Practical Privacy for the Library: An Intellectual Freedom Issues Briefing

Galen Charlton
Bill Marden
Michael Robinson

American Library Association
2017 Annual Conference, Chicago
McCormick Place, Room W178a
26 June 2017
Galen Charlton
Equinox Open Library Initiative
gmc@equinoxinitiative.org
@gmcharlt

Bill Marden
The New York Public Library
williammarden@nypl.org
@mardenbill

Michael Robinson
Consortium Library, University of Alaska Anchorage
Chair, ALA Intellectual Freedom Privacy Subcommittee
mcrobinson@alaska.edu
Experts and influencers

• Matt Beckstrom
• Deborah Caldwell-Stone, ALA
• Seeta Peña Gangadharan
• Jeff Godin, Traverse Area District Library

• Alison Macrina (Library Freedom Project, @flexlibris)
• Bobbi Newman (@bobbinewman)
• Bonnie Tijerina (@bonlth)
• Jessamyn West (@jessamyn)
The Privacy Spectrum

- Completely public
- Openly-discoverable

- Completely private
- Known only to the originator of the data

“No person can stand outside of the network. It is part of the fabric of contemporary society, from which one cannot remove oneself.”

-- Peter Brantley, Univ. of California - Davis
The Visibility of Data

There are things we know we know.

We know there are some things we do not know.

There are also the unknown unknowns...

...the ones we don’t know we don’t know.
The library, as the unique sanctuary of the widest possible spectrum of ideas, must protect the confidentiality of its records in order to insure its readers' right to read anything they wish, free from the fear that someone might see what they read and use this as a way to intimidate them...

-- Supporting documentation for 1981 passage of New York State statute CPLR §4509
....Without such protection there would be a chilling effect on our library users, as inquiring minds turn away from exploring varied avenues of thought because they fear the potentiality of others knowing their reading history.

-- Supporting documentation for 1981 passage of New York State statute CPLR §4509
Types of Data Controls
(Regulating the Collection, Storage, Access & Transmission of Data)

1. Legal controls
   a. International laws (incl. GDPR)
   b. U.S. Federal laws
      i. Legislative
      ii. Judicial
   c. State
      i. Legislative (incl. specific library-related statutes)
      ii. Judicial
   d. Local
   e. Subpoenas, warrants & other legal requests
Types of Data Controls (cont’d)

2. Contractual controls (e.g., third-party vendors)

3. Behavioral controls
   a) Personal choice / personally-modifiable controls
      i. Point-in time choices (e.g., opt-in / opt-out)
   b) What one does, when and where

4. Technical controls
   a) Software
      i. Locally installed
      ii. Network installed
   b) Information security (firewalls, intrusion detection, etc.)
   c) Technical standards (ISO 27001, PCI, TRUSTe, etc.)
Configuring the ILS to enhance patron privacy
Policy and controls

• Technical controls necessarily must be guided by policy

• Policy in turn is guided by turning up as many "known knowns" as possible
Know what you store

• Patron directory information

• Circulation transactions
  — Loans
  — Hold requests

• Patron passwords
  — Passwords should be salted and hashed using a standard algorithm, but if not, an exposure can quickly snowball to systems not controlled by the library

• Credit card information
Know where you store it

• Database
  – Tables that store PII in obvious ways
  – ... and which do not — e.g., MARC records

• Filesystem
  – Logs

• Reports

• Backups
Know who can access it

• Staff access
  – Down with shared-role accounts!
  – Down with simple passwords!
  – Down with accounts that outlast staff members' tenure at the library

• Consortial and shared systems

• System and server access
  – Including by whoever hosts the ILS
Know your options

• Circulation-history retention
  — Many systems allow the connection between a loan and a patron to be severed; but often this is an option that has to be turned on

• Hold-slip printing

Image: Adapted from CC-BY 2.0 image by Tammy on Wikimedia Commons
Know how you exchange data

• As will be discussed, HTTPS for all web access

• Online protocols
  – SIP2: isn't inherently encrypted, but can be tunnelled over SSH or TLS
  – Z39.50: can also be tunnelled over TLS, although not commonly done

• Bulk data transfers
  – Upload patron data to a collections agency or a union catalog? Ensure the connection is encrypted
Know how security best practices are encouraged

- Patron password resets
- Interaction with third-party services
- Public catalog workstations

Please enter your user name or barcode to identify your library account and request a password reset

Note: You must have a valid email address associated with your library account. If not, please contact your local library for further assistance.

Barcode: 
User name: 
Submit
Know what you're monitoring

- Log analysis
- Intrusion detection
- Firewalls
Know that the whole organization is involved

• Software and network security mean little without the active cooperation of library staff

• The policy work — and getting buy-in — is as important as the technical measures
Know that protecting patron privacy is a journey

Continuous improvement
  • Best practices change
  • Threats change
  • Security technology changes
  • Staff change
  • Hosting environments and software change

Image: Diagram by Karn G. Bulsuk (www.bulsuk.com) (CC-BY 3.0)
Know some additional resources

• *Protecting Patron Privacy: A LITA Guide*
  Bobbi Newman and Bonnie Tijerina, eds.

• *Protecting Patron Privacy with Evergreen*
  Galen Charlton (Equinox)
  Jeff Godin (Traverse Area District Library)
  bit.ly/EGPatronPrivacy

• ALA/LITA Privacy Checklists
  www.ala.org/lita/advocacy
HTTPS
HTTPS & Let’s Encrypt

- Why HTTPS?
- Encrypt the web movement
- Free & easy HTTPS with Let’s Encrypt
- Examples of using Let’s Encrypt
Why HTTPS?

HTTP vs HTTPS

User

Insecure connection

Normal HTTP (80)

User

Encrypted Connection

Secure HTTPS (443)

SSL Certificate
Eavesdrop on HTTP
# Eavesdrop on HTTPS

![WireShark screenshot](image)

The screenshot shows a packet capture using WireShark for a HTTPS session. The packet details are as follows:

- **Time**: The timestamps of the packets.
- **Source**: The source IP address.
- **Destination**: The destination IP address.
- **Protocol**: The protocol used for the session.
- **Info**: Additional information about the packet, such as SSL/TLS handshake details.

### Packets Captured

- **Frame 11**: Shows a packet with a payload that is encrypted. The payload is shown as hexadecimal values.

### Key Details

- **Encrypted Application Data**: The application data is encrypted and shown in hexadecimal format.
- **Packet Details**: The screenshot includes a detailed view of the packet capture and analysis features available in WireShark.

This screenshot illustrates how to monitor and capture HTTPS traffic, highlighting the importance of securing such connections to prevent eavesdropping and data theft.
Encrypt-the-Web Movement

- Electronic Frontier Foundation launches Encrypting the Web campaign
- Federal government sites are required to be HTTPS
- Google boosts ranking of HTTPS sites in search results
- Firefox & Chrome begin to warn that HTTP sites are insecure
- Secure the News project to move news sites to HTTPS
- Library Digital Privacy Pledge encourages libraries & content providers to adopt HTTPS
Free & Easy HTTPS

- Certificate authority
- Tools for easy install
- Free
Let’s Encrypt - Sponsors

Platinum

- mozilla
- Akamai
- CISCO
- EFF
- OVH
- chrome

Gold

- IdenTrust
- gemalto
- FORD FOUNDATION

Silver

- facebook
- AUTOMATTIC
- ALA American Library Association
- shopify
- CYON
- infomaniak
- HOSTPOINT
- SiteGround
- Sucuri
- VULTR
- PlanetHoster
- netsparker
Let’s Encrypt - Adoption

- **Rapid adoption** since debut in Nov 2015
- Jan 2016: 248,000 certificates
- Jan 2017: 28 million certificates
- 50% of web is now HTTPS
Let’s Encrypt - Get Certificate

- Install on web server with modern OS in minutes using certbot client
- Available on over 100 web hosting platforms
Certbot - Package

• Certbot available as package in many repositories

• Most recent versions of Redhat, CentOS, Debian, Ubuntu, Fedora...

Install
Certbot is packaged in EPEL (Extra Packages for Enterprise Linux). To use Certbot, you must first enable the EPEL repository.

After doing this, you can install Certbot by running:

```bash
$ sudo yum install python-certbot-apache
```

Get Started
Certbot has a fairly solid beta-quality Apache plugin, which is supported on many platforms, and automates both obtaining and installing certs:

```bash
$ certbot --apache
```

If you’re feeling more conservative and would like to make the changes to your Apache configuration by hand, you can use the `certonly` subcommand:

```bash
$ certbot --apache certonly
```

To learn more about how to use Certbot read our documentation.

Automating renewal
Certbot can be configured to renew your certificates automatically before they expire. Since Let’s Encrypt certificates last for 90 days, it’s highly advisable to take advantage of this feature. You can test automatic renewal for your certificates by running this command:

```bash
certbot renew --dry-run
```

If that appears to be working correctly, you can arrange for automatic renewal by adding a `cron` or `systemd` job which runs the following:

```bash
certbot renew
```
Certbot - Download

- Certbot available for download in older distributions
- Must support Python 2.6 or better
- Certbot auto updates once installed

Install
Certbot is packaged in EPEL (Extra Packages for Enterprise Linux). To use Certbot, you must first enable the EPEL repository.

After doing this, you can install Certbot by running:

```bash
$ sudo yum install python-certbot-apache
```

Get Started
Certbot has a fairly solid beta-quality Apache plugin, which is supported on many platforms, and automates both obtaining and installing certs:

```bash
$ certbot --apache
```

If you're feeling more conservative and would like to make the changes to your Apache configuration by hand, you can use the `certonly` subcommand:

```bash
$ certbot --apache certonly
```

To learn more about how to use Certbot, read our documentation.

Automating renewal
Certbot can be configured to renew your certificates automatically before they expire. Since Let's Encrypt certificates last for 90 days, it's highly advisable to take advantage of this feature. You can test automatic renewal for your certificates by running this command:

```bash
certbot renew --dry-run
```

If that appears to be working correctly, you can arrange for automatic renewal by adding a `cron` or `systemd` job which runs the following:

```bash
certbot renew
```
Third-Party Clients

• Third-party clients can get certificates from Let’s Encrypt
• Available for various programming languages and environments
• Useful for Windows or servers without Python

Rust
• letsencrypt-rs

Windows
• ACMESharp (.NET, PowerShell)
• letsencrypt-win-simple (.NET)
• Certify GUI (.NET, WinForms)
• oocx/acme.net (.NET)
• kelunik/acme-client (PHP)

Libraries

Go
• Lego
• hlandau/acme

Java
• zero11it/acme-client
• shred/acme4j

Node.js
Web Hosting & Let’s Encrypt

- Available on over 100 web hosts
- Some with one-click installation
- WordPress plugin
- Browser-based 3rd party clients
HTTPS & Let’s Encrypt

- HTTPS does not solve all privacy problems for libraries
- But its a good start
- Recipes for Let’s Encrypt on a variety of library servers
- Library content & service providers should integrate Let’s Encrypt into their products
Privacy in Web Browsing
Privacy in Web Browsing

- Threats to privacy
- Browser best practices
- Browser extensions
- VPNs
- Tor browser
Threats to Browsing Privacy

- Activity tracking
- Fingerprinting
- Cross-site Scripting
- Open Wi-Fi
- Network Monitoring
  - Work
  - School
  - ISP
Browser Best Practices

- Use current browsers
- Use privacy mode
- Frequently delete cache, cookies
- Alternative Search Engines
  - DuckDuckGo
  - StartPage
- Use multiple browsers
  - one for anonymous surfing
  - 2nd for personalized/login
  - 3rd for banking/financial
Browser Extensions

- **HTTPS Everywhere**
  - Encrypts communications with many major websites

- **Privacy Badger**
  - Blocks spying ads & invisible trackers

- **Disconnect.me / Ghostery**
  - Visualize & block advertising, analytics & social media trackers.
VPNs

How a VPN Protects Your Data

Without a VPN:
- Data that is not sent via HTTPS is not secure
- Your IP address is visible to all websites you visit
- Data is not encrypted and is logged by your ISP

With a VPN:
- Data is encrypted to the VPN server
- Your IP changes to the one given by your VPN
- All data is encrypted and cannot be logged

Key:
- Happy ISP capturing your data
- Sad ISP that cannot capture your data
- The world wide web
- Your beautiful computer
- A mighty fine VPN server

All traffic is visible by your ISP
Which makes them happy.

All traffic is routed through your VPN
Which makes you happy!
VPNs (cont’d)

- Protects from network monitoring by ISP, work, school
- Provides some browsing anonymity
- Not just browser, covers other apps
- Not all VPNs created equal, be selective
Tor Browser

- Free browser based on Firefox
- Uses Tor Network
- Obscures IP address
- Keeps no history
- Blocks ads
- Blocks fingerprinting
- Prevents cross-site scripting

www.torproject.org/
Tor Network

My TOR client contacts TOR network which creates a random path to my destination server.

AES-encrypted http request to and from TOR

Anonymized routing through TOR network virtual circuits

Relay node

Guard node

Exit node

Public servers

Un-encrypted http packets to/from destination

Internet

Encrypted tunnel

Un-encrypted link

TOR network node
Tor Browser Issues

- Should not add Flash & other extensions
- Some scripts do not work, e.g. Captcha
- Can be slow
- Only protects browser not other apps
- Associated with the darknet
Tails – Tor’s B@d@ss Cousin

- OS for privacy & anonymity
- Based on Debian
- Boot & run from USB or CD ROM
- Leaves no trace on PC
- Uses Tor network
- Robust suite of apps

tails.boum.org
Privacy Browsing Resources

Choose Privacy Week
chooseprivacyweek.org

Library Freedom Project
libraryfreedomproject.org

SJPL Virtual Privacy Lab
www.sjpl.org/privacy
Experts and influencers

- Matt Beckstrom
- Deborah Caldwell-Stone, ALA
- Seeta Peña Gangadharan
- Jeff Godin, Traverse Area District Library
- Alison Macrina (Library Freedom Project, @flexlibris)
- Bobbi Newman (@bobbinewman)
- Bonnie Tijerina (@bonlth)
- Jessamyn West (@jessamyn)
Galen Charlton
Equinox Open Library Initiative
gmc@equinoxinitiative.org
@gmcharlt

Bill Marden
The New York Public Library
williammarden@nypl.org
@mardenbill

Michael Robinson
Consortium Library, University of Alaska Anchorage
Chair, ALA Intellectual Freedom Privacy Subcommittee
mcrobinson@alaska.edu