Project Blacklight

Timothy Pasquel

CSC341 - 020

Abstract

This individual project aims to investigate how websites collect personal information that the public may not have been notified of. Students will run Blacklight against three websites, research the findings, and write up or present the outcomes. This project will reinforce Course Objectives A through D, and possibly E, from the course syllabus.

Keywords: Markup group, Blacklight, CSC341, Fall 2023

About Blacklight

About the Markup group

The Markup group is a nonprofit organization newsroom that investigates how companies and institutions are using technology to change society. Their purpose is to share with the public what they find in their investigations with these companies using data to prove their finding. Markup's approach is scientific, which means that they use datasets that they create from the ground up, and show their work with transparency by making everything public. Markup says they have embraced the idea of a Show Your Work philosophy which means that whenever and wherever possible, they will publish the data sets they use, the code they use or make, as well as describe the analysis that they have conducted during their investigation. Finally, they have a privacy compromise for their readers which means that they will not expose their viewers to third-party tracking products, they collect as little personal information as possible, and will not monetize any data they collect from users.

Introducing Blacklight

The purpose of Blacklight, a tool created by the Markup group, is to provide the public with a way to uncover how their data is being collected by websites and institutions as they are on the internet. Blacklight does a test on seven different surveillance methods when investigating a website. The first method is looking for third-party cookies (A cookie is data that is saved onto the user's device that can uniquely identify the user. It can only be read by the user who set it) which are used commonly to create a profile of a user based on their internet usage to be used in marketing activities. Next they look for ad trackers which are scrips and beacons that collect a multitude of information about a user. The third method is key logging in which third parties monitor the text that is typed in a webpage before any sort of submission is taken. Furthermore, there is session recording which is a technology that records a user's behavior on a webpage which includes mouse scrolling, mouse clicking, touch screen taps, mouse movement, and anything a user types. The fifth method is canvas fingerprinting which is an attempt to identify a user's browser without setting a cookie, this can identify a user even if they block all cookies. In addition, there is Facebook tracking which is code that enables other websites to target a user later with ads on Facebook. Lastly, there is Google Analytics ‘Remarketing Audiences’ which is a tracker that allows websites to provide targeted ads to users across the scope of the internet. The Blacklight tool matters because it provides the public with free tools that can help educate them in plain English on how websites and institutions are tracking and monitoring them so they can make more educated decisions while browsing through the internet. The tool was created by Aaron Sanking and Surya Mattu who are both a part of the Markup team. Blacklight was built using the NodeJS Javascript environment, and the Puppeteer Node library which is a browser that gives the tool more control over a Chromium browser. The tool works by taking the URL the user has entered, opening a headless web browser with a new profile, and visiting that site's homepage as well as another randomly selected page that is deeper inside the same website. While the tool does this, it runs a custom software that looks for scrips and network requests to see how a user's data is being collected. To look at the scripts, Blacklight changes some of the browser's Window Application Programming Interface (API) which allows Blacklight to keep track of which scripts made what function calls using Stacktrace-js. Network requests are logged using a tool included in Puppeteer’s API. This software is used to look for the seven surveillance methods that were described above and generate an instant report.

How to deal with your findings

There were many findings when Markup did a study using Blacklight. One of the key findings was that in an analysis of 53 web browsers in 2018, every browser had at least one loophole that allowed for tracking to still occur for their users. Another finding was that users who use Safari, a web browser that blocks cookies by default, 73 percent of ad shown to these users, the users were unable to associate with those ads. However, for people who used Chrome, only 17 percent of ads shown to Chrome users, users were unable to associate with those ads. Luckily, there are additional add-ons and actions that a user can do to protect browser privacy. One example is using web browsers such as Mozilla's Firefox, Apple Safari, and Brave as they all block third-party cookies by default. If a user likes to use Chrome then they could use extensions such as Privacy Badger and Ghostery as tracker blockers. Furthermore, a way to get rid of major trackers such as Facebook and Google trackers is to not use those two sign-ins when making or signing into an account of another website. The Facebook and Google sign-in saves users a few seconds when signing in, however, it links accounts with that website and gives these companies information as to what someone is doing. While most tracking done by websites and companies is not beneficial to the user, they can have some benefit in some niche settings. For example, many cookies on a website are called first-party cookies and are usually required to make a site function normally and are therefore not blocked by default by browsers. Therefore, these cookies would be beneficial to the user as otherwise, the website would not be as effective. Also, some users like what they call the convenience of being tracked. This means that they do not mind the idea of their information being used to help cater to a more personalized experience. For example, if a user was looking for a new kitchen knife and was searching it up a lot, getting advertisements about kitchen knives may be considered a good thing for that user as it makes it easier for them to find a new kitchen knife.

Blacklight Investigation

Website #1: https://www.walgreens.com/

1. 36 Ad trackers found on this site. This is more than the average of seven that we found on popular sites

Websites containing advertising tracking technology load JavaScript code or small invisible images that are used to either build your advertising profile or to identify you for ad targeting on this site. These techniques are often used in addition to cookies to profile you. Blacklight detected trackers on this page sending data to companies involved in online advertising. Blacklight detected scripts belonging to Oracle Corporation, OpenX Technologies Inc, and twenty-seven other companies.

2. 59 Third-party cookies found. This is more than the average of three that we found on popular sites

These are commonly used by advertising tracking companies to profile you based on your internet usage. Blacklight detected 59 third-party cookies on this site. Blacklight detected cookies set for Conversant LLC, iSpot.tv and and twenty-two others.

3. We found this website capturing user keystrokes.

Key logging is when a website captures the text that you type into a webpage before you hit the submit button. This technique has been used to identify anonymous web users by matching them to postal addresses and real names. This technique was used by four percent of popular websites when we scanned them in September 2020.

On the site you are inspecting, information entered in the name, family-name, given-name fields were logged. Blacklight detected a script belonging to the company Walgreen Co. doing this on this site. However… There are other reasons for key logging, such as providing autocomplete functionality. Blacklight cannot determine the intent behind the inspected website's use of this technique.

Website #2: www.summitracing.com

1. 8 Ad trackers found on this site. This is more than the average of seven that we found on popular sites

Websites containing advertising tracking technology load JavaScript code or small invisible images that are used to either build your advertising profile or to identify you for ad targeting on this site. These techniques are often used in addition to cookies to profile you. Blacklight detected trackers on this page sending data to companies involved in online advertising. Blacklight detected scripts belonging to **Microsoft Corporation**, **Reddit Inc.**, and two other companies.

2. 7 Third-party cookies found. This is more than double the average of three that we found on popular sites

These are commonly used by advertising tracking companies to profile you based on your internet usage. Blacklight detected **7** third-party cookies on this site. Blacklight detected cookies set for **Oracle Corporation**, **Alphabet, Inc.** and and two others.

3. This website loads trackers on your computer that are designed to evade third-party cookie blockers.

Canvas fingerprinting was detected on this website. This technique is designed to identify users even if they block third-party cookies. It can be used to track users' behavior across sites. This technique was used by six percent of popular sites when we scanned them in September 2020. Blacklight detected a script loaded from **summitracing.com** doing this on this site. It secretly draws the following image on your browser when you visit this website for the purpose of identifying your device.

4. When you visit this site, it tells Facebook.

The Facebook pixel is a snippet of code that sends data back to Facebook about people who visit this site and allows the site operator to later target them with ads on Facebook. A Facebook spokesperson told The Markup that the company set up this system so that a user doesn’t have to be “simultaneously logged into Facebook and viewing a third-party website for our business tools to function.” Common actions that can be tracked via pixel include viewing a page or specific content, adding payment information, or making a purchase. The Facebook pixel appeared in thirty percent of popular websites when we scanned them in September 2020.

5. This site allows Google Analytics to follow you across the internet.

This site uses Google Analytics and seems to use its ”remarketing audiences” feature that enables user tracking for targeted advertising across the internet. This feature allows a website to build custom audiences based on how a user interacts with this particular site and then follow those users across the internet and target them with advertising on other sites using Google Ads and Display & Video 360. A Google spokesperson told The Markup that site operators are supposed to inform visitors when data collected with this feature is used to connect this browsing data with someone’s real-world identity. You know when those shoes you were looking at follow you around the internet? This is one of the trackers leading to that. This feature appeared in fifty percent of popular websites when we scanned them in September 2020.

Website #3: https://apnews.com/

1. 88 Ad trackers found on this site. This is more than the average of seven that we found on popular sites

Websites containing advertising tracking technology load JavaScript code or small invisible images that are used to either build your advertising profile or to identify you for ad targeting on this site. These techniques are often used in addition to cookies to profile you. Blacklight detected trackers on this page sending data to companies involved in online advertising. Blacklight detected scripts belonging to **Adelphic, Inc.**, **Collective Roll**, and sixty-nine other companies.

2. 131 Third-party cookies found. This is more than the average of three that we found on popular sites

These are commonly used by advertising tracking companies to profile you based on your internet usage. Blacklight detected **131** third-party cookies on this site. Blacklight detected cookies set for **Nativo, Inc**, **Unruly Group Limited** and and fifty-eight others

3. This site allows Google Analytics to follow you across the internet.

This site uses Google Analytics and seems to use its ”remarketing audiences” feature that enables user tracking for targeted advertising across the internet. This feature allows a website to build custom audiences based on how a user interacts with this particular site and then follow those users across the internet and target them with advertising on other sites using Google Ads and Display & Video 360. A Google spokesperson told The Markup that site operators are supposed to inform visitors when data collected with this feature is used to connect this browsing data with someone’s real-world identity. You know when those shoes you were looking at follow you around the internet? This is one of the trackers leading to that. This feature appeared in fifty percent of popular websites when we scanned them in September 2020.

InfoSec Applications

Course Objectives C and A:

The security of information systems should be one of a business's main priorities whenever possible, with respect to business. Without security, information such as customer data, business secrets, intellectual property, user names and passwords, and critical business infrastructures could all be at risk. If a business would lose or damage would occur to any of these categories, the image of the business would decrease and they would lose customers, shareholders, and profit. One specific security threat that was found in the analysis listed above that could affect the security of information systems would be key loggers. Key loggers are a piece of software that captures a user's keystrokes while on a site. This means that if a user is putting their username a password in on a site and that site is also using a key logger to monitor the user's keystrokes, the user's username and password are now exposed and are unprotected. This security threat is a huge issue from both the infosec perspective as well as from the user’s perspective. From the infosec preservative it is a threat because, if a site has a key logger and that user’s username and password are compromised by it, there is now a large breach in security for the company. For example, if this was done to a company executive such as the Chief Executive Officer (CEO), whoever owns the key logger now has access to the CEO’s username and password. Even if the person who created the key logger is within the company, that person should not have access to that kind of data as they can then act as the CEO without proper permission. Furthermore, this is a security issue for the user because on average, people do not have usernames or passwords that deviate much often. This means that if a key logger swipes one of these passwords, it is possible that they just stole every one of that user's passwords, or have a close match to other usernames and passwords that they have. Overall, key logging is a huge threat to security from both the infosec side as well as the user side.

Course Objective B:

When it comes to information security, the topic can get fairly complicated quickly. Many of these complicated problems have solutions, however, for some companies it can be hard to get these solutions into practice. To do this properly, it is important to enact policies so that people follow them and information can remain safe and secure. An example would be that key logging as listed above can be a major threat to a company. Key logging is the third method of surveillance measured by the Markup group as stated on page 3 of this document. This problem can also be seen directly in the Walgreens example listed on page 6 when the Markup tool states they capture user’s keystrokes. As an acting Chief Information Security Officer (CISO), the person creating the information security policies, it would be in good taste to enact a policy that is similar to the one developed by System Admin, Audit, Network, and Security (SANS). A good password policy will make sure that if key loggers are present, they can only affect a user for so long. An adjusted version based on passwords is below:

1. Overview

Passwords are a critical aspect of computer security. A weak or compromised password can result in unauthorized access to our most sensitive data and/or exploitation of our resources. All staff, including contractors and vendors with access to Pasquel LLC systems, are responsible for taking the appropriate steps, as outlined below, to select and secure their passwords.

2. Purpose

The purpose of this policy is to establish a standard for the secure use and protection of all work related passwords.

3. Scope

The scope of this policy includes all personnel who have or are responsible for an account (or any form of access that supports or requires a password) on any system that resides at any Pasquel LLC facility, has access to the Pasquel LLC network, or stores any non-public Pasquel LLC information.

4. Policy

4.1 Password Creation and Use

4.1.1 All user-level and system-level passwords must conform to the Password Construction Guidelines.

4.1.2 Users must use a separate, unique password for each of their work related accounts. Users may not use any work related passwords for their own, personal accounts.

4.1.3 Staff are allowed to use authorized, approved password managers to securely store and manage all their work related passwords.

4.1.4 User accounts that have system-level privileges granted through group memberships or programs such as sudo must have a unique password from all other accounts held by that user to access system-level privileges. In addition, it is highly recommended that some form of multi-factor authentication is used for any privileged accounts

CONSENSUS POLICY RESOURCE COMMUNITY

4.2 Password Change

4.2.1 Passwords should be changed when there is reason to believe a password has been compromised or fails to meet our Password Creation Requirements.

4.2.1 Passwords should be changed regularly. For example, every 3-4 months.

4.3 Password Protection

4.3.1 Passwords must not be shared with anyone, including supervisors and coworkers. All passwords are to be treated as sensitive, Confidential Pasquel LLC information. Corporate Information Security recognizes that legacy applications do not support proxy systems in place. Please refer to the technical reference for additional details.

4.3.2 Passwords must not be inserted into email messages or other forms of electronic communication, nor revealed over the phone to anyone.

4.3.3 Passwords may be stored only in password managers authorized by the organization.

4.3.4 Do not use the "Remember Password" feature of applications (for example, web browsers).

4.3.5 Any individual suspecting that their password may have been compromised must report the incident and change all relevants passwords.

4.4 Application Development

Application developers must ensure that their programs contain the following security precautions:

4.4.1 Applications must support authentication of individual users, not groups.

4.4.2 Applications must not store passwords in clear text or in any easily reversible form.

4.4.3 Applications must not transmit passwords in clear text over the network.

4.4.4 Applications must provide for some sort of role management, such that one user can take over the functions of another without having to know the other's password.

4.5 Multi-Factor Authentication

4.5.1 Multi-factor authentication is require for work related accounts and encouraged for personal accounts

5. Policy Compliance

5.1 Compliance Measurement

CONSENSUS POLICY RESOURCE COMMUNITY

The Infosec team will verify compliance to this policy through various methods, including but not limited to, business tool reports, internal and external audits, and feedback to the policy owner.

5.2 Exceptions

Any exception to the policy must be approved by the Infosec team in advance.

5.3 Non-Compliance

An employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment.

6. Related Standards, Policies and Processes

- Password Construction Guidelines

7. Revision History

Date of Change	Responsible	Summary of Change
Sept. 2023	CISO Pasquel	Updated and Converted to New Format

Course Objective D:

As discussed above, key logging is a major threat to both the individual as well as to an organization. Again, page 3 defines how key logging is a surveillance method, and page 6 shows how key logging is a real problem with the Walgreen example. Luckily, there are ways to help combat this threat using the CIA triad. Again, that triad is composed of Confidentiality, Integrity, and Availability with each section serving their specific purpose in security. As an acting CISO, the company could use confidentiality against key logging by using a two-factor authentication along with a password. This means that even if a password is stolen by a key logger, the adversary still could not get into an account without the two-factor authentication that would be held interdependently by the user. This would be functional and easy to apply to both an entire organization as well as for an individual. Two-factor authentications are usually just simple text messages through either a Short Messaging Service (SMS) or through a curated application. For Integrity, it would be important to ensure that the password logs kept by the business do not change by any means necessary. This is vital because if someone or something were to change a password in a password log, then the integrity of the data has changed and that could be an organization’s employee or an independent user would not be able to log into the system to conduct their role. Lastly, there is Availability. Availability could be implemented by only allowing system admins or high tier employees to access password logs. The reason for this is in case an employee or an individual forgets their password, there still needs to be a person there to change it so that they can continue to have access to their account for their personal function. However, only specific, high-level, trustworthy people should have access to this kind of data. In the end, key logging is a threat that should not be taken lightly for either an organization or an individual, but there are ways to counter act such as using the CIA triad to ensure password security.

References

SANS. 2022, October. Security Policies Templates. https://www.sans.org/information-security-policy/?category=application-security

The Markup. 2020, September 22. The Markup. https://themarkup.org/archive