On the Other Side of the Table: Hosting Capture-the-Flag (CTF) Competitions

An Investigation from the CTF Organizer's Perspective

Benjamin Carlisle, Michael Reininger, Dylan Fox,
Daniel Votipka, and Michelle L. Mazurek
University of Maryland
bcar15@terpmail.umd.edu, {mreining, dylfox21, dvotipka, mmazurek}@cs.umd.edu

Abstract

Cybersecurity competitions, such as Capture the Flag events (CTF), create a space for safe and legal cybersecurity education and practice. CTF participants must solve challenge sets designed by organizers to earn points, honing their problem-solving skills in defense and attack scenarios. Competition participation provides a clear educational benefit to competitors; however, the impact on, and process used by, challenge creators is less understood. This paper provides a preliminary exploration of how CTF organizers design and vet challenges, as well as their motivation for becoming a CTF organizer. Using chat data from one CTF as a case study, we analyze the challenge-development life cycle identifying the steps required to create and run the CTF. Our results show that organizers follow a general set of steps for challenge development but struggle in several areas, including how to evaluate challenge difficulty, how much vetting is required for challenges, and how to create challenges regardless of experience.

1 Introduction

CTF competitions promote cybersecurity education and interest through creative, time-sensitive puzzle solving. CTFs can be hosted online or in-person, offering flexibility for classroom or large-scale event use. Traditionally, CTFs are divided into multiple challenges, each asking the competitor to demonstrate a specific cybersecurity skill (e.g., binary exploitation, decrypting ciphertext) in order to reveal a secret value called a *flag*. The competitor submits the flag to confirm exploitation success and receives a predetermined number of points.

CTFs' primary educational value comes by giving students a creative way to apply knowledge [1, 11]. Challenge-solving techniques often build on fundamental cybersecurity concepts taught in undergraduate computer security courses. Also, challenges frequently extend what is taught in the classroom, leading competitors to conduct further research to solve problems. When paired, this fusion of learning environments can advance students' cybersecurity education [3, 10].

Moreover, CTF challenges are often inspired by real-world software vulnerabilities, some even leading to the discovery of new vulnerabilities in existing software [16]. This realism provides further inspiration, as students know they are developing useful skills [20]. Students are also likely to transfer concepts into a more concrete setting [12]. It is important to note that challenges typically require some modification from their real-world presentation to make them solvable in a shorter period of time and focus student attention on key concepts.

Finally, CTFs operate as friendly competitions. Students receive points for solving challenges and can see how they rank relative to their peers. This provides motivation for competing through the game's excitement [20] and can guide students toward challenges of appropriate difficulty based on their levels of experience [7, 18]. Drawbacks to the competition format include issues with challenge difficulty labeling [6] and discouragement for new competitors joining competitions late with lower scoreboard standings [2, 13, 19–21].

In all cases, the onus on creating an effective CTF is on the organizers: everything from challenge creation to platform management falls under their responsibility. In this work, we take the first step toward understanding the organizers and their process through a case study of UMDCTF 2020. Specifically, we aim to identify the methods the organizers employ to develop and vet challenges in order to meet the expectations of CTF competitors and the motivations and social dynamics within these groups. To this end, our work seeks to study and answer the following research questions:

- Q1: What is the challenge development life cycle?
- Q2: How are challenges vetted by the CTF organizers?
- Q3: What motivates a CTF organizer to help in organization and what do they gain?
- Q4: What topics do organizers discuss, and at what point in the CTF process?

In our case study, we observed the UMDCTF organizers take their responsibility seriously and collaborate as a team to produce a successful competition. We identified multiple challenges organizers face, as well as their organizing and vetting process. This includes the struggle for organizers to vet their challenges given their low resources, score their challenges appropriately despite knowing how to solve it, and the steps required to successfully create a challenge. We also include how the organizers shift from challenge creation to logistical discussion during the CTF.

2 Related Work

In an effort to teach real-world skills to job-seeking students and to address the shortage of cybersecurity new hires, CTFs are generally accepted as an effective teaching tool to help train the next generation of cybersecurity professionals [9]. In 2004, Eagle et al. identified class lectures as either generally constructionist: focused on creating better systems, or protectionist: focused on learning to protect current systems. In contrast, CTFs offer *destructionist* learning to teach students to learn security by behaving like malicious actors.

Prior work outlines field studies conducted on CTF participants to understand the educational impact of the event and the challenges [5, 8, 23]. While this work introduces key insights for improving cybersecurity education, existing work has yet to fully engage with the CTF organizer population, and thus leaves out a necessary perspective. In this paper, we investigate CTF organizing through the UMDCTF 2020 case study and provide detailed accounts from the organizers behind this collegiate CTF as a first step toward understanding and supporting this population.

There are known issues that hinder learning in CTFs. Considering that participants have widely varying levels of experience, many may feel left out when first participating in CTFs [17]. This may be due to a participant's limited security background, or more generally, to miscommunication while developing a CTF challenge. Likewise, organizers face understated complications assigning difficulty levels to new problems: CTF problems must be challenging enough to offer new insights to the participants, but not so challenging that novice participants are alienated [6]. Existing work considers how CTFs can be used as an educational tool, lowering the barrier of entry to CTFs [14]. Prior work also details CTFs' potential benefits to instructors as learning tools; Mirkovic et al., for example, simplify the CTF administration process for teachers [15]. However, very little work exists focusing on the CTF organizers, e.g., their educational benefits, development process, or how to most effectively run a CTF.

3 CTF Case Study

Our case study, UMDCTF 2020 (https://umdctf.io/), was scheduled as an in-person, all-day event, on a university campus. The competition was transitioned online due to concerns related to the COVID-19 pandemic. This affected communication between organizers, as they could no longer meet physically. Simultaneously, this allowed individuals outside of the CTF's general vicinity to participate. UMDCTF 2020 was held successfully as a 24-hour event beginning on 18 April, 2020 at 10:00 AM EST. The competition was "jeopardy-style"—participants solve independent challenges across several categories for increasing points—with 136 challenges within the categories of Cryptography, Exploitation, Forensics, Steganography, Reverse Engineering, Web, RF (Radio Frequency), and Misc. UMDCTF 2020 attracted more than *1,000* competitors from over 50 countries, with only 11 organizers. Each organizer was either a undergraduate or graduate student at UMDC. We suspect that the vast majority of participants were college students on the East Coast, given the primarily intra-college advertising method. However, many teams consisted of college students from outside of the region, high school teams, or professional teams. The organizers were a mix of graduate and undergraduate students at the University of Maryland.

Platform	Non-Redacted	Redacted	Total
Slack	409	245	654
Discord	263	58	321
Zoom	99	2	101

Table 1: Number of messages (or sentences, for Zoom) per platform.

4 Methods

Our study is divided into three phases: collecting data from chat sources, anonymizing and redacting the records, and analyzing the data with qualitative coding techniques. This study was approved by the University of Maryland's Institutional Review Board (IRB).

4.1 Data Collection

Our research team was granted access to internal chat records and meeting notes from the UMDCTF 2020 organizers. Two of the researchers were also UMDCTF 2020 organizers, which facilitated the level of access. This robust source of data provided unique insight into the CTF organizers' dynamics and behaviors prior to and throughout UMDCTF 2020.

We began by collecting internal communications data from three sources: Slack, Zoom, and Discord. Slack and Discord are group messaging platforms that provide different channels for communication, while Zoom is a group video calling platform. Each platform was used by the UMDCTF organizers for different purposes, and as a result, required individualized analysis approaches. The number of messages per source can be found in Table 1.

Slack. The main coordination medium for UMDCTF organizers was Slack. This source was frequently used by nearly all members leading up to UMDCTF, though was relatively unused during the event itself. Our dataset was obtained from the chat messages sent via a private channel for organizers within a larger invitation-only cybersecurity club Slack, `#umdctf2020-challdev`. For the purposes of our study and time constraints, we collected and analyzed the most recent 654 slack messages from this group beginning on March 9, 2020. Earlier messages were not present due to Slack's 20,000

message limit. Earlier Slack communication contained planning for an in person CTF (as opposed to online) and initial challenge initialization details.

Zoom. As a replacement for in-person meetings, which were canceled due to the COVID-19 pandemic, two separate Zoom calls were held and recorded by the UMDCTF organizers in order to discuss infrastructure decisions and give updates on challenge development progress. Due to time constraints associated with IRB approval, only data from the second Zoom meeting, which was held on April 17, was used for analysis. This data was collected in an audio-only format and was transcribed manually and then anonymized.

Discord. In contrast to Slack and Zoom, which were used in the planning stages, Discord was used during UMDCTF. Although the CTF organizers used multiple Discord channels, our dataset includes 321 messages from `#staff-text-chat` channel, which was the channel used solely by CTF organizers for event coordination purposes.

Survey. A short, anonymous survey was sent to all UMDCTF organizers who participated in the study. The survey questions are listed in Appendix 6 and concern the organizers' demographics and challenge development process.

4.2 Anonymization and Redaction

Prior to collection and anonymization, on April 7th, we asked all organizers of UMDCTF for consent to use their data in this study and informed them that their messages would be removed and all PII related to them would be redacted. Willing organizers had their names and their known aliases/chat handles recorded. Each participant and all their aliases were mapped to a uniquely generated participant ID.

Slack/Discord. Upon obtaining the Slack and Discord datasets, we converted the raw messages into a standard JSON format with the following key types: `author`, `message`, `attachment`, `timestamp`, `channel`, and `app`.

Next, we anonymized any relevant PII in other messages by replacing the organizer's name and associated aliases with the participant ID. We verified the automated anonymization process manually. All organizers who did not agree to participate had their messages changed to 'REDACTED'. Any relevant identifiers were also removed.

Zoom. The transcribed Zoom calls were divided into sentences to match the formatting of the Slack and Discord chats. We then applied the same anonymization techniques as described above, replacing participant names and aliases with their unique identifier and redacting sentences spoken by non-participating organizers.

4.3 Data analysis

We used a qualitative open-coding process [22, pg. 101-122] to analyze the data. A single researcher built a codebook based on a preliminary chat dataset review. Next, two researchers analyzed Slack and Discord data and coded each message using Table 2's tags. During coding, the researchers continually updated the codebook to fit new observed message types. Any disagreements were resolved by a third researcher independently. The codebook resulted in a Cohen's Kappa (κ) of 0.64 for primary tags and 0.73 for secondary tags. These relatively low values were largely due to discrepancies in researcher interpretation for messages lacking context (e.g., due to redaction) since messages that followed context gaps were interpreted differently between the two coders. As our final κs indicate limited reliability, we present our following results as preliminary work requiring further analysis. We discuss this further in Section 4.4. The Zoom and survey data have yet to be coded and are only used here anecdotally due to time constraints (i.e., since this is currently a work in progress) for the Zoom data and low number of responses for the survey data.

After coding, we examined the results to identify key themes and select specific exemplar interactions.

Tag	Name	Description
	Primary Tags	
SF	Social Filler	Unproductive messages and interjections; ex: lol, hm, oh, memes, etc
EL	Event Logistics	Logistics surrounding the event itself, shirts, sponsors, food, meetings between organizers, etc
PL	Personnel Logistics	Logistics of an organizer (personal events)
TL	Technical Logistics	Messages related to technical infrastructure of challenges, such as GitHub issues, server issues, and which challenges should be in what categories
PC	Problem Creation	Related to a challenge's creation
PS	Problem Scoring	Related to how a challenge's scoring
PI	Participant Discussion	Interacting with or discussing participants of the CTF
PT	Problem Testing	Relating to testing of fully / semi developed challenges
UNK	Unknown	Researchers could not decide because of context
NA	Not Applicable	Redacted messages
	Secondary Tags	
P+	Positive	Supportive or in agreement
N=	Neutral	Neither Positive or Negative
N-	Negative	Combative, aggressive, or in disagreement

Table 2: The codebook used for discord and slack chats.

4.4 Limitations

Due to the fixed-time nature of the event under investigation and unexpected delays in the ethics approval process, data collection did not start until after CTF organization began. This limited data collection from the organizers' various sources. In particular, a Zoom call one week prior to the event was not analyzed for this reason. Further, during UMDCTF, organizers continuously communicated using a voice call in Discord that was not recorded. As a result, gaps may exist in our data. This was addressed by using an "UNK" tag in the codebook for cases in which context could not be established while coding.

Second, not all UMDCTF 2020 organizers gave consent to collect and use their messaging data. Out of the 11 organizers, one elected to remove themselves from our data. While we expected this, as CTF organizers are often privacy-conscious, the participant authored a significant number of messages and as a result, had a significant impact on data analysis. The lack of context made it difficult to achieve code agreement for many messages. Ultimately, this hindered the extent to which we could draw general conclusions. We were able to compensate for data redaction difficulties by iteratively reevaluating our code book, and by having a third researcher resolve any discrepancies between the two primary coders. However, because of our low reliability score, our results should be considered preliminary, pending further analysis.

Since organizers were only observed through group communication, the entirety of the planning process was not present in the recorded chat. Organizers were likely to create complete challenges without mentioning them in the group chats, preventing a complete understanding of the challenge creation process from the collected data. Similarly, group chats, as opposed to direct messages, add a social desirability component that might affect how organizers communicate, since chats are observed by all group chat members. We attempted to limit the effect of this through the use of a survey to gain closer insight into the process.

Two of the researchers were also UMDCTF 2020 organizers, this could have potentially led to some bias in the data and its analysis. This was mitigated by having all data collected regardless of researcher status. Additionally, neither of the organizer-authors participated as coders, as not to introduce bias into the coding process.

Every CTF is organized differently. UMDCTF 2020 organizers differ from other CTF organizers and even UMDCTF organizers from prior years. This year's event was particularly unique because of the transition to an online event due to COVID-19. Future work should evaluate the organizational process for other CTFs.

5 Results

In this section, we present the findings from our UMDCTF case study, organized by our research questions (Section 1).

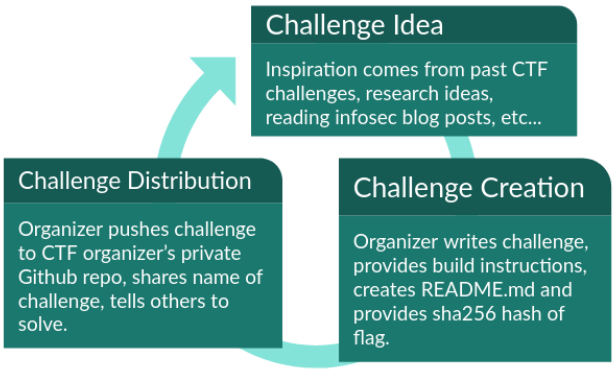

Figure 1: The challenge development lifecycle.

5.1 Challenge Development Lifecycle (RQ1)

Well formed challenges are central to a CTF's success. As a result, the challenge creation process is important. In this section, we analyze the challenge creation process to explore the techniques employed by CTF organizers. Furthermore, we visualize the challenge development lifecycle in Figure 1. The challenge development life cycle outline we highlight comes predominantly from the Slack data we collected, where challenge development was most frequent.

Challenge Idea. Every challenge begins with an idea. Anecdotally, we observed that challenges are often inspired by a past challenge. Sometimes, a challenge can be inspired by a research topic (i.e., from an organizer's internship experience), or by learning about a cybersecurity concept online.

Challenge ideas are kept secret from the other organizers. Although this is not always the case, we rarely saw instances of organizers discussing challenge ideas before the start of the competition. We believe this occurs to prevent similar challenges from being created. However, some organizers believe in the benefits of sharing challenge ideas. For example, one organizer stated, "working on challenges tonight if anyone wants to bounce ideas."

Once inspiration is determined, the organizer then begins challenge creation.

Challenge Creation. At this stage, the organizer creates their challenge and embeds the flag into the problem. Once the challenge development is finished, the organizer commits any associated challenge and README files that contain the problem description and the flag's SHA256 hash into the organizer's code sharing platform of choice. In our case

study, the UMDCTF 2020 organizers used a private Github repository for challenge collaboration. We note that our study does not focus on the specific challenge creation process or how an organizer goes through the process of creating a challenge, but rather on how challenges are incorporated into the overall competition and CTF development.

Challenge Infrastructure. Many challenges such as exploitation challenges must be run on a server. As a result, challenges creators likely have to consider the infrastructure of the CTF when creating them. Although we anticipated more messages to discuss challenge infrastructure, and its effect on challenge creation, little information was found. As a result, future research is needed to determine its effect.

After challenge creation, the challenge development like cycle then moves on to testing, vetting, and scoring.

5.2 Challenge Distribution and Vetting (RQ2)

Once an organizer adds their challenge to the repository, they share it on Slack so others can try to solve it. For instance, we noticed that for nearly every new CTF challenge added to the private repository, the organizer broadcasted a message such as: "@channel New Challenge! [Challenge Name]."

Challenges are mostly reviewed by multiple organizers. Once released, other organizers are tasked with attempting to solve the challenge. For most challenges, this meant testing by multiple organizers. However, this is not always possible. Organizers have limited time due to the voluntary nature of CTF organizing. Therefore, some challenges were only tested by one other organizer prior to release. This can lead to some challenges having bugs, being unsolvable, or too easy.

Challenge Write-ups. Organizers must decide whether to include challenge "write-ups". We observed an interesting debate between organizers about whether to include this documentation for every challenge. One organizer proposed that write-ups should include a brief challenge description, a recommended solution, and any appropriate hints. The same organizer further suggested these write-ups would assist in answering participant questions, could be released after UMDCTF 2020 for others to learn from, and would help the challenge author document their idea, simplifying challenge testing. In response, other organizers argued challenge *creation* should be prioritized over challenge *evaluation*. In fact, one organizer said, "if [writing write-ups] were mandatory, I'm not [creating write-ups]." We suggest organizers, while open to new ideas, often act independently and in a voluntary capacity. In this manner, many CTF organizers view added work (such as self-evaluating CTF problems) as unnecessary.

Challenge Scoring. The final phase in the challenge creation process is assigning a difficulty to the CTF problem. Unfortunately, there is no clear evidence from our dataset that indicates a consistent method for difficulty rating. We identified specific instances of organizers struggling to assign their challenge's point values. In one case, an organizer said that although they created a challenge, they had "no idea the difficulty." Ultimately, the organizer estimated the problem's difficulty to the best of their ability, assigning a high point value.

Further, there appears to be an inherent bias in challenge score assignment; the organizer created the challenge and therefore knows exactly how to solve it. Their peers may not necessarily have the same skills and experience, which can vastly change the difficulty of a challenge. Also, challenge creators may not be aware of techniques which could make a challenge easier. For example, one moderate difficulty challenge was solved in under a minute using a technique unknown to the creator. Testing could be used to determine the challenge difficulty, but this raises several issues. First and foremost, a volunteer group such as the UMDCTF organizers do not have the resources for large-scale testing to adequately assess challenge difficulty. This is evident as the organizers constantly reminded each-other to test challenges even as a significant number of challenges were left untested, leaving many challenges with less reliable scores.

Scoring System Options. Organizers have two options for scoring CTFs: static and dynamic. In static scoring, a challenge is worth a constant number of points, while in dynamic scoring the points may change. Suggested factors for a dynamic score are the flag submission accuracy rate, the time between opening a challenge and solving it, and the number of teams who have solved the challenge. The first two factors add more skill refinement to scoring, whereas the third democratizes the difficulty rating. However, this can cause challenge scores to rapidly decrease, as less experienced participants will attempt mostly lower scored challenges and when they are solved, the scores will decrease. Similarly, participants with less experience will not know which challenges to attempt at the beginning of the CTF since they all have the same score. This was a strong source of debate for the organizers, who carefully considered these options. Additionally, since the CTF was transitioned to online, team size limits were removed for the first time in the CTF's history. This meant that the decision on which scoring method to implement required further thought since it could provide larger teams unfair advantages. As a result, organizers must weigh the pros and cons of each option in order to decide how their CTF will be scored.

5.3 Organizer Motivation and Learning Experiences (RQ3)

Organizer Motives. CTF organizers want to teach and support the organization. Before addressing the challenge creation process, it is important to understand the motivation behind CTF organizing. In our survey, we identify two explicit motivations. First, organizers desire to share their knowledge. 60% of respondents reported this as their motivation for orga-

nizing, pointing to the relationship between CTF organizing and teaching. CTF organizers are also motivated to assist in challenge development to improve their organization's CTF. Our survey shows that 40% of organizers were initially driven to help organize after requests from other CTF organizers.

Organizers Learning Technical Concepts. The educational value of challenge creation is evolutionary as opposed to revolutionary. Organizers reported in our survey that they preferred selecting challenge topics within their area of knowledge and interest. Then, they learn new information about that research area. One respondent said, "... creating challenges makes you [learn] some nuanced skills and ideas." Another stated, "... [challenge creating] gives a different perspective and allows me to think about how a program should function rather than just breaking it." This experience allows authors to gain a deeper understanding of security concepts through multiple perspectives [4]. Similarly, since challenge authors must have knowledge on both how to build and break a specific idea, this requires a level of understanding that is difficult to acquire for unfamiliar topics. Instead, they use their existing knowledge to create challenges, resulting in an expansion on already understood topics.

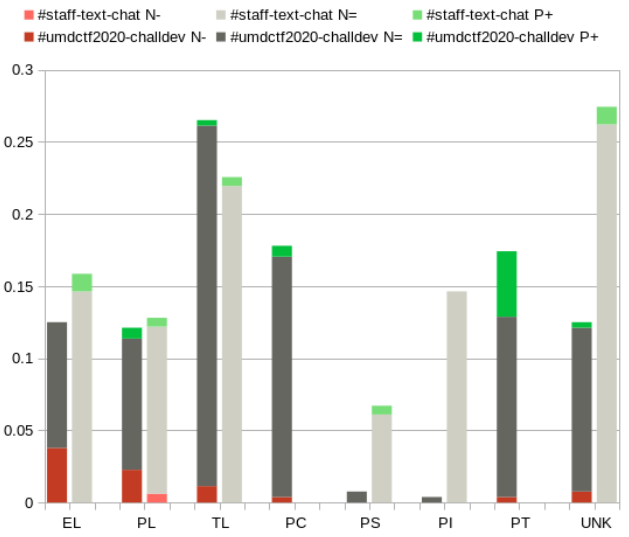

Figure 2: This chart shows the frequency of primary and secondary tags for each tag for the channels #umdctf2020-challdev coding (Slack, before the CTF) and #staff-text-chat coding (Discord, during the CTF).

5.4 Organizers' Focus Shift During the CTF (RQ4)

After coding, frequency of primary and secondary tags for each communication medium are shown in Figure 2. Since Slack was used primarily before the CTF and Discord during, the two mediums effectively show the chat content prior to and throughout the CTF, respectively. Before the CTF, chats were more serious, with more focus on creation and testing. During the CTF, chats were less likely to be negative and more likely to be concerned with scoring and participants. This matches the organizer's likely most pressing concerns at the time. Since organizers are performing the challenge development lifecycle to output as many quality challenges as possible before the CTF, they are likely going to partake in challenge focused discussion as opposed to CTF focused discussion. This instead occurs while the CTF is happening since organizers were shown to more frequently discuss scoring and participants. Interestingly, instead of discussing problem scoring before the CTF, its increased rate of discussion during the CTF shows organizer's have trouble with scoring challenges, since the challenge would have been already submitted at this point. Overall, organizers show a focus shift from before to during the CTF that supports the challenge development lifecycle.

6 Conclusion and Future Work

The educational merit of CTFs is well established in academic and industry research. Until now, the practice of creating and organizing CTFs has not been thoroughly examined. We present preliminary results of a baseline analysis of CTF organizational communications. Our work illustrates how a CTF is developed and showcases discussions, decisions, and behavior patterns from the CTF organizer perspective. Additionally, we uncover multiple problems that organizers face including how to mitigate creation bias in scoring challenges, perform the steps of challenge creation, and ensure that challenges are appropriately vetted despite limiting testing and resources.

This study was limited in scope; to expand on these concepts, more research is welcome on creating, organizing, and hosting better, more effective CTFs. More complete access to chat channels and meetings would also help to provide more context and allow more in-depth findings. To this end, we plan to carry out similar studies both on future iterations of UMDCTF and other CTFs including those hosted both academically and professionally . We envision active collaboration with other CTF organizers to gain a more diverse and generalizable understating of the CTF creation process.

References

[1] Susan A Ambrose, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman. How learning works: Seven research-based principles for smart teaching. John Wiley & Sons, 2010.

[2] Lecia Jane Barker, Kathy Garvin-Doxas, and Michele Jackson. Defensive climate in the computer science classroom. SIGCSE Bull., 34(1):43–47, February 2002.

[3] S. Bratus. What hackers learn that the rest of us don't: Notes on hacker curriculum. IEEE Security Privacy, 5(4):72–75, 2007.

[4] Ann L. Brown. The development of memory: Knowing, knowing about knowing, and knowing how to know. volume 10 of Advances in Child Development and Behavior, pages 103 – 152. JAI, 1975.

[5] Peter Chapman, Jonathan Burket, and David Brumley. Picoctf: A game-based computer security competition for high school students. In Proceedings of the 1st USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14), San Diego, CA, August 2014. USENIX Association.

[6] Kevin Chung and Julian Cohen. Learning obstacles in the capture the flag model. In Proceedings of the 1st USENIX Summit on Gaming, Games, and Gamification in Security Education, 3GSE '14, San Diego, CA, 2014. USENIX Association.

[7] Paul Cobb, Erna Yackel, and Terry Wood. A constructivist alternative to the representational view of mind in mathematics education. Journal for research in mathematics education, pages 2–33, 1992.

[8] G. Conti, T. Babbitt, and J. Nelson. Hacking competitions and their untapped potential for security education. IEEE Security & Privacy, 9(03):56–59, may 2011.

[9] Chris Eagle, L. Clark, John, and School (U.S.) Naval, Postgraduate. Capture-the-flag: Learning computer security under fire, 2004.

[10] K Anders Ericsson and Neil Charness. Expert performance: Its structure and acquisition. American psychologist, 49(8):725, 1994.

[11] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. Psychological review, 100(3):363, 1993.

[12] Gary A Klein. Sources of power: How people make decisions. MIT press, 2017.

[13] J. Margolis and A. Fisher. Unlocking the Clubhouse: Women in Computing. The MIT Press. MIT Press, 2002.

[14] L. McDaniel, E. Talvi, and B. Hay. Capture the flag as cyber security introduction. In Proceedings of the 49th Hawaii International Conference on System Sciences, pages 5479–5486, 2016.

[15] Jelena Mirkovic and Peter A. H. Peterson. Class capture-the-flag exercises. In Proceedings of the 1st USENIX Summit on Gaming, Games, and Gamification in Security Education, San Diego, CA, August 2014. USENIX Association.

[16] Tyler Nighswander. Building a competitive hacking team. San Francisco, CA, January 2016. USENIX Association.

[17] Ernest T Pascarella and Patrick T Terenzini. How college affects students: Findings and insights from twenty years of research. ERIC, 1991.

[18] Jean Piaget. Success and understanding. Routledge, 1978.

[19] Andrew Ruef, Michael Hicks, James Parker, Dave Levin, Michelle L. Mazurek, and Piotr Mardziel. Build it, break it, fix it: Contesting secure development. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16, pages 690–703, New York, NY, USA, 2016. ACM.

[20] Elaine Seymour and Nancy M. Hewitt. Talking About Leaving: Why Undergraduates Leave the Sciences. Westview Press, 2000.

[21] Casey A Shapiro and Linda J Sax. Major selection and persistence for women in stem. New Directions for Institutional Research, 2011(152):5–18, 2011.

[22] Anselm Strauss and Juliet Corbin. Basics of qualitative research, volume 15. Newbury Park, CA: Sage, 1990.

[23] Daniel Votipka, Michelle L. Mazurek, Hongyi Hu, and Bryan Eastes. Toward a field study on the impact of hackingcompetitions on secure development. 8 2018.

Survey Text

1. How many (roughly) CTFs have you taken part in as an organizer?
2. How many CTFs did you participate in before you decided to become an organizer?
3. What caused you to start taking part in CTF challenges as an organizer rather than a participant?
4. CTFs often require an infrastructure for the event to run successfully. Have you assisted in building or setting up the infrastructure for a CTF you have organized? If so, what did you do?
5. Do you believe that organizing CTF events has strengthened your technical computer security knowledge? If so, how has it strengthened this knowledge?
6. In general terms, what is your workflow when creating individual challenges for a CTF event, from conception to implementation?
7. Is this workflow altered depending on the type of the challenge you are trying to create?
8. Is this workflow altered depending on the difficulty of the challenge you are trying to create?
9. When generating challenges for CTF events, how often are challenges from previous CTFs copied, modified, or adapted?
10. How do you select technical topics for individual challenges?
11. Do you often generate challenges that you would be able to solve beforehand, or do you find yourself often doing research on technical topics in order to generate challenges? Please explain.
12. At what point in the challenge creation workflow is the "difficulty" of the challenge established?
13. Do you believe that your understanding of a question you created can bias your "difficulty" measure of that challenge?
14. What is your gender?
    - Male
    - Female
    - Prefer not to say
15. What is your age?
16. What is your educational status?
    - In High School

- High School Degree attained
- In Undergrad
- Undergrad degree attained
- In Graduate School
- Graduate degree attained
- Prefer not to answer
- Other