Web Audit Project

This is a group project; you will work in teams of three to four. Each group will submit a single solution.

The work your group submits must be entirely your group's own work. You may consult with other students about the conceptualization of the project and the meaning of the questions, but you may not look at any part of someone else's solution or collaborate with anyone outside your group. You may consult published references, provided that you appropriately cite them (e.g., with program comments), as you would in an academic paper. The team captain is responsible for submitting the group deliverables.

Your submission MUST include the following information:

1. List of students in your group
2. List of people you discussed the project with (outside your group)
3. List of references used (online material, course nodes, textbooks, wikipedia, etc.)

If any of this information is missing, at least 20% of the points for the assignment will automatically be deducted from your assignment. See also discussion on plagiarism and the collaboration policy on the course syllabus.

1 Introduction

In this project you will choose a website and analyze the client-server communication from the website, in order to understand what information is collected and recorded when users surf the web. Your primary tool will be the developer tools bundled with the Chrome browser. You will be analyzing and interacting with production systems in this project. Please make sure you read and understand the rules given below before beginning the assignment. Failure to follow the rules will result in an automatic F in this course.
The project has these following deliverables:

1. Writeup with choice of website, and analysis of the website's responsible disclosure policy. Due: Jan 31, 2024.

2. Midterm
    a. Disclosure, to Professor Tian, of any potential vulnerability that you might have stumbled upon as part of your work. This should be done verbally during Prof Tian's office hours. Due: Feb 21, 2024 (or earlier)
    b. Presentation for the result you find and write a midterm report. Due: Feb 21, 2024
3. Final presentation and report by March 13, 2024.

2 Story

The Intelligence and Security Apparatus (ISA) of the fictional Republic of Lacedaemonia wishes to spy on its own citizens. The ISA may use the following methods (in order of increasing costs):

1. Eavesdrop on any unencrypted network traffic, by acting as a Man In The Middle (MITM) on the network.
2. Gain access to the server and learn everything the server knows.
3. Gain access to the client and learn what the client knows.
4. Eavesdrop on *encrypted* internet traffic (by hacking into the site's certificate authority, stealing its keys, issuing a bogus certificate for the target website, and using this certificate in a man-in-the-middle attack that eavesdrops on encrypted network traffic).

As the ISA does not want to get caught (for obvious legal and political reasons). Thus, information that it can learn from passively observing traffic is the most dangerous and should be the primary focus of your analysis.

Additionally, but far less importantly, you should also consider what information the site collects from the user and who the this information is shared with, as the Republic of Lacedaemonia has treaty and legal agreements with many countries that would allow us to subpoena or request these records.

Your job will be to choose a website to analyze. The ultimate goal of this analysis is to determine the information leaked/stored by the web application, the value of this information to the ISA, and finally to catalog the methods that could be employed to access it.

3 Websites

You should select a website from the following list, or any website that is listed on the following list of websites that support bug bounty programs:

https://bugcrowd.com/list-of-bug-bounty-programs

You may only choose a site that has a responsible disclosure policy and/or a bug-bounty program. If you really want to review a web application that is not on the list below, or on the bugcrowd list, please discuss this with Professor Tian during her office hours; if you do this, you should be ready with information about the sites' responsible disclosure program.

4 First deliverable: Introduction Presentation.

Your first deliverable is containing the following information:

1. Your group members.
2. The name and url of the main page of the website you will analyze.
3. A link to the site's responsible disclosure policy or bug-bounty program.
4. A presentation describing the process for disclosing a vulnerability that you might have found on that site, according to the site's responsible policy/bug bounty program. Make sure you describe

    (1) how you would go about disclosing the vulnerability

    (2) what the site commits to do once the vulnerability has been disclosed to them (for example, to fix the vulnerability within X days, to not take actions against the person disclosing the vulnerability, to disclose the vulnerability on their website or blog, etc.).

        Note*:

        Please check on the constraints of the website and if they will sue the researcher. If the website does not have this kind of information cleared, please change your target website. (You can still talk about the loophole of all websites you have checked as a part of the presentation for the First deliverable)

        If you feel like the responsible disclosure policy contains any loopholes that can be used against the security researcher, you should also discuss this in your presentation. Please talk to professor Tian if you are worried about any of these rules. If they are particularly worrying, we encourage you to choose a new site to work on for your project. All statements you make should be supported using direct quotes from the site's terms of service, responsible disclosure policy, bug bounty program, etc.

5 Analysis.

Your task is to prepare:

(1) a midterm presentation following the guideline on syllabus

(2) a midterm report summarizes your initial findings on what information the ISA can glean from the website, using each of the four snooping techniques discussed in Section 2; you should focus especially on the first technique – eavesdropping on unencrypted network traffic.

This is an open-ended project, so, !!!as long as you follow the rules in Section 5.1!!!, your video can focus on any issues in Section 5.2, or other issues that you find interesting. As usual in this class, we will be looking for in-depth research, rather than superficial information.

5.1 The rules.

IMPORTANT!	In the process of your analysis you may discover surprising results and vulnerabilities. Do not discuss them with anyone outside of your group without first consulting with Professor Tian. Failure to do this will result in an automatic F. We have this policy also to protect you from potential law suits.

As this task involves interacting with a private party's computer systems it is very important that you avoid anything outside of what a normal user would do.

THAT MEANS, YOU CAN LOOK BUT YOU CAN'T TOUCH.

The following rules are guidelines for what we consider a "normal user" would do. But, in doing this lab, you should err on the side of caution. Just because the rules do not say not to do something, this does not mean that you can do it.[1]

1. DO: Use the site as a normal user would, follow links, click buttons, interact.
2. DO: Watch and record what actions the site takes, what it saves to your disk, what URLs it requests, what information it asks from the user.
3. DO: Analyze what data the site has about a user and how it saves this data. Is everything stored on the server, or is some of the data recorded locally in cookies? What sorts of vulnerabilities might this create?
4. DO: Read the responsible disclosure policy for your web applications and make sure not to violate it.

   [1]In Airbud a dog is allowed to play basketball because "there is no rule on the books saying that a Dog can´t play basketball", we will not accept such an argument. The absence of a rule forbidding a particular action does not imply that it is allowed or condone it in any way.

5. DO: Read the site's privacy policy and any associated news media on site privacy, and consider how it affects user privacy.
6. DO NOT: Edit URLS.[2],
7. DO NOT: Change cookies, post javascript or strange characters into forms.
8. DO NOT: Do anything which violates the law, other users' privacy, the user agreement of the web application, or the code of computing ethics and of Boston University.
9. DO NOT: Attempt to attack the server or client in any way, including but not limited to XSS, CSRF and SQL injection.
10. DO NOT: Save the webpage to disk, and then alter it and load it.
11. DO NOT: Post online or discuss your results with anyone outside your group without first consulting with Professor Tian

    [2]This may seem harmless, but sometimes it is not. We have seen many examples in class where editing URLs can result in successful SQL or XSS attacks. There have also been instances of production systems being crashed by a user deleting a single field from a URL and requesting it. At least one person has been sentenced to more than 3 years in federal prison for generating malicious URLs and accessing them (weev).

5.2 Things to Look For

1. Cookies: Does the site use cookies, if so what data is stored in the cookies? Who can access these cookies? Are these Secure Cookies? Persistent or Session cookies? What could you learn about a user if you had access to their cookies?
2. Are ads shown to the user? Where are these ads loaded from? Who can display these ads? Can the ads contain javascript? What can an advertiser learn about the user?
3. What URLs are fetched when the site loads? Are any of these URLs offsite? Are these URLs protected with HTTPS? What are these URLs used for and what are the privacy risks? Do these URLs contain any sensitive data?
4. Does the site use mixed HTTP and HTTPS content?
5. How does the site track the user? Cache? Cookies?
6. Are there any side channels? For instance does the site make a request each time a user hits a key? Even if these requests were encrypted, what could an eavesdropper learn?
7. What technologies is the site using? What (if any) front-end frameworks is it using (e.g. React.js, Angular.js, Vue.js)? What (if any) back-end frameworks is it using (e.g. Flask, Ruby on Rails)? What server software is it running (e.g. Apache, Nginx, HAProxy)? *Note:* tools like Wappalyzer can help make this easier. You may also want to look elsewhere to try to deduce the technologies being used. For instance, you can check whether the site has a dev blog, or see if it has job postings that are looking for experience with certain web technologies.
8. What is the site's privacy policy? Who do they share data with? Could ISA purchase the user data from the web application's company or affiliate? How does the site comply with GDPR?
9. Does the site violate any user privacy laws?

5.3 Browser development tools.

Your main tool for this project will be the Chrome browser's developer tools:

https://developers.google.com/chrome-developer-tools/

or Firefox's browser developer tools:
https://developer.mozilla.org/en-US/docs/Tools

You can access these tools by right clicking on a page and selecting "Inspect Element"; this will open an interface that allows you to inspect the requests, cookies, network traffic, etc. You must not not use the CONSOLE tab as part of your analysis.

5.4 Second deliverable: Report and presentation

You should submit the report as youWebsiteName_midterm.pdf. Your report should be following the requirements of our midterm report. Your presentation should follow the guideline mentioned on the syllabus and incorporate all four group mates in the presentation.

6 Third deliverable: Final Report and presentation

Each group will submit a final report. The final report should include your findings for the website you choose and think about common themes you observe about web security, tracking and privacy. Your report should concentrate on common trends in web security and privacy.

You should seek to answer these questions:

- What vulnerabilities or privacy issue have you found
- How do you find these issues?
- Potential solutions to the problems you identified
- How you understand the privacy issue of websites and what is your steps of checking if the websites have violated GDPR or other standards
- How much of web traffic is encrypted or not? Is end-to-end encryption being used, or can the provider of the site read all your messages etc?
- How much of web traffic is encrypted or not? Is end-to-end encryption being used, or can the provider of the site read all your messages etc?
- What are common tracking practices? Who's doing the tracking? How is tracking executed (cookies, pixels, etc)?
- Common themes in privacy policies?
- What surprises you (or does not surprise you) about privacy and security on the web?

Submissions. You should submit this as

group_number_WebsiteName_Final.pdf.