Project Blacklight

Ryan Livinghouse

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

Blacklight is a nonprofit newsroom that delves into the ways powerful entities use technology to influence society. Their main purpose is to carry out meaningful, data-driven investigations into how technology impacts society. They aim to serve the public through their journalism. Their approach is described as scientific. They will construct datasets from the ground up, verify their reporting, and operate under "The Markup Method," emphasizing transparency throughout their investigative process.

Their work philosophy is "Show Your Work." They strive for complete transparency, often publishing the datasets, code, and methodologies used in their investigations. Blacklight also distributes their stories through various channels. Finally, they always offer their work to be republished free under the Creative Commons license.

Their privacy promise states that they guarantee not to subject their readers to third-party tracking. They commit to collection minimal personal information from their readers. Blacklight vows to never monetize any collected data, emphasizing that readers' privacy is of utmost importance, even if it means increased operation complexities and costs for them.

Note that if you run the Blacklight website through the Blacklight software, absolutely nothing shows up aside from a small message. The message goes on to explain how the markup has ensured that their website is free from trackers in line with their privacy pledge to collect minimal data from readers. They avoid cookies and don't share user data with advertisers at ALL. They then go on to explain that building a tracker-free site was no easy feat, but they prioritize user privacy over monetary gain.

The Markup represents a new kind of media organization, prioritizing data-driven investigations and placing a strong emphasis on transparence and the privacy of their readers.

Introducing Blacklight

Blacklight serves as a real-time website privacy inspector. Its main objective is to inform users about the various tracking technologies websites use upon their visit, providing insights into data beign collected without the users' knowledge. The seven surveillance methods include:

1. Small pieces of data that tracking companies store in your web browser when you visit a website. These cookies identify users when they visit other websites containing tracking code from the same company.
2. Ad trackers will keep an eye on users' online behavior for targeted advertising.
3. Text monitors that will report whatever users type back to a server are key loggers.
4. Surveillance that monitors and records all of a user's behavior on a webpage, including mouse movements, clicks, scrolling, and anything typed into a form is a type of session recording.
5. Canvas fingerprinting is a method websites use to uniquely identify and track visitors without using cookies.
6. When sites track users using Facebook's capabilities is called Facebook tracking.
7. Google Analytics "Remarking Audiences" refers to Google Analytics' feature that allows advertisers to target ads to specific users based on their past interactions with a website.

So why does such a tool matter? As concerns about online privacy increase and surveillance technologies multiply, tools such as Blacklight offer transparency to sites that hide the fact that they track us. It gives users knowledge about what personal data websites might be collecting, giving the power back to the users. Without checks like these, the difficulty in finding out which websites will track us would be harder to pinpoint.

Its creators, the Markup, a nonprofit news entity specializing in data-centric journalism, developed Blacklight. It works by utilizing a real-time inspection of any URL. By simply pasting the link in the search box, it will mimic user behaviours (to a point) on the site and log which tracking techniques are being used. It will then present its findings in easy-to-understand sections based on what types of tracking techniques were used. Blacklight uses an automated website scraping technology that will do all of this for free.

How to deal with your findings

Websites use various tracking technologies that can monitor users behavior in real time. This includes watching every mouse movement, recording keystrokes, and even playing back user interactions as real time videos. Many websites use third part cookies and tracking pixels to build

profiles about users, and then combine this with their browsing history for targeted advertising. Some websites use techniques like canvas fingerprinting, which can identify users even if cookies are blocked. Key logging and session recording are used by some sites, which can capture sensitive information like medical conditions, credit cards, and other highly personal data.

Trackers are both convenient and controversial at the same time, making them have strengths and weaknesses. Tracking technologies allow websites to offer personalized experiences and targeted advertisements. Also, first party cookies can enhance user experience by remembering user preferences and login sessions. Some weaknesses, however, include infringing on user privacy, collecting data for malicious purposes, and users being unaware of what kind of data they are giving away.

What additional add-ons/actions can protect browser privacy? Using privacy-enhanced tools like FourthParty, Privacy Badger and FP Detective is a good start. Opting for browsers like Firefox, Brave, Edge and Safari which offer stronger privacy protections by default compared to browsers whose goal it is to sell as much as they can, like Chrome. Browsers like Tor provide the highest level of privacy protection. As a user, you can also clear cookies periodically, adjust the browser's settings to block third-party cookies.

Can tracking be benificial to users? Yes, but situationally so. Tracking allows websites to offer a tailored browsing experience, showing content and ads relevant to the user's interests and history. Also, first party cookies remember user preferences like language, session, dark/light mode etc.. Finally, users might receive ads for products or services they are genuinely interested in, making online shopping more efficient.

Blacklight Investigation

Website #1: us.coca-cola.com

Blacklight Inspection Result

Blacklight works by visiting each website with a headless browser running custom software built by The Markup. To learn more, read our methodology.

12

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 **Adobe Inc.**, **The Trade Desk Inc**, and six other companies.

17

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 **17** third-party cookies on this site. Blacklight detected cookies set for **Verizon Media**, **Snap Inc.** and **Adobe Inc.**, **ByteDance Ltd.**, **Pinterest, Inc.**, **Tapad, Inc.**, **Alphabet, Inc.**, **The Trade Desk Inc**, **Magnite, Inc.**, and **Verizon Media**

How We Define ThisSurvey of Popular Websites

Tracking that evades cookie blockers wasn't found.

Canvas fingerprinting was not 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 across sites. It secretly draws an image on your browser when you visit websites that use it, for the purpose of identifying your device. This technique was used by six percent of popular sites when we scanned them in September 2020.

How We Define This

Session recording services not found on this website.

Blacklight did not detect the use of a session recorder, which tracks user mouse movement, clicks, taps, scrolls, or even network activity. Websites that use the technique compile this data into videos and heat maps that website owners can watch to see how users interact with the site. Research has shown these practices are insecure and make sensitive user data such as passwords and credit card information more vulnerable to leaks. This technique was used by fifteen percent of popular websites when we scanned them in September 2020.

How We Define This

We did not find this website capturing keystrokes.

Key logging is when a website captures the text that you type into a web page 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.

How We Define This

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.

How We Define This

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 #2: www.microsoft.com/en-us/?ql=2

Blacklight Inspection Result

Blacklight works by visiting each website with a headless browser running custom software built by The Markup. To learn more, read our methodology.

22

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 **Exponential Interactive Inc.**, **Flashtalking Inc**, and thirteen other companies.

How We Define ThisSurvey of Popular Websites

41

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 **41** third-party cookies on this site. Blacklight detected cookies set for **Microsoft Corporation**, **Alphabet, Inc.** and and twelve others

How We Define ThisSurvey of Popular Websites

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 belonging to the company **Microsoft Corporation** 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.

However...

While Blacklight accurately detects the presence of canvas fingerprinting on a website, it cannot determine if the purpose is user behavior monitoring or for fraud prevention or bot detection.

How We Define This

Session recording services not found on this website.

Blacklight did not detect the use of a session recorder, which tracks user mouse movement, clicks, taps, scrolls, or even network activity. Websites that use the technique compile this data into videos and heat maps that website owners can watch to see how users interact with the site. Research has shown these practices are insecure and make sensitive user data such as passwords and credit card information more vulnerable to leaks. This technique was used by fifteen percent of popular websites when we scanned them in September 2020.

How We Define This

We did not find this website capturing keystrokes.

Key logging is when a website captures the text that you type into a web page 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.

How We Define This

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.

How We Define This

Google Analytics' "remarketing audiences" feature not found.

The Google Analytics "remarketing audiences" feature 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: www.tesla.com

Blacklight Inspection Result

Blacklight works by visiting each website with a headless browser running custom software built by The Markup. To learn more, read our methodology.

7

Ad trackers found on this site.*This is the* ***same as*** *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 the companies **Twitter, Inc.**, **Alphabet, Inc.** and **4Cite Marketing**.

How We Define ThisSurvey of Popular Websites

5

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 **5** third-party cookies on this site. Blacklight detected a cookie set for **Twitter, Inc.**

How We Define ThisSurvey of Popular Websites

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 **tesla.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.

However...

While Blacklight accurately detects the presence of canvas fingerprinting on a website, it cannot determine if the purpose is user behavior monitoring or for fraud prevention or bot detection.

How We Define This

Session recording services not found on this website.

Blacklight did not detect the use of a session recorder, which tracks user mouse movement, clicks, taps, scrolls, or even network activity. Websites that use the technique compile this data into videos and heat maps that website owners can watch to see how users interact with the site. Research has shown these practices are insecure and make sensitive user data such as passwords and credit card information more vulnerable to leaks. This technique was used by fifteen percent of popular websites when we scanned them in September 2020.

How We Define This

We did not find this website capturing keystrokes.

Key logging is when a website captures the text that you type into a web page 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.

How We Define This

Facebook Pixel not found on this website.

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.

How We Define This

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:

Importance of Infosec

Security for information systems is a multi-hundred-billion-dollar business (Kapko). From the perspective of infosec, it is important to maintain business reputation. Companies will invest significant resources in building trust with their customers. A security breach can result in a loss of customer trust and damage to a company’s brand, which can be costly to recover from. Since numerous industries are regulated by laws that mandate specific security as well as privacy measures, not complying with these measures can lead to substantial legal repercussions. Also, customer data and intellectual property are often stored on private company servers. These trade secrets, patents, and proprietary software that keep a company competitive can be detrimental to sales. Additionally, cybersecurity threats can have significant monetary impact from threats such as ransomware.

User Perspective on Security

From a user's point of view, it is only normal to expect that their own personal and sensitive data, including payment and medical records, will be safeguarded. Users also do not want their history, preferences, or online habits to be tracked without their consent. This information can be misused for targeted advertisements or other malicious intentions. They expect that their personal privacy is maintained. Furthermore, users have a standard expectation when it comes to digital transactions. Since 82% of transactions in America use some form of digital payment, users need to trust that their transactions are secure (Goel).

Blacklight findings on Tesla

Blacklight detected the presence of 5 third-party tracking cookies on Tesla's website, which is higher than the average of 3 found on popular sites according to Blacklight. Among these, a notable one was set for Twitter, Inc.

Implications for Infosec on www.tesla.com

From an infosec's perspective, the presence of third-party tracking cookies can introduce vulnerabilities to the website. Malicious entities might exploit these cookies which could lead to a data breach. This is a potential vulnerability to any data that is sent between the two servers. Also, users might view Tesla as "less trustworthy" if they become aware of the extra use of third-party trackers. This can harm the company's reputation and user trust. Similarly, relying on third-party entities such as Twitter, Inc., means that Tesla's user data handling is also dependent on the security and privacy practices of these third parties. If any of these third parties were to experience a breach, Tesla might indirectly face backlash, too.

User Concerns Regarding Third-Party Cookies on www.tesla.com

Third-party cookies, especially those from major platforms like Twitter, can track user activity not only on Tesla's site, but also across other websites. This can create a detailed profile of a user's online

behavior, which could be a “make or break” for any user worried about their privacy. From a user’s perspective, improperly managed security from either the main party or any third-party server can give attackers a way into their personal information or even worse—manipulate their bank information. Moreover, if the privacy policies differ between the two parties, a user could have their data exploited, despite the original party’s site telling them they would not sell the user’s information. Users might feel that their privacy is being invaded if they did not explicitly consent to their data being tracked and sold.

Course Objective B:

Web Application Security Policy

1. OVERVIEW
    a. With the increasing integration of third-party services and the substantial growth of web-based applications, it has become critical to address potential vulnerabilities. Web application vulnerabilities, especially those introduced through third-party tracking cookies and similar integrations, can expose an organization to significant risks. To mitigate these risks, it is important that we rigorously assess ANY web application for vulnerabilities PRIOR to use.
2. PURPOSE
    a. This policy sets guidelines for web application security assessments within the company. This includes evaluations for potential weaknesses arising from third-party integrations, tracking cookies, without user consent, misconfigurations, and weak authentication. The primary focus is to limit the potential damage to this company as well as follow global privacy standards.
3. SCOPE
    a. This policy encompasses all web application security assessments conducted within our company, focusing on maintaining security, compliance, risk management, and change control. The policy extends to third-party integrations and cookie usage, especially tracking cookies, guaranteeing user data privacy. All findings are treated as CONFIDENTIAL, with distribution limited to a “need to know” basis.
4. POLICY
    a. 4.1: Assessment Criteria:
        i. New Application Release – Subject to a full assessment before approving change control documentation or release.
        ii. Third Party or Acquired Web Application – Undergo a full assessment, post required to align with this policy.
        iii. Point Releases – Subject to assessment based on the risk of changes to application functionality.
        iv. Emergency Releases – Can bypass assessments temporarily, assuming associated risks until a formal evaluation occurs.
        v. Annual Review – Comprehensive assessment to address potential risks.

vi. Third-Party Cookie Integration – Before integrating any third-party tracking cookie, it must undergo risk assessment to evaluate potential security implications.

b. 4.2 Issue Remediation – Issues identified are classified based on the OWASP Risk Rating Methodology. Each risk level requires certain mitigation steps:
   i. High – Immediate mitigation before deployment. Applications may be taken offline or denied release if vulnerabilities persist.
   ii. Medium – Review, determine, and schedule necessary mitigation. Applications might be taken offline or release denied if vulnerabilities pose significant risk.
   iii. Low – Schedule an appropriate review and correction procedure.
c. 4.3 Assessment Levels:
   i. Full – Comprehensive tests for all known vulnerabilities based on the OWASP Testing Guide. This includes manual penetration testing.
   ii. Quick – Automated scans focusing on the OWASP Top Ten web application security risks.
   iii. Targeted – Performed to verify vulnerability remediation or new application features.
d. 4.4 Tools – the following are APPROVED for security assessments:
   i. Approved tool 1
   ii. Approved tool 2

5. RELATED STANDARDS, POLICIES, AND PROCESSES
   a. OWASP Top Ten Project
   b. OWASP Testing Guide
   c. OWASP Risk Rating Methodology

(Original web application security policy by SANS. Cited as SANS below.)

Course Objective D:

In the ever-changing world of web security, third-party tracking cookies present unique challenges for both organizations and individual users. Addressing these challenges through the lense of the CIA triad (Confidentiality, Integrity, Availability) provides a comprehensive approach to safeguarding interests. Confidentiality means keeping data private. Integrity ensures the data stays correct and untouched. Availability makes sure people can always access their data when they need to. Along with ideas such as “zero trust” and using tools to stop data leaks, this approach helps keep both company and user data safe. The following list explains the CIA triad further from an organization and user perspectives in a much easier to read format:

1. Confidentiality:
   a. Organization’s Liability Perspective:

i. Data Encryption – Ensure that all data (especially those that involve third-party integrations, use strong encryption. This prevents unauthorized access to sensitive data and shields the organization from potential legal problems due to data breaches.
ii. Access Control – Implement strict access control mechanisms to determine who can access what data. This includes setting privileges for third-party integrations and ensuring the only access necessary data.

b. Personal Privacy Perspective:
i. Clear Cookie Policies – Clearly inform users about the types of cookies in use, their purpose, and how long they stay active. This level of transparency ensures users are aware of how their data is being used.
ii. Opt-in/out Mechanisms – Allow users to opt-in or opt-out of non-essential cookies. Giving users control over their data enhances trust and ensures their privacy.

2. Integrity:
a. Organization's Liability Perspective:
i. Data Validation – Ensure that any data reveived from third-party cookies or sent to third-party services is validated to prevent injection hacks.
ii. Regular Reviews – Conduct regular security audits of third-party services to ensure they follow the organization's securty standards and do not introduce vulnerabilities.
b. Personal Privacy Perspective:
i. Notices for Altered Data – If data from users is altered in any way, users should be promptly notified. This ensures that users always have accurate knowledge about their data.

3. Availability:
a. Organization's Liability Perspective:
i. Service Level Agreements – When integrating with third-party services, have clear service level agreements to ensure their services are consistently available. This will prevent disruptions to users (Coursera).
ii. Backup Solutions – Regularly back up data to ensure that in the event of a failure of any kind, the essential services remain available.
b. Personal Privacy Perspective:
i. Easy Access: Ensure users can easily access and delete their data.
ii. Support – If users have concerns about their data, offer prompt and efficient support.

The CIA triad covers most of the aspects of the issue, however there are certainly additional technologies to low the chances of liability and protect privacy. Assuming no trust and verifying everything, even if it comes from trusted sources, will mitigate liability. The idea of "zero trust" is to double check everything. Double checking ensures that the company does not see reproductions due to other companies' faults. Also, using data loss prevention software to monitor and control data transferring across the organization's network. Logging everything helps to minimize risk. Finally, using a

multi factor authentication system to access data will add another layer of security on top of simple passwords and usernames.

References

Kapko, Matt. (2023, March 17). Global cybersecurity spending to top $219B this year: IDC. *Cybersecurity Dive.* URL.

Goel, Vaibhav. (2021, October 26). New trends in US consumer digital payments. *McKinsey & Company*. URL.

SANS Policy Team. (2022, October). Web Application Security Policy, *SAANS*. URL.

Coursera. (2023, June 16). What Is a Service-Level Agreement (SLA)? And How To Write One, *Coursera*. URL.

Last Name, A. B. (Year). Article Title. *Journal Title*, Pages #-#. URL. URL.

Last Name, C. D. (Year). *Book TitleBook Title* URL.

Last Name, D. E., Last Name, F. G. (Year). *Report TitleReport Title* URL.

Last Name, H. I. (Year, Month Day). Article Title/Headline. *Periodical.Periodical.*

Organization Name. (Year, Month Day). *Webpage Title.* URL.