Student Observatory in Cybersecurity Education

Yijun Liu
Emory University

James Mattei
Tufts University

Daniel Votipka
Tufts University

Abstract

Capture The Flag competitions and games have been employed in cybersecurity in addition to formal education [3] [17]. The goal of this study is to examine students' browsing activities in completing CTF challenges. To achieve the goal, we developed a student observatory that allows us to monitor their activities. We can use students' browsing activities to understand students' learning process for CTF challenges while completing them and, therefore, bring insights into Computer Security education. The student observatory developed in the study is a Chrome Extension, CTF browsing monitor, that collects the URLs visited, the percentages of webpages viewed by participants, and the time stamps on each visit. There will also be a short survey containing three questions about the webpage just viewed for non-search engine URLs to further gain participants' perspectives on the webpage they just view.

1 Introduction

Capture The Flag (CTF) competitions and games have been employed in cybersecurity in addition to formal education since 1996, when it first originated [3] [17]. CTF has gained popularity and significance in the field of cybersecurity. There are over 200 events recorded on ctftime.org in 2022, and even major tech companies such as Google and Meta hold CTF competitions each year [1] [13] [2]. In CTF competitions, participants need to utilize their knowledge and skills in cybersecurity to find and exploit vulnerabilities in programs to answer the challenges in the competition [15]. CTF participants practice challenges and study knowledge and skills throughout their careers as CTF participants, where those knowledge and skills in vulnerability discovery are a significant part of cybersecurity education [17]. Universities have also successfully employed CTF challenges in their existing cybersecurity curriculums [16].

1.1 Goals of the Study

The goal of this study is to examine students' browsing activities in completing CTF challenges. To achieve the goal, we developed a student observatory that allows us to monitor their activities. Students' browsing activities will allow us to understand their learning process for CTF challenges while completing them and, therefore, bring insights into Computer Security education. Specifically, the four goals we have are the following:

1. What are students searching for when completing CTF challenges?
2. What online resources from the Web are students using to solve CTF challenges?
3. Are the available resources understandable and transferable to the challenge context? Do they support better learning?
4. What elements are missing from the online CTF resources, and what could we add?

As the study has just begun, we will mainly focus on the student observatory we developed in this paper.

1.2 Contribution of the Study

Most research in CTF and cybersecurity education focuses on CTF challenges and applicable scenarios for CTF challenges. Some research touch on CTF promotes learning for students. Yet, no research aims to investigate how and what students learn from completing CTF challenges. By answering the research questions above, this paper will be beneficial to the following stakeholders:

- *CTF content creators and members in cybersecurity* may be able to add missing critical CTF resources online
- *CTF organizers* can potentially promote platforms and sites that participants find the most useful for CTF challenges to prepare for CTFs

- *Teachers* may know what kinds of information (images, texts, examples) students can best process for CTF challenges.
- *Students and CTF participants* can utilize search engines more efficiently and focus more on reliable sites to gather information.

2 Related Works

The research investigates students' browsing behaviors in completing CTF challenges to study their learning process for CTF challenges and, therefore, bring insights into Computer Security education. Thus, the related works for this paper lie in two different areas: 1) CTF as an educational tool in Computer Security Education and 2) Web as an informational resource for education. In this section, this paper will show how this research is informed by the previous studies and differs from the state-of-the-art.

2.1 CTFs Are Educational

Numerous studies in the field have proven that CTFs are valuable educational resources for Computer Security education over decades. The paper written by Švábenský et al. analyzes the cybersecurity topics taught in CTFs and finds that CTF covers all eight knowledge areas in cybersecurity [17]. Thus, CTF challenges comprehensively include the knowledge for cybersecurity education [17]. Moreover, the CTF challenges also allow for incorporating pedagogical principles through personalization and providing incremental feedback (scaffolding) [15]. When students are solving a challenge, they can obtain feedback on whether they are correct or not immediately after they submit their answers: this allows students to understand their knowledge and understanding of the vulnerability assessed in the challenge [15] [3].

In addition to CTF challenges to be educational, the formats of CTF also provide educational benefits to students. Different from traditional lectures and exercises, CTFs provide a gamified learning environment that is more fun and attractive and exercises that has grander diversities in questions [3] [10]. Vykopal et al. have shown that CTF challenges allow students to learn cybersecurity skills in hands-on and more entertaining ways; college students also prefer CTF challenges more than their regular homework assignments in Computer Security [16]. In Owens et al.'s study, students are shown to be more self-motivated and more willing to learn new knowledge and approaches to vulnerability challenges while participating in CTFs [10]. Students also have deepened their expertise in cybersecurity by solving the CTF challenges [10]. Votipka et al. describe CTF challenges that help students to find online communities that allow peer learning [15]. Not only for students to learn new knowledge, but the vulnerability discovery challenges in CTFs also significantly improve people's vulnerability discovery process in real-life settings such as work [14].

Previous studies have proven that CTF challenges are educational and encourage students' learning. Nevertheless, no study has yet focused on students' learning processes and behaviors while completing CTF challenges. In short, no studies have focused on how students learn and what they learn from Web search engines while completing the CTF challenges. Additionally, CTF challenges are usually not beginner-friendly and sometimes even discourage beginners from participating [3] [17] [8] [9] [16] [10]. This study will recognize a particular focus on beginners, capturing their behaviors and investigating how CTF challenges have become intimidating for beginners.

2.2 Learning from the Web

The Web serves as an essential information resource for students. Various studies in information science and HCI focus on how students capture information on the Web via search engines, a universal way of acquiring new information online. Griffiths and Brophy find that the use of search engines has also become the dominant strategy for students seeking information [5]. Therefore, in this study, we focus on the use of search engines to gather online information. Nevertheless, they also discover that students often find it challenging to locate information from the search engine result page and have to spend a great deal of time finding the right keywords that direct them to the desired results [5].

Kuiper et al. have conducted experiments on K-12 students, investigating their behaviors in solving questions via the Web [7]. They found that students tend to find an exact answer rather than gathering information online that they can use to deduce the answers to the questions [7]. Moreover, students behave poorly in evaluating the content information as they tend to believe that good page designs imply reliable information. Students are also confused about "quantitatively good" and "qualitatively good": they tend to trust websites that contain more information, texts, and images [7]. Kuiper et al. point out that students with different levels of prior knowledge, genders, and ages approach information differently [7]. Thus, their learning results in domain-specific knowledge, search skills, and vital thinking skills vary after they complete the experiment in answering questions [7]. Inspired by Kuiper et al., we decided to capture the pages that students have visited via web scrapers to examine the page contents in detail and discover potential correlations between how students rate the usefulness of the pages they visit and the contents. In addition, we also ask students to complete background survey questions to gain

information about their skill levels and demographics.

Gwizdka and Spence research the correlations between task difficulty and searching behaviors [6]. They found that for objective tasks, the unique pages visited for the task, time spent on the task, and the total number of pages visited are the three independent variables that positively correlate to the level of difficulties of tasks [6]. We will also consider these factors as indicators in our research to assess the difficulties of the challenges participants have done.

All previous studies on students' behaviors in information retrieval for academic purposes do not touch on the keywords they search on search engines. Additionally, previous studies do not identify web pages and domains that students find important and valuable. In other words, there are no research on how student assess each site they visit for educational purposes and the kinds of information students can access on the Web. This study will investigate these factors.

3 Goals

This section will discuss the goals of the study in detail. The fundamental purpose of this study is to understand how students learn and what they learn from Web search engines while doing CTF challenges. We want to find out how students search for, look at, learn, and finally, complete CTF challenges.

As browsers serve as the primary tool for students accessing CTF challenge resources, we decide to investigate students' browsing activities while completing CTF challenges. Based on the previous study, we have also identified the importance of search engines for students gathering information from the Web. Thus, our key focuses lie on these two parts: the web pages they visited and the search engines. To further address browsing activities in detail, we have proposed the following four research questions:

RQ1: What are students searching for when completing CTF challenges? The dominant way for students to collect new knowledge from the Web is through search engines [5]. However, search terms and the choices of search engines vary among students. Our first step is to capture the keywords students are searching for and the search engines they are using, as well as to understand the underlying reasons for choosing the keywords and search engines. After students use search engines, the search engines will display a list of resources for the given keywords. We are also interested in the websites that students eventually choose and gain insights into the potential changes in visited websites for different hints, keywords, and challenge types.

RQ2: What online resources from the Web are students using to solve CTF challenges? After looking into the search terms and pages clicked from the search result page, our next step is to examine the actual pages participants visited. In examing the web pages, a few vital features of web pages we have identified include:

- The types of the pages (for example, official documentation or crowdsourced forum/blogs).
- Connectivities between the pages.
- The frequencies that participants visit the domains or websites.

We are also interested in the potential differences in the above measurements among participants with different skill levels in CTF challenges.

RQ3: Are the available resources understandable and transferable to the challenging context? Do they support better learning? This research question focuses on the learning component while completing the CTF challenges, particularly the usefulness of the available resources. Specifically, we want to emphasize the media elements on web pages that can potentially impact students' learning. Nevertheless, the usefulness of the web contents may also vary between students with different skill levels in CTF challenges.

RQ4: What elements are missing from the online CTF resources, and what could we add? We also aim to identify the current deficiencies in learning through the existing online CTF communities and find potential solutions to improve students' learning.

4 Design

In this section, we explain the design of the experiment investigating students' browsing activities while completing CTF challenges. The process of the experimentation includes participant recruitment, hosting selected CTF challenges in a CTF platform (website), designing a Chrome Extension to monitor browsing activities, and utilizing a post-study survey to understand students' browsing activities further. Moreover, we plan to develop a web scraper to explore the contents that participants view in addition to browsing activities.

4.1 Participants

Experience with programming is a fundamental prerequisite for students to understand the questions, thus, being able to participate in CTF competitions. Therefore, we require participants to pass screening questions regarding their programming skills [4]. This study aims to include students with various CTF experiences, with a particular focus on

recruiting beginners, the group that is identified as being discouraged by the CTF challenges the most. College students, particularly students with computer-science-related majors, have programming backgrounds, but their exposures to CTF vary, which may exist a large number of CTF beginners. Additionally, participants from past CTF competitions will likely have more experience and higher skill levels with CTF challenges. Thus, we decide to recruit through both colleges' and CTF competitions' mailing lists.

4.2 CTF Challenges

To best understand students' behaviors in completing CTF challenges, this study aims to create a similar environment as the CTF competition for participants and tests participants on typical CTF topics. Thus, we decided to host CTF challenges on an established CTF platform. In total, there will be 4 CTF challenges assigned to each participant, and the four challenges will be randomly sampled from a total of 16 challenges. The challenges will be varied by problem types, depths of vulnerability in code (one primary function or 2-3 functions deep), and the presence of mitigation. The four problem types we identified as typical and, therefore, will be employed in the study are format string, buffer overflow, integer overflow, and heap overflow. In addition, we will progressively reveal hints by time spent on each question.

4.3 Chrome Extension

The easiest way to track browsing activities is through a browser extension. In the study, we decide to develop a Chrome Extension, CTF Browsing Monitor, that collects the URLs visited, the percentages of webpages viewed by participants, and the time stamps on each visit. There will also be a short survey containing three questions about the webpage just viewed for non-search engine URLs to further gain participants' perspectives on the webpage they just view. The questions are: 1) Was this a website related to the CTF challenge? 2) Was it helpful for solving the problem? 3) Was this website understandable? After tracking, the extension will direct participants to the post-study survey. The design details will be included in the implementation section.

4.4 Post-study Survey

The post-study survey provides a greater understanding of students' browsing behaviors. The three main sections of the post-study survey are questions related to each challenge, background questions on students' previous experience/knowledge in the materials associated with CTF challenges, and demographics. For each challenge, we ask students to rate the usefulness of keywords they searched for, their background in the particular vulnerability type, and whether they use other non-browser resources. The academic background questions focus on their past CTF experience, vulnerability discovery skills, and computer security education.

4.5 Web Scraper

We decide to design a web scraper to capture critical information on a page after we capture the URLs. We want to know the types of pages and the connectivity between pages. However, we plan to develop the web scraper further after we have a pilot study done.

5 Implementation

When this paper was written, the project had finished the Chrome Extension and the post-study survey. Therefore, for implementation, this paper will focus on implementing the Chrome Extension, CTF Browsing Monitor. The Chrome Extension has already been published in the Chrome Web Store.

5.1 Design Principle

A browser plug-in is an easiest and lightest way to track browsing activities. Therefore, to fulfill the primary purpose of the study, investigating students' browsing activities, we decide to develop a browser extension. The extension should be user-friendly in terms of its interface, usability, and installation. The interface is designed to be simple yet colorful to identify different buttons, as shown in figure 1. Moreover, tracking browsing activities may be sensitive when participants accidentally visit sites they may not wish to share their activities. As a result, we decided to develop an extension that tracks browsing activities locally, as well as constantly reminding participants about this tracking extension's presence.

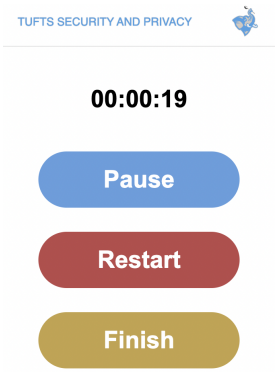

Figure 1. Interface of CTF Browsing Monitor

5.2 Implementing via Chrome Web Store

Different browsers have distinct rules and principles for extension development. Therefore, we must choose a platform to develop the extension. The browser we have chosen is Chrome due to its extremely high market share, 69.28% [12]. In addition, Chrome has a well-established extension store that allows only participants to install the extension easily: Chrome allows an "unlisted" publishing method that only participants can download the extension via an invitation link.

5.3 Tracking

The main functions of the CTF Browsing Monitor can be separated into two parts: tracking browsing activities and prompting a short survey after each visited page. In this section, I will describe the measurements we track.

CTF Browsing Monitor records the time stamps when participants view and stop viewing a URL. Stop viewing refers to the scenarios when participants close the page, switch between tabs, or open other applications. Moreover, when the participants stop viewing a page, the extension calculates the percentage of the webpage that is viewed by participants. The calculation is done by getting the window height (Winh), the webpage height (Webh), and the height of the scrolled contents (no longer visible in the current window but viewed, Scolh). Therefore, we calculate the percentage of the webpage viewed by the function*(Winh + scolh) / Webh * 100*. Tracking the URL fulfills RQ1 in collecting the websites participants visit. We can also extract keywords that participants search from the URLs of the search engine result pages. Tracking URLs also allows us to calculate the frequencies of students visiting a webpage or a domain, which we can use to answer RQ2 partially. Moreover, tracking the time stamps and the percentage of pages viewed can help us to gain quantitative data in understanding RQ3.

5.4 Short Survey

The survey interface is shown in figure 2. We cannot tell whether a page is helpful or not purely through the time spent on a page and the percentage of the page that participants view. Nevertheless, asking students about the page visited after the experiment may be difficult as they might be unable to remember the details. Therefore, we choose to popup a survey that asks participants about the page they just visited right after participants leave the page. The questions are: 1) Was this a website related to the CTF challenge? 2) Was it helpful for solving the problem? 3) Was this website understandable? All these questions are rated on a scale of 5, and the stages are: not at all, slightly, somewhat, moderately, and extremely. The survey page displays the heading and URL of the page visited to clarify the page we ask participants to rate. The result of the survey will be recorded.

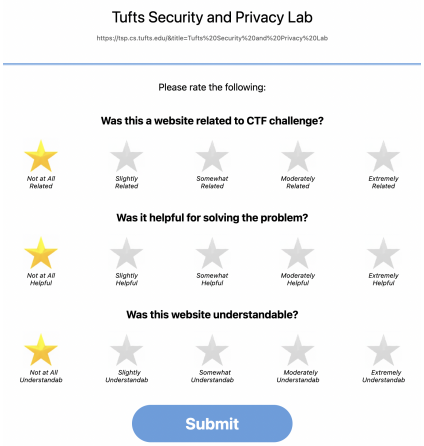

Figure 2. Interface of the Pop-up Short Survey

We decide not to pop up the survey page for participants visiting search engines. Moreover, the survey only appears once, which means that when participants revisit the site, the survey does not pop up again. We decide not to let participants redo the survey for the URLs they have already done before because we believe that redoing the same survey for the same URL will make the surveys more fatiguing.

5.5 Upload Post-Study Survey

When the experiment has finished, participants will need to click the "Finish" button on the extension interface. A Microsoft Excel form will be downloaded automatically, recording their activities on participants' computers. We choose to use Excel because Excel is the most frequently used spreadsheet format, and participants will be more likely to be able to open it without special instructions on their computers. In addition to the downloaded form, a popup on the extension interface will direct students to the post-study survey they need to fill in, as shown in figure 3. The URL that directs students to the post-study survey also contains the keywords they searched for each question during the experiment as a way to pass variables, where participants will be asked to rate the keywords they searched in the post-study survey. Rating keywords mainly serve to answer RQ1 and RQ3.

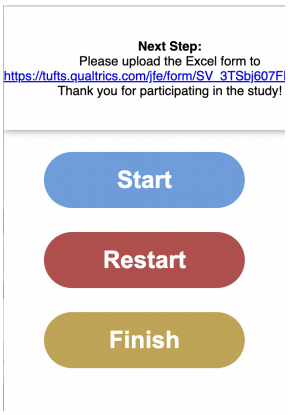

Figure 3. Interface of the Pop-up Link to the Post-study Survey

5.6 Privacy Awareness

Browsing activities are highly private. Even in a closed study setting, participants may accidentally visit sites where they may not prefer to share the data with us. As a result, we decide to track the activities locally. Data will be stored locally on participants' computers. After participants hit the finish button on the extension interface, a Microsoft Excel form will download automatically to the participants' computers. Participants may open the Excel form on their computers and check on their activities and decide to upload the results to us or not.

In addition, to remind participants that the extension is tracking their browsing activities, a red bar is fixed on the top of each page, showing that "Your activities are being monitored" when the extension is on. The web page with the extension is shown in figure 4. Moreover, a built-in stopwatch is clearly shown on the extension page. Participants will also acknowledge that the extension is on when they see the stopwatch is on.

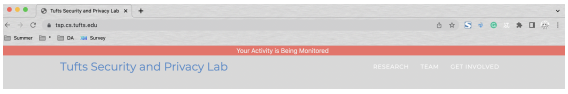

Figure 4. Red Bar when Extension is On

6 Discussion

Since the study has not been piloted yet, we do not have any data or experiments to evaluate the study. The discussion section will mainly focus on the development of the Chrome Extension, CTF Browsing Monitor. I (Yijun Liu) developed this Chrome Extension independently with the help of my mentor Dr. Dan Votipka and teammate James Mattei.

A key takeaway for me in developing area-specific programs was reading through the official documentation carefully. It was my first time designing a Chrome Extension, and I was not familiar with the languages, specific requirements, or APIs for developing Chrome Extension. There were only limited resources online that I could learn from. Even though I read through a lot of forum posts and QA posts, most of my questions and issues were eventually solved by the official documentation. Moreover, I made a mistake in developing an old Chrome Extension version that was just replaced by the new version when I started developing it. Reading official documentation carefully would also help me to avoid the mistake of restructuring and re-developing the extension.

A handy API I found and used for the extension is *sheetjs* [11], an open-source API that can extract data from spreadsheets and generate new spreadsheets. With the API, I was able to export an Excel form easily from the local database.

The extension is not perfect in achieving the percentage of a page viewed. The current method works perfectly for static websites. Nevertheless, suppose the website is dynamic, and the entire page will not load until users scroll down (for example, the Twitter feed). In that case, the percentage will be inaccurate as we cannot obtain the height of the whole web page. I have not yet found any method that can solve this problem. However, after several self-testing, I believe that most of the websites students may visit to seek answers are static, and this problem will not likely occur frequently.

Moreover, due to the limitations of Chrome, the extension's in-app stopwatch is unable to run smoothly when the extension page is not opened at the front. It means that when users open the extension page again while they are using the extension, they will first see the stopwatch showing the same time as they closed the extension page before. Only after a few seconds will the stopwatch update to the current time.

7 Conclusion

We have developed the Chrome Extension CTF Browsing Monitor to track students' browsing activities. CTF Browsing Monitor can 1) collect the URLs visited, 2) calculate the percentages of webpages viewed by participants, 3) record the time stamps on each visit, and 4) pop up a short survey containing three questions about the webpage just viewed. The extension helps us to record the essential data to examine students' browsing activities while completing CTF challenges and bring insights into improving students' learning in cybersecurity.

Acknowledgments

We thank Google LLC for supporting this research.

References

[1] Facebook ctf official website.

[2] Google ctf.

[3] Kevin Chung and Julian Cohen. Learning obstacles in the capture the flag model. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, San Diego, CA, August 2014. USENIX Association.

[4] Anastasia Danilova, Alena Naiakshina, Stefan Horstmann, and Matthew Smith. Do you really code? designing and evaluating screening questions for online surveys with programmers. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 537–548, 2021.

[5] Jill Griffiths and Peter Brophy. Student searching behavior and the web: Use of academic resources and google." library trends 53 (4. *Library Trends*, 53, 03 2005.

[6] Jacek Gwizdka and Ian Spence. What can searching behavior tell us about the difficulty of information tasks? a study of web navigation. volume 43, 10 2007.

[7] Els Kuiper, Monique Volman, and Jan Terwel. The web as an information resource in k–12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, 75(3):285–328, 2005.

[8] Antonio Martorana. *Investigating the Experiences of Female CTF Players*. PhD thesis, 2022.

[9] Makoto Nakaya, Takayuki Abe, and Hiroyuki Tominaga. Implementation and trial practices for hacking competition ctf as introductory educational experience for information literacy and security learning. 2016.

[10] Kentrell Owens and Alex Fulton. pico-boo!: How to avoid scaring students away in a ctf competition. 2019.

[11] SheetJS. Sheetjs/sheetjs:sheetjs community edition – spreadsheet data toolkit.

[12] statcounter. Browser market share worldwide.

[13] Ctftime Team. Ctftime.org / all about ctf (capture the flag).

[14] Daniel Votipka, Rock Stevens, Elissa Redmiles, Jeremy Hu, and Michelle Mazurek. Hackers vs. testers: A comparison of software vulnerability discovery processes. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 374–391, 2018.

[15] Daniel Votipka, Eric Zhang, and Michelle L. Mazurek. Hacked: A pedagogical analysis of online vulnerability discovery exercises. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1268–1285, 2021.

[16] Jan Vykopal, Valdemar Švábenský, and Ee-Chien Chang. *Benefits and Pitfalls of Using Capture the Flag Games in University Courses*, page 752–758. Association for Computing Machinery, New York, NY, USA, 2020.

[17] Valdemar Švábenský, Pavel Čeleda, Jan Vykopal, and Silvia Brišáková. Cybersecurity knowledge and skills taught in capture the flag challenges. *Computers Security*, 102:102154, 2021.