CS50's Introduction to Cybersecurity

OpenCourseWare

Donate (https://cs50.harvard.edu/donate)

David J. Malan (https://cs.harvard.edu/malan/)
malan@harvard.edu
(https://www.facebook.com/dmalan) (https://github.com/dmalan)
(https://www.instagram.com/davidjmalan/)
(https://www.linkedin.com/in/malan/)
(https://www.reddit.com/user/davidjmalan)
(https://www.threads.net/@davidjmalan) (https://twitter.com/davidjmalan)

Lecture 4

Preserving Privacy

- This is CS50's Introduction to Cybersecurity.
- Today, let's consider what information we are sharing without our knowledge and how we can restrict that sharing.

Web Browsing History

- Your browsing history is both a feature and a potential threat to privacy.
- You may not want someone to have access to what websites you have visited.
- You can clear your browser history. However, you may be logged out of all services.
- Servers typically have *logs* that track user activities. Therefore, even when you clear your browser history, servers keep track of what you have accessed.
- A server log may appear as follows:

  ```
  log_format combined '$remote_addr - $remote_user [$time_local] '
                      '"$request" $status $body_bytes_sent '
                      '"$http_referer" "$http_user_agent"';
  ```

  Notice that this includes your IP address, your local time, and other details are shared in these digital envelopes being shared between computers.
- How can we exert some sort of control over what we can share?

HTTP Headers

- As we discussed, HTTP Headers are key-value pairs sent between your computer and a server.
- Consider the following URL that may be visited using the link shown in the following HTML file.

  ```
  <a href="https://example.com">cats</a>
  ```

  This HTML presents a link called cats that directs the user to `example.com`.
- When you visit a website, the browser shares by default the link that referred you there.
- When you click a link, the browser shares with websites what website referred you. Hence, the following header is shared from the browser to the server:

  ```
  Referer: https://www.google.com/search?q=cats
  ```

  Notice that this header shares what you were searching for.

- Would it not be nice to be able to suppress what is being shared? Consider the following:

  ```
  Referer: https://www.google.com/
  ```

  Notice the following only shares the origin: Not the specific search you were doing.

- The following meta tags can be added to your website to restrict sharing only the origin of traffic.

  ```
  <meta name="referrer" content="origin">
  ```

  Notice the `content` attribute is set to `origin`.

- One can restrict further by adding the following to your website to provide no referrer information.

  ```
  <meta name="referrer" content="none">
  ```

  Notice the `content` attribute is set to `none`.

Fingerprinting

- Each browser presents more or less information about your identity and behavior than others.
- Regardless of the browser you choose, servers log your activities.
- *Fingerprinting* is a way by which third parties can identify you based upon clues that are available, even when you have restricted your browser from sharing as much information about you as possible.
- One such piece of information is the *User-Agent* request header, which describes your device as follows:

  ```
  Mozilla/5.0 (Linux; {Android Version}; {Build Tag etc.})
  AppleWebKit/{WebKit Rev} (KHTML, like Gecko)
  Chrome/{Chrome Rev} Mobile Safari/{WebKit Rev}
  ```

  Notice that your browser, OS version, and device are identified.

- Web servers can also locate your IP address and log it.
- Web servers can also discover your screen resolution, extensions installed, fonts installed, and other information.
- When this information is gathered together over time, it can make you more and more identifiable.

Session Cookies

- Recall cookies are like a virtual hand stamp to track you individually.
- *Session cookies* are a piece of information that servers place on your computer to identify you.
- A session cookie may appear as follows:

  ```
  HTTP/3 200
  Set-Cookie: session=1234abcd
  ```

  Notice that this cookie tells the server that your session is `1234abcd`.
- Every sequence of session numbers or characters will be unique for each user.
- Session cookies typically expire after a period of time determined by the server.

Tracking Cookie

- *Tracking cookies* are designed to track you.
- Third parties use such cookies to track your behavior on a website. Consider the following:

  ```
  Set-Cookie: _ga=GA1.2.0123456789.0; max-age=63072000
  ```

  Notice that this Google Analytics cookie lasts two years and tracks your activity by presenting itself to each new site you visit.

Tracking Parameters

- Where cookies are hidden "under the hood" of your browser, *tracking parameters* are visible in the links you access.
- Consider the following URL:

  ```
  https://example.com/ad_engagement?click_id=YmVhODI1MmZmNGU4&campaign_
  ```

  Notice that the value for `click_id`, `YmVhODI1MmZmNGU4`, tracks you specifically.
- While cookies are tracked in the background, you can see how links you visit (based on the URL) can track you.
- More and more, browsers are tending toward sanitizing tracking parameters. Consider the following URL:

```
https://example.com/ad_engagement?campaign_id=23
```

Notice that this link *only* tracks the campaign to which you are responding. There is no longer a value for `click_id`.

Third-Party Cookies

- Another type of cookie is a *third-party cookie*.
- Third parties (i.e., other servers or companies) want to understand how you travel between websites. Consider the following HTTP request:

  ```
  GET /ad.gif HTTP/3
  Host: example.com
  Referer: https://harvard.edu/
  ```

  Notice that this request specifically asks to `GET` a file called `ad.gif` from `example.com`.
- Automatically, the server responds with the following headers:

  ```
  HTTP/3 200
  Set-Cookie: id=1234abcd; max-age=31536000
  ```

  The `Set-Cookie` response header sets a cookie called `id` that lasts three years.
- If you browse to another website that utilizes the same ad, `example.com` now knows you are browsing both `harvard.edu` and `yale.edu`. Say you later make the following HTTP request:

  ```
  GET /ad.gif HTTP/3
  Cookie: id=1234abcd
  Host: example.com
  Referer: https://yale.edu/
  ```

  Notice that the third-party cookie from earlier, `id=1234abcd`, is now being shown to `example.com` again, thus revealing that you've later visited `yale.edu`.
- Third-party cookies can be used to track us and monetize information about us.

Private Browsing

- One method by which to help protect your activity is *private browsing*.
- In a private browsing window or tab, past cookies are eliminated.
- Still, the web still works as the web does! New cookies can still be formed in the ecosystem of a private browsing window.

- Even more poignant, servers can still track your activity within your single browsing session.

Supercookies

- Whoever provides your internet service can always inject their own cookies into your HTTP headers without your knowledge.
- You may be able to opt out of supercookies with your internet provider.

DNS

- The *Domain name system* or *DNS* is a service by which website names, like `harvard.edu`, are resolved to specific IP addresses.
- By convention, traffic to DNS is entirely unencrypted. Hence, you are announcing to the world what website you are attempting to visit.
- Your internet service provider and DNS services know exactly where you are attempting to visit.
- An alternative called *DNS over HTTPS* or *DoH*, as well as *DNS over TLS* or *DoT*, are services by which you can encrypt your DNS requests.

Virtual Private Network (VPN)

- VPNs, recall, are a way to connect the internet in such a way that it appears you are doing so from a different device.
- VPNs establish an encrypted connection between your own computer (`A`) and a trusted server (`B`). Server `B` then sends your request to the internet, so it appears as if your traffic is coming from `B` and not `A`.
- VPNs do not protect you if your computer is infected with malware.
- VPNs will make it appears as though your traffic is coming from the VPN's IP address instead of your own computer's.

Tor

- *Tor* is a piece of software that redirects your traffic to a node of Tor servers.
- Traffic is directed through many encrypted nodes.
- By design, the software does not remember much of your activity.

- Utilizing such a service provides a high likelihood that little can be identified about you.
- However, do note that if you are the only person utilizing Tor on a local network at your place of work or school, it is quite possible through other means to identify who you are.
- No technology provides you with absolute protection.

Permissions

- Operating systems are, by growing convention, asking to utilize certain permissions on your device.
- *Location-based services*, by default, provide your geographic location. Best to be mindful that Apple Maps and Google Maps are very much aware of where you are at any given time if you provide them with such permissions.

Summing Up

In this course, we discussed…

- Many lessons that have, hopefully, raised your awareness about what information is provided to third parties;
- How vulnerabilities arise in computers, servers, software, and your overall privacy;
- How you can mitigate these vulnerabilities with your increased awareness; and
- How you can better manage your privacy and those of others you serve.

This was CS50's Introduction to Cybersecurity.