Looking Forward

While this study shows important progress toward library websites configured to provide stronger protections for the privacy for those that use them, much room remains for improvement. The deadline has already passed for securing web-based services, with browsers now flagging nonencrypted library websites as unsecure and not trustworthy. When considering other factors such as redirection to achieve mandatory HTTPS, the current status is not as impressive. The widespread use of tracking agents, especially when available anonymization is not implemented, means even more concern regarding privacy protections.

There will not be an easy or fast track in the deployment of properly secured HTTPS on the websites of the remaining libraries still using unencrypted communication. Libraries have generally seen slow transitions away from obsolete technologies in favor of modern alternatives. Those remaining represent a long tail of libraries with very sparse resources that also have a low level of awareness about the technical issues involved. Given the current rate of transition, I would anticipate that the number of library websites that do no use HTTPS will be less than a few percent by the end of 2020.

Privacy by Design

In the future, privacy will need to be one of the key considerations in the design of library websites if they are to be consistent with library values and meet the strategic objectives of libraries. In the same way that library websites should be responsive, work with all types of devices, and meet requirements for persons with disabilities, they should also conform to requirements for privacy protection.

Strategies for Achieving Privacy-Respecting Services

Several actions could be taken to accelerate the achievement of full compliance of privacy on library websites and related services:

- Those in leadership positions in libraries should be involved in this issue. It should not be up to the discretion of technologists. Administrators should rather hold technologists accountable to provide standard privacy protections in all systems deployed by the library.
- Professional bodies, such as the American Library Association, could further strengthen their guidance for the encryption of all web-based services used by libraries to provide access to information.
- Organizations providing or distributing funding for the implementation of library websites should require that those resources support HTTPS-only communications. I observe that many of the library websites without HTTPS encryption are funded through IMLS grants.
- Technology providers, including commercial and nonprofit, should ensure that their products are developed with the ability to operate with HTTPS-only communications and that this configuration option is enabled except in the case of unusual circumstances where such a configuration would not be possible because of local dependencies. This requirement would be especially relevant to any content management systems used to manage library websites as well as online catalogs and discovery services.
- Libraries should stipulate requirements for secure communications on all technology-related services they purchase. This requirement should apply to both browser-based interfaces and behind-the-scenes communications using standard protocols like SIP, NCIP, or Z39.50 as well as APIs.

Reducing the exposure of personal information of persons visiting websites due to the placement of tracking agents will be much more difficult to achieve. There appears to be limited awareness of privacy issues related to the tracking agents for analytics and for those related to social networks or the advertising ecosystem. Libraries are well motivated to move into the realm of big data and analytics to assess and refine their services. Libraries increasingly see personalized services and targeted marketing as ways to improve engagement with their community members and to combat the existential threats to funding and support.

Progress in mitigating the threats to privacy related to the use of tracking agents can be achieved through these measures:

- **Self-auditing of websites and related resources:** Libraries should at least be aware of the tracking agents present on their web-based services. Library personnel should use tools such as Ghostery to confirm which tracking agents have been installed. In many cases, these agents may have appeared on the library site inadvertently. Libraries often borrow scripts or widgets from other libraries or from commercial sources to achieve desired visual effects or functionality. These components may in turn invoke tracking agents. An audit of the tracking agents would inform a process to identify the specific code that invokes the agent and a review regarding which agents are viewed as tolerable within the library's privacy policies and which should be eliminated.
- **Comprehensive anonymization of tracking data:** This study shows a low rate of IP anonymization in the configuration of Google Analytics. This report provides information that the anonymization configuration of Google Analytics is more consistent with protecting the privacy of the individuals that use library-provided resources. Administrators and policy makers in the library community should make recommendations, if not mandates, that anonymization of IP addresses be implemented on any service that involves tracking agents and transmission of user activity to a third party.
- **Alternative privacy-respecting services:** Libraries have a significant interest in promoting their services to their communities. As libraries work to implement marketing strategies, they should ensure that the technologies that support these efforts do not intrude on the privacy of their users in ways that may not be intended or that are inconsistent with stated policies. While it's tempting to make use of tools and frameworks provided for free by the leading technology giants, libraries must assess any compromises that these tools require relative to user privacy and pursue or develop alternatives when needed.

Ongoing Research and Analysis

This issue of *Library Technology Reports* describes the author's ongoing project in the exploration of the trends and technologies related to the security and privacy of library websites and related systems. In this phase of the work, the study has expanded beyond a focus on the largest libraries, such as the members of the Association of Research Libraries and Urban Library Council, to the comprehensive sets of public and academic libraries in the United States. This expanded scope was made possible through the development of automated tools to identify pertinent characteristics. Identifying the proportion of libraries using HTTPS or those implementing tracking agents would not be feasible through methods based on manual inspection.

The next phase of work in this area will include refinement of the automated tools to more definitively identify tracking agents and to expand the body of libraries studied to other countries. Additional work is also needed to analyze the technical interactions between tracking agents placed on library websites and the advertising ecosystem. A clearer understanding of how traces of online information-seeking behavior performed on library websites can leak into ad networks will help inform future recommendations on what tracking agents can be allowed relative to patron privacy concerns.

A complete transition to HTTPS-only communications on library websites can be considered as basic table stakes in the struggle to protect user privacy on library websites. Enforcing encryption provides protection against hypothetical intruders that might be interested in capturing the interactions of individuals with library-provided information services. In the realm of tracking agents, the adversaries are well known. The advertising-based web ecosystem seems to continually expand its appetite for personal information. Libraries will need to be ever more vigilant in the future to ensure an impermeable firewall between their services and the surrounding ad-based commercial infrastructure.

Additional References and Resources

Brantley, Peter, Marshall Breeding, Eric Hellman, and Gary Price. "Swords, Dragons, and Spells: Libraries and User Privacy." Project briefing, CNI's December 2014 member meeting. Online video, 44:23, posted January 23, 2015. https://www.cni.org/news/video-libraries-and-user-privacy.

Breeding, Marshall. "Privacy and Security for Library Systems." *Library Technology Reports* 52, no. 4 (May/June 2016).

Breeding, Marshall. "Protecting Patron Privacy." *Smart Libraries Newsletter* 36, no. 5 (May 2017).

National Information Standards Organization. *NISO Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software-Provider Systems (NISO Privacy Principles)*. White paper. Baltimore, MD: NISO, December 10, 2015. https://www.niso.org/publications/privacy-principles.

O'Brien, Patrick, Scott W. H. Young, Kenning Arlitsch, and Karl Benedict. "Protecting Privacy on the Web: A Study of HTTPS and Google Analytics Implementation in Academic Library Websites." *Online Information Review* 42, no. 6 (2018): 734–51, https://doi.org/10.1108/OIR-02-2018-0056.

Santa Cruz County Civil Grand Jury. "Patron Privacy at Santa Cruz Public Libraries: Trust and Transparency in the Age of Data Analytics." June 24, 2019. http://www.co.santa-cruz.ca.us/Portals/0/County/GrandJury/GJ2019_final/SantaCruzPublicLibrariesReport.pdf.