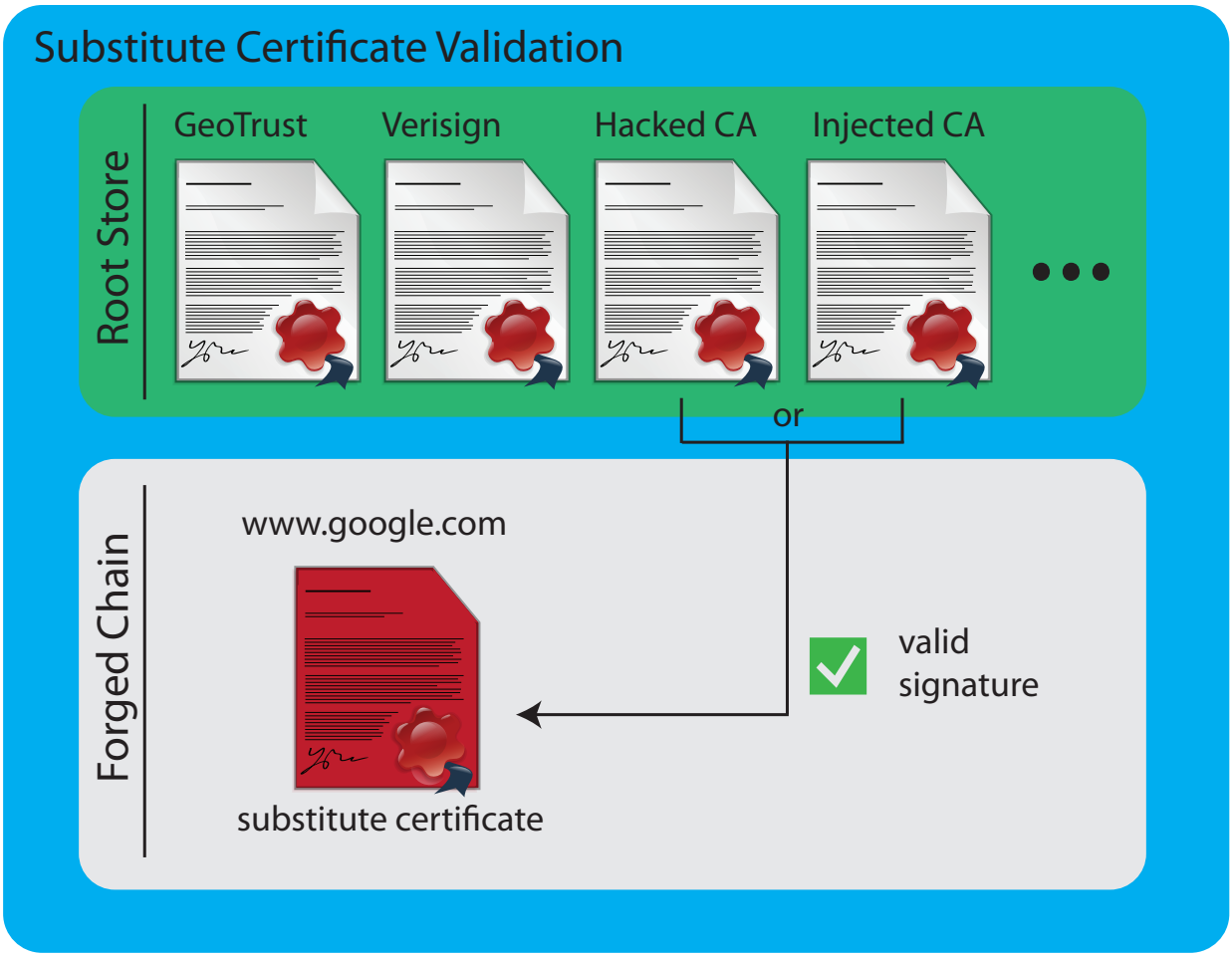

Figure 1.4: Example of substitute certificate acceptance by the client, due to hacked or injected CAs

TLS connections (e.g., [1, 10, 14, 54, 59]). Although these solutions are intended to assist organizations in the protection of their infrastructure and intellectual property, it is unclear to what extent they follow secure practices or if they allow notification and consent by the affected users. The use of these products is controversial because browser software still shows a lock icon during such sessions, misleading users and compromising the end-to-end security promises made by TLS. Given that there appear to be both benevolent and malicious uses of the technique of using substitute certificates, we refer to this approach as a *TLS proxy*.

1.2 Our Study

Unfortunately, very little is known about the prevalence and nature of TLS proxies on the Internet. There is also no work done to understand what users think about the use of TLS proxies and under what circumstances their use is acceptable. In this work we use a two-part study to provide the security community with a better understanding in both of these areas. In the first part, we provide the community with the results of a measurement study conducted to obtain quantitative data on TLS proxies on the Internet today. This work has been published in the following article:

> O'Neill, M., Ruoti, S., Zappala, D., & Seamons, K. *TLS Proxies: Friend or Foe?* In proceedings of ACM Internet Measurement Conference (IMC), 2016.

In the second part, we report on the results of two surveys conducted to better understand user attitudes toward TLS proxies. This work has also been published, in the following article:

> Ruoti, S., O'Neill, M., Zappala, D., & Seamons, K. *User Attitudes Toward the Inspection of Encrypted Traffic.* In proceedings of Symposium on Usable Privacy and Security (SOUPS), 2016.

For the measurement portion of our work, we study TLS proxies from the perspective of the client. A variety of studies have examined the certificate ecosystem by scanning secure

servers from a single point of view [18, 20, 30] or using passive monitors from several vantage points[7, 20]. However, detecting proxies necessitates measurements at the client, and less work has been done in this space.

Two recent works have found some evidence for TLS proxies by measuring certificates received by clients. Concurrent with some of our work, Huang et al. measure the prevalence of TLS proxies that intercept traffic from clients connecting to Facebook [33], finding that 1 in 500 TLS connections are proxied, mostly by corporate Internet filters and personal antivirus software. In addition, a small number of connections were found to be intercepted by malware. Because this study uses Flash to detect a certificate mismatch, it does not detect proxies affecting most mobile devices. The Netalyzer project measured certificates received by Android apps, assessing 15,000 sessions and identifying just one case of a TLS proxy running in an analytics app [63]. Though this is a very low rate of prevalence (30 times less than Huang's study), the app was found to whitelist several sites, including Facebook. This indicates that measurements of proxies should examine low-profile sites that are unlikely to be whitelisted.

We conduct measurements of TLS proxy prevalence using a Flash app deployed with a Google AdWords campaign. Like Huang, our measurements use Flash to detect a certificate mismatch without any user interaction. However, we deploy our tool using a Google AdWords campaign, which affords a number of advantages. First, we are able to target our measurements toward a server that ordinarily does not receive significant traffic. This enables us to detect proxies that may intentionally whitelist a popular site such as Facebook in order to avoid detection. Second, we are able to actively measure clients, based on how much money we spend on the advertisement, enabling us to collect as many as 12 million measurements in one week by spending $750 per day. Third, we are able to target our measurements at any country, so that we can measure proxy prevalence in distinct areas of the world. Fourth, we are able to measure any site that has a permissive Flash socket policy file. Together, these characteristics give us a broader view of TLS proxies on the Internet.

We report on a two-part measurement study of TLS proxies using a Google AdWords campaign. The first study measures proxies broadly, wherever Google places our advertisement, comprising 2.9 million certificate tests, with proxied users in 142 countries. This study measures proxies intercepting traffic to a new server on our campus, which is highly unlikely to be whitelisted by any proxy. The second part of the measurement study specifically targets users in five countries (China, Ukraine, Russia, Egypt, and Pakistan) in addition to the world at large. This covers 12.3 million certificate tests, finding proxied users in 147 countries. In addition, the second study measures proxies intercepting traffic to twelve sites on the Alexa top 1 million, in addition to our own server.

Our basic findings are as follows:

- The first part of our measurement study found 11,764 proxied connections out of 2.9 million total measurements (0.41% or approximately 1/250 of all connections) spanning 142 countries. This rate is double that reported by Huang. We found that most substitute certificates claim to be from benevolent TLS proxies, with 70.87% claiming to be generated by a firewall software and 12.66% claiming to be generated by a corporate network.
- The second measurement study, which queried multiple secure hosts, found 50,761 proxied connections out of 12.3 million total measurements (again, 0.41% of all connections) spanning 147 countries. It is surprising that the overall prevalence is identical in both studies, which seems to indicate that none of the sites we tested was whitelisted by proxies.
- Our second measurement study targeted specific countries with the Google AdWords placement. We find that proxy rates vary significantly with respect to the origin country of the user. China has an exceptionally low rate of TLS proxies whereas the United States and other western nations tend to have much higher rates of TLS proxies. Targeted countries also have a greater rate of unclassifiable TLS proxies that disclose little to no information about their nature.

- In both measurement studies we found numerous instances of negligent and malicious behavior. Our analysis of one parental filter finds that it masks forged certificates, allowing an attacker to easily perform a MITM attack against the firewall's users. In addition, we found eight malware products affecting over 3,600 connections that install a new root certificate and act as a TLS proxy to dynamically insert advertisements on secure sites. We also found evidence that spammers are using TLS proxies in their products and that botnets may be using this technique. We found numerous other suspicious circumstances in substitute certificates, such as a null Issuer Organization, falsified certificate authority signatures, and downgraded public key sizes.

Our surveys examine user understanding and attitude toward the use of TLS proxies. We surveyed 1,976 people across two surveys regarding their opinions of TLS proxies and their use in inspecting encrypted traffic. The results of the first survey of 1,049 individuals showed a surprising willingness by participants to accept the inspection of encrypted traffic, provided they are first notified. Based on the results of the first survey, we conducted a second survey of 927 individuals to further explore user attitudes towards inspection of encrypted traffic in specific situations.

Our contributions from these surveys include the following insights:

- User opinions toward TLS proxies and the inspection of encrypted traffic are nuanced. Many express concerns about privacy and identity theft from hackers (75.8%) or government surveillance (70.9%). Yet there is broad, general acceptance of TLS proxies when used by employers, schools, etc. (71.7%).
- Most participants indicated support for the inspection of encrypted traffic as long as they were first notified of it (90.7%). Likewise, participants indicated strong support for legislation requiring notification or consent (83.2%).
- When asked about specific situations in which TLS proxies might be used (e.g., at work, at school, at a café, or at home), support for TLS proxies ranges from 65% to 90% of

participants (including those who want notification or consent). Support for inspection of encrypted traffic without notification or consent is strongest at elementary schools (45.9%) and at businesses when employees are using company-provided computers (47.9%). Participants generally favor consent in cases when they feel in control (at home, free WiFi, their own device at work) versus notification when an organization is in control (public library, school, company computer). In nearly all the scenarios we posed, only a small minority of the participants indicated that using TLS proxies is not acceptable. The one exception is government surveillance, in which case 47.5% say that this is not acceptable.

- Many users would have a negative opinion if they discovered that the owner of their network used a TLS proxy without prior notification and/or consent (60.8%), though for some (34.2%) it would depend on who the owner was and how they were using the technology. Some would change their behavior on the network, either discontinuing to use it (17.2%) or changing which sites they visited (6%).
- We identify personas based on participants' responses regarding TLS proxies: pragmatic (76.5%), privacy fundamentalist (17.0%), jaded (5.0%), and unconcerned (1.0%). Jaded participants are interesting in that their opinions regarding privacy and security align with the privacy fundamentalist persona, but their practices align with the unconcerned persona. This dichotomy stems from the fact that these users feel that regardless of what steps they take, they are powerless to prevent compromise of their online information, and so choose to not do anything to protect themselves.

While several of our findings might seem intuitive, it is important to ground intuitions in data, and this work provides the first survey of user opinions on this topic. In addition, participants showed a high level of engagement in the survey, notwithstanding the complexity of the topic. Many users shared in-depth analysis of trade-offs in open responses, demonstrating that they care deeply about this issue. User attitudes toward TLS proxies

provide an important data point along the spectrum of discussion that is currently taking place regarding who should have access to encrypted information.

Part II

Measurements

Chapter 2

The Measurement Tool

We have developed a tool to measure the prevalence of TLS proxies using existing, widely-deployed technologies. The tool runs silently from the perspective of the user; no user action is required, either to install any software or to interact with the tool. This is a significant advantage as compared to other work that requires client-side software installation [4–6, 31, 46, 65].

2.1 Design

To meet our objective of using existing browser technologies without requiring further client installation, we take advantage of the widespread deployment and transparency afforded by the Adobe Flash runtime. By hosting a Flash application on a web page the server can upload it to a visiting client, which runs it without any user interaction. Our tool works by sending a `ClientHello` message to a TLS-enabled server and recording the `ServerHello` and `Certificate` messages received in response. The retrieved certificates are then sent to a reporting server for analysis. This process is handled in three steps, illustrated in Figure 2.1:

1. **Retrieve measurement tool.** The client browser connects to the Distribution Server, where the Flash application is hosted. The application need not be physically visible to the user and can merely be embedded in the background of an otherwise normal web page or be imported through some means such as an HTML `iframe`. The web page data, along with the embedded Flash application, is then downloaded by the client.

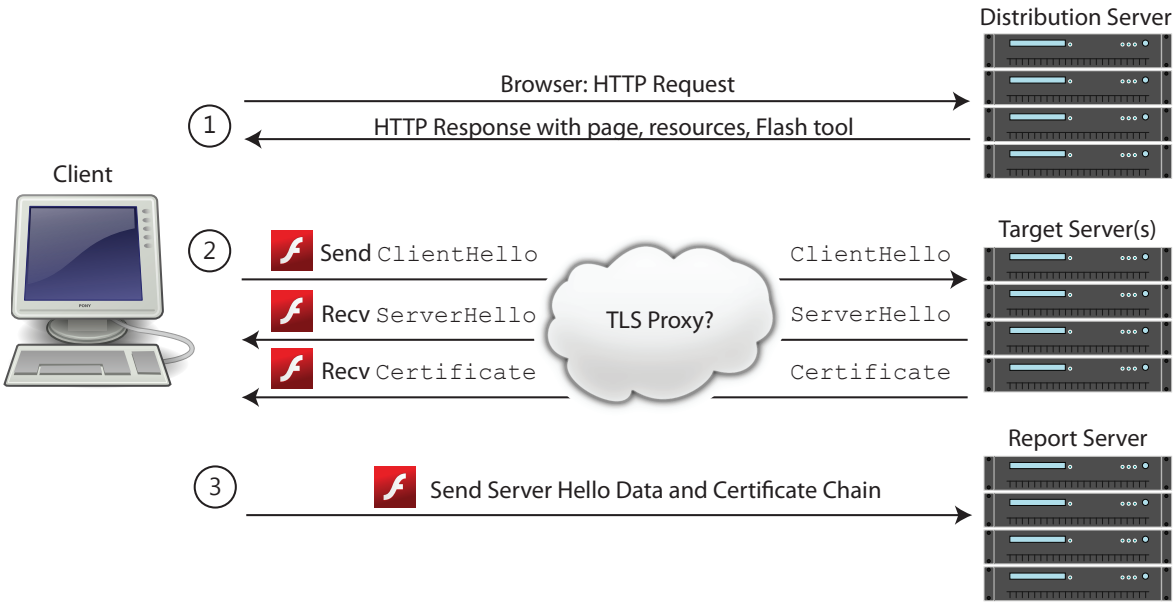

Figure 2.1: Flash TLS Proxy Measurement Tool

2. **Record certificate.** The Flash tool is run automatically by the browser. The tool is designed to perform a series of experiments one for each Target Server to be queried. Target Servers are the hosts to be probed when detecting the presence of a TLS proxy between hosts and the client machine. For each experiment, the application issues a `ClientHello` message to the relevant Target Server to initiate a TLS handshake, which triggers a TLS handshake from either the actual Target Server or a TLS proxy, if present.

3. **Report results.** As the Target Servers (or TLS proxy) respond to this TLS session initiation, the application records the `ServerHello` and `Certificate` messages received. Each partial TLS connection is terminated before the handshake completes and before any actual data is transmitted between the two endpoints. For each completed experiment (Target Server), the tool reports these results to the Report Server using an HTTP POST request. The Reporting Server then compares the certificate received with the expected original, the date, and the IP address of the POST request.

The Distribution Server, Target Servers, and Report Server need not be distinct for the tool to function properly. In fact, a single host may take on all three roles if desired. For example, a website administrator seeking to detect TLS proxies between her website and her end users need only do the following:

1. **Host the Flash application on the web server** and invoke it from desired webpages. The tool can be deployed transparently within existing web pages with no visible changes. In addition, the tool can be imported through other means in HTML from a single location, which may be useful for administrators with large applications spread across multiple servers or via a content-distribution network seeking to detect TLS proxies.
2. **Host a simple *socket policy file* on the Target Server.** For security reasons the Flash runtime (since Flash 9.0) requires that applications attempting to establish a TCP connection with a remote host first obtain permission from that host via a simple policy file. The request for this file is sent automatically by the Flash runtime. The software to host such a file is extremely simple and easily deployed. This particular security feature of Flash prohibits the tool from testing client connections to arbitrary Target Servers; all hosts tested must first grant permission through their respective socket policy files. For our website, we serve our socket policy file on the same port used by our web server (80 or 443). This reduces the effect of captive portals, which often block traffic targeting ports other than those used by HTTP and HTTPS (e.g., airport public access WiFi). Our socket policy server, implemented as an Apache module, is provided on our website along with links to other standalone implementations.

The contents of a sample socket policy file are shown in the listing below. We note for best practices that system administrators should take care to ensure that their socket policy files are not too permissive. That is, they should only allow connections from applications served by hosts they recognize and only to ports on which they are prepared to accept the associated traffic. For administrators of secure websites, this means that the `domain` attribute

on line 4 can be set to the domain of the site itself (or wherever the Flash application is hosted) and that the `to-ports` attribute on the same line can be set to 443 (HTTPS). If additional TLS services are run on the host the `to-ports` value can be extended using comma-delimited port numbers of the other services. Additional domains can also be granted connect permissions by adding lines similar to line 4 with the additional hostnames for their `domain` values. For both the `domain` and `to-port` attributes, a wildcard value of `*` is also allowed (though not recommended).

```
1 <?xml version="1.0"?>
2 <cross-domain-policy>
3    <site-control permitted-cross-domain-policies="master-only"/>
4    <allow-access-from domain="example.com" to-ports="443" />
5 </cross-domain-policy>
```

In addition to the recommendations for the socket policy file, we also suggest that website administrators wishing to use this TLS proxy detection long-term enhance it with modern obfuscation techniques. As this method of detection becomes more well-known, the likelihood increases of a TLS proxy accounting for its use and either blocking its queries or allowing them to transit without modification. We leave these obfuscation techniques to the interested implementer, as their usefulness is inversely proportional to their notoriety.

2.2 Implementation

To implement our tool it was necessary to retrieve the certificate used during a TLS handshake. It would have been preferable to use JavaScript or HTML5 to retrieve the certificate used as part of a current TLS connection, but unfortunately there is no API available for this. Firefox allows a plugin to request the certificate, but plugins require manual client installation. This left us with the alternative of establishing a plain TCP connection with the target server and then initiating a TLS handshake. Unfortunately, the ability to use a plain TCP connection rules out the use of HTML5 `WebSockets`.

Due to these constraints, we opted to use the Adobe Flash platform. Java web plugins also fit the requirements, but lacked the widespread install base of Adobe Flash. Beginning with version 11.0 of the Flash runtime, Adobe made available a `SecureSockets` API that allows developers to access certificate data from a TLS connection. However, these versions of Flash were too recent to enjoy the reported 98.9% desktop market penetration of Flash 9.0 [2]. Thus we implemented our tool in ActionScript using only libraries supported by the Flash 9.0 runtime. Using the `Socket` API provided by Flash 9.0 we implemented functionality required to perform a partial TLS handshake. After receiving the full `Certificate` message from the Target Server the handshake is aborted and the connection is closed. The Flash application records and parses all certificates in the chain received from the `Certificate` message (as some hosts offer certificate chains) and stores them locally until it parses the final one. All certificate data, in PEM format, is concatenated and then sent as an HTTP POST request to Report Server for analysis. The Reporting Server hosts a MySQL database and simple PHP application responsible for receiving the POST request and storing the data contained within it. This request is performed using the `Socket` API as well.

Code for the Flash measurement tool, socket policy server, and collected datasets are available for download at `tlsresearch.byu.edu`.

2.3 Limitations

Our tool is unable to measure TLS proxies being used against most mobile devices. An overwhelming majority of mobile platforms do not support Flash and Adobe has discontinued their development of Flash for mobile devices. Our tool is likewise unable to measure TLS proxies being used against browsers that use ad blocking technology.

In addition, it is possible that TLS proxies could be engineered to circumvent our measurements. At the time of our study, our measurement methodology was not well known, so it is unlikely that any attacker was evading detection or tampering with our reports. However, in the case that this methodology becomes well-known, it would be difficult to

prevent dedicated attackers from modifying their TLS proxies to avoid our measurements. Interested parties are referred to modern obfuscation practices if they wish to make a more stealthy detection tool.

Finally, we note that since the Target Servers must host a permissive socket policy file to function, only hosts that are controlled by the experimenter or those that have wildcard permissive socket policies are able to be scanned. However, we do note that in our experience most of the hosts that serve a socket policy file have very permissive policies already.

Chapter 3

Google AdWords Campaigns

To achieve rapid and widespread deployment of our measurement tool we leveraged the Google AdWords platform. This strategy for using an advertising campaign to conduct an end-user measurement study has previously been used to study CSRF attacks [9], DNS rebinding attacks [36], and DNSSEC deployment [34, 35, 44]. Our study is the first to use this same method to measure the deployment of TLS proxies. The results from this study shed light on the legitimate demand for TLS proxies as well as several suspicious or duplicitous practices.

For deployment on an advertisement network, the ad servers themselves constitute the Distribution Server in our architecture found in Figure 2.1. The Target Server was set as our own website, `tlsresearch.byu.edu`, for the first study and a variety of other hosts for the second study, listed later. Finally, the Report Server was hosted on the same machine as our website and used a simple PHP application and a MySQL database for storing the certificate chains reported back from clients.

To accommodate placement in advertisements, our measurement tool was modified to contain a visible canvas on which we place a simplistic advertisement for our research lab. Figure 3.1 shows the advertisement as it appeared to web users during our measurement study. In this configuration, the Flash tool also bears a `clickTAG`, code that links an Flash advertisement to the advertiser's landing page, required by many advertisement networks. For this work, the `clickTAG` referred users who opted to click our ad back to our website, `tlsresearch.byu.edu`. We also replaced the random numbers used by the TLS handshake

Figure 3.1: Appearance of tool via Google AdWords

code with static values, as some ad networks block Flash advertisements that use random number generators. Our measurement tool was run as soon as the browser loaded the advertisement, and required no interaction from users.

For our ad campaigns we leveraged the CPM (cost-per-impression) bidding model for our campaign, which maximizes the number of unique clients presented with our ad. We set the Max. CPM to $10 USD. To help us reach a global audience we indicated that our ad should be served to all locations and languages. Additionally, since ads are shown only on websites that match a set of designated keywords we selected our keywords based on phrases that were currently trending globally on Google Trends [29]. We set our ad to show uniformly throughout the day so as to collect data from users in a variety of locations and situations (e.g., home, commuting, work).

Along with the certificate, we also recorded the IP address of the client tested. This IP address was then used to query the MaxMind GeoLite [48] database to gather geolocation information.

3.1 Campaign Setup for First Measurement Study

Our initial Google AdWords advertising campaign ran from January 6, 2014 to January 30, 2014. During the first 17 days of the study we varied the amount of money allocated to the ad campaign, but for the last week we kept it at $500/day. In this study we gathered certificate data for our own website, `tlsresearch.byu.edu`, using it as the Target Server. We used the following keywords for the study: *Nelson Mandela*, *Sports*, *Basketball*, *NSA*,

Internet, *Freedom*, *Paul Walker*, *Security*, *LeBron James*, *Haiyan*, *Snowden*, *PlayStation 4*, *Miley Cyrus*, *Xbox One*, and *iPhone 5s*.

This campaign generated 4,634,386 impressions and 3,897 clicks (not required to complete the measurement) at a cost of $4,911.97. In total we completed 2,861,244 successful measurements.

Chapter 4 discusses the results from this study.

3.2 Campaign Setup for Second Measurement Study

To increase the number of measurements collected and to better understand the nature of TLS proxies we conducted a second set of measurements approximately eight months after the first study.

One question unanswered by the first study was whether TLS proxies were intercepting all traffic, or whether they selectively intercepted traffic according to white or blacklists. To shed light on this subject, we decided to gather measurements for different types of sites:

- **Popular:** Sites from the Alexa top 25,000. Six websites were included in this category.
- **Business:** Commercial sites unlikely to be blocked by places of business. Five websites were included in this category.
- **Pornographic:** Pornographic websites (expected to be blocked by parental filters and places of business). Five websites are included in this category.
- **Authors':** The single website operated by the authors and also used in the first measurement study.

The policy restrictions of the Flash runtime prohibit establishment of socket connections to arbitrary hosts. Thus all sites used in our study had to host permissive socket policy files that allowed connections to port 443 from any domain. We scanned for the presence of permissive socket policy files on the entirety of the Alexa top 1 million websites, and selected

Website		
Top 25,000	Business	Porn
qq.com	airdroid.com	pornclipstv.com
promodj.com	webhost1.ru	porno-be.com
idwebgame.com	restaurantesecia.com.br	pornbasetube.com
parsnews.com	speedtest.net.in	pornozip.net
idgameland.com	iprank.ir	pornorasskazov.net
vcp.ir		

Table 3.1: Website types probed in second measurement study

the highest ranked such websites for each type to use in the second measurement. Table 3.1 lists the additional hosts we probed.

At most 12 of these sites were queried by a single served instance of our Flash measurement tool. Due to differences in Internet connectivity quality and hardware and software performance, not all clients served with our ad were able to successfully perform experiments with all hosts. The tool was configured to first test the connection to the authors' website, before attempting to perform other experiments on other hosts.

In this second study, we also targeted specific countries by creating an additional ad campaigns. The ad image used in both the global and country specific ad campaigns was the same. Some of the countries we wanted to target were unavailable in Google AdWords (e.g., Iran, Syria) and after discussion we settled on the following five countries: China, Egypt, Pakistan, Russia, and Ukraine.

The second study ran from October 8, 2014 4:00 PM MDT to October 15, 2014 4:00 PM MDT. The budget for the global campaign was $500/day and the country-specific campaigns were $50/day. We used the following keywords for the study: *Nelson Mandela*, *Sports*, *Internet Security*, *Basketball*, *Football*, *Freedom*, *NCAA*, *Paul Walker*, *Boston Marathon*, *Election*, *North Korea*, *Harlem Shake*, *PlayStation 4*, *Royal Baby*, *Cory Monteith*, *iPhone 6*, *iPhone 5s*, *Samsung Galaxy S4*, *iPhone 6 Plus*, and *TLS Proxies*.

The breakdown of costs and results are given in Table 3.2. In total we completed 12,314,756 successful measurements.

Chapter 5 discusses the results from this second study.

Campaign	Impressions	Click	Cost
Global	3,285,598	5,424	$4,021.78
China	689,233	652	$401.41
Egypt	232,218	1,777	$378.17
Pakistan	183,849	2,536	$378.26
Russia	230,474	203	$401.36
Ukraine	364,868	294	$390.69
Total	5,079,298	11,077	$6,090.19

Table 3.2: Second Study Statistics

3.3 Limiting Risk

When utilizing the Flash measurement tool to perform experiments on different Target Servers, there are some risks to users that need to be addressed. First we note that the existence of permissive Flash socket policies on Target Hosts, which are required for the tool to function, are an explicit indication from administrators that connections from our tool are allowed. However, since the tool transparently queries the Target Server from users' machines, we must consider protecting users themselves from potential harm. For example, since browsing gambling and pornographic websites is a forbidden practice on many company campuses, and even against the law in some jurisdictions, proper care must be taken to avoid unnecessary risk to users when Target Servers fall into these categories. Given that interaction with users is not possible when distributing the tool through an advertisement network (often both technically infeasible as well as against terms of service), it is not possible to obtain user consent in such a situation. This is common in many Internet measurement scenarios. To mitigate the risk to users in our Ad Campaign study, we employ the following precautions:

- **Breadcrumbing:** We force the tool to first connect to our Target Server before attempting any additional experiments. Our Target Server provides a webpage for curious users and network monitors to obtain information on our experiments. If network

administrators were to investigate this traffic by visiting our web server, they would be shown a description of our research and provided with our contact information[1].

- **Connection Termination:** Since the tool is only interested in certificates sent by the Target Server (or TLS proxy intercepting the connection), we abort the TLS connection after we have received this data. Furthermore, we abort the TLS handshake itself, making it clear to network monitors that no connection was actually established to the Target Server.
- **No Data Transfer:** As a result of our early TLS handshake termination, no website content of any kind is ever transferred from the Target Server to the client. This further illustrates to network monitors that users are not actually visiting the Target Server, as browsers automatically download content once a connection is established.

The primary benefit of this methodology is that the security community can become rapidly aware, at a global scale, of the nature and prevalence of TLS proxies. Knowledge of TLS proxies in the wild, both benevolent and malicious, is traditionally very difficult to obtain. Distribution of our tool through an advertisement network is a cost-effective, large-scale, measurement option to shed light on practices that subvert authentication guarantees of the most common security protocol on the Internet. This and Huang's work [33] has informed the community about these practices and motivated a subsequent study that identified weaknesses in personal firewalls [17].

Finally, our tool complies with Google AdWords' terms of service. Use of our tool through other advertisement networks or distributions should be preceded by a thorough investigation of applicable terms of service.

[1] We were never contacted.

3.4 Ethical Considerations

When first attempting to publish the results of our measurement studies in the 2015 ACM Internet Measurement Conference, we unfortunately neglected to include a discussion of ethics. As a result, the discussion of our risk minimization techniques was also omitted. This omission was met with some concern by two responsible anonymous reviewers (RA and RB), who issued the following two remarks.

> the critical problem with this paper is the lack of an ethics statement and the lack of IRB approval. I'm willing to accept that the first experiment was harmless, i.e. having clients connect to a benign server controlled by the researchers. However, the second experiment has the potential for harm, since users were sending packets to pornographic sites without their knowledge. –RA

> I have ethical concerns over how the study was conducted. While there likely isn't much harm in visiting an innocuous website, or a website associated with the research group, the ad does make it look like the user visited the website, in the case of a censored website or pornographic site, the proxy might be in place for the purpose of checking for this activity. For example, imagine a company policy that prohibits viewing pornographic material in the workplace. This user now has connected to a prohibited site from their browser. Unfortunately the authors have no discussion of these risks. –RB

A subsequent submission included a discussion of ethics and removed the results from the second measurement study as an additional precaution. This second submission was accepted. However, we wished to gain further insight into the ethical issues encountered by the work in our second measurement study. In response to the initial reviewer criticism we held informal talks with our IRB. The IRB indicated that our study would likely be exempt from review because 1) our methodology probed servers rather than downloading content

from them, 2) we took steps to minimize risk, and 3) there was no possibility to allow for user consent. The inability to obtain user consent is common in Internet measurement work.

The networking community is still in the process of developing ethical standards related to Internet measurement broadly. A recent paper by Burnett et al. was controversial because they measured online censorship using a method that caused users to visit sites likely to be blocked by censoring technology [13]. In this case, measurements were not limited to probes; the measuring software tried to download content, without user consent, to see if the attempt would be blocked. The paper was published with a note from the SIGCOMM program committee indicating that they did not approve of the methodology. Subsequently, a paper by Jones et al. [37] discussed these ethical issues, finding that Internet censorship measurements "fall into an ethical grey area", due to the lack of true relation to human subjects research, and the difficulty in evaluating the degree of risk. The paper includes the opinion that measurements that use probes below the application layer are clearly preferable to those that download content.

In dicussing these ethical issues with colleagues, we conclude that the networking community's best current guideline is that risk is effectively mitigated if measurements consist of probes that are sent at a low rate. We believe our methodology sufficiently mitigates risk by only probing, rather than completing a TLS handshake or downloading content. We note that our tool operates very similarly to typical advertisements on the web, in the sense that they force browsers to connect to websites not solicited by the user. Given that this is common on the web, network observers are less likely to hold users accountable for every host their machines contact. However, we acknowledge that unconstrained use of our methodology may not fall within the realm of ethical behavior. As such, we provide the following list of recommendations for future researchers embarking on similar studies:

- **Careful Target Selection:** When selecting the target hosts for a measurement, carefully evaluate the nature of each host in the context of the societal norms and civil regulations of affected users. In this step, it is important to not include a host that may

adversely affect the legal standing or reputation of users affected by the measurement. Note specifically that in some jurisdictions with less civil rights, some users may be considered guilty until proven innocent. Furthermore, even seemingly innocuous hosts in one jurisdiction may be illegal or inappropriate to visit in others. This extends to different environments within a single jurisdiction as well (e.g., company policies may differ from personal ones). We note that, if doing our study again, our second measurement study would have omitted the probes to pornographic websites, especially in jurisdictions where legal action is potentially taken against those who view such material.

- **Brevity:** Abort connections as quickly as possible. Once the data pertinent to the study has been obtained, minimize potential risks to users by severing the connection immediately. This reduces user risk by minimizing the window of opportunity for measurement tools to trigger network alarms, while also informing any network observers that the connection was not likely caused by normal human interaction. For example, the network trace for an employee visiting a prohibited website and immediately closing the browser is drastically different from a local measurement tool aborting a connection to that same website and prematurely terminating the TLS handshake.
- **Publicity:** Where possible, provide network observers with easy paths to identify the purpose and true origin of the traffic patterns generated by the measurement tool. This allows network monitors to place blame for any prohibited connections on the tool rather than the user. Note that, unless it is strictly necessary for the experiment, we do not recommend attempting to camouflage measurement traffic in any way, as it may subject users to undue suspicion if uncovered.

Chapter 4

First Measurement Study

Our first measurement study was targeted at a general, global audience. During the duration of this ad campaign, we served 4.36 million ads and successfully completed 2.86 million measurements. Of those tests, 11,764 returned a different X.509 certificate than was served by our secure web server, indicating the presence of a TLS proxy.

The users behind a proxied connection that were identified by our campaign originated in 142 countries and from 8,589 distinct IP addresses. Due to the targeting algorithms used by Google AdWords, our tool's exposure to these countries is not uniformly distributed. Table 4.1 shows the countries with the most proxied connections in our study. For each country, the table lists the total number of proxied connections, the total number of connections, and the percentage of total connections to that country that were proxied. Some countries have significantly higher percentages of proxied connections than the average, including France (1.09%), Canada (0.87%), Belgium (0.81%), the United States (0.79%), and Romania (0.74%). Together, connections from the United States and Brazil account for 36% of detected proxies.

4.1 Analysis of Issuer Organization

We first analyze the contents of the Issuer Organization in the substitute certificates we collected. We use `openssl` to decode the certificates and store them in a database, where we can run queries. We also manually inspect the contents of the relevant fields to identify the issuing organization and their software products, using web searches to determine their identity. We emphasize that our results in this section are based on the intercepting proxy

Rank	Country	Proxied	Total	Percent
1	France	812	74,789	1.09%
2	Canada	303	34,695	0.87%
3	Belgium	136	16,816	0.81%
4	US	2,252	285,078	0.79%
5	Romania	696	94,116	0.74%
6	Brazil	2,041	298,618	0.68%
7	Portugal	185	29,799	0.62%
8	India	302	51,348	0.59%
9	Turkey	303	65,195	0.46%
10	S.Korea	196	46,660	0.42%
11	Russia	224	58,402	0.38%
12	Spain	226	62,569	0.36%
13	Japan	111	31,751	0.35%
14	Netherlands	104	31,938	0.33%
15	UK	759	259,971	0.29%
16	Germany	499	187,805	0.27%
17	Ukraine	160	61,431	0.26%
18	Taiwan	101	61,195	0.17%
19	Poland	182	110,550	0.16%
20	Italy	200	129,358	0.15%
	Other (215)	1,972	869,096	0.23%
	Total	11,764	2,861,180	0.41%

Table 4.1: Proxied connections by country, ordered by percentage proxied

Rank	Issuer Organization	Connections
1	Bitdefender	4,788
2	PSafe Tecnologia S.A.	1,200
3	Sendori Inc	966
4	ESET spol. s r. o.	927
5	Null	829
6	Kaspersky Lab ZAO	589
7	Fortinet	310
8	Kurupira.NET	267
9	POSCO	167
10	Qustodio	109
11	WebMakerPlus Ltd	95
12	Southern Company Services	62
13	NordNet	61
14	Target Corporation	52
15	DigiCert Inc	49
16	ContentWatch, Inc.	42
17	NetSpark, Inc.	42
18	Sweesh LTD	39
19	IBRD	26
20	Cloud Services	23
	Other (332)	1,121

Table 4.2: Issuer Organization field values

self-identifying itself in the certificate. It is certainly possible that malicious proxies have hidden their tracks by masquerading as a legitimate organization in the Issuer Organization field, and we cannot detect this.

Table 4.2 shows the values for the Issuer Organization field of the substitute certificates. Table 4.3 provides a breakdown of values present in the Issuer Organization field of the substitute certificates. The majority of certificates from proxied connections have an Issuer Organization field matching the name of a personal or enterprise firewall (69.54%). Another 12.66% have the name of an organization set as the Issuer Organization (e.g., Lawrence Livermore National Laboratory, Lincoln Financial Group). Additionally, 7% (829) of the substitute certificates have null values for the Issuer Organization field.

Proxy Type	Connections	Percent
Business/Personal Firewall	8,101	68.86%
Organization	1,394	12.66%
Malware	1,112	8.65%
Unknown	840	7.14%
Parental Control	156	1.33%
Business Firewall	69	0.59%
Certificate Authority	49	0.42%
School	32	0.27%
Personal Firewall	11	0.09%
Telecom	0	0%

Table 4.3: Classification of claimed issuer, ordered by proxy share

The most suspicious activities discovered were revealed by certificates with an Issuer Organization that matched the names of malware. *Sendori, Inc*, *WebMakerPlus Ltd*, and *IopFailZeroAccessCreate* appeared in 966, 95, and 21 Issuer Organization fields, respectively. Sendori poses as a legitimate enterprise, however they produce software that compromises the DNS lookup of infected machines, allowing them to redirect users to improper hosts. A TLS proxy component is used to bypass host authenticity warnings in the browser. The substitute certificates generated by the TLS proxy are signed by a root authority that was added to the root store of the local machine at the time of infection. Substitute certificates issued by Sendori originated from 30 distinct countries.

The WebMakerPlus malware is primarily associated with inserting advertisements into Web pages. We hypothesize that WebMakerPlus uses a TLS proxy to simulate that their advertisements are served from a secure connection and to modify secure pages in transit to include such content. Substitute certificates containing markings for WebMakerPlus originated from 16 distinct countries.

Manual Internet queries revealed that malware was responsible for an Issuer Common Name field value of *IopFailZeroAccessCreate*. The certificates containing this value originated

from 14 distinct countries. Disturbingly, each certificate contained the same 512-bit public key. This malware was also reported by [33].

It is somewhat surprising that these malware programs self-identify in the substitute certificates they generate, as an attacker can arbitrarily select values for the fields in a substitute certificate.

In addition to malware discoveries, we found that the names of two companies highly associated with spam were also present in numerous Issuer Organization fields. The names *Sweesh LTD*, and *AtomPark Software Inc* were found in 39 and 20 substitute certificates, respectively. AtomPark offers tools for spammers including "email extractors" and "bulk mailers". Sweesh offers services to spammers to overcome "hurdles" faced by advertisers and publishers. Internet searches reveal that Sweesh may be responsible for the development of WebMakerPlus.

Not all of the root certificates found in the collected substitute chains were unique. In the 11,764 substitute chains 8,341 distinct roots were found. For example, 310 leaf certificates signed by *Fortinet* all used the same root certificate, and these were obtained from 155 distinct IP addresses. This behavior was consistent across many of the popular issuers identified (e.g., POSCO, Southern Company Services, Target Corporation). These organizations are likely using a single root to sign intermediate certificates and then deploying these at various endpoints where they operate TLS proxies.

4.2 Negligent Behavior

Where possible, we installed and characterized personal firewall software from many of the most common companies whose names were provided in the Issuer Organization, Issuer Organizational Unit, and Issuer Common Name fields of our collected certificates. We characterized the behavior of these solutions when running behind our own TLS proxy which issued certificates signed by an untrusted CA. While most solutions properly rejected our forged certificates, Kurupira, a parental filter that is responsible for 267 proxied connections

in our dataset, did not. When visiting `google.com` and `gmail.com`, Kurupira replaced our untrusted certificate with a signed trusted one, thus allowing attackers to perform a transparent man-in-the-middle attack against Kurupira users without having to compromise root stores. In contrast, BitDefender not only blocked this forged certificate, but also blocked a forged certificate that resolved to a new root we installed in the victim machine's root store.

Among the negligent behavior we found are TLS proxies that generate substitute certificates with weak cryptographic strength. Our original certificate has a public key size of 2048 bits. However, we find that 5,951 (50.59%) substitute certificates have public key sizes of 1024 bits and 21 certificates have public key sizes of 512 bits. 23 (0.20%) TLS proxies generated substitute certificates that used MD5 for signing, 21 (0.18%) which were also 512 bit keys. Interestingly, some TLS proxies generated certificates that have better cryptographic strength than our certificate. Seven (0.06%) used certificates with a key size of 2432 and five (0.04%) used SHA-256 for signing.

In addition to problems with cryptographic strength, we discovered that 49 (0.42%) substitute certificates claim to be signed by DigiCert, though none of them actually are. The original certificate from our secure web server is issued by DigiCert High Assurance CA-3, indicating the TLS proxy likely copied this field when creating the substitute. It is alarming that a TLS proxy would opt to copy this field, as it signifies a masquerading as the legitimate authority. It is possible that these proxies are operated by malicious individuals doing their best to not be detected by the user.

Finally, we note that 110 substitute certificates have modifications to the subject field. For 51 (0.43%) certificates, the subject did not match our website's domain. In many cases a wildcarded IP address was used that only designated the subnet of our website. In two cases the substitute certificate is issued to the wrong domain entirely: `mail.google.com` and `urs.microsoft.com`. These certificates appear to be legitimate for those domains and properly validate back to GeoTrust and Cybertrust roots, respectively.

Chapter 5

Second Measurement Study

During the second ad campaign we successfully completed 12.3 million measurements targeting five specific countries as well as the world in general. Of those tests, 50,761 returned a different X.509 certificate than was served by the authoritative host.

5.1 Analysis of Issuer Organization

Table 5.1 contains the breakdown of Issuer Organization fields from our second measurement study. As in our first measurement study, we find that the majority of TLS proxies claim to be features of firewall solutions (74.42%). Organization and school names are also prevalent, accounting for another 6.01% of Issuer Organization values. However, we see an increase in the relative popularity of the Unknown category as compared to the first study (10.75% from 7.14%). The Unknown category comprises certificates with null or blank issuer fields, or otherwise uncategorizable values. In tandem with this finding we note that the Malware category has decreased in relative popularity from 8.65% to 5.06%. These results were obtained largely from our five targeted countries. The increase in the Unknown category of TLS proxies in these countries is particularly alarming. It is possible that malware using TLS proxy features in these regions is more opaque than its earlier counterparts, opting not to disclose its identity through Issuer Organization fields. Even if this is not the case, it is alarming to note that TLS proxies may be decreasing their already-poor visibility to users in those countries.

Proxy Type	Connections	Percent
Business/Personal Firewall	36,005	70.93%
Unknown	5,455	10.75%
Organization	3,531	6.96%
Malware	2,571	5.06%
Business Firewall	1,231	2.43%
Personal Firewall	536	1.06%
School	482	0.95%
Telecom	447	0.88%
Parental Control	428	0.84%
Certificate Authority	68	0.13%

Table 5.1: Classification of claimed issuer, ordered by proxy share (2nd study)

Another distinguishing feature of our second study's Issuer Organization fields is the presence of the names of Telecom companies in the dataset. We found 375 proxied connections from IP addresses owned by a Korean telecom company, LG UPLUS. Another four telecom company names were reported from an additional 72 connections. It is unclear whether LG UPLUS and the other companies are using TLS proxies within their own office buildings or using them to intercept the traffic of their own users. The latter behavior is not without precedent; Nokia has recently come under fire for such an operation [51].

5.2 Proxy prevalence by specific country

Our second measurement study via AdWords contained six mini-campaigns. Five of these targeted the countries of China, Ukraine, Russia, Egypt, and Pakistan. These countries were selected for their contemporary civil struggles and their respective governments' stance on free speech. The final mini-campaign targeted the world in general. Table 5.2 shows a breakdown of the number of connections tested per country. We immediately see that all five specific countries targeted by our campaign lie within the top six most-prevalent countries, showing the dependability of Google AdWords' country targeting feature. We also note the relatively low percentage of TLS-proxied traffic from China. Before this study the authors

Rank	Country	Proxied	Total	Percent
1	Romania	2,210	185,749	1.19%
2	US	3,327	385,811	0.86%
3	Brazil	1,889	232,454	0.81%
4	UK	2,056	266,873	0.77%
5	Japan	2,033	273,532	0.74%
6	India	716	102,869	0.70%
7	Germany	1,091	177,586	0.61%
8	Egypt	3,720	660,937	0.56%
9	Italy	737	145,438	0.50%
10	Turkey	1,975	411,962	0.48%
11	Indonesia	798	181,971	0.44%
12	Pakistan	1,890	456,792	0.41%
13	Russia	4,532	1,116,341	0.40%
14	Greece	516	130,613	0.40%
15	Poland	456	127,806	0.36%
16	Czech Rep.	343	110,170	0.31%
17	Taiwan	530	186,942	0.28%
18	Ukraine	4,329	1,575,053	0.27%
19	Korea	1,722	836,556	0.21%
20	China	563	2,549,301	0.02%
	Other (209)	15,328	2,200,000	0.70%
	Total	50,761	12,314,756	0.41%

Table 5.2: Proxied connections by country, ordered by percentage proxied (2nd study)

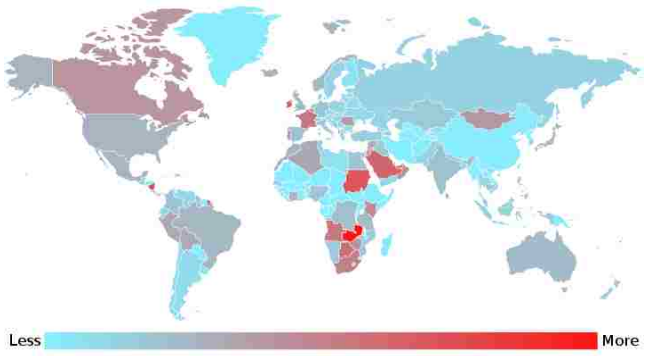

Figure 5.1: Heat-map of TLS proxy prevalence by country. Highest = 12% proxy rate, lowest = 0% proxy rate

Website Type	Connections	Proxied	Percent Proxied
Popular	5,132,342	20,965	0.41%
Business	1,787,875	7,494	0.42%
Pornographic	3,004,996	12,458	0.41%
Authors'	2,353,717	9,844	0.42%

Table 5.3: Proxied connection breakdown by host type

hypothesized that this country would have a high amount of TLS-proxied traffic due to its stance on civil liberties and government surveillance. We also find that Ukraine, Russia, Egypt and Pakistan all have a lower TLS-proxy percentage than western nations such as the United States (0.86%) and the UK (0.77%). This may be due to the fact that most detected firewall solutions are of western origin and western install base. Thus it is possible that the general lower proxy rates in our target countries is due more to consumer choice and buying power. If true, this would also give some validity to the claimed issuer fields, which include the names of firewall solutions.

The relative prevalence of TLS proxies by country is visualized in Figure 5.1. Low TLS-proxy rates are signified by blue and gradually transition to red with increasing proxy rate. The map includes connection data from our second study in 228 countries and territories.

5.3 Proxy behavior by type of host

The augmented measurement tool used in our second study connected to various types of hosts: popular, business, pornographic, and our own. The breakdown of the prevalence of TLS proxies with respect to each host type is shown in Table 5.3 (note that the number of connections to each type of host varies due to the variances in network conditions and computer performance of our users and of the hosts themselves). The percentage of proxied traffic to each type of host is nearly identical. We also find that individual TLS proxies are also indiscriminate in their behavior toward these types of hosts. These results suggest that TLS proxies do not employ blacklists when deciding which traffic to intercept. Given Huang et al.'s finding of a 0.20% TLS-proxy rate for Facebook connections, there is some evidence to suggest that at least some TLS proxies are employing whitelists when determining whether to intercept a connection. Facebook's popularity far exceeds the popularity of our chosen popular hosts (we were constrained to hosts which served permissive Flash socket policy files), so sites akin to it are in a class of their own. It is possible that many benevolent TLS proxies are configured to ignore extremely popular websites run by reputable organizations, perhaps to preserve some privacy and reduce performance overhead.

5.4 More malware

Our second study revealed a continued presence of malware in the TLS proxy space. All of our previously discovered malware was also present in our second study, with an additional five discoveries: issuer fields containing the values Objectify Media Inc (1069 connections), Superfish, Inc. (610 connections), WiredTools LTD (131 connections), Internet Widgits Pty Ltd (67 connections), and ImpressX OU (16 connections). Web research indicated that all these are malware products and one of them, Internet Widgits Pty Ltd, has ties to a botnet. These malware, combined with new instances of the malware exposed previously, accounted for 2,571 of proxied connections in our second study.

One suspicious Issuer Organization field was *kowsar*. Certificates with this identifier appeared 268 times, and were retrieved by 266 unique IP addresses. Unlike other Issuer Organizations we found, this identifier did not appear to be associated with a large organization (which would indicate a corporate firewall) or a personal firewall product. The IP addresses given this certificate are from numerous countries around the world and from many different ISPs. Contrast this with the Certificate Issuer *DSP*, which appeared 204 times, but with only one IP. In this case, *DSP* is being used by the Department of Social and Family Affairs (also called the Department of Social Protection), Ireland, and thus likely represents a corporate firewall. The pattern for *kowsar* is indicative of either a popular personal firewall or an active attack such as a botnet.

Similar oddities appear, but on a smaller scale. For example, the Certificate Issuer field *Information Technology* appeared 33 times, covering 3 IP addresses. These IPs were from a Japanese chemical company, an ISP in Netherlands, and a chapter of the American Red Cross. These are such disparate organizations that this looks like suspicious behavior, though it is possible that each of these organizations set up a corporate firewall and chose the same name for the certificate they generated. The Issuer field *MYInternetS* appeared 36 times from 6 different ISPs. Five of these are in Denmark, from a variety of ISPs and a university, yet one is from a cable subscriber in the United States. It is difficult to determine whether cases like this are examples of malware.

Even more worrisome are the 5,455 instances where we could not identify the issuer, and the 1,518 where the issuer field is null or blank. Whoever set up the TLS proxy in these cases did not want to be identified.

Part III

Surveys

Chapter 6

Methodology of First Survey

The diversity of TLS proxy behavior and prevalence found in our measurement study prompted us to conduct a survey on user attitudes toward TLS proxies in the general public. The security community at large has generally considered TLS proxies to be malicious, yet our measurements suggested that many organizations and individuals were utilizing them for protective measures. Reconciling these two attitudes required additional insight from users themselves. We were especially curious whether users were aware of TLS proxies, and how they felt about TLS proxy use in different institutions and circumstances.

To better understand users' perspectives on these issues, we surveyed 1,976 people across two surveys regarding their opinions of TLS proxies and their use in inspecting encrypted traffic. Both surveys were conducted using the Amazon Mechanical Turk (MTurk) crowdsourcing service. Participants were given $1 USD as compensation for completing each survey, and both surveys were approved by our Institutional Review Board.

In February 2014, we conducted the first online survey using the Amazon Mechanical Turk (MTurk) crowdsourcing service. We gathered responses on Wednesday, February 12, 2014 between 7:50 AM and 5:22 PM (PST). Each participant could take the survey once and received $1 USD as compensation upon completing the survey. In total 1,262 people completed the online survey. The survey was approved by our Institutional Review Board and is contained in Appendix A.1. The data from both surveys is accessible at `tlsresearch.byu.edu`.

6.1 Instructing Participants

Before conducting this survey, we felt it was unlikely that most people would be aware of TLS proxies (an assumption that was upheld by our results). This presented a dilemma: either we would need to only survey individuals who were already aware of TLS proxies or we would need to instruct participants about TLS proxies. Both of these options have significant drawbacks. Limiting the survey to individuals with pre-existing knowledge regarding TLS proxies would likely limit us to participants with highly technical backgrounds, thus failing to gather information about broader opinions related to the inspection of encrypted traffic. On the other hand, instructing participants on TLS proxies has the risk of unintentionally biasing them one way or another, and requires them to answer questions about a subject they potentially just learned about.

Because our research goal was to survey broad opinions regarding the inspection of encrypted traffic, we preferred not to limit our population to the small fraction of users who are already aware of this issue. Instead, we chose to accept the limitations related to instructing participants about TLS proxies and survey as many participants as possible. For our goals, this was preferable to ignoring the opinions of a large portion of users.

To address the risks related to instructing participants on an issue and then surveying them, we spent considerable effort and time crafting our description of TLS proxies. Our goals were to (1) give a simple and concise overview of how TLS proxies are used to inspect encrypted traffic, and (2) present participants with a fair and unbiased description of how the inspection of encrypted traffic could be used for both benevolent and malicious purposes.

In preparation for writing the description of TLS proxies, we examined the literature and observed that existing descriptions of TLS proxies were not neutral in tone and would unduly bias participants. We talked with businesses that sell proxies (i.e., Blue Coat, Symantec) and read opinions from privacy advocates to better understand both sides' opinions. Based on the information in these sources, we composed a draft of our description of TLS proxies, focusing on using language that was informative and neutral in tone, allowing

participants to form their own opinions. Our team of researchers, which included members who are fundamentally opposed to TLS proxies and members who accept their benevolent uses, iterated on this description until all members were satisfied with its wording.

We then tested this description using a convenience sample of six individuals from our university who were not a part of our research group to ensure it was balanced and understandable. Based on feedback from the convenience sample, we made minor edits to the description.

Finally, we tested this revised description using MTurk to ensure that participants felt that the description was sufficiently understandable. Of the 80 participants in this pilot survey, nearly all participants (73; 91%) indicated that the description of TLS proxies helped them understand what TLS proxies are and how they are used (2 participants indicated the description was not helpful (2; 3%), with the remainder being undecided (5; 6%)). We also examined participant responses to free response questions and found that, as reported, most participants' answers reflected an accurate understanding of TLS proxies. As such, we included this version of the description in both surveys, as shown in Figure 6.1.

6.2 Survey Contents

The survey begins by gathering demographic information. It then instructs participants about TLS proxies and their use in the inspection of encrypted traffic. Next, participants are asked to share their opinions regarding the use of TLS proxies and the inspection of encrypted traffic. These questions survey participant opinions as to whether TLS proxies are a breach of their privacy and whether there are acceptable uses for TLS proxies. Participants are also asked their reasoning for why TLS proxies should or should not be allowed. Also, participants are asked which parties they are concerned about using TLS proxies and what, if any, measures should be used to regulate their use.

The survey then asks participants about how they would personally react to having a TLS proxy on a network they use to connect to the Internet. This section includes two

When you connect to the Internet you do so through some organization's network. For example, at home you connect to your Internet service provider's (ISP) network, while at work you connect to your employer's network. To protect your information from others on the network you can create secure connections to the websites you use (HTTPS). This is done automatically for you when you log into a website. The secure connection encrypts your Internet traffic so that no one else can view or modify your communication with the website (see Figure A).

Figure A

The network you use to connect to the Internet can also be set up to use a system called a TLS proxy. TLS proxies sit in the middle of your secure connection to the websites you view (see Figure B). At the TLS proxy your Internet traffic is decrypted and the web proxy can view and modify it. Afterwards, the TLS proxy will then re-encrypt your traffic and forward it along. This is done silently and without the knowledge of you or the website you connect to.

Figure B

TLS proxies can be set up by the organization that controls your Internet (for example, your ISP, school, or employer) and also by malicious attackers. TLS proxies have many different uses:

Protective	**Malicious**
Blocking malware and viruses	Stealing passwords
Protecting company secrets	Identity theft
Blocking harmful websites	Tracking government dissidents
Catching malicious individuals	Spying (for example the NSA)
	Censorship

Figure 6.1: TLS Proxy Description

open-ended questions, the first asking them what concerns they might have and the second asking them how it would affect their opinion of the organization running the TLS proxy. Finally, participants are given a chance to express any remaining comments they might have.[1]

6.3 Survey Development

Before running our survey, we conducted a pilot survey using MTurk to ensure that we would get meaningful and thoughtful results. This pilot survey was IRB approved and included 80 participants. Based on our analysis of participants' answers in this pilot survey, it was clear that participants generally understood the description of TLS proxies presented to them, and so we proceeded to launch the full survey. Responses from the pilot survey are not included in our results.

6.4 Qualitative Data Analysis

To better understand participants' opinions regarding TLS proxies and to avoid biasing their responses, we included several open-ended questions in the survey. For each question, we created a codebook to categorize participant responses. One researcher reviewed all the participant responses and created the initial codebooks. The codebooks were then modified through discussion with the coders.

After coding was completed, all of the coders met together to discuss the data. As part of this discussion they were encouraged to identify themes that they had seen in the data. Particular attention was paid to the themes that they felt the codebook did not adequately cover. Coders also shared responses that they felt best represented the various viewpoints expressed by participants.

In total, there were seven coders that analyzed the data. We validated the consistency of the coders using Fleiss' Kappa [26]. Coders' agreement ranged from "substantial agreement"

[1] As shown in the Appendix, questions are grouped onto several pages. After questions on one page are answered and the user continues with the survey, they are unable to return and modify their answers.

to "almost perfect agreement" (with kappa values ranging from .687 to 1, mean of .865 and median of .833).

6.5 Amazon Mechanical Turk

We used Amazon Mechanical Turk (MTurk) to recruit survey participants. MTurk has become an increasingly popular method for gathering participant data for usability studies and user surveys. Buhrmester et al. found that MTurk participants are significantly more diverse than typical American College samples and that data obtained from MTurk studies is at least as reliable as those obtained via more traditional methods [12]. Kittur et al. used MTurk participants to classify Wikipedia entries and found that that they could produce results equivalent to expert raters [42]. While MTurk has known limitations, it is still a mostly reliable platform for rapidly obtaining results related to user sentiment [40, 61].

6.6 Quality Control

To ensure participants provided valid data, we accepted only participants that had previously completed 1,000 tasks on MTurk with an overall task approval rate of 95% or higher. Second, the seven coders examined participants' responses to open-ended questions in order to ensure that participants had both understood the description of TLS proxies and remained on topic. We validated the consistency of the coders' choice to exclude participants' responses using Fleiss' Kappa [26] and found that coders were in perfect agreement (kappa value of 1). During the coding process, a participant's responses were discarded if their answers were clearly spam (i.e., copying the text of a Wikipedia page), or they did not understand the questions being asked (i.e., their answers discussed HTTP proxies). In total, we excluded 153 participants' responses (12.1%) as spam and 60 participants (4.8%) as misunderstandings. The remaining 1,049 participants' responses constitute the results of our first survey.

6.7 Demographics

The demographics for the participants are shown in Table 6.1. Most participants were from the United States (87%), with the rest primarily from India (11.5%). Although results from a previous paper suggested that MTurk participants from India are less concerned with privacy [38], the results from our first survey found that they were more likely to report privacy concerns than their counterparts from the United States of America ($\chi^2[2, N = 1049] = 12.35, p < 0.01$).

Participants were skewed towards males (61%), and ages were centered around 25–32 (46%). Most participants were single (60%) and had no children (62%). Nearly all participants had completed high school, with the majority having completed some level of higher education (57%).

Participants were asked to self-report their level of knowledge of Internet security, with most rating somewhere between somewhat knowledgeable and mildly knowledgeable (78%).

After reading the description of TLS proxies, participants were asked whether they had prior knowledge of TLS proxies. Most participants reported having little to no awareness of TLS proxies before the survey: unaware (66.5%), unsure (8.1%), aware (25.4%). We speculate that due to the effects of illusory superiority, the number of participants that were unaware of TLS proxies before the survey was even higher than reported [28, 32]. Additionally, participants may have conflated knowledge of traditional web proxies with knowledge of TLS proxies.

6.8 Limitations

In our survey, participant demographics were slightly skewed towards a younger male population and nearly all participants were from the US and India. Additional work could be done to replicate our results with different populations. Cross-cultural, international surveys

	Survey 1 (N=1,049)	Survey 2 (N=927)
Country		
United States	86.9%	94.3%
India	11.5%	5.7%
Other	0.3%	N/A
Gender		
Male	61.1%	60.6%
Female	38.6%	38.9%
Prefer not to answer	0.3%	0.4%
Age		
18–24 years old	18.7%	17.8%
25–34 years old	47.0%	45.8%
35–44 years old	19.6%	21.8%
45–54 years old	8.6%	7.9%
55+ years old	5.8%	6.3%
Prefer not to answer	0.3%	0.4%
Relationship		
Single	59.5%	60.9%
Married	35.5%	35.6%
Other	4.7%	2.7%
Prefer not to answer	0.6%	0.8%
Children		
Yes	36.6%	32.5%
No	62.3%	67.2%
Prefer not to answer	0.9%	0.3%
Education		
No diploma	1.0%	0.6%
High school	12.4%	11.0%
Some college or university credit	28.9%	29.3%
College or university degree	49.9%	50.5%
Post-Secondary Education	7.6%	8.4%
Prefer Not To Answer	0.2%	0.1%
Knowledge		
No Knowledge	4.6%	2.6%
Somewhat Knowledgeable	35.7%	32.4%
Mildly Knowledgeable	42.4%	47.8%
Highly Knowledgeable	14.4%	15.2%
Expert	2.4%	1.8%
Prefer Not To Answer	0.2%	0.2%

Table 6.1: Participant Demographics

would be especially interesting, but these should be conducted by researchers that can engage participants in their native language and have an understanding of participants' cultural perceptions.

As shown in prior work, participants' reported security preferences and desires do not always align with their actual behaviors [67]. Often users will report being more privacy minded than they are in practice. Interestingly, in our survey participants indicated a high level of acceptance for TLS proxies, which could suggest that real-world acceptance of TLS proxies is even higher than we measured. On the other hand, many participants reported wanting to have their consent obtained, or at least be notified of, the inspection of encrypted traffic; in practice, it is possible that fewer participants would actually be interested in being notified.

Finally, while we spent considerable effort to craft a fair and unbiased description of TLS proxies and the inspection of encrypted traffic, there is still the possibility that it had a significant effect on some participants' responses. For example, in the real world, users often learn about security issues from the news, which is often sensational and biased. In contrast, our description strove for neutrality, and as such may have led to users taking a more rational view of the inspection of encrypted traffic than would occur in the wild. While we chose to accept these limitations in order to obtain opinions from as many participants as possible, an open avenue for future research is to find a way to gather equally widespread opinions in a way that has fewer limitations.

Chapter 7

Results from First Survey

In this chapter we discuss the results of our survey in three areas: acceptable uses for TLS proxies, general concerns toward their use, and the reaction participants would have if they discovered a network they use employed a TLS proxy.

7.1 Acceptable Uses of TLS Proxies

Figure 7.1 shows participant attitudes toward proxies. A somewhat surprising result is that participants largely (752; 71.7%) felt that there were acceptable uses for TLS proxies. This feeling prevailed even though nearly half of the participants (522; 49.8%) indicated that TLS proxies are an invasion of privacy, and only one-eighth of participants (185; 17.6%) felt they presented no invasion of privacy. There is a strong correlation between thinking TLS proxies were an invasion of privacy and believing that there were not acceptable uses for them ($\chi^2[4, N = 1049] = 141.50, p < 0.001$). Nevertheless, over a quarter of participants (297; 28.0%) felt that TLS proxies were an invasion of privacy, but still had acceptable uses.

To better understand what uses might be acceptable, we asked participants who felt there were acceptable uses to enumerate those uses in an open-ended question. The results from our coded responses are shown in the top part of Table 7.1. The acceptable uses are largely concentrated on three use cases:

1. **Protecting organizations (493; 65.6%).** Many participants felt that organizations (e.g., businesses, government agencies, schools, libraries) had a right to protect their own intellectual property and security. This included protecting the company from viruses

Opinion	Participants
Acceptable Uses	
Protect organizations	51.4% (n=539)
Protect individuals	34.8% (n=365)
Law enforcement and surveillance	8.9% (n=93)
Censor content	7.1% (n=75)
Never censor content	3.1% (n=32)
Acceptable at work, not at home	2.9% (n=30)
Concerns	
Hackers and spying	60.5% (n=635)
Privacy and identity theft	55.4% (n=581)
Done without knowledge or consent	13.2% (n=138)
Reactions	
Negative	60.8% (n=638)
Positive	5.0% (n=52)
Depends	34.2% (n=359)
Suspicious	25.8% (n=271)
Discontinue use	17.2% (n=180)
Change behavior (besides discontinue)	6.2% (n=65)

Table 7.1: Qualitative Response Categorization (N=1,049)

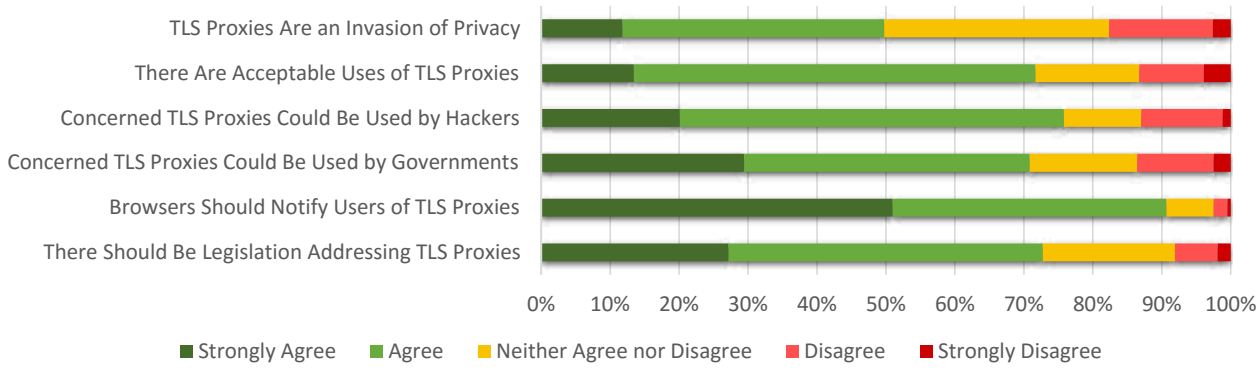

Figure 7.1: Participant Attitudes Toward TLS Proxies (N=1,049)

and hackers, filtering inappropriate or potentially malicious websites, and preventing the leak of sensitive information. Participants mentioned that since these organizations provide the Internet for their employees or constituents they had a right to use TLS proxies on their own networks.

2. **Protecting individuals (339; 45.1%).** Participants saw value in businesses using TLS proxies to protect their customers. This protection came in one of two forms:

   - **Direct.** Antivirus applications and firewalls could use TLS proxies to filter malware and viruses. Similarly, ISPs could use TLS proxies to detect and prevent phishing attackers and block other inappropriate or malicious websites.
   - **Indirect.** Participants recognized that they have a significant amount of private information stored externally on the web (e.g., at Amazon or Google). In order to protect this data, participants hoped that the companies storing their private data would employ TLS proxies internally to ensure the safety of the customer's data.

3. **Law enforcement and surveillance (65; 8.6%).** Nearly a tenth of participants expressed that law enforcement agencies should also be allowed to use TLS proxies. This includes use by local or federal agencies to track criminal or terrorist activity. Several participants also expressed that while this was a legitimate use it should only be done with a valid warrant or if there was an imminent threat to national security.

7.2 Concerns

Even though many participants in the first survey saw acceptable uses for TLS proxies, they were not without concerns or reservations. Based on our coding, we grouped these concerns into the categories shown in the middle part of Table 7.1. Three-quarters of the participants (795; 75.8%) mentioned they worried about hackers and nearly as many were concerned about the possibility for governmental spying (743; 70.9%). There was also a strong correlation between the concern that hackers could use TLS proxies and that the government could use them ($\chi^2[4, N = 1049] = 194.57, p < 0.001$).

The most visceral concerns were related to the breach of privacy. One of the open response questions asked participants to list what possible concerns they had regarding the use of TLS proxies. Over half of participants (581; 55.4%) mentioned they were concerned with a loss of privacy and personal information. Nearly a tenth of participants (104; 9.91%) mentioned having their identity stolen, and even more participants had answers that addressed the issue of identity theft generally.

A non-negligible number of the participants freely shared that either they, a family member, or other acquaintance had been the victim of account compromise. Similar to the finding of Shay et al. [58] this was a traumatic experience and it left participants especially concerned that TLS proxies could be used to perpetrate identity theft. R208 shared,

> *"A major concern that I would have would be the security of my personal and financial information. I have many friends who have been victims of identity theft and fraud, and would hate to have to go through what they did."*

Participants were also concerned that TLS proxies could be used without their knowledge. One-eighth of participants (138; 13.2%) mentioned in the open response question that they were concerned with privacy. Furthermore, when directly asked about notification, an overwhelming majority of participants (951; 90.7%) asserted they wanted to be notified by their browsers of the presence of TLS proxies. Similarly, participants largely (942; 89.8%) felt

that there should be legislation concerning TLS proxies. Most (782; 74.5%) wanted legislation to require notification, and nearly as many (701; 66.8%) wanted legislation to require consent.

7.3 Reactions

Participants in the first survey had varied responses on how they would react to learning that they currently use a network that employs TLS proxies. Based on our coding, we grouped these concerns into the categories shown in the bottom part of Table 7.1. Over half of participants (638; 60.8%) mentioned that it would negatively affect their opinion of the owner of that network. For example, R77 stated,

> *"I would be angry and would feel that organization violated my trust. I would wonder what information that organization had been collecting on me and what they planned to do with it. If it was my employer, I also would think that organization did not trust me and would consider working somewhere else."*

Still, a third of participants (359; 34.2%) said that their reaction would depend on who the owner of the network was and how they were using the proxy. For example, if the owner of the network was their employer they would not have a negative reaction, but if it was their ISP or government they would be very unhappy. Participants also mentioned that their approval would rest on whether or not any personal information was collected and/or sold and whether their consent had first been obtained. R960 explained,

> *"It would be on a case by case basis. I can see some instances where it would be understandable, but if it was going on without my consent, I would be wary of dealing with them in the future."*

Participants also mentioned ways in which their behavior would change if they learned a network was employing a TLS proxy. A quarter of participants (271; 25.8%) said that it would make them suspicious of the owner of that network. A quarter of participants (245; 23.4%) also mentioned that they would change their behavior on that network. For some

participants (180; 17.2%) this included discontinuing use of the network and its services, while others (65; 6.2%) mentioned they would change the content they looked at on the Internet or be more careful about entering personal information, including but not limited to e-commerce transactions. At the extreme, some participants mentioned they would quit their job if they found that their employer's network used a TLS proxy. For example, R127 expressed,

> *"If my employers were secretly spying on my private data, I would sue them if legally possible, and quit the job regardless."*

7.4 Personas

As our research group discussed the answers to open response questions in the first survey, it became clear that the participants could generally be classified into one of four personas: *pragmatic*, *privacy fundamentalist*, *jaded*, and *unconcerned*. After recognizing this, two members of the research group re-evaluated 90 participant responses and categorized participants into one of these four personas. The Fleiss' Kappa for this classification was 1 (i.e., perfect agreement). One researcher then classified the rest of the responses. The breakdown of participants into these categories is given in Table 7.2.[1]

Even though three of these personas have similar names to personas formulated by Westin [66], our categories are in no way based on the research of Westin. Instead, our methodology for creating personas more closely relates to that of Woodruff et al. [67], i.e., analyzing how participants indicate they would act in various privacy-related situations in order to determine their persona. Moreover, we do not intend these personas to be a definitive list of privacy personas, but rather view them as a helpful way to identify trends within our data.

[1]There were ten participants whose answers were vague enough that we did not feel comfortable classifying them as any of the personas.

Persona	Number	Percent
Pragmatic majority	802	76.5%
Privacy fundamentalist	178	17.0%
Jaded	48	4.6%
Unconcerned	11	1.0%
Unclassified	10	1.0%

Table 7.2: Participant Persona Categorization (N=1,049)

Pragmatic Majority, N=802

The pragmatic majority weighed consumer benefits and protections of public safety against costs of intrusive practices, believed that organizations should earn the public's trust, and wanted to have the opportunity to opt-out of intrusive practices. This group was strongly correlated with being more likely to feel that there were acceptable uses for TLS proxies ($\chi^2[4, N = 1028] = 230.48$, $p < 0.001$). R93 stated,

> *"I think it is perfectly acceptable for organizations (companies, schools, libraries, etc.) to use TLS proxies because it protects their computers. It keeps hackers from getting to sensitive or confidential information of the organization. In addition, it blocks harmful viruses that can cause a lot of damage and expense in repair. It can also keep individuals from accessing websites (employees from playing online games or minors from accessing pornography). It is perfectly reasonable for companies to employee[sic] this device for these purposes when an individual is using their computer. We should not expect privacy when we are using someone else's computer."*

Though the pragmatic majority all weighed consumer benefits versus intrusive practices, they were not uniform in their conclusions about where and how TLS proxies should be used. Some recognized the right of employers to use them, while others believed they should only be allowed in narrow cases such as with a warrant.

Privacy Fundamentalist, N=178

The privacy fundamentalist was generally distrustful of organizations that ask for personal information, in favor of legislation enhancing privacy, and chose privacy controls over consumer benefits when a trade-off existed between the two. These participants were strongly correlated with being more likely to feel TLS proxies were an invasion of privacy ($\chi^2[4, N = 1028] = 114.81, p < 0.001$). These participants were also more likely to support legislation of TLS proxies ($\chi^2[2, N = 1028] = 14.40, p < 0.001$).

The defining feature of the privacy fundamentalist was that they viewed privacy as so important that it could not be traded for any benefit, no matter how great. As emphatically stated by R1119, *"I believe privacy is sacrosanct and one could argue that it's a Constitutional right."*

They were also likely to relate the use of TLS proxies to more traditional methods of surveillance such as wiretapping and intercepting mail.

Jaded, N=48

Jaded individuals were aware that violations of privacy happen regularly, believed that governments conduct surveillance on the general public, and had lost hope that they can have privacy online. These participants felt that "the system" was rigged to remove any real chance of them having privacy. For example, R713 expressed,

> *"I know that it is my choice to use the internet; however, since I live in a remote area with no transportation to the nearest city (30 miles away) I am 'stuck' working and banking and doing business on the internet. I feel it is unfair to be made to choose between being 'safe' and having privacy freedom. I am especially disgusted by our government's spying behaviors and the rhetoric about it being necessary for national defense."*

Likewise, when asked about concerns regarding the use of TLS proxies, R831 shared,

"None. The government (via the NSA) is already reading everything we do and share online. So no surprises there."

Other jaded participants felt they had no choice in the matter because in the United States Internet service providers often have a monopoly.

Unconcerned, N=11

Unconcerned participants were generally trustful of organizations that ask for personal information, willing to sacrifice personal privacy to obtain consumer benefits, and not in favor of legislation to protect or enhance privacy. In our survey, we found very few unconcerned participants (1%). It is possible that the recent news regarding widespread government surveillance caused participants to be more privacy aware and sensitive. In addition, our use of qualitative data to classify participants allowed us to recognize that participants were part of the pragmatic majority even when their Likert responses might seem to indicate otherwise.

Chapter 8

Methodology of Second Survey

Our first survey revealed that participants' opinions related to TLS proxies were closely tied to the situation in which TLS proxies were being used. To better clarify user feelings in this area, we formulated a second survey in which we ask participants about a series of specific scenarios where inspection of encrypted traffic could be used. This second survey serves to give quantitative backing to the qualitative data gathered in the first survey.

We collected data for our second survey on Tuesday, February 24, 2015 between 11:02 AM and 1:06 PM (PST). Each participant could take the survey once and received $1 USD as compensation upon completing the survey. The survey begins exactly as the first survey by gathering demographic information and then instructing participants about TLS proxies and their uses, both benevolent and malicious. Participants are then asked their opinions regarding the use of TLS proxies in various circumstances. In total 1,005 people completed the online survey. The survey was also approved by our Institutional Review Board and is contained in Appendix A.2.

8.1 Survey Description

The first portion of the second survey includes the same description of TLS proxies as the first one. It then asks several questions repeated from the first survey: whether TLS proxies are an invasion of privacy and whether there are acceptable uses for TLS proxies.

The main portion of this survey asks participants their opinion regarding different situations where TLS proxies may be used to inspect encrypted traffic, such as by an employer,

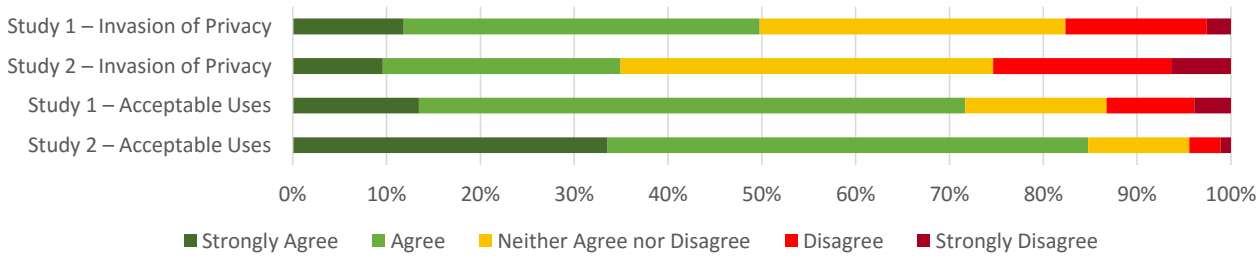

Figure 8.1: Participant Attitudes Toward TLS Proxies (Survey 1 – N=1,049, Survey 2 – N=927)

at a school, or a café with free WiFi. The full list of scenarios is given in Figure 9.1. For each situation, participants are asked whether the organization should be allowed to run a TLS proxy, with responses taken from (1) *No*, (2) *Only if I consent*, (3) *Only if I am notified (consent not required)*, (4) *Yes (neither notification nor consent required)*, or (5) *Unsure*. To choose the situations, we used responses from open-ended questions in the first survey, along with suggestions from our research team to fill out the list. Finally, we had a single open-ended question where participants could share any opinions they still had remaining at the end of the survey.

We note that this survey had the same limitations as our first survey.

8.2 Quality Control

To ensure participants provided valid data, we accepted only participants that had previously completed 1,000 tasks on MTurk with an overall task approval rate of 95% or higher. Second, we limited participants to the United States and India. This was done because with the first survey coders struggled to understand answers to free response questions from outside the United States and India.[1] Third, we looked at the single open-ended question to determine if participants had entered spam (e.g., copied an answer from Wikipedia). Finally, we used two validation questions in the second survey because there were not enough open responses to always distinguish spam entries.

[1]Moreover, these represent a small enough portion of participants that their responses had no significant effect on the data.

	Survey 1 (N=1,049)	Survey 2 (N=927)
Prior Knowledge of TLS Proxies		
Strongly Agree	4.1%	8.4%
Agree	21.3%	27.9%
Neither Agree nor Disagree	8.1%	13.2%
Disagree	48.1%	34.3%
Strongly Disagree	18.4%	16.2%

Table 8.1: Participants' Knowledge of TLS Proxies

In total, we excluded 78 participant's responses (7.8%). The remaining 927 participant's responses constitute the results of our second survey.

8.3 Demographics

The demographics for the participants were summarized earlier in Table 6.1. There were no significant differences in the demographics of the first and second surveys.

Chapter 9

Results from Second Survey

In this chapter we discuss results from our second survey. First we compare results from the three questions that were the same between both surveys. We then discuss the quantitative data regarding participants' opinions regarding different deployment scenarios for TLS proxies.

9.1 Comparison

In both surveys, after reading the description of TLS proxies, participants were asked whether they had prior knowledge of TLS proxies. These are shown in Table 8.1. In the first survey, most participants reported having little to no awareness of TLS proxies before the survey: aware (25.4%), unsure (8.1%), unaware (66.5%). In the second survey, more participants reported being aware of proxies beforehand (the difference is statistically significant, $\chi^2[4, N = 1976] = 60.003, p < 0.001$), though over half still reported having little to no awareness of TLS proxies before the survey: unaware (50.5%), unsure (13.2%), aware (36.3%). We note that 172 of our respondents in our second survey also took part in our original survey, which probably accounts for this difference. These repeat respondents were not explicitly solicited. It should also be noted that the primary purpose of this secondary survey was to gather opinions on different deployment scenarios for TLS proxies. The questions posed to gather this information were unique to the second survey. It is our belief that only two questions on the second survey could have their responses influenced by the first survey: 1) asking about prior knowledge of TLS proxies and 2) asking how well the TLS

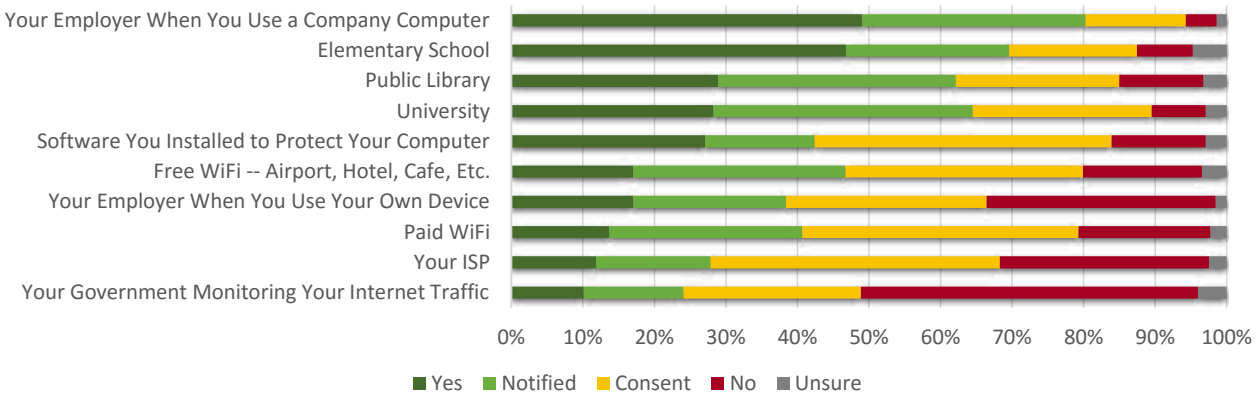

Figure 9.1: Participant Responses on Scenarios—Should the Organization Be Allowed To Run a TLS Proxy? (N=927)

proxy description assisted in helping users understand them. All other questions posed were merely demographic, spam countermeasures, or solicitations of opinions.

We also compared responses relating to whether participants in both surveys felt that TLS proxies were an invasion of privacy, and whether TLS proxies had acceptable uses (see Figure 8.1). Participants in the second survey were less likely to view TLS proxies as an invasion of privacy (first survey – 50%, second survey – 35%), with the difference being statistically significant ($\chi^2[4, N = 1976] = 54.228, p < 0.001$). Similarly, participants in the second survey were also more likely to feel that there were acceptable uses for TLS proxies (first survey – 72%, second survey – 85%), with this difference also being statistically significant ($\chi^2[4, N = 1976] = 140.654, p < 0.001$).

It is important to note that in both surveys, after participants answered each group of questions (see Appendix) participants were unable to return to earlier groups of questions and alter their answers. As such, the above reported differences are not due to differences in the survey, as up to this point the surveys were identical.

9.2 Scenarios

We asked participants regarding their opinions towards the inspection of encrypted traffic in specific scenarios. For each scenario, participants indicate whether they were comfortable with

the traffic being intercepted ("Yes"), whether they wanted to be notified ("Notified"), whether they wanted their consent to be obtained ("Consent"), or whether they were uncomfortable with it. The results for these questions are summarized in Figure 9.1.

Participants in our second survey are generally willing to accept the use of TLS proxies in most situations, with acceptance ranging from 65% to 90% of participants, when summing together those who accept it, those who desire notification, and those who desire both notification and consent. For both employers (when you use your own computer) and elementary schools, the support for using TLS proxies without notification or consent from users is surprisingly strong (455; 49.1% and 434; 46.8%). This may be due to a belief in employer rights in the first case and a desire to protect children in the second case. In both cases there is still strong support for either notification or consent (419; 45.2% and 377; 40.7%).

The strongest objections to any kind of TLS proxy are for government monitoring (437; 47.1%), using your own device at work (297; 32.0%), or using your own ISP (271; 29.2%). Note these latter two map to situations where the user has paid for the device or for network access. Users have stronger objections to TLS proxies when they pay for network access through a home ISP than when they pay for WiFi when they are away from home.

When examining the differences among opinions for notification versus consent, we see that the preference for consent is higher for personal firewalls (software you installed to protect your computer), your ISP, free WiFi, paid WiFi, and using your own device at work. The preference is higher for notification for a public library, university, elementary school, and using a company computer at work. This seems to be a clear split that favors consent in cases where the user feels in control versus notification when an organization is in control. The strongest support for consent is with a personal firewall (385; 41.5%), your ISP (375; 40.5%), and paid WiFi (358; 38.6%).

Chapter 10

Survey Open Responses

In this chapter we discuss interesting themes we saw as we analyzed participants' responses to the open-ended questions.

10.1 Informed Participants

Most of the participants showed a high level of engagement in the survey. At the end of the survey when asked if they had any additional comments, a large number of participants mentioned that they were thankful that we had informed them of this information. Some even asked where they could get more information on the topic of TLS proxies. Additionally, we were impressed with the in-depth analysis of trade-offs that many users shared, which often went far beyond the scope of any information provided to them in the survey.

Participants clearly understood that there were trade-offs involved with the use of TLS proxies and the inspection of encrypted traffic, weighing the benevolent uses for schools or workplaces and the danger of misuse by insiders or by hackers. As they struggled with this trade-off, participant responses indicated confusion, doubt, worry, equivocation, and reasoned conclusions. Confusion regarding how to resolve the conflict was evident when participants labeled it a "grey area." R988 considered both good and bad uses and worried, *"How are you supposed to know which is happening?"*

Some participants weighed the trade-offs and resolved the dilemma by deciding that proxies should only be used by consent. For example, R827 expressed:

"I believe that TLS proxies are an invasion of privacy, as is anything that monitors my internet usage without my permission. However if you are using someone else's (like a company's) network, they have every right to make the rules of use... This is one of those doubled-edged swords – it can be used for your good and security and it can be used to harm and spy on you. Because of the distinct possibility of lost privacy, this type of proxy should [not be] used, except by your agreement, not by anyone else."

Others wanted companies or schools to be able to use TLS proxies for security purposes, but also wanted to prevent them from being used for government surveillance or by hackers. Still others felt TLS proxies should *only* be used by the government to catch terrorists or criminals.

Similarly, of the participants who were against the use of TLS proxies, the reasons for opposing TLS proxies were not amorphous, but concrete and rational. For example, R666 stated:

"I think TLS proxies don't sound very safe because it sounds like an invasion of privacy. I don't think organizations should be able to decrypt your internet traffic and modify it and re-encrypt it. Perhaps they are just trying to protect against viruses and the like but it doesn't sound safe for the person using the internet. What if this technology was misused? Someone could get [h]old of your financial information for example. It sounds to[o] risky. I wouldn't want to buy something online and risk someone having access to my credit card number."

10.2 Notification and Consent

Numerous participants expressed a desire for notification and consent when TLS proxies were being used on a network. A typical response as given by R413 was,

"Well for some things it would be understandable, I'd just like to be informed so I know the risk I'm taking."

R313 expressed,

"If I encrypt something no one has the right to unencrypt it unless I give them the right to - simple as that."

Participants expressed extreme distrust for those who would use TLS proxies without informing users, going so far as to say they *"would hate them," "would wonder what they are looking for,"* and *"would assume they were up to no good."*

Others stated they would change their behavior if notified about a proxy, such as avoiding commercial transactions, using a VPN to circumvent a proxy, or self-censorship of their Google searches and other online communication.

10.3 Jaded Participants

We were surprised to find that 4.5% of participants were "jaded" towards the current state of privacy online. They felt that currently it is largely impossible to have any expectation of privacy or security. Many felt that the government was already spying on the population at large, and that even without TLS proxies the government could find a way to gain access to their private information. Others felt that even if they discovered that their traffic was being intercepted, they would have no recourse as their access to the Internet is controlled by a monopoly.

We find this group concerning, as this is not a group of individuals unconcerned with security and privacy. Rather they are a group that still cares about privacy, but has lost all hope that they can actually achieve digital privacy. This is a troubling trend, as such individuals are unlikely to adopt solutions that could actually benefit them. As such, work needs to be done to determine how this type of user's trust can be regained.

10.4 Changing Opinions

Between our two surveys, we noticed differences in the way participants viewed TLS proxies. This demonstrates that users' perceptions towards security and privacy are not static. As such, it is important that work such as this be done on a regular basis, helping the security community stay abreast of current opinions and attitudes.

One interesting difference is that in the second survey fewer participants viewed inspection of encrypted traffic as an invasion of privacy, and more participants felt that there were acceptable uses for this practice. One possible explanation for this difference is that news stories have been discussing how encryption and other privacy preserving technologies could be used by terrorist organizations. Still, additional research is needed to better understand this shift in attitudes towards security and privacy.

Part IV

Related Work and Conclusion

Chapter 11

Related Work

11.1 TLS MITM Mitigation

A large body of work seeks to detect and prevent TLS proxies, generally regarding them as MITM attacks. Clark and van Oorschot [15] provide an extensive survey of this area and provide one of the few research papers that acknowledges the existence of benevolent TLS proxies. Below we survey the various mitigation approaches in the field.

Multi-path probing allows clients to determine whether the certificate they have been given for a server is different from those seen by most other clients. Representative systems include Perspectives [65], Convergence [46], and DoubleCheck [4]. Crossbear [31] goes a step further to use traceroute to localize the origin of the attack. Other systems use existing Certificate Authorities or centralized notaries to vouch for the authenticity of a certificate [5, 6, 22]. Each of these systems may suffer from false alarms due to benign changes to certificates and the presence of multiple valid certificates for a given site [7].

There are several proposals to leverage the *shared password* between the client and server to prevent a MITM attack. Direct Validation of Certificates (DVCert) [16] permits the server to attest to the authenticity of all the certificates used in a session with the web application, including certificates from other domains. TLS Session-Aware User Authentication [53] thwarts TLS MITM attacks through user authentication tokens based on client credentials and TLS session information. The proposed TLS-SRP protocol [60] would extend the TLS handshake to support mutual authentication based on a shared password.

Several proposals leverage *DNS* to prevent MITM attacks. ConfiDNS [55] utilizes the temporal and spatial redundancy of the existing DNS system to assess agreement for IP resolution. The DNS-Based Authentication of Named Entities (DANE) [57] protocol enables administrators to bind hostnames to their certificates. This permits public keys to be transmitted via DNSSEC without involving a CA.

Certificate pinning [23] is a Google proposal for the web server to limit all future HTTPS connections to a limited set of server certificates. Pinning is a trust-on-first-use technology. The Google Chrome browser comes pre-configured with some Google certificates already pinned in advance so the user does not have the TOFU issue with those sites. Chrome also trusts any locally installed trusted roots, so benevolent proxies and malware can circumvent the pinning process. Trust Assertions for Certificate Keys (TACK) [47] is a TLS extension for the server to pin a signing key that must sign all other keys in the domain.

Another approach is to use an *audit log* of valid certificates issued by Certificate Authorities. Representative work in this area is Certificate Transparency (CT) [43, 56] and the EFF sovereign keys (SK) project [21]. The Accountable Key Infrastructure (AKI) [41] is a proposal for new infrastructure to validate public keys and reduce the reliance on CAs. This system also includes public log servers that support public validation of certificate integrity.

A different approach acknowledges that there is an industry need for TLS inspection to detect malware or protect intellectual property. Several proposals to the IETF from industry introduce mechanisms that would make proxies visible to the other participants in a chain of TLS connections and could include user notification and consent [45, 50]. Another proposal [52] introduces an extension to TLS that enables a client to share decryption keys with a TLS proxy. A potential drawback is that the key-sharing mechanism represents a new attack surface that hackers could attempt to exploit to acquire the decryption keys for a TLS session.

11.2 Measurements

The most closely related work in this field is a recent paper, published during our own work, by Huang et al., which independently develops a measurement tool that is very similar to ours and conducts a measurement study of TLS proxies that intercept the Facebook website [33]. Generally speaking, the advantage of Huang's methodology is that they find proxies specifically targeting Facebook, whereas the advantage of our methodology is that we can target our measurements for selected countries and for selected websites that have permissive Flash socket policy files. This enables us to actively collect a broader measurement of proxies.

In comparing our results to Huang, the prevalence of proxies in our study is roughly twice what was measured by Huang (0.41% versus 0.20%). In addition, we find a wider array of malware, deceptive practices, and suspicious circumstances. Both of these results are likely due to the more comprehensive measurements we make, avoiding a site such as Facebook that is likely on the whitelist for many proxies. Our measurements of WebMakerPlus, Objectify Media, Superfish, WiredTools, Internet Widgits Pty, ImpressX, and kowsar all represent malware found only in our study. Likewise, the presence of spam infections from Sweesh and AtomPark are unique to our study, as is the evidence of botnets using TLS proxies. We are the first to identify a parental filter replacing an untrusted certificate with a trusted one. Our country-specific measurement campaigns add additional data to the field.

We note that there are also some differences between the characteristics of the substitute certificates detected in our study and Huang. For instance, we find that chain depths of two or more certificates are more common. Chains with a depth of two or more certificates accounted for 20% of our substitute chains and 9% of Huang's. Note that the legitimate chains in both studies had a chain depth of two. In addition, 68 of our proxy results contained a chain depth of 5, compared to only 2 reported by Huang. Due to these depth differences, we also found more certificate chain sizes larger than 1000 bytes (20% vs 9%). We also see differences in the public key sizes of substitute certificates when comparing

our results to those of Huang. In particular, we find less certificates using 512-bit key lengths (us: 21, Huang: 119) and the presence of keys larger than 2048 (us: 7, Huang: 0).

The only other paper to find evidence of TLS proxies is the work from The Netalyzer project, which analyzes the root store of Android devices [63]. Their primary findings include the use of manufacturer and vendor-specific certificates, the presence of unusual root certs, and third party apps that manipulate the root store. In addition, they find one case of a TLS proxy, out of 15,000 assessed TLS sessions. The app whitelists several sites, including Facebook, Twitter, and several Google sites, but intercepts mail from Yahoo, Google, and traffic to several major banks. It is difficult to compare the prevalence (1 in 15K) to rates found by Huang and this paper because the sample is from users choosing to download the Netalyzer App.

Another closely related paper is the Crossbear system [31], which is designed for volunteer hunters to work together to detect and localize real-world TLS MITM attacks. After the client establishes a TLS connection to a website, the client sends the received certificate chain to a central Crossbear server. The Crossbear server establishes its own secure connection with the website and also queries Convergence for additional data about the website's certificate. This information is recorded in a database on the server and is also sent to the client. If the cumulative data received by the client suggest a MITM is present, the client performs a traceroute operation to the malicious server and sends that information to the Crossbear server. The Crossbear server attempts to localize the origin of the MITM attacker by using traceroute data from many Crossbear clients. Crossbear was deployed in 150 locations on the PlanetLab testbed and had not detected any attacks (or benevolent TLS proxies) at the time of the report.

Finally, a number of surveys collect and analyze TLS certificates and certificate authorities on the Internet [3, 7, 18, 20, 30]. These studies do not examine the use of substitute certificates by TLS proxies, but focus on issues such as TLS errors, properties of certificates and the PKI system, and poor security practices.

11.3 Surveys

There have been prior studies that survey user's attitudes about their online security and privacy. Still, no prior study has looked specifically at user attitudes toward the inspection of encrypted traffic.

McDonald and Cranor [49] used interviews and a survey to explore user's knowledge and perception of online behavioral advertising practices. They discuss the potential chilling effect of these practices based on 40% of the users that self-reported they would change their behavior if they learned advertisers were collecting data. Similarly, users reported in our survey that they would change their behavior if they learned that their encrypted data was being inspected.

Ur et al. [62] also studied user opinions about online behavioral advertising by conducting 48 semi-structured interviews with non-technical users. Similar to our work, they found users had nuanced opinions about the trade-offs for a technology that was both useful and privacy invasive. They determined that users were not receiving effective notice and choice mechanisms. Our surveys reveal a strong desire for notification and choice regarding the inspection of encrypted traffic.

Shay et al. [58] surveyed users via Amazon Mechanical Turk about their attitudes and experiences with compromised email or social networking sites. They found that many respondents gave high quality responses to open response questions and discussed implications for security mechanism designers. Likewise, our work has significance for the designers of mechanisms to inspect encrypted traffic.

Anton et al. [8] surveyed users in 2008 to see if their attitudes on privacy concerns had changed from the same survey administered in 2002. They found that the top three concerns of U.S. users were information transfer, notice/awareness, and information storage. While the top three concerns had not changed, their level of concern had risen. The top three concerns for European users were the same but in a different order; notice/awareness

came in third place. Concerns for notice/awareness are important to both groups, and was a prominent factor in our surveys.

Woodruff et al. [67] examined how well users' classification by the Westin Privacy Segmentation Index predicted their actual behavior. They found that although many participants were classified as privacy fundamentalists, their actions in hypothetical situations were not consistent with this classification. Similarly, while we group participants into personas with names similar to the Westin categories, we do so by looking at how participants indicate they would react to hypothetical situations and not using any of Westin's several privacy indexes.

Chapter 12

Conclusion

Our work provides new perspectives on TLS proxies to the security community, both from the quantitative technical viewpoint of our measurement tool as well as the qualitative viewpoint of our survey participants. Our measurement studies, obtained from clients in well over a hundred countries, exposed a variety of TLS proxies worldwide. Analysis of substitute certificate fields shows that most TLS proxies claim to be acting on behalf of users, behaving as firewalls for both personal and business use. However, since TLS proxies violate the normal hierarchy of trust, it is impossible to verify the identities professed in such fields. Despite this, we have found eight distinct, self-identifying malware which proxied over 3,600 of our total connections. Even more TLS proxy instances chose to remain entirely anonymous by providing indiscernible or no information in substitute certificate data. Our additional findings of telecom-run TLS proxies, null issuer fields, and falsified certificate authority signatures further highlight the need for transparency in this area. We find that overall, 0.41% of all connections tested are behind a TLS proxy. Given that both benevolent and malicious uses of TLS proxies use similar if not identical methodologies, distinguishing between the two is a difficult task.

The prevalence of malware using TLS proxying techniques illustrates the need for stronger controls over the root stores of browsers and operating systems. Modifying the root store should require administrative privileges, and monitoring software should be used to remove certificates from the store that are considered malicious or that are run by untrustworthy organizations. We also stress the need for better systems in the browser

and operating system to assist in both user awareness of proxy presence and distinguishing between benevolent and malicious uses of TLS proxies.

In addition, better measurement tools are needed to understand the prevalence and nature of TLS proxies. The method used by Huang is still viable, but only works to detect proxies affecting a single server. Our measurements indicate that this may undercount proxies when that server is well-known, and yet measuring at well-known servers is the only way to get large amounts of data with this method. Using a Flash advertisement provides a more scalable and robust method for detecting proxies, but this does not work if a user has an ad blocker installed. Moreover, we have found that most advertising networks no longer allow these types of advertisements. In the future, a community-driven, voluntary measurement platform would significantly help to collect these types of measurements.

The diversity of TLS proxy behavior and prevalence found in our measurement study prompted us to survey user attitudes toward TLS proxies. Our surveys, which constitute the first surveys of general (i.e., non-expert) user attitudes toward TLS proxies, contain responses from 1,976 people. Responses indicate that participants hold nuanced opinions on security and privacy trade-offs, with most recognizing legitimate uses for the proxies, but are concerned about threats from hackers or government surveillance. A significant concern about malicious uses of TLS inspection is identity theft, and many would react negatively and some would change their behavior if they discovered inspection occurring without their knowledge. We also find that a small but significant number of participants are jaded by the current state of affairs and have lost any expectation of privacy.

User attitudes toward TLS proxies provide an important data point along the spectrum of discussion that is currently taking place regarding who should have access to encrypted information. The results of our survey demonstrate that participants were generally aware of the trade-offs between privacy and security, and that most participants were willing to sacrifice some privacy for additional security. Nevertheless, participants strongly supported

notification and consent for when encrypted traffic is being inspected. We stress the importance of considering these views as the security community rises to the task of tackling TLS proxies.

References

[1] AccessData Group. AD locksmith. `http://www.accessdata.com/products/cyber-security/ad-locksmith`. Accessed: 9 January, 2014.

[2] Adobe. Adobe Flash Player PC penetration. `http://www.adobe.com/products/player_census/flashplayer/PC.html`. Accessed: 22 March, 2013.

[3] Devdatta Akhawe, Bernhard Amann, Matthias Vallentin, and Robin Sommer. Here's my cert, so trust me, maybe?: Understanding TLS errors on the web. In *International Conference on World Wide Web (WWW)*, pages 59–69, 2013.

[4] Mansoor Alicherry and Angelos D. Keromytis. Doublecheck: Multi-path verification against man-in-the-middle attacks. In *IEEE Symposium on Computers and Communications (ISCC)*, pages 557–563. IEEE, 2009.

[5] Bernhard Amann, Matthias Vallentin, Seth Hall, and Robin Sommer. Extracting certificates from live traffic: A near real-time SSL notary service. Technical report, TR-12-014, ICSI, 2012.

[6] Bernhard Amann, Matthias Vallentin, Seth Hall, and Robin Sommer. Revisiting SSL: A large-scale study of the Internet's most trusted protocol. Technical report, TR-12-015, ICSI, 2012.

[7] Bernhard Amann, Robin Sommer, Matthias Vallentin, and Seth Hall. No attack necessary: The surprising dynamics of SSL trust relationships. In *Computer Security Applications Conference (ACSAC)*, pages 179–188, 2013.

[8] Annie I. Antón, Julia B. Earp, and Jessica D. Young. How Internet users' privacy concerns have evolved since 2002. *IEEE Symposium on Security and Privacy (SP)*, pages 21–27, 2010.

[9] Adam Barth, Collin Jackson, and John C. Mitchell. Robust defenses for cross-site request forgery. In *ACM Conference on Computer and Communications Security (CCS)*, pages 75–88, 2008.

[10] Blue Coat. Proxysg. http://www.bluecoat.com/products/proxysg. Accessed: 9 January, 2014.

[11] Chad Brubaker, Suman Jana, Baishakhi Ray, Sarfraz Khurshid, and Vitaly Shmatikov. Using frankencerts for automated adversarial testing of certificate validation in SSL/TLS implementations. In *IEEE Symposium on Security and Privacy (SP)*, pages 114–129, 2014.

[12] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[13] Sam Burnett and Nick Feamster. Encore: Lightweight measurement of web censorship with cross-origin requests. *ACM SIGCOMM Computer Communication Review*, 45(4): 653–667, 2015.

[14] Tim Chiu. The growing need for SSL inspection. http://www.bluecoat.com/security/security-archive/2012-06-18/growing-need-ssl-inspection/, 2011. Accessed: 27 February, 2014.

[15] Jeremy Clark and Paul C. van Oorschot. SoK: SSL and HTTPS: Revisiting past challenges and evaluating certificate trust model enhancements. In *IEEE Symposium on Security and Privacy (SP)*, pages 511–525. IEEE, 2013.

[16] Italo Dacosta, Mustaque Ahamad, and Patrick Traynor. Trust no one else: Detecting MITM attacks against SSL/TLS without third-parties. In *European Symposium on Research in Computer Security (ESORICS)*, pages 199–216. Springer, 2012.

[17] Xavier de Carné de Carnavalet and Mohammad Mannan. Killed by proxy: Analyzing client-end TLS interception software. In *Network and Distributed System Security Symposium (NDSS)*. Internet Society, 2016.

[18] Zakir Durumeric, James Kasten, Michael Bailey, and J Alex Halderman. Analysis of the HTTPS certificate ecosystem. In *ACM Internet Measurement Conference (IMC)*, pages 291–304, 2013.

[19] Peter Eckersley and Jesse Burns. An observatory for the SSLiverse. http://www.eff.org/files/DefconSSLiverse.pdf, 2010.

[20] Peter Eckersley and Jesse Burns. The (decentralized) SSL observatory. In *USENIX Security Symposium*, 2011.

[21] Electronic Frontier Foundation (EFF). The Sovereign Keys Project. http:/www.eff.org/sovereign-keys/, 2011.

[22] Kai Engert. MECAI - mutually endorsing CA infrastructure. http://kuix.de/mecai. Accessed: 21 September, 2016.

[23] Chris Evans and Chris Palmer. Certificate Pinning Extension for HSTS. Internet-Draft draft-evans-palmer-hsts-pinning-00, Internet Engineering Task Force, November 2011. URL https://tools.ietf.org/html/draft-evans-palmer-hsts-pinning-00. Work in Progress.

[24] Sascha Fahl, Marian Harbach, Thomas Muders, Lars Baumgärtner, Bernd Freisleben, and Matthew Smith. Why Eve and Mallory love Android: An analysis of Android SSL (in) security. In *ACM Conference on Computer and Communications Security (CCS)*, pages 50–61, 2012.

[25] Sascha Fahl, Marian Harbach, Henning Perl, Markus Koetter, and Matthew Smith. Rethinking SSL development in an appified world. In *ACM Conference on Computer and Communications Security (CCS)*, pages 49–60, 2013.

[26] Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378, 1971.

[27] Martin Georgiev, Subodh Iyengar, Suman Jana, Rishita Anubhai, Dan Boneh, and Vitaly Shmatikov. The most dangerous code in the world: validating SSL certificates in non-browser software. In *ACM Conference on Computer and Communications Security (CCS)*, pages 38–49, 2012.

[28] Arthur M. Glenberg, Alex Cherry Wilkinson, and William Epstein. The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10(6): 597–602, 1982.

[29] Google. Google trends globally trending keywords. http://www.google.com/trends/?geo. Accessed: 1 January, 2014.

[30] Ralph Holz, Lothar Braun, Nils Kammenhuber, and Georg Carle. The SSL landscape: a thorough analysis of the X.509 PKI using active and passive measurements. In *ACM Internet Measurement Conference (IMC)*, pages 427–444, 2011.

[31] Ralph Holz, Thomas Riedmaier, Nils Kammenhuber, and Georg Carle. X.509 forensics: Detecting and localising the SSL/TLS men-in-the-middle. In *European Symposium on Research in Computer Security (ESORICS)*, pages 217–234. Springer, 2012.

[32] Vera Hoorens. Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality*, 63(4):793–817, 1995.

[33] Lin-Shung Huang, Alex Rice, Erling Ellingsen, and Collin Jackson. Analyzing forged SSL certificates in the wild. In *IEEE Symposium on Security and Privacy (SP)*, pages 83–97, 2014.

[34] Geoff Huston. Counting DNSSEC. `https://labs.ripe.net/Members/gih/counting-dnssec`. Accessed: 26 February, 2014.

[35] Geoff Huston and George Michaelson. Measuring DNSSEC performance. `http://potaroo.net/ispcol/2013-05/dnssec-performance.html`. Accessed: 26 February, 2014.

[36] Collin Jackson, Adam Barth, Andrew Bortz, Weidong Shao, and Dan Boneh. Protecting browsers from DNS rebinding attacks. *ACM Transactions on the Web (TWEB)*, 3(1):2, 2009.

[37] Ben Jones, Roya Ensafi, Nick Feamster, Vern Paxson, and Nick Weaver. Ethical concerns for censorship measurement. In *ACM SIGCOMM Workshop on Ethics in Networked Systems Research*, pages 17–19, 2015.

[38] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. Privacy attitudes of Mechanical Turk workers and the US public. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 37–49, 2014.

[39] Gregg Keizer. Hackers spied on 300,000 Iranians using fake Google certificate. `http://www.computerworld.com/article/2510951/cybercrime-hacking/hackers-spied-on-300-000-iranians-using-fake-google-certificate.html`. Accessed: 27 October, 2015.

[40] Patrick G. Kelley. Conducting usable privacy & security studies with Amazon's Mechanical Turk. In *Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[41] Tiffany Hyun-Jin Kim, Lin-Shung Huang, Adrian Perring, Collin Jackson, and Virgil Gligor. Accountable key infrastructure (AKI): A proposal for a public-key validation infrastructure. In *International Conference on World Wide Web (WWW)*, pages 679–690, 2013.

[42] Aniket Kittur, Ed H. Chi, and Bongwon Suh. Crowdsourcing user studies with Mechanical Turk. In *Conference on Human Factors in Computing Systems (SIGCHI)*, pages 453–456, 2008.

[43] Ben Laurie, Adam Langley, and Emilia Kasper. Certificate Transparency. RFC 6962, October 2015. URL https://rfc-editor.org/rfc/rfc6962.txt.

[44] Wilson Lian, Eric Rescorla, Hovav Shacham, and Stefan Savage. Measuring the practical impact of DNSSEC deployment. In *USENIX Security Symposium*, 2013.

[45] Salvatore Loreto, Robert Skog, Hans Spaak, John Mattsson, Dan Druta, and Mohammad Hafeez. Explicit Trusted Proxy in HTTP/2.0. Internet-Draft draft-loreto-httpbis-trusted-proxy20-01, Internet Engineering Task Force, August 2014. URL https://tools.ietf.org/html/draft-loreto-httpbis-trusted-proxy20-01. Work in Progress.

[46] Moxie Marlinspike. SSL and the future of authenticity. *Black Hat USA*, 2011.

[47] Moxie Marlinspike and Trevor Perrin. Trust assertions for certificate keys. http://tack.io/, 2013.

[48] MaxMind. Geolite. http://dev.maxmind.com/geoip/legacy/geolite/#IP_Geolocation. Accessed: 27 February, 2014.

[49] Aleecia M. McDonald and Lorrie Faith Cranor. Americans' attitudes about Internet behavioral advertising practices. In *ACM Workshop on Privacy in the Electronic Society*, pages 63–72, 2010.

[50] Dr. David A. McGrew, Dan Wing, and Philip Gladstone. TLS Proxy Server Extension. Internet-Draft draft-mcgrew-tls-proxy-server-01, Internet Engineering Task Force, January 2013. URL https://tools.ietf.org/html/draft-mcgrew-tls-proxy-server-01. Work in Progress.

[51] David Meyer. Nokia: Yes, we decrypt your HTTPS data, but don't worry about it. http://gigaom.com/2013/01/10/nokia-yes-we-decrypt-your-https-data-but-dont-worry-about-it/. Accessed 10 January, 2013.

[52] Yoav Nir. A Method for Sharing Record Protocol Keys with a Middlebox in TLS. Internet-Draft draft-nir-tls-keyshare-02, Internet Engineering Task Force, September 2012. URL https://tools.ietf.org/html/draft-nir-tls-keyshare-02. Work in Progress.

[53] Rolf Oppliger, Ralf Hauser, and David Basin. SSL/TLS session-aware user authentication–or how to effectively thwart the man-in-the-middle. *Computer Communications*, 29(12): 2238–2246, 2006.

[54] Palo Alto Networks. Decryption. https://www.paloaltonetworks.com/products/features/decryption.html. Accessed: 27 February, 2014.

[55] Lindsey Poole and Vivek S. Pai. ConfiDNS: Leveraging scale and history to improve DNS security. In *Workshop on Real, Large, Distributed Systems (WORLDS)*, volume 6, 2006.

[56] Mark D. Ryan. Enhanced certificate transparency and end-to-end encrypted mail. In *Network and Distributed System Security Symposium (NDSS)*. Internet Society, 2014.

[57] Jakob Schlyter and Paul E. Hoffman. The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. RFC 6698, October 2015. URL https://rfc-editor.org/rfc/rfc6698.txt.

[58] Richard Shay, Iulia Ion, Robert W Reeder, and Sunny Consolvo. My religious aunt asked why I was trying to sell her Viagra: experiences with account hijacking. In *Conference on Human Factors in Computing Systems (SIGCHI)*, pages 2657–2666. ACM, 2014.

[59] Symantec. Web gateway. http://www.symantec.com/web-gateway. Accessed: 9 January, 2014.

[60] David Taylor, Trevor Perrin, Thomas Wu, and Nikos Mavrogiannopoulos. Using the Secure Remote Password (SRP) Protocol for TLS Authentication. RFC 5054, March 2013. URL https://rfc-editor.org/rfc/rfc5054.txt.

[61] Blase Ur, Patrick Gage Kelley, Saranga Komanduri, Joel Lee, Michael Maass, Michelle L. Mazurek, Timothy Passaro, Richard Shay, Timothy Vidas, Lujo Bauer, et al. How does your password measure up? The effect of strength meters on password creation. In *USENIX Security Symposium*, pages 65–80, 2012.

[62] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 4:1–4:15. ACM, 2012.

[63] Narseo Vallina-Rodriguez, Johanna Amann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. A tangled mass: The Android root certificate stores. In *Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, pages 141–148. ACM, 2014.

[64] Security Week. StartSSL flaw allowed attackers to obtain SSL cert for any domain. http://www.securityweek.com/

`startssl-flaw-allowed-attackers-obtain-ssl-cert-any-domain`. Accessed: 5 September, 2016.

[65] Dan Wendlandt, David G. Andersen, and Adrian Perrig. Perspectives: Improving SSH-style host authentication with multi-path probing. In *USENIX Annual Technical Conference*, pages 321–334, 2008.

[66] Alan F. Westin. Harris-Equifax consumer privacy survey. *Atlanta, GA: Equifax Inc*, 1991.

[67] Allison Woodruff, Vasyl Pihur, Sunny Consolvo, Lauren Schmidt, Laura Brandimarte, and Alessandro Acquisti. Would a privacy fundamentalist sell their DNA for $1000... if nothing bad happened as a result? The Westin categories, behavioral intentions, and consequences. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 1–18, 2014.

Appendices

Appendix A

Surveys

A.1 First Survey

Page 1

We are conducting an academic research survey about public opinions on Internet security. The survey will take approximately 5 minutes.

We will not collect any personally identifying information. If you do not complete the survey we will not store any of your responses. If you have any questions or concerns about the information collected, please contact us at [email redacted].

Page 2

What is your gender?

- *Male*
- *Female*
- *I prefer not to answer*

What is your age?

- *18 – 24 years old*
- *25 – 34 years old*
- *35 – 44 years old*
- *45 – 54 years old*
- *55 years or older*
- *I prefer not to answer*

What is the highest degree or level of school you have completed?

- *Some school, no high school diploma*
- *High school graduate, diploma or the equivalent (for example: GED)*

- *Some college or university credit, no degree*
- *College or university degree*
- *Post-secondary education*
- *I prefer not to answer*

What is your marital status?

- *Married*
- *Single*
- *Other*
- *I prefer not to answer*

Do you have children?

- *Yes*
- *No*
- *I prefer not to answer*

In which country do you reside?

Page 3

Where are taking this survey?

- *Home*
- *Work*
- *School*
- *Library*
- *Retail (coffee shop, internet cafe, etc.)*
- *Other*
- *I prefer not to answer*

What type of Internet connection are you using?

- *Wired*
- *WiFi*
- *Cellular (3G, 4G, etc.)*
- *Other*
- *I don't know*
- *I prefer not to answer*

How knowledgeable are you about Internet security?

- *Expert*
- *Highly knowledgeable*
- *Mildly knowledgeable*
- *Somewhat knowledgeable*
- *No Knowledge*
- *I prefer not to answer*

When connecting to a website securely, for example when doing online shopping or banking, who should be able to see the contents of your Internet traffic? (Choose all that apply)

- *Me*
- *My Internet provider*
- *The website*
- *Malicious individuals*
- *Everyone*

Page 4

When you connect to the Internet you do so through some organization's network. For example, at home you connect to your Internet service provider's (ISP) network, while at work you connect to your employer's network. To protect your information from others on the network you can create secure connections to the websites you use (HTTPS). This is done automatically for you when you log into a website. The secure connection encrypts your Internet traffic so that no one else can view or modify your communication with the website (see Figure A).

Figure A

The network you use to connect to the Internet can also be set up to use a system called a TLS proxy. TLS proxies sit in the middle of your secure connection to the websites you view

(see Figure B). At the TLS proxy your Internet traffic is decrypted and the web proxy can view and modify it. Afterwards, the TLS proxy will then re-encrypt your traffic and forward it along. This is done silently and without the knowledge of you or the website you connect to.

Figure B

TLS proxies can be set up by the organization that controls your Internet (for example, your ISP, school, or employer) and also by malicious attackers. TLS proxies have many different uses:

Protective	**Malicious**
Blocking malware and viruses	Stealing passwords
Protecting company secrets	Identity theft
Blocking harmful websites	Tracking government dissidents
Catching malicious individuals	Spying (for example the NSA)
	Censorship

`Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree`

- *The above description of TLS proxies helped me to clearly understand what TLS proxies are and how they are used.*

Page 5

`Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree`

- *Prior to taking this survey, I was aware that organizations were using TLS proxies.*
- *TLS proxies are an invasion of privacy.*
- *There are acceptable uses for TLS proxies.*

`Only seen if selected "Agree" or "Strongly Agree" to acceptable uses for TLS proxies.`

Please explain which organizations should be allowed to use TLS proxies and for what purpose. (only shown on an Agree or Strongly Agree answer from above)

`Only seen if selected "Disagree" or "Strongly Disagree" to acceptable uses for TLS proxies.`

Please explain why TLS proxies should never be allowed.

Page 6

`Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree`

- *I am concerned that TLS proxies could be used by hackers to compromise my Internet security.*
- *I am concerned that TLS proxies could be used by the government to collect my personal information.*
- *Browsers should notify users if there is a TLS proxy intercepting and decrypting their Internet traffic.*
- *There should be legislation that addresses TLS proxies.*

`Only seen if selected "Agree" or "Strongly Agree" to legislation that addresses proxies.`

What should legislation that addresses TLS proxies do? (Choose all that apply)

- *Prevent their use*
- *Require organizations to obtain consent before using a TLS proxy*
- *Require organizations to inform users when a TLS proxy is being used*
- *I don't believe that legislation is required*
- *Other*

Page 7

The following statements and questions are about how you would personally react to having a TLS proxy on a network you use to connect to the Internet.

`Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree`

- *I believe TLS proxies are in use on a network I use to connect to the Internet.*

Please explain what concerns you have about a TLS proxy being used on a network you personally use to connect to the Internet.

Please explain how it would change your opinion of an organization if you discovered that they were using a TLS proxy.

If you have any other thoughts, please share them with us below:

A.2 Second Survey

Page 1

What is your gender?

- ○ *Male*
- ○ *Female*
- ○ *I prefer not to answer*

What is your age?

- ○ *18 – 24 years old*
- ○ *25 – 34 years old*
- ○ *35 – 44 years old*
- ○ *45 – 54 years old*
- ○ *55 years or older*
- ○ *I prefer not to answer*

What is the highest degree or level of school you have completed?

- ○ *Some school, no high school diploma*
- ○ *High school graduate, diploma or the equivalent (for example: GED)*
- ○ *Some college or university credit, no degree*
- ○ *College or university degree*
- ○ *Post-secondary education*
- ○ *I prefer not to answer*

What is your marital status?

- ○ *Married*
- ○ *Single*
- ○ *Other*

- *I prefer not to answer*

Do you have children?

- *Yes*
- *No*
- *I prefer not to answer*

In which country do you reside?

- *United States*
- *India*
- *Other*

How knowledgeable are you about Internet security?

- *Expert*
- *Highly knowledgeable*
- *Mildly knowledgeable*
- *Somewhat knowledgeable*
- *No Knowledge*
- *I prefer not to answer*

Page 2

When you connect to the Internet you do so through some organization's network. For example, at home you connect to your Internet service provider's (ISP) network, while at work you connect to your employer's network. To protect your information from others on the network you can create secure connections to the websites you use (HTTPS). This is done automatically for you when you log into a website. The secure connection encrypts your Internet traffic so that no one else can view or modify your communication with the website (see Figure A).

Figure A

The network you use to connect to the Internet can also be set up to use a system called a TLS proxy. TLS proxies sit in the middle of your secure connection to the websites you view (see Figure B). At the TLS proxy your Internet traffic is decrypted and the web proxy can view and modify it. Afterwards, the TLS proxy will then re-encrypt your traffic and forward it along. This is done silently and without the knowledge of you or the website you connect to.

Figure B

TLS proxies can be set up by the organization that controls your Internet (for example, your ISP, school, or employer) and also by malicious attackers. TLS proxies have many different uses:

Protective	**Malicious**
Blocking malware and viruses	Stealing passwords
Protecting company secrets	Identity theft
Blocking harmful websites	Tracking government dissidents
Catching malicious individuals	Spying (for example the NSA)
	Censorship

Ordering of questions randomized.

`Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree`

- *The above description of TLS proxies helped me to clearly understand what TLS proxies are and how they are used.*
- *Stealing passwords and identity theft are in the list of malicious uses shown above.*
- *Blocking malware and viruses are in the list of malicious uses shown above.*
- *Prior to taking this survey, I was aware that organizations were using TLS proxies.*
- *TLS proxies are an invasion of privacy.*
- *There are acceptable uses for TLS proxies.*

Page 3

For each scenario listed below, provide your opinion on whether or not the organization should be allowed to run a TLS proxy.

Ordering of questions randomized.

`No, Only if I consent, Only if I am notified (consent not required), Yes (Neither notification nor consent required), Unsure`

- *Your employer when you use a company computer*
- *Your employer when using your own device (cell phone, tablet, laptop)*
- *Elementary school*
- *Public Library*
- *University*
- *Paid WiFi – Airport, Hotel, Cafe, etc.*
- *Free WiFi – Airport, Hotel, Cafe, etc.*
- *The company that provides Internet access at your home*
- *Personal firewall – software that you have installed to protect your computer*
- *Your government monitoring your Internet traffic*

Page 4

Please feel free to write any thoughts you have on the subject of TLS proxies. We will use this information to help guide future research. (Optional)