Application Level Security in a Public Library: A Case Study

Richard Thomchick and
Tonia San Nicolas-Rocca

ABSTRACT

Libraries have historically made great efforts to ensure the confidentiality of patron personally identifiable information (PII), but the rapid, widespread adoption of information technology and the internet have given rise to new privacy and security challenges. Hypertext Transport Protocol Secure (HTTPS) is a form of Hypertext Transport Protocol (HTTP) that enables secure communication over the public internet and provides a deterministic way to guarantee data confidentiality so that attackers cannot eavesdrop on communications. HTTPS has been used to protect sensitive information exchanges, but security exploits such as passive and active attacks have exposed the need to implement HTTPS in a more rigorous and pervasive manner. This report is intended to shed light on the state of HTTPS implementation in libraries, and to suggest ways in which libraries can evaluate and improve application security so that they can better protect the confidentiality of PII about library patrons.

INTRODUCTION

Patron privacy is fundamental to the practice of librarianship in the United States (U.S.). Libraries have historically made great efforts to ensure the confidentiality of personally identifiable information (PII), but the rapid, widespread adoption of information technology and the Internet have given rise to new privacy and security challenges. The USA PATRIOT Act, the rollback of the Federal Communications Commission rules prohibiting internet service providers from selling customer browsing histories without the customer’s permission, along with electronic surveillance efforts by the National Security Agency (NSA) and other government agencies, have further intensified privacy concerns about sensitive information that is transmitted over the public internet when patrons interact with electronic library resources through online systems such as an online public access catalog (OPAC).

Hypertext Transport Protocol Secure (HTTPS) is a form of Hypertext Transport Protocol (HTTP) that enables secure communication over the public internet and provides a deterministic way to guarantee data confidentiality so that attackers cannot eavesdrop on communications. HTTPS has been used to protect sensitive information exchanges (i.e., e-commerce transactions, user authentication, etc.). In practice, however, security exploits such as man-in-the-middle attacks have demonstrated the relative ease with which an attacker can transparently eavesdrop on or hijack HTTP traffic by targeting gaps in HTTPS implementation. There is little or no evidence in the literature that libraries are aware of the associated vulnerabilities, threats, or risks, or that researchers have evaluated the use of HTTPS in library web applications. This report is intended to shed light on the state of HTTPS implementation in libraries, and to suggest ways in which libraries can evaluate and improve application security so that they can better protect the
confidentiality of PII about library patrons.

The structure of this paper is as follows. First, we review the literature on privacy as it pertains to librarianship and cybersecurity. We then describe the testing and research methods used to evaluate HTTPS implementation. A discussion on the results of the findings is presented. Finally, we explain the limitations and suggest future research directions.

LITERATURE REVIEW

The research begins with a survey of the literature on the topic of confidentiality as it pertains to patron privacy; the impact of information technology on libraries; and the use of HTTPS as a security control to protect the confidentiality of patron data when it is transmitted over the public internet. While there is ample literature on the topic of patron privacy, there appears to be a lack of empirical studies that measure the use of HTTPS to protect the privacy of data transmitted to and from patrons when they use library web applications.

The Primal Importance of Patron Privacy
Patron privacy has long been one of the most important principles of the library profession in the U.S. As early as 1939, the Code of Ethics for Librarians explicitly stated, “It is the librarian’s obligation to treat as confidential any private information obtained through contact with library patrons.” The concept of privacy as applied to personal and circulation data in library records began to appear in the library literature not long after the passage of the U.S. Privacy Act of 1974.

Today, the American Library Association (ALA) regards privacy as “fundamental to the ethics and practice of librarianship,” and has formally adopted a policy regarding the confidentiality of personally identifiable information (PII) about library users, which asserts, “confidentiality exists when a library is in possession of personally identifiable information about users and keeps that information private on their behalf.” This policy affirms language from the ALA Code of Ethics, and states that “confidentiality extends to information sought or received and resources consulted, borrowed, acquired or transmitted including database search records, reference questions and interviews, circulation records, interlibrary loan records, information about materials downloaded or placed on ‘hold’ or ‘reserve,’ and other personally identifiable information about uses of library materials, programs, facilities, or services.” With the advent of new technologies used in libraries to support information discovery, more challenges arise to protect patron privacy.

The Impact of Information Technology on Patron Privacy
Researchers have studied the impact of information technology on patron privacy for several decades. Early research by Harter and Machovec discussed the data privacy challenges arising from the use of automated systems in the library, and the associated ethical considerations for librarians who create, view, modify, and use patron records. Fouty addressed issues regarding the privacy of patron data contained in library databases, arguing that online patron records provide more information about individual library users, more quickly, than traditional paper-based files. Agnew and Miller presented a hypothetical case involving the transmission of an obscene email from a library computer, and an ensuing FBI inquiry, as a method of examining privacy issues that arise from patron internet use at the library. In addition, Merry pointed to the potential for violations of patron privacy brought about by tracking of personal information attached to electronic text supplied by publishers.
The consensus from the literature, as articulated by Fifarek, is that technology has given rise to new privacy challenges, and that the adoption of technology in the library has outpaced efforts to maintain patron privacy. This sentiment was echoed and amplified by John Berry, former ALA president, who commented that there are “deeper issues that arise from the impact of converting information to digitized, online formats” and critiqued the library profession for having “not built protections for such fundamental rights as those to free expression, privacy, and freedom.” ALA affirmed these findings and validated much of the prevailing research in a report from the Library Information Technology Association, which concluded, “User records have also expanded beyond the standard lists of library cardholders and circulation records as libraries begin to use electronic communication methods such as electronic mail for reference services, and as they provide access to computer, web and printing use.”

In more recent years, library systems have made increasing use of network communication protocols such as HTTP and focus of the literature has shifted towards internet technologies in response to the growth of trends such as cloud computing and Web 2.0. Mavodza characterizes the relevance of cloud computing as “unavoidable” and expounds on the ways in which Software-as-a-Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) and other cloud computing models “bring to the forefront considerations about . . . information security [and] privacy . . . that the librarian has to be knowledgeable about.” Levy and Bérard caution that next-generation library systems and web-based solutions are “a breakthrough but need careful scrutiny” of security, privacy, and related issues such as data provenance (i.e., where the information is physically stored, which can potentially affect security and privacy compliance requirements).

**Protecting Patron Privacy in the “Library 2.0” Era**

“Library 2.0” is an approach to librarianship that emphasizes engagement and multidirectional interaction with library patrons. Although this model is “broader than just online communication and collaboration” and “encompasses both physical and virtual spaces,” there can be no doubt that “Library 2.0 is rooted in the global Web 2.0 discussion,” and that libraries have made increasing use of Web 2.0 technologies to engage patrons. The Library 2.0 model disrupts many traditional practices for protecting privacy, such as limited tracking of user activity, short-term data retention policies, and anonymous browsing of physical materials. Instead, as Zimmer states, “the norms of Web 2.0 promote the open sharing of information—often personal information—and the design of many Library 2.0 services capitalize on access to patron information and might require additional tracking, collection, and aggregation of patron activities.” As ALA cautioned in their study on privacy and confidentiality, “Libraries that provide materials over websites controlled by the library must determine the appropriate use of any data describing user activity logged or gathered by the web server software.” The dilemma facing libraries in the Library 2.0 era, then, is how to appropriately leverage user information while maintaining patron privacy.

Many library systems require users to validate their identity through the use of a username, password, PIN code, or another unique identifier for access to their library circulation records and other personal information. However, several studies suggest the authentication process itself spawns a trail of personally identifiable information about library patrons that must be kept confidential. There is discussion in the literature about the value of using HTTPS and SSL certificates to protect patron privacy and build a high level of trust with users, and general awareness about importance of encrypting communications that involve sensitive information, such as “payment for fines and fees via the OPAC” or when “patrons are required to enter personal...
details such as addresses, phone numbers, usernames, and/or passwords.” 22 However, as Breeding observed, many OPACs and other library automation software products “don’t use SSL by default, even when processing these personalization features.” 23 These observations call library privacy practices into question, and are concerning since “hackers have identified library ILSs as vulnerable, especially when libraries do not enforce strict system security protocols.” 24

One of the challenges facing libraries is the perception that “a library’s basic website and online catalog functions don’t need enhanced security.” 25 As a matter-of-fact, one of the most common complaints against HTTPS implementation in libraries has been: “we don’t serve any sensitive information.” 26 These beliefs may be based on the historical practice of using HTTPS selectively to secure “sensitive” information and operations such as user authentication. But in recent years, it has become clear that selective HTTPS implementation is not an adequate defense. The Electronic Frontier Foundation (EFF) cautions, “Some site operators provide only the login page over HTTPS, on the theory that only the user’s password is sensitive. These sites’ users are vulnerable to passive and active attacks.” 27 Passive attacks do not alter systems or data. During a passive attack, a hacker will attempt to listen in on communications over a network. Eavesdropping is an example of a passive attack. 28 Active attacks alter systems or data. During this type of attack, a hacker will attempt to break into a system to make changes to transmitted or stored data, or introduce data into the system. Examples of active attacks include man-in-the-middle, impersonation, and session hijacking. 29

**HTTP Exploits**

Web servers typically generate unique session token IDs for authenticated users and transmit them to the browser, where they are cached in the form of cookies. Session hijacking is a type of attack that “compromises the session token by stealing or predicting a valid session token to gain unauthorized access to the web server,” often by using a network sniffer to capture a valid session ID that can be used to gain access to the server. 30 Session hijacking is not a new problem, but the release of the Firesheep attack kit in 2010 increased awareness about the inherent insecurity of HTTP and the need for persistent HTTPS. 31 In the wake of Firesheep’s release and several major security breaches, Senator Charles Schumer, in a letter to Yahoo!, Twitter, and Amazon, characterized HTTP as a “welcome mat for would-be hackers” and urged the technology industry to implement better security as quickly as possible. 32 These and other events prompted several major site operators, including Google, Facebook, PayPal, and Twitter, to switch from partial to pervasive HTTPS. Today these sites transmit virtually all web application traffic over HTTPS. Security researchers from these companies, as well as from several standards organizations such as Electronic Frontier Foundation (EFF), Internet Engineering Task Force (IETF), and Open Web Application Security Project have shared their experiences and recommendations to help other website operators implement HTTPS effectively. 33 These include encrypting the entire session, avoiding mixed content, configuring cookies correctly, using valid SSL certificates, and enabling HSTS to enforce HTTPS.

**TESTING TECHNIQUES USED TO EVALUATE HTTPS IMPLEMENTATION**

There is little or no evidence in the literature that libraries are aware of the associated vulnerabilities, threats, or risks, or that researchers have evaluated the use of HTTPS in library web applications. However, there are many methods that libraries can use to evaluate HTTPS and
SSL/TLS implementation, including automated software tools and heuristic evaluations. These methods can be combined for deeper analysis.

**Automated Software Tools**

Among the most widely used automated analysis software tools is SSL Server Test from Qualys SSL Labs. This online service “performs a deep analysis of the configuration of any SSL web server on the public internet” and provides a visual summary as well as detailed information about authentication (certification and certificate chains) and configuration (protocols, key strength, cipher suites, and protocol details). Users can optionally post the results to a central “board” that acts as a clearinghouse for identifying “insecure” and “trusted” sites. Another popular tool is SSLScan, a command-line application that, as the name implies, quickly “queries SSL services, such as HTTPS, in order to determine the ciphers that are supported.” However, these tools are limited in that they only report specific types of data and do not provide a holistic view of HTTPS implementation.

**Heuristic Evaluations**

In addition to automated software tools, librarians can also use heuristic evaluations to manually inspect the gray areas of HTTPS implementation, either to validate the results of automated software or to examine aspects not included in the functionality of these tools. One example is HTTPSNow, a service that lets users report and view information about how websites use HTTPS. HTTPSNow enables this activity by providing heuristics that non-technical audiences can use to derive a relatively accurate assessment of HTTPS deployment on any particular website or application. The project documentation includes descriptions of, and guidance for identifying, HTTP-related vulnerabilities such as use of HTTP during authenticated user sessions, presence of mixed content (instances in which content on a webpage is transmitted via HTTPS while other content elements are transmitted via HTTP), insecure cookie configurations, and use of invalid SSL certificates.

**RESEARCH METHODOLOGY**

A combination of heuristic and automated methods was used to evaluate HTTPS implementation in a public library web application to determine how many security vulnerabilities exist in the application and assess to the potential privacy risks to the library’s patrons.

**Research Location**

This research project was conducted at a public library in the western US that we will call West Coast Public Library (WCPL). This library was established in 1908 and employs ninety staff and approximately forty volunteers. In addition, it has approximately 91,000 cardholders. As part of its operations, WCPL runs a public-facing website and an integrated library system (ILS) that includes an OPAC with personalization for authenticated users.

**Test**

To conduct the test, a valid WCPL library patron account was created and used to authenticate one of the authors for access to account information and personalized features of WCPL’s OPAC. Next, the Google Chrome web browser was used to visit WCPL’s public-facing website. A valid patron name, library card number, and eight-digit PIN number were then used to gain access to online account information. Several tasks were performed to evaluate HTTPS usage. A sample search
query for the keyword “recipes” was performed in the OPAC while logged in. The description pages for two of the resources listed in the search engine result page (one printed resource and one electronic resource) were clicked on and viewed. The electronic resource was added to the online account’s “book cart” and the book cart page was viewed.

During these activities, HTTPSNow heuristics were applied to individual webpages and to the user session as a whole. The web browser’s URL address window was inspected to determine whether some or all pages were transmitted via HTTP or HTTPS. The URL icon in the browser’s address bar was clicked on to view a list of the cookies that the application set in the browser. Each cookie was inspected for the text, "Send for: Encrypted connections only," which indicates that the cookie is secure. Individual webpages were checked for the presence of mixed (encrypted and unencrypted) content. Information about individual SSL certificates was inspected to determine their validity and encryption key length. All domain and subdomain names encountered during these activities were documented. The Google Chrome web browser was then used to access the Qualys SSL Server Test tool. Each domain name encountered was submitted. Test results were then examined to determine whether any authentication or configuration flaws exist in WCPL’s web applications.

RESULTS AND DISCUSSION

Given the recommendations suggested by several organizations (e.g., EFF, IETF, OWASP), we evaluated WCPL’s web application to determine how many security vulnerabilities exist in the application, and assess the potential privacy risks to the library’s patrons. The results of tests, as discussed below, suggest that WCPL’s web application processes a number of vulnerabilities that could potentially be exploited by attackers and compromise the confidentiality of PII about library patrons. This is not surprising given the lack of research on HTTPS implementation, as well as the general consensus in the literature that technology adoption has outpaced efforts to maintain patron privacy.

Based on the results of these tests, WCPL’s website and ILS span across several domains. Some of these domains appear to be operated by WCPL, while others appear to be part of a hosted environment operated by the ILS vendor. Based on this information, it is reasonable to conclude that WCPL’s ILS utilizes a “hybrid cloud” model. In addition, random use of HTTPS is observed in the OPAC interface during the testing process. This is discussed in the following sections.

Use of HTTP During Authenticated User Sessions

Library patrons use WCPL’s website and OPAC to access and search for books and other material available through the library. Given the results of the tests, WCPL does not use HTTPS pervasively across its entire web application. During the test, we found that WCPL’s website is transmitted via HTTP by default. This was after manually entering in the URL with an “https” prefix, which resulted in a redirect to the unencrypted “http” page. We continued to test WCPL’s website and OPAC by performing a query using the search bar located on the patron account page. We found that WCPL’s OPAC transmits some pages over HTTP and others over HTTPS. For example, when a search query is performed in the search bar located on the patron account page, the search engine results page is sometimes served over HTTPS, and sometimes over HTTP (see figure 1). This behavior is not limited to specific pages; rather it appears to be random. This security flaw leaves library patrons vulnerable to passive and active attacks that exploit gaps in HTTPS implementation, which allows an attacker to eavesdrop on and hijack a user-session providing the attacker with access to private information.
Figure 1. Results of the Library’s use of HTTPS.

Presence of Mixed Content
When a library patron visits a webpage served over HTTPS, the connection with the web server is encrypted, and therefore, safeguarded from attack. If an HTTPS webpage includes content retrieved via HTTP, the webpage is only partially encrypted, leaving the unencrypted content vulnerable to attackers. Analysis of WCPL’s website did not reveal any explicit use of mixed content on the public-facing portion of the site. Test results, however, detected unencrypted content sources on some pages of the library’s online catalog. This, unfortunately, puts patron privacy at risk as attackers can intercept the HTTP resources when an HTTPS webpage loads content such as an image, iFrame or font over HTTP. This compromises the security of what is perceived to be a secure site by enabling an attacker to exploit an insecure CSS file or JavaScript function, leading to disclosure of sensitive data, malicious website redirect, man-in-the-middle attacks, phishing, and other active attacks.36

Insecure Cookie Management
Cookies are small text files, sent from a web server and stored on user computers via web browsers. Cookies can be divided into two categories: Session and Persistent. Persistent cookies are stored on the user’s hard drive until they are erased or expire. Unlike persistent cookies, session cookies are stored in memory and erased once the user closes their browser. Provided that computer settings allow for it, cookies are created when a user visits a website. Cookies can be set up such that communication is limited to encrypted communication, and can be used to remember login credentials, previous information entered into forms, such as name, mailing address, email address, and the like. Cookies can also be used to monitor the number of times a user visits a website, the pages a user visits, and the amount of time spent on a webpage.
The results of the tests suggest that WCPL’s cookie policies are inconsistent. We found two types of cookies present. Within one domain, the web application uses a JSESSION cookie that is configured to send for “secure connections only.” This indicates that the session ID cookie is encrypted during transmission. Another domain uses an ASP.NET session ID that is configured to send for any connection, which means the session ID could be transmitted in an unencrypted format. Cookies transmitted in an unencrypted format could be intercepted by an attacker in order to eavesdrop on or hijack user sessions. This leaves user privacy vulnerable given the type of information contained within cookies.

**FLAWED ENCRYPTION PROTOCOL SUPPORT**

Transport Layer Security (TLS) is a protocol designed to provide secure communication over the web. Websites using TLS, therefore, provide a secure communication path between their web servers and web browsers preventing eavesdropping, hijacking, and other active attacks. This study employed the SSL Server Test from Qualys SSL Labs to perform an analysis of WCPL’s web applications. Results of the Qualys test (see figure 2) indicate that the site does not support TLS 1.2, which means the server may be vulnerable to passive and active attacks, thereby providing hackers with access to data passed between a web server and web browser accessing the server. In addition, the application’s server platform supports SSL 2.0, which is insecure because it is subject to a number of passive and active attacks leading to loss of confidentiality, privacy, and integrity.

![Figure 2. Qualys Scanning Service Results.](image)

The vulnerabilities discovered during the testing process may be a result of uncoordinated security. This is concerning because it is a by-product of the cloud computing approach used to operate WCPL’s ILS. While libraries may have acclimated to the challenge of coordinating security measures across a distributed application, they now face the added complexity of coordinating
security measures with their vendors, who themselves may also utilize additional cloud-based offerings from third parties. As cloud technology adoption increases and cloud-based infrastructures become more complex and distributed, attackers will likely attempt to find and exploit systems with inconsistent or uneven security measures, and libraries will need to work closely with information technology vendors to ensure tight coordination of security measures.

Unencrypted communication using HTTP affects the privacy, security, and integrity of patron data. Passive attacks such as eavesdropping, and active attacks such as hijacking, man-in-the-middle, and phishing can reveal patron login credentials, search history, identity, and other sensitive information that, according to ALA, should be kept private and confidential. Given the results of the testing done in this study, it is clear that WCPL needs to revisit and strengthen their web application security measures by, according to organizations within the security community, using HTTPS pervasively across the entire web application, avoiding mixed content, configuring cookies limited to encrypted communication, using valid SSL certificates, and enabling HSTS to enforce HTTPS. Implementing improvements to HTTPS will mitigate attacks by strengthening the integrity of WCPL’s web applications, which in turn, will help protect the privacy and confidentiality of library patrons.

LIMITATIONS AND FUTURE RESEARCH

This research was performed at a public library in the western U.S. Therefore, future research is needed to study the implementation of HTTPS to increase patron privacy at other public libraries, libraries in other parts of the U.S. and in other countries. It would also be valuable to conduct similar research at libraries of different types, including academic, law, medical, and other types of special libraries. SSL Server Test from Qualys SSL Labs and HTTPSNow were used to evaluate the use of HTTPS at WCPL. The use of other evaluation techniques may generate different results.

While a major limitation of this study is the evaluation of a single public library and the implementation of HTTPS to ensure patron privacy, a next phase of research should further investigate the policies in place that are used to safeguard patron privacy. These include security education, training, and awareness programs, as well as access controls. Furthermore, Library 2.0 and cloud computing are fundamental to libraries, but create risks that could impact the ability to keep patron PII safeguarded. As such, future research should evaluate the impact Library 2.0 and cloud computing applications have on maintaining the confidentiality of patron information.

CONCLUSION

The library profession has long been a staunch defender of privacy rights, and the literature reviewed indicates strong awareness and concern about the rapid pace of information technology and its impact on the confidentiality of personally identifiable information about library patrons. Much work has been done to educate librarians and patrons about the risks facing them and the measures they can take to protect themselves. However, the research and experimentation presented in this report strongly suggest that there is a need for WCPL and other libraries to reassess and strengthen their HTTPS implementations. HTTPS is not a panacea for mitigating web application risks, but it can help libraries give patrons the assurance of knowing they take security and privacy seriously, and that reasonable steps are being taken to protect them. Finally, this report concludes that further research on library application security should be conducted to assess the overall state of application security in public, academic, and special libraries, with the
long-term objective of enabling ALA and other professional institutions to develop policies and best practices to guide the secure adoption of Library 2.0 and cloud computing technologies within a socially connected world.

REFERENCES


18 Zimmer, “Patron Privacy in the ‘2.0’ Era,” p. 44.


36 Palmer and Zhu, “How to Deploy HTTPS Correctly.”