Please cite this article as follows: V. Švábenský, P. Čeleda, J. Vykopal, and S. Brišáková, *Cybersecurity Knowledge and Skills Taught in Capture the Flag Challenges*, Elsevier Computers & Security, 2020, ISSN 0167-4048, DOI: 10.1016/j.cose.2020.102154, URL: https://www.sciencedirect.com/science/article/pii/S0167404820304272

Cybersecurity Knowledge and Skills Taught in Capture the Flag Challenges

Valdemar Švábenský (corresponding author)

Masaryk University, Brno, Czech Republic

Pavel Čeleda

Masaryk University, Brno, Czech Republic

Jan Vykopal

Masaryk University, Brno, Czech Republic

Silvia Brišáková

Masaryk University, Brno, Czech Republic

Abstract

Capture the Flag challenges are a popular form of cybersecurity education, where students solve hands-on tasks in an informal, game-like setting. The tasks feature diverse assignments, such as exploiting websites, cracking passwords, and breaching unsecured networks. However, it is unclear how the skills practiced by these challenges match formal cybersecurity curricula defined by security experts. We explain the significance of Capture the Flag challenges in cybersecurity training and analyze their 15,963 textual solutions collected since 2012. Based on keywords in the solutions, we map them to well-established ACM/IEEE curricular guidelines to understand which skills the challenges teach. We study the distribution of cybersecurity topics, their variance in different challenge formats, and their development over the past years. The analysis showed the prominence of technical knowledge about cryptography and network security, but human aspects, such as social engineering and cybersecurity awareness, are neglected. We discuss the implications of these results and relate them to contemporary literature. Our

results indicate that future Capture the Flag challenges should include non-technical aspects to address the current advanced cyber threats and attract a broader audience to cybersecurity.

Keywords: cybersecurity education, security training, capture the flag, curricular guidelines

1. Introduction

Training security professionals is a slow but steady solution to the global cybersecurity workforce gap [1]. Educational institutions, computing societies, government organizations, and private companies are aware of this situation and introduce new curricula, study programs, and courses. Cybersecurity is an integral part of ACM/IEEE Computing Curricula 2020 (CC2020) [2], and specialized cybersecurity curricula, such as CSEC2017 [3], have been emerging in recent years.

Along with formal education, an increasingly popular method of practicing cybersecurity skills is via informal *Capture the Flag* (CTF) games and competitions. In these events, small teams of participants exercise their cybersecurity skills by solving various tasks in an online learning environment. CTF tasks, called *challenges*, feature diverse assignments from exploiting websites, through cracking passwords, to breaching unsecured networks. A successful solution of a challenge yields a text string called a *flag* that is submitted online to prove reaching the solution.

CTF originated among cybersecurity enthusiasts at a hacker conference DEF CON in 1996 [4]. However, CTF is no longer the niche of exclusive hacker groups. This educational game format quickly gained popularity, and now, teachers across the world are using it to complement education. CTF has been used successfully in university classes [5, 6] and in undergraduate security competitions [7, 8]. Even tech giants like Google and Facebook host CTFs [9, 10] that attract hundreds of attendees every year. Unlike traditional teaching formats, such as lectures and homework assignments, CTFs are more

Email addresses: svabensky@ics.muni.cz (Valdemar Švábenský (corresponding author)), celeda@ics.muni.cz (Pavel Čeleda), vykopal@ics.muni.cz (Jan Vykopal), brisakova@mail.muni.cz (Silvia Brišáková)

casual and often include competitive or game elements. However, because of their informality, it is unclear how they fit into cybersecurity curricula.

CTF participants publish their solutions to the challenges online. They do it to demonstrate solving the tasks and to share their knowledge with others. The solutions, called *writeups*, are useful mainly in two ways. First, they are a learning resource that describes how the challenge was solved, which can prove useful in future CTFs and allow others to discover new solutions. Second, the writeups inspire CTF creators since they provide insight into the challenge assignment, even if the assignment is no longer available. We will investigate a third possible yet unexplored use of writeups. In our research, we regard them as a dataset and mine information about cybersecurity topics from them.

1.1. Goals of This Paper

By analyzing the content of writeups, we examine how the informal CTF challenges map to formal CSEC2017 curricular guidelines defined by security experts. We seek to uncover the breadth of the cybersecurity topics that CTF can teach to enhance education and training. Specifically, we pose the following three research questions.

1. *What is the distribution of cybersecurity topics in CTF challenges?*
2. *How does the distribution of topics differ between various CTF formats?*
3. *How has the distribution of topics evolved over the past decade?*

The first question explores dominant, typical, and underrepresented cybersecurity topics within the analyzed writeups. The second question divides the writeups according to the two most popular CTF formats and compares them, allowing educators to choose a suitable format for their learning activities. The third question splits the writeups based on the year of the corresponding CTF event and searches for trends over the years.

1.2. Contributions of This Paper

Online CTF challenges feature practical assignments, scale to hundreds of students, and include game elements. They are suitable for secondary, tertiary, professional, and extracurricular education. This paper aims to support their further transfer into the practice of teaching and learning security. Answering the research questions will be valuable for various stakeholders.

- *Cybersecurity experts* will know which cybersecurity skills they or their team members can practice via CTF.
- *Educational managers* and *curricular designers* can see how informal education via CTF helps fulfill formal cybersecurity learning outcomes. Moreover, at a higher level, they can consider which cybersecurity topics can be supplemented by CTF in their study programs.
- *Teachers* and *CTF content creators* may focus on more common cybersecurity knowledge to help students interested in CTF. Alternatively, they can teach the fields uncovered by CTF.
- *Students* and *CTF participants* can better understand the content of previous challenges and prepare for future challenges.

1.3. Paper Structure

This paper is organized into seven sections. Section 2 explains the key terms to familiarize the readers with CTF challenges, writeups, and cybersecurity curricula. Section 3 describes primary and secondary studies related to writeups, curricular design, and educational text analysis. Section 4 details our methods for the collection and analysis of writeup data. Section 5 presents the findings and answers the three research questions. Section 6 offers practical insights and lessons learned from this research. Finally, Section 7 concludes and summarizes our contributions.

2. Background and Terminology

This section defines the key terms used throughout the paper: Capture the Flag in Section 2.1, writeups and their web catalogs in Section 2.2, and Cybersecurity Curricular Guidelines in Section 2.3.

2.1. Capture the Flag and Its Formats

The term Capture The Flag originally refers to an outdoor game for two teams. Each team must simultaneously defend a (physical) flag in their base and steal the other team's flag. Since the 1990s, this playground has also moved to cyberspace. In cybersecurity, the term CTF denotes a broad spectrum of events with different scope and format. These include online competitions in attacking only, educational games with the instructional support of learners, or games played just for entertainment at popularization events.

CTF can be hosted on various technical platforms [11]. The infrastructure is usually online, making CTF suitable for distance and blended learning. In fact, most CTFs are held remotely, although the participants can also meet on a physical location (see Table 1). The most common formats of CTF are *jeopardy* and *attack-defense*.

In a jeopardy CTF, the participants choose from challenges divided into categories, such as cryptography, reverse engineering, or gaining ownership of a service (*pwn* in hacker jargon). Each challenge is usually worth a different number of points, imitating the format of the famous television show *Jeopardy!*. The participants solve the challenges locally at their computers or interact with a remote server. An example task is that the participants receive a binary file containing an encrypted flag, which they have to recover.

In an attack-defense CTF, each team of participants controls and maintains an identical instance of a computer network, whose hosts run vulnerable services. The goal is to patch the services and protect the network assets while exploiting the vulnerabilities in the services of other teams. The scoring is based on a combination of successful exploits and defensive countermeasures while maintaining the services' availability.

2.2. CTF Web Catalogs

CTF participants publish their writeups on different websites, such as GitHub, YouTube, or personal blogs. We focus on CTFtime.org [12], which is the most prominent web portal about CTF. Since its foundation in 2012, this community-run project has been collecting information about past CTF events, planned future events, team rankings, and challenge writeups. Figure 1 shows an example of a writeup (a step-by-step solution) posted on the CTFtime website. It refers to the jeopardy challenge from Section 2.1.

Apart from jeopardy and attack-defense, CTFtime lists a third format of CTF: Hack Quest. However, this term is rarely used [13], and according to the available information, we did not find any difference from jeopardy. Therefore, we consider these two terms a synonym.

Table 1 shows that the number of CTFs listed at CTFtime increases each year. This growing popularity indicates that more people participate in CTF events. As a result, more data are generated that can be analyzed. Jeopardy is undoubtedly the dominant format, followed by attack-defense. Moreover, approximately two-thirds of CTFs are available remotely. The rest takes place at a physical location (typically the finals). Remote CTFs are more accessible to a wide range of participants.

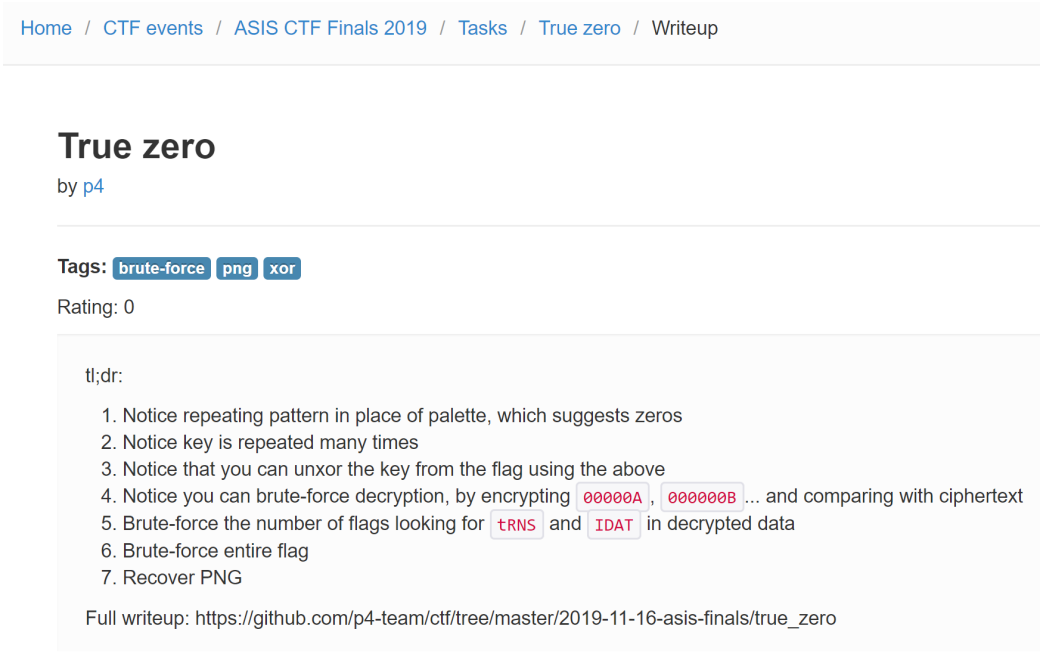

Figure 1: Example of a writeup. Source: `https://ctftime.org/writeup/17308`.

2.3. Cybersecurity Curricular Guidelines

Curricula are formal documents that describe the knowledge and skills taught at an educational institution. Prominent curricula in computing, such as the ACM/IEEE 2013 computing curricula [14] and CC2020 [2], include only broad cybersecurity topics. That is why specialized cybersecurity curricula started to emerge (see [15, 16] for a detailed overview). Among these, we chose the *Cybersecurity Curricular Guidelines* (CSEC2017) developed by The Joint Task Force on Cybersecurity Education [3] because they are widely established in the field.

CSEC2017 defines eight *Knowledge Areas* (KAs) in cybersecurity. Each KA encompasses different skills and knowledge.

1. *Data security* includes cryptography, forensics, data integrity, and authentication.
2. *Software security* focuses on secure programming, testing, and other aspects of software development.
3. *Component security* deals with the security of components integrated into larger systems, including their design and reverse engineering.

	Division by Game Format			*Division by Location*		
Year	**Jeopardy**	**Attack-Defense**	**Hack Quest**	**Remote**	**On-site**	**Total**
2012	23	10	2	19	16	35
2013	41	13	1	35	20	55
2014	49	8	1	36	22	58
2015	65	12	2	48	31	79
2016	90	14	3	67	40	107
2017	125	14	2	102	39	141
2018	136	16	1	102	51	153
2019	175	20	3	145	53	198
2020	126	13	4	130	13	143
Total	830 (86%)	120 (12%)	19 (2%)	684 (71%)	285 (29%)	969

Table 1: The numbers of all CTF events posted on CTFtime.org [12] from January 1, 2012, to October 9, 2020, divided according to the game format and location. Since the year 2020 is incomplete, it has fewer events so far. The small number of on-site events in 2020 is probably due to worldwide COVID-19 restrictions.

4. *Connection security* encompasses network services, defense, and attacks.
5. *System security* aims at securing a system as a whole, including access control and penetration testing.
6. *Human security* is about protecting individuals' identity, data, and privacy. It includes social engineering and cybersecurity awareness.
7. *Organizational security* governs risk management, security policies, and incident handling at the level of organizations.
8. *Societal security* examines cybersecurity at the national or global level. It concerns cybercrime, cyber law, and governmental policies.

Each KA is divided into *Knowledge Units* (KUs), which are further split into *Knowledge Topics*. (Although CSEC2017 uses *Topic*, we changed it to *Knowledge Topic* to introduce the abbreviation KT.) Figure 2 provides an example of Data security KA. Its first KU, Cryptography, forms the first column of the table. The second column contains the names of the subordinate

KTs. The third column describes the knowledge and skills that belong to the KTs. Overall, there are 8 KAs, 55 KUs, and 287 KTs. This paper identifies the distribution of the KAs and KUs in the writeups of CTF challenges.

DATA SECURITY		
Essentials - Basic cryptography concepts, - Digital forensics, - End-to-end secure communications, - Data integrity and authentication, and - Information storage security.		
Knowledge Units	**Topics**	**Description/Curricular Guidance**
Cryptography		
	Basic concepts	This topic covers basic concepts in cryptography to build the base for other sections in the knowledge unit. This topic includes: • Encryption/decryption, sender authentication, data integrity, non-repudiation, • Attack classification (ciphertext-only, known plaintext, chosen plaintext, chosen ciphertext), • Secret key (symmetric), cryptography and public-key (asymmetric) cryptography, • Information-theoretic security (one-time pad, Shannon Theorem), and • Computational security.
	Advanced concepts	This topic includes: • Advanced protocols:

Figure 2: Excerpt from the CSEC2017 curricular guidelines [3, Chapter 4, page 24].

3. Related Work

This section presents the related publications and explains how this research differs from state of the art.

3.1. Analysis of CTF Writeups

The closest research publication to this paper is by Burns et al. [17]. They analyzed CTF writeups to find essential skills and knowledge to study when preparing to participate in CTF events. The work focused on 160 events

with about 3,600 solutions posted on GitHub in the years 2011–2016. They analyzed the data to develop challenges for beginners, which are grouped into six categories: *crypto*, *web*, *reverse*, *forensic*, *pwn*, and *misc*. Although the article did not specify the exact details of analysis methods applied to CTF writeups, the authors likely read and classified the writeups manually. We aim to automate the process to achieve the results faster and more reliably by reducing human errors. Moreover, we map the results to formally defined Knowledge Areas and Units of a cybersecurity curriculum.

In a paper by Chothia et al. [18], students of a cybersecurity course submitted writeups to CTF challenges they solved. The researchers manually graded these writeups and examined the correlations of the resulting grades with the number of submitted flags. Moreover, they considered the writeups as an indicator of whether students had understood the learning content. Similarly, Schreuders et al. [19] and Leune et al. [20] instructed their students to submit writeups of CTF challenges. However, the two papers did not mention further analysis of the writeups. Overall, there was no published attempt to map CTF writeups to a cybersecurity curriculum.

3.2. Analysis of Cybersecurity Skills and Curricula

Cabaj et al. [21] analyzed cybersecurity topics in 21 master's degree programs at top-ranking universities worldwide. They informed how the programs cover cybersecurity topics and how the topics are distributed among the taught courses. For the topic analysis, they chose ACM/IEEE 2013 computing curricula [14] complemented with the CSEC2017 [3]. Their results revealed the increasing importance of non-technical cybersecurity areas: *Human*, *Organizational*, and *Societal security*. In related work, Hallett et al. [22] mapped the content of various cybersecurity curricula onto the knowledge areas from the United Kingdom's Cybersecurity Body of Knowledge [23].

In a technical report from a European Union cybersecurity project [24], 96 universities were surveyed about which KAs from CSEC2017 they cover. They identified that *Data* and *Connection security* are covered the most, while *Organizational security* is covered the least.

Carroll [25] investigated the skills required in cyber warfare, specifically for developing a workforce in offensive and defensive cyberspace operations. He surveyed 23 cyberspace professionals from the military, civilian, and private sector about their core knowledge, skills, and abilities. As a result, he recommends actions to improve the preparation for cyberspace operations.

Jones et al. [26] examined the essential knowledge, skills, and abilities for cybersecurity jobs. They surveyed 44 cybersecurity professionals attending the major hacker conferences Black Hat and DEF CON. As a result, they suggest prioritizing knowledge about networks, vulnerabilities, programming, and interpersonal communication. In related work, Haqaf and Koyuncu [27] used the Delphi method [28] (a structured communication technique) to discover the key skills for information security managers. Finally, Brooks et al. [29] analyzed IT security job advertisements to determine the skills that employers are interested in.

In our previous work [30], we performed a literature review of 71 cybersecurity education papers. As a part of the review, we mapped the content of the papers to the CSEC2017. The dominant KA was *Data security*. Other technical KAs, *Software*, *Connection*, and *System security* were strongly present too. Nevertheless, *Human*, *Organizational*, and *Societal security* were not neglected either. *Component security* was the least common.

3.3. Analysis of Topics in Other Textual Data to Support Education

Latent Dirichlet allocation (LDA) is a method for probabilistic topic modeling [31]. Given a set of text documents, LDA discovers their underlying topics, which are probability distributions over the words (terms) in the documents. LDA is a commonly used method that addresses the limitations of probabilistic latent semantic indexing [32]. Marçal et al. [33] used LDA to identify computing topics in questions asked on Stack Overflow. Similarly, Rouly et al. [34] used LDA to analyze course descriptions in course catalogs of different universities.

Nadeem et al. [35] developed a method for recommending relevant reading materials to software developers based on vulnerabilities in their code. They statically analyzed program source code to discover the vulnerabilities. Then, they computed the cosine similarity between the description of the vulnerabilities and public articles in the Common Weakness Enumeration (CWE) database. To model the articles, they used a standard approach of Term Frequency – Inverse Document Frequency (TF–IDF) [36, p. 302]. It assigns weights to the words in text based on their frequency and importance.

4. Methods

This section explains the methods we chose to answer the research questions posed in Section 1.1.

4.1. Extracting Cybersecurity Keywords from the CSEC2017 Guidelines

To identify cybersecurity topics covered by CTFs, we searched for cybersecurity keywords in the writeups. We began by extracting these keywords from the CSEC2017 [3]. For each of the eight tables with Knowledge Areas (see the example in Figure 2), one author manually extracted keywords from the third column, "Description/Curricular Guidance". Another author revised the extraction. Then, we repeated the process to minimize errors and ensure the inclusion of all relevant keywords.

We gathered 1,623 keywords and organized them in a JavaScript Object Notation (JSON) file [37]. Its excerpt is shown in Figure 3, and its structure follows Figure 2. KAs are JSON objects that contain other objects, KUs. Similarly, KUs contain an array of KT objects, which then contain individual keywords. To verify the file's correctness and syntax, we wrote a JSON validation schema [38] for it.

```
"Knowledge Area": "Data Security",
"Knowledge Units": [
    {
        "name": "Cryptography",
        "Knowledge Topics": [
            {
                "name": "Basic Concepts",
                "keywords": [
                    "encryption",
                    "decryption",
                    "sender authentication",
                    ...
```

Figure 3: Excerpt from the JSON file with the cybersecurity keywords we searched for in the writeups. Notice how the content corresponds to the curricula excerpt in Figure 2.

4.2. Downloading the Writeups

More than 969 CTF events with a total of 12,952 challenges (tasks) and 23,517 writeups have been posted on CTFtime since its foundation in 2012. We focused on the writeups of events that took place from January 1, 2012, to October 9, 2020, since 2012 is the first year to contain any writeups.

```
for each event posted on CTFtime.org:
    for each task in the event:
        for each writeup of the task:
            download the text of the writeup
```

Figure 4: Pseudocode for downloading the writeups.

To download the content of the writeups, we used the Python `requests` library [39] to implement the algorithm sketched in Figure 4.

Downloading the writeup text differed based on the content of the writeup. There were three possibilities:

1. The writeup text was present directly on the CTFtime webpage. In this case, we simply scraped it.

2. The writeup included only a link to an external website. If the website was GitHub, which was the most common case, we followed the link. Then, if the text contained at least one keyword, we scraped the raw file content. For other websites, such as the authors' blogs, we ignored the link. The reason was that each website had a different structure, so automating the download would be time-consuming.

3. The writeup included a combination of the text and an external link. If the external link was not on GitHub, we scraped the text as in option 1. If the external link was on GitHub, we scraped both the CTFtime text (option 1) and the GitHub text (option 2). Then, we counted the cybersecurity keywords in both files and kept the file with more keywords since we considered it more representative. Often, CTFtime included a sketch of the writeup (its subset) and GitHub its full text.

Then, we performed data cleaning, such as removing HTML tags and links to external websites. After the cleaning, we discarded writeups shorter than two characters because our shortest keyword was two characters long. The remaining writeups were categorized by year and format and saved as a text file. Altogether, these files act as the input for the analysis (see Section 4.3).

Table 2 shows the resulting number of downloaded writeups categorized by years and format. All the events we worked with are jeopardy or attack-defense. Although 19 events were Hack Quests, none contained any writeup, so we excluded them.

Year	Jeopardy	Attack-Defense	Total
2012	90	9	99
2013	145	3	148
2014	118	1	119
2015	419	0	419
2016	1,927	4	1,931
2017	2,499	15	2,514
2018	3,919	21	3,940
2019	3,153	17	3,170
2020	3,609	14	3,623
Total	15,879 (99%)	84 (1%)	15,963

Table 2: The numbers of downloaded and subsequently analyzed writeups posted on CTFtime.org [12] from January 1, 2012, to October 9, 2020.

Several factors caused the difference between the total number of writeups (23,517) and the downloaded writeups (15,963). The most common one was that the writeup was linked on an external website or written in a PDF, which we did not parse. In rare cases, the writeup was empty, deleted by the author, or did not pass through the data cleaning process.

4.3. Analyzing the Downloaded Writeups to Identify the Keywords

Figure 5 shows an overview of all entities that take part in the analysis. The analysis script takes two inputs: the keywords file and the downloaded writeups. Each writeup is represented using a *Bag of words* model [40], so the order of the words is disregarded, and we count the keywords in each writeup. Formally, we define the analysis as follows.

4.3.1. Input for the Analysis Script

The script for analyzing the writeups has two inputs:

- $K = \{k_1, \ldots, k_n\}$, a set of N keywords. We defined $N = 1{,}623$ keywords. Each k_i belongs to exactly one KT, which belongs to exactly one KU, which belongs to exactly one KA.
- $W = \{w_1, \ldots, w_m\}$, a set of M writeups. Overall, $M = 15{,}963$. W is partitioned into subsets for the second and third research questions.

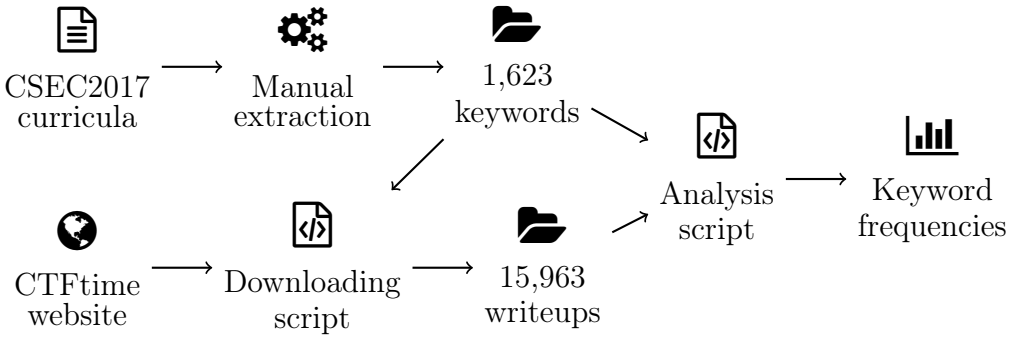

Figure 5: Overview of the data collection and analysis pipeline.

4.3.2. Counting the Keyword Frequencies

Given W and K, the goal is to compute $C = (c_{ij})$, an $M \times N$ matrix, where each c_{ij} is the count of occurrences of keyword k_j in writeup w_i. The value c_{ij} is further referred to as *Term Frequency* (TF). Figure 6 shows an illustrative example of counting the keywords in the writeups.

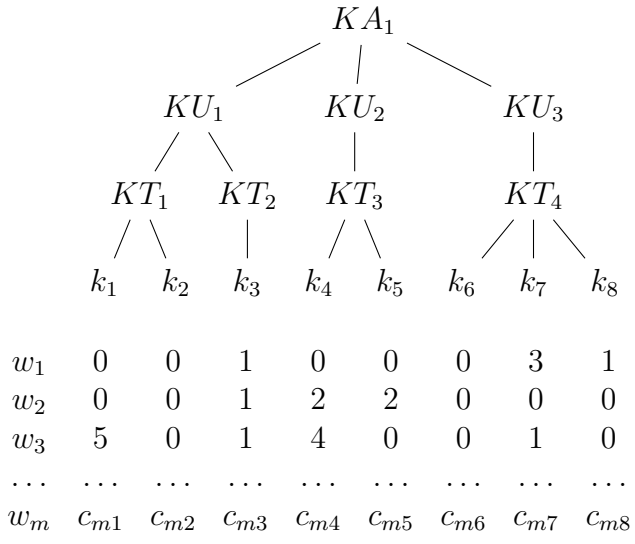

Figure 6: Example of counting the keyword frequencies in the writeups.

Note that the matrix will be sparse; it will contain a zero value for each keyword that does not occur in the writeup w_i. Since each writeup explains

the solution to a single challenge, it is extremely unlikely that many of the 1,623 keywords would be present in a single writeup.

When counting the keywords in writeups, the matching is always case insensitive because the writeups are often informal and written in lowercase. Moreover, two additional rules are applied. If the keyword is an abbreviation, such as LAN (for Local Area Network), we seek its exact match. However, if the keyword is not an abbreviation, we seek only its stem [41]. For example, if our defined keyword is "encryption", and the writeup contains the word "encrypting", we consider it a match.

4.3.3. Normalizing the Term Frequencies

The writeups have different lengths. Naturally, it is more likely for longer writeups to include more keywords. Therefore, to eliminate the bias of longer writeups over the shorter ones, we need to normalize the TF values. We could normalize by dividing each TF by the length of the writeup, but this would give us impractically small numbers. Therefore, we decided to divide all TF values by their sum within each row. For example:

$$w_1\colon\ [0,0,1,0,0,0,3,1] \xrightarrow{\text{normalization}} [0,0,\tfrac{1}{5},0,0,0,\tfrac{3}{5},\tfrac{1}{5}]$$

This normalization yields a value we call the *Normalized Term Frequency* (NTF), which has multiple benefits. First, it maps all the values in the matrix C to a common range $[0,1]$. Second, it preserves the relative differences between the original TF values. Third, the NTF values sum to 1 for each writeup w_i, and so can be easily represented by percentages.

4.3.4. Assigning the Writeups to KUs and KAs

The process of assigning a writeup w_i to a KU is as follows. Each KU is assigned the sum of NTFs of its respective keywords. For example, suppose that KU_1 contains keywords k_1, k_2, k_3. Then $c_{i1} + c_{i2} + c_{i3}$ is assigned to KU_1. Returning to the example matrix in Figure 6, given only the writeup w_1, KU_1 would receive the value $\frac{1}{5}$, KU_2 the value 0, and KU_3 $\frac{4}{5}$.

This calculation is applied to all writeups and all 55 KUs. Afterward, the assigned values are normalized by dividing them by M, the total number of writeups. We do this to achieve the same benefits as for the normalization above. Finally, the process is analogous for grouping KUs into KAs. The output of the analysis is the distribution of KUs/KAs in the writeups W.

We publish the analysis script with the supplementary materials for this paper [42]. Its basis was created in a previous project [43] and updated for this work. See also Section 7 for details.

5. Results and Discussion

This section answers the three research questions (RQ) about the distribution of cybersecurity topics overall, in the two CTF formats, and throughout the years 2012–2020.

5.1. RQ1: Distribution of Knowledge Areas and Units in CTF Writeups

We now answer the first research question by looking into the distribution of cybersecurity topics in the 15,963 analyzed writeups. In total, we identified 232,160 keyword matches, corresponding to about 14.5 keywords per writeup on average. Out of the 1,623 keywords, 1,012 were not found in any writeup. The ten most common keywords were: *log*, *password*, *exploit*, *encrypt*, *class*, *pwn*, *http*, *decrypt*, *crypto*, and *reverse*.

5.1.1. Overall Distribution of Knowledge Areas

Figure 7 shows the distribution of cybersecurity KAs in the writeups. We can see that the analyzed writeups incorporate each KA to some extent. The most prominent is *Data security*: among all keyword matches, more than 27% corresponded to it. The second place is taken by *Connection security*, and *System security* is the third.

These three KAs are above the average of 12%. They are the most popular probably due to containing skills and knowledge that are suitable to test in CTF challenges. *Data security* includes knowledge about cryptography, authentication, and secure communication. *Connection security* comprises network services and defense. Finally, *System security* involves penetration testing and multi-stage attacks.

Next, KAs *Software, Organizational, Component*, and *Human security* have similar values of around 8–10%, which are below the average. The third least is *Component security*, possibly because it often involves using physical devices. This requires the device to be physically present next to the participant, which is complicated and costly for remote CTFs.

The least frequent KA is *Societal security*, with only 3% value. This is not surprising since societal aspects of cybersecurity, such as privacy or cyber law, are not usually covered by CTF challenges. The creators of CTF primarily focus on technical knowledge.

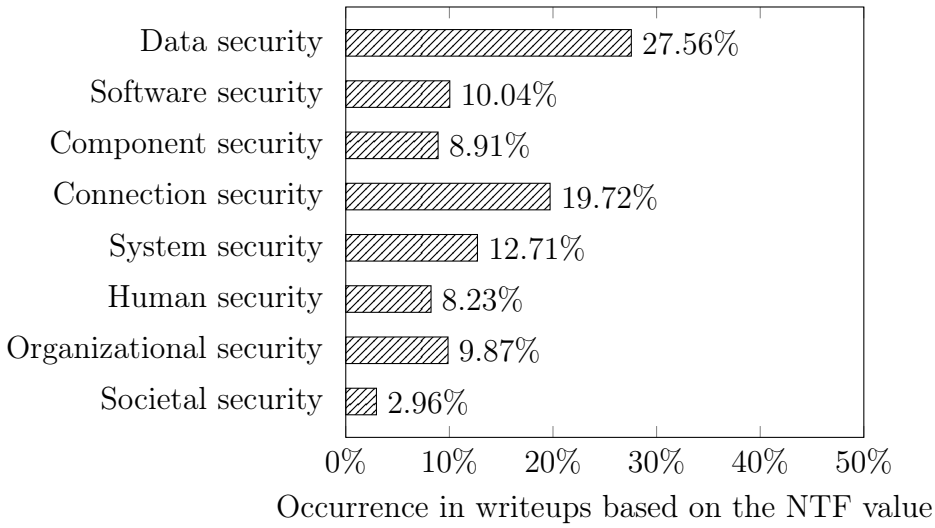

Figure 7: Overall distribution of KAs in the downloaded writeups.

5.1.2. Overall Distribution of Knowledge Units

Table 3 lists the top ten prominent KUs we identified, along with their key KT and a parent KA to provide context. *Cryptography* is the most prominent KU, arguably because of dealing with ciphers. Tasks requiring a participant to decrypt data are popular, and may also be used as a subtask in various bigger tasks. Moreover, symmetric and asymmetric cryptography are often core content in cybersecurity courses.

On the other hand, one of the 55 KUs was not present at all: *Physical interfaces and connectors*. We anticipated this result since most CTFs are remote and rarely involve the hacking of physical devices. Other least prominent KUs are again mostly about physical devices or belong to the societal parts of cybersecurity, such as *Cyber law*.

5.2. RQ2: Distribution of Knowledge Areas and Units in Jeopardy and Attack-Defense CTF Formats

This section answers the second research question. We compare the distribution of KAs and KUs in jeopardy and attack-defense formats separately.

5.2.1. Distribution of Knowledge Areas in Jeopardy and Attack-Defense CTF

Figure 8 shows that the distribution of jeopardy writeups closely resembles the overall distribution shown in Figure 7. This is natural, since jeopardy

%	Knowledge Unit	Key Topic	Parent Knowledge Area
14.9	Cryptography	Encryption	Data security
8.8	Component design	Reverse engineering	Component security
7.1	Implementation	Secure programming	Software security
7.0	System control	Penetration testing	System security
6.5	Digital forensics	Artifact analysis	Data security
5.8	Distributed systems architecture	HTTP(S), web attacks	Connection security
5.3	Network services	Network protocols and attacks on them	Connection security
5.3	Network implementations	TCP/IP, network attacks	Connection security
3.9	Identity management	Authentication	Human security
3.8	Business continuity, disaster recovery, and incident management	Incident response	Organizational security

Table 3: Ten most prominent KUs overall sorted by the NTF value.

writeups constitute more than 99% of the total (see Table 2). However, when looking at attack-defense writeups alone, it is apparent that the distribution differs. The top three jeopardy KAs are *Data*, *Connection*, and *System security*. For attack-defense, however, *Connection security* dominates, followed by *Data* and *Software security*.

Connection security is prominent in attack-defense because this format heavily relies on networking skills. By definition, attack-defense focuses on attacking systems of opposing teams via a network. Typically, this requires the participants to exploit network services, analyze traffic, or obtain credentials for a remote connection. On the other hand, jeopardy often includes standalone cryptographic challenges, giving rise to *Data security*.

5.2.2. Distribution of Knowledge Units in Jeopardy and Attack-Defense CTF

Again, the distribution of KUs for the jeopardy format alone follows the overall trends in Table 3. Because attack-defense CTFs have few writeups, they had almost no impact on the most prominent KUs overall. Out of the

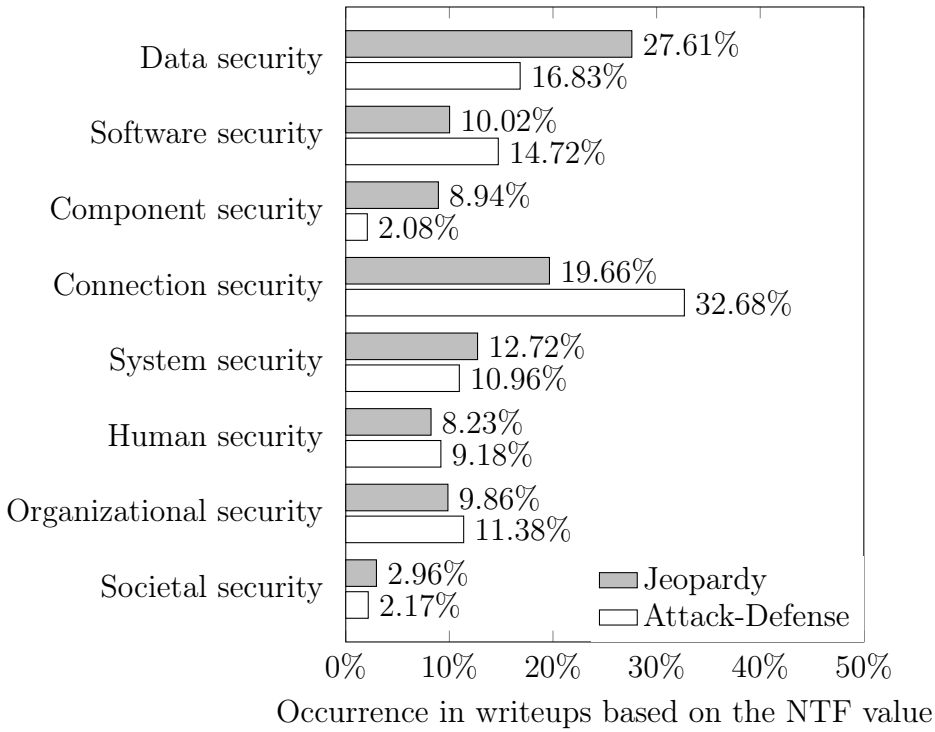

Figure 8: Distribution of KAs in 15,879 jeopardy and 84 attack-defense writeups.

top ten KUs in jeopardy, three of them belong to KA Connection Security: *Distributed systems architecture*, *Network services*, and *Network implementations*. Data Security contains *Cryptography* and *Digital forensics*, and the remaining KAs, excluding Societal security, have one representative each.

For the attack-defense format, Connection security has three KUs again among the top ten (see Table 4). Data security contains *Cryptography* and *Digital forensics*. Software security contains *Implementation* and *Deployment and maintenance*. Again, the remaining KAs, excluding Component and Societal security, have one representative each.

Regarding the least frequent KUs, we observed one KU not present in jeopardy writeups, but 22 KUs were not identified in any of the attack-defense writeups. An example in both formats is *Physical Interfaces and Connectors*, and we add the full list in the supplementary materials [42]. These may inspire cybersecurity educators to design new types of CTF.

We statistically compared whether the KU distribution difference between

%	Knowledge Unit	Key Topic	Parent Knowledge Area
12.3	Network implementations	TCP/IP, network attacks	Connection security
10.1	Distributed systems architecture	HTTP(S), web attacks	Connection security
8.9	Implementation	Secure programming	Software security
8.3	Digital forensics	Artifact analysis	Data security
8.2	Network services	Network protocols and attacks on them	Connection security
7.2	Business continuity, disaster recovery, and incident management	Incident response	Organizational security
6.7	System control	Penetration testing	System security
5.7	Social engineering	Deception	Human security
5.1	Deployment and maintenance	Software configuration and patching	Software security
5.1	Cryptography	Encryption	Data security

Table 4: Ten most prominent KUs for the attack-defense format sorted by the NTF value.

jeopardy and attack-defense is significant. Before choosing a statistical test, we ran a normality test on the whole computed matrix C. We chose the Anderson–Darling test implemented in the Python `scipy` library [44], which strongly rejected the hypothesis of the data having a normal distribution.

Due to the skewed data distribution, we considered only non-parametric statistical tests. We chose the Mann-Whitney U test implemented in the Python `scipy` library [45]. Each of the two CTF formats represents one test sample with 55 NTF values corresponding to each KU. The test indicated that the distribution difference of KUs between jeopardy and attack-defense format is significant ($U = 1184$, $p = 0.02$). However, we could not use this test to examine the distribution of KAs between the formats. This is because the Mann-Whitney U test requires at least 20 observations in each sample [45], but we have only eight observations (KA values).

5.3. RQ3: Distribution of Knowledge Areas and Units from 2012 to 2020

This section answers the third research question. We look at the variance in the distribution of KAs and KUs in the CTF writeups divided by year.

5.3.1. Distribution of Knowledge Areas per Year

Figure 9 shows that the distribution of KAs varies only slightly over the years. However, the chart highlights some deviations. *Data security* had a lower occurrence in 2012 and 2013. Similarly, *Software security* had the smallest presence in 2013 and 2014. However, *Component security* was the most popular in 2013. *Connection security* is steadily between 17–22% since 2015. *System security* peaked in 2014 and stayed at 7–15% in other years, with *Human* and *Organizational security* repeating a similar trend. As for the *Societal security*, it reached its highest percentage, 4.5%, in 2014 too.

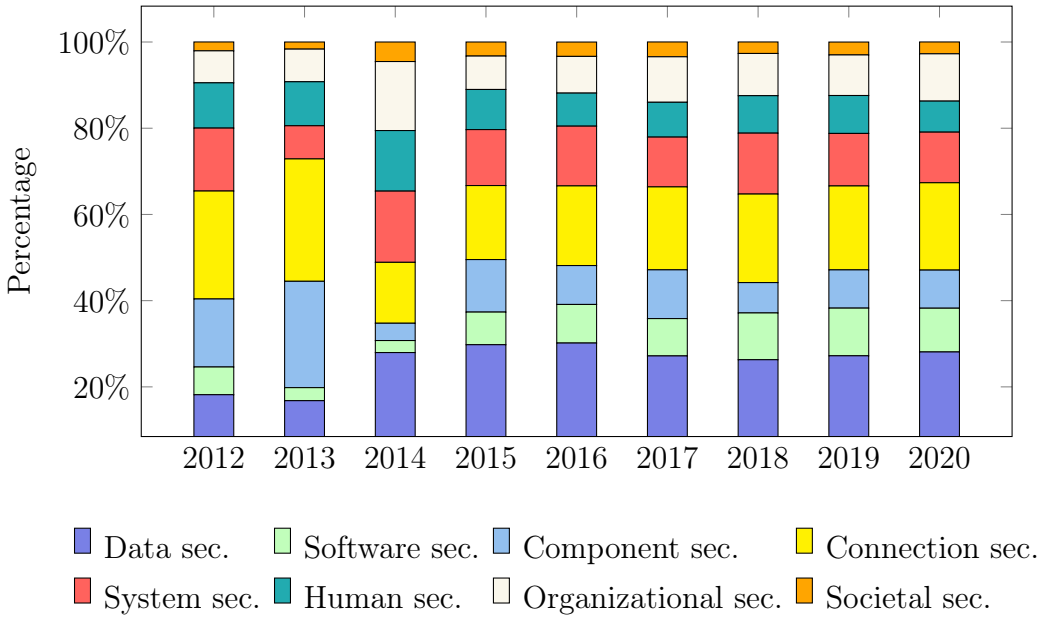

Figure 9: Distribution of KAs in years 2012–2020.

5.3.2. Distribution of Knowledge Units per Year

Top ranking KUs do not vary much. Between the years 2012–2020, 15 different KUs appeared at the top ten spots, and 12 of them appeared more than once. The steady ones appearing each year are *Cryptography*, *Component design*, and *Network implementations*. Those that appeared every year

except one are *System control*, *Digital forensics*, *Distributed systems architecture*, *Network services*, and *Systems administration*. Finally, *Implementation* appeared all years except 2013 and 2014.

We tested the null hypothesis that the medians of KU values throughout 2012–2020 are equal to see if there are statistically significant differences. We used Kruskal-Wallis test [46] to examine the differences in the distribution of KUs, running it on nine samples (years 2012–2020) with 55 observations (KUs). The test concluded that the difference was significant ($\chi^2 = 25.88$, $p = 0.0011$). When running the test for KA distributions, the test stated that the distribution differences are not significant ($\chi^2 = 0.26$, $p = 0.9999$). Based on Figure 9, this was expected since the differences appear small.

5.4. Limitations of This Study

The main limitation explained in Section 4.2 is that we downloaded only 15,963 out of 23,517 available writeups. Moreover, out of the 15,963 analyzed writeups, only 8,688 (54%) contained at least one keyword. This effectively reduced our dataset, although it remained substantially large. The median length of a writeup was 309 characters, and the average length was 3,979 characters.

Another limitation is the possibility of false positives in the keyword matches. However, we randomly selected 80 writeups (about 1% of the 8,688) and manually searched for the cybersecurity keywords in their content. Then, we compared these findings with the results of the automated keyword analysis. The results were the same: the desired keywords were matched, and no false positives occurred. If more writeups were validated like this, we would gain greater assurance that no false positives are in the sample, but the process would be time-consuming and error-prone.

Finally, some may find it problematic that the writeups rarely include the precise challenge assignments, which would allow us to double-check the relevance of the writeup to the challenge. However, the writeups were written by experts and enthusiasts who solved the CTF challenge, and so we consider them reliable. In the future, writeup databases could include a separate record of the assignment.

6. Lessons Learned and Future Work

We now share the educational implications of the results, their comparison with previous work, and practical insights stemming from this research. Finally, we propose ideas for future work.

6.1. Implications for Cybersecurity Professionals and Educators

Cybersecurity topics are represented unevenly in CTF challenges. This is understandable since some topics, such as attacks on encryption or reverse engineering of code, have been a staple in CTF challenges for years. These technical challenges are popular among CTF creators and participants, who are usually cybersecurity enthusiasts from a private or academic sector. The chosen topics may reflect their opinion on which topics are the most important, are feasible to implement, or are simply fun.

Nevertheless, this opens a new possibility for CTF creators to incorporate human aspects of cybersecurity into the CTF format. Although phishing attacks are a severe threat, as they are the biggest malware infection vector [47], our results indicate that the CTF format does not address this topic. Therefore, preparing challenges that teach these aspects can be a valuable and engaging experience, even for those without a deep technical background. This can open up the CTF format for beginners and other users who are not full-time computing experts.

Cybersecurity teachers can use CTF challenges as a suitable hands-on complement to their traditional classes, as was previously tried in [5]. Their regular classes may then focus on areas not covered by CTFs. Moreover, CTFs are excellent for all forms of online learning, including distance and blended learning, due to their remote accessibility. Finally, educators can help their students prepare for CTF challenges. Prominent universities such as Carnegie Mellon or the University of California, Santa Barbara, have their prestigious CTF teams [48, 49].

Starting to participate in CTF is often frustrating for newcomers [50, 51] since the challenges tend to be difficult. Knowing which topics are the most frequent may help beginners to prepare and avoid disappointment. It may even provide the basis for a CTF training program that would generate recommendations for personalized training paths.

6.2. Comparison with the Results of Related Publications

We discovered that *Cryptography* is the most prominent KU overall. Similarly, Burns et al. [17] also identified *crypto* as the top-ranking category of

CTF challenges, even though they categorized the writeups differently, not using the cybersecurity curricula.

Cabaj et al. [21] found that most cybersecurity master's degree programs include *Data security* (e.g., KU Cryptography), *Software security* (e.g., KU Programming robustly), *Connection security* (e.g., KU Network defense), and *System security* (e.g., KU Penetration testing). This corresponds to the most prominent topics we found in CTF, and also to those most researched in cybersecurity education papers [30]. Among the non-technical aspects, *Organizational* and *Societal security* prevail.

The survey of 104 European master's programs in cybersecurity [24] revealed that Data and Connection security dominate. KUs covered in mandatory courses by more than half of the surveyed institutions were *Cryptography*, *Secure communication protocols*, and *Network defense*. When including also optional courses or subtopics of other courses, these KUs were prominent as well: *Data integrity and authentication*, *Access control*, and *System access*. The least covered KA was *Component security*. Again, these results are similar to ours.

For cyber warfare operations, Carroll [25] prioritizes networking, fundamental security principles, telecommunications, network defense, and management of vulnerabilities and risks. Most of these topics belong to the *Connection security* KA, followed by *Data security* and *Organizational security*. According to Jones et al. [26], cybersecurity professionals also prioritize *Connection* and *System security*.

Overall, *Data* and *Connection security* dominated both in related work and CTFs. These areas include essential technical foundations of cybersecurity, which supports the fact that CTF aligns well with formal study programs. However, educators should not forget about the importance of non-technical aspects, such as human security, privacy, ethics, and law.

6.3. Legal Aspects of Research That Involves Third Party Data

CTFtime website states that all writeups are copyrighted by their authors. Still, it is allowed to analyze the data for research purposes and present aggregate results as in this paper. In the USA, this is granted by the Copyright Law [52]. In the European Union, the same exception for research holds [53].

The best practice in research is to publish the analyzed data and software as supplementary materials with the paper. However, this would require obtaining permission from the author of each writeup, which is practically impossible. Without this permission, we would be re-publishing the content,

which the copyright does not allow. Therefore, we publish the writeup folder structure, but the writeup files include only the link to the original writeup. Similarly, we cannot publish the Python script used to download the writeups because it would create a local copy of the data unauthorized by the authors. However, it is a simple web scraper that can be replicated based on Section 4. We publish only the analysis script, which is more specific for this work.

As a guideline for other researchers, we recommend carefully reviewing the conditions under which it is permissible to publish third party materials. In these cases, the support of replicable scientific research and the right to open access to information clashes with the protection of intellectual property. However, if the writeup authors and portals such as CTFtime used a Creative Commons license instead of the traditional copyright, this would simplify the future (re)use of their work.

6.4. Future Work

Future researchers can address the limitations of this study, such as downloading the remaining writeups, and thus improve the accuracy of our results. Probabilistic methods can be employed to match the keywords in writeups stochastically. Another possibility is to apply machine learning to the dataset. In [34], the authors used clustering to "identify groupings of similar documents according to their term frequency vector Euclidean distance". The same method can be applied to writeups. If it reveals clusters of writeups on a single topic, it can support the validity of our results. Finally, classification algorithms can categorize the writeups and compare them with the CSEC2017 Knowledge Areas. This fine-grained classification would allow mapping the CTF topics onto specific learning outcomes.

7. Conclusion

This work is a pioneering attempt to connect two different aspects of cybersecurity education: (i) popular hands-on challenges prepared by security experts and (ii) formal study programs facilitated by professional educators. If the goal is to exercise cybersecurity skills, CTF challenges suitably complement traditional formats of education delivered by schools and universities. They allow hundreds of students to practice a wide variety of cybersecurity skills online in a hands-on and engaging way.

We analyzed the cybersecurity topics in almost 16,000 written solutions (writeups) of CTF challenges held in the recent decade. The goal of the

analysis was to determine how the topics defined in the current Cybersecurity Curricular Guidelines (CSEC2017) are represented in CTF challenges. The analysis showed that topics such as cryptography and network security dominate. Interestingly, the same topics are prevalent in the current study programs and reflect the contemporary literature.

Although CTF challenges are excellent for practicing technical skills, they do not address topics such as phishing and general cybersecurity awareness. However, these topics are of utmost importance for mitigating the current advanced cyber threats. CSEC2017 defines cybersecurity as "a computing-based discipline involving technology, people, information, and processes to enable assured operations" [3], but the interdisciplinary "people" aspect is currently missing in CTF. This opens up a new opportunity to design CTFs that would reach a broader, non-technical audience and perhaps attract more people into cybersecurity.

Our paper provides numerous contributions for cybersecurity professionals, teachers, educational managers, CTF participants, and CTF designers. First, we gathered insights into cybersecurity topics practiced via CTF. Next, we discussed the implications of these results and connected them to the state-of-the-art curricular development in cybersecurity. Finally, we provided recommendations for other researchers, along with the directions for future work to motivate further research.

Supplementary materials for the paper are freely published on Zenodo [42]. The archive includes mainly the URLs of the analyzed writeups and the analysis script. This will support other researchers in replicating our work and building upon the analysis of CTF writeups.

Acknowledgements

This research was supported by the ERDF project CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence (No. CZ.02.1.01/0.0/0.0/16_019/0000822).

References

[1] (ISC)^2, Strategies for Building and Growing Strong Cybersecurity Teams, Technical Report, Cybersecurity Workforce Study, 2019.

[2] ACM/IEEE, Computing Curricula 2020, 2020. `https://cc2020.nsparc.msstate.edu/` [Last accessed on October 23, 2020].

[3] Joint Task Force on Cybersecurity Education, Cybersecurity Curricular Guideline, 2017. http://cybered.acm.org/ [Last accessed on October 23, 2020].

[4] DEF CON, CTF Archive, 2020. https://www.defcon.org/html/links/dc-ctf.html [Last accessed on October 23, 2020].

[5] J. Vykopal, V. Švábenský, E.-C. Chang, Benefits and Pitfalls of Using Capture the Flag Games in University Courses, in: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 752–758.

[6] J. Mirkovic, P. A. H. Peterson, Class Capture-the-Flag Exercises, in: 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14), USENIX Association, San Diego, CA, 2014, pp. 1–8.

[7] G. Vigna, K. Borgolte, J. Corbetta, A. Doupé, Y. Fratantonio, L. Invernizzi, D. Kirat, Y. Shoshitaishvili, Ten Years of iCTF: The Good, The Bad, and The Ugly, in: 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14), USENIX Association, San Diego, CA, 2014, pp. 1–7.

[8] N. Backman, Facilitating a Battle Between Hackers: Computer Security Outside of the Classroom, in: Proceedings of the 47th ACM Technical Symposium on Computing Science Education, SIGCSE '16, ACM, 2016, pp. 603–608.

[9] Google, Capture the Flag, 2020. https://capturetheflag.withgoogle.com/ [Last accessed on October 23, 2020].

[10] G. Singh, Announcing Facebook CTF 2019, 2019. https://www.facebook.com/notes/facebook-bug-bounty/announcing-facebook-ctf-2019/2629218463759030/ [Last accessed on October 23, 2020].

[11] S. Kucek, M. Leitner, An Empirical Survey of Functions and Configurations of Open-Source Capture the Flag (CTF) Environments, Journal of Network and Computer Applications 151 (2020).

[12] CTFtime, CTFtime.org / All about CTF (Capture The Flag), 2020. https://ctftime.org/ [Last accessed on October 23, 2020].

[13] M. Gondree, Z. N. Peterson, P. Pusey, Talking about talking about cybersecurity games, ;login: (2016).

[14] M. Sahami, A. Danyluk, S. Fincher, K. Fisher, D. Grossman, E. Hawthorne, R. Katz, R. LeBlanc, D. Reed, S. Roach, et al., Computer Science Curricula 2013: Curriculum guidelines for undergraduate degree programs in Computer Science, Association for Computing Machinery (ACM)-IEEE Computer Society (2013).

[15] A. Parrish, J. Impagliazzo, R. K. Raj, H. Santos, M. R. Asghar, A. Jøsang, T. Pereira, E. Stavrou, Global Perspectives on Cybersecurity Education for 2030: A Case for a Meta-discipline, in: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE 2018 Companion, ACM, New York, NY, USA, 2018, pp. 36–54.

[16] D. Mouheb, S. Abbas, M. Merabti, Cybersecurity curriculum design: A survey, in: Transactions on Edutainment XV, Springer, Berlin, Heidelberg, 2019, pp. 93–107.

[17] T. J. Burns, S. C. Rios, T. K. Jordan, Q. Gu, T. Underwood, Analysis and Exercises for Engaging Beginners in Online CTF Competitions for Security Education, in: 2017 USENIX Workshop on Advances in Security Education (ASE 17), USENIX Association, Vancouver, BC, 2017, pp. 1–9.

[18] T. Chothia, C. Novakovic, An Offline Capture The Flag-Style Virtual Machine and an Assessment of Its Value for Cybersecurity Education, in: 2015 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 15), USENIX Association, Washington, D.C., 2015, pp. 1–8.

[19] Z. C. Schreuders, T. Shaw, M. Shan-A-Khuda, G. Ravichandran, J. Keighley, M. Ordean, Security Scenario Generator (SecGen): A Framework for Generating Randomly Vulnerable Rich-scenario VMs for Learning Computer Security and Hosting CTF Events, in: 2017

USENIX Workshop on Advances in Security Education (ASE 17), USENIX Association, Vancouver, BC, 2017, pp. 1–10.

[20] K. Leune, S. J. Petrilli, Using Capture-the-Flag to Enhance the Effectiveness of Cybersecurity Education, in: Proceedings of the 18th Annual Conference on Information Technology Education, SIGITE '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 47–52.

[21] K. Cabaj, D. Domingos, Z. Kotulski, A. Respício, Cybersecurity education: Evolution of the discipline and analysis of master programs, Computers & Security 75 (2018) 24–35.

[22] J. Hallett, R. Larson, A. Rashid, Mirror, Mirror, On the Wall: What are we Teaching Them All? Characterising the Focus of Cybersecurity Curricular Frameworks, in: 2018 USENIX Workshop on Advances in Security Education (ASE 18), USENIX Association, Baltimore, MD, 2018, pp. 1–9.

[23] University of Bristol, The Cyber Security Body Of Knowledge, 2020. https://www.cybok.org/ [Last accessed on October 23, 2020].

[24] N. Dragoni and A. L. Lafuente and A. Schlichtkrull and L. Zhao, Addressing the Shortage of Cybersecurity Skills in Europe, 2020. Deliverable 6.2 of the Cyber Security for Europe project, https://cybersec4europe.eu/addressing-the-shortage-of-cybersecurity-skills-in-europe/ [Last accessed on October 23, 2020].

[25] J. Carroll, Offensive and Defensive Cyberspace Operations Training: Are we There yet?, in: European Conference on Cyber Warfare and Security, Academic Conferences International Limited, pp. 77–86.

[26] K. S. Jones, A. S. Namin, M. E. Armstrong, The Core Cyber-Defense Knowledge, Skills, and Abilities That Cybersecurity Students Should Learn in School: Results from Interviews with Cybersecurity Professionals, ACM Trans. Comput. Educ. 18 (2018) 11:1–11:12.

[27] H. Haqaf, M. Koyuncu, Understanding key skills for information security managers, International Journal of Information Management 43 (2018) 165 – 172.

[28] H. A. Linstone, M. Turoff, et al., The Delphi method, Addison-Wesley Reading, MA, 1975.

[29] N. G. Brooks, T. H. Greer, S. A. Morris, Information systems security job advertisement analysis: Skills review and implications for information systems curriculum, Journal of Education for Business 93 (2018) 213–221.

[30] V. Švábenský, J. Vykopal, P. Čeleda, What Are Cybersecurity Education Papers About? A Systematic Literature Review of SIGCSE and ITiCSE Conferences, in: Proceedings of The 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20, ACM, New York, NY, USA, 2020, pp. 2–8.

[31] D. M. Blei, Probabilistic topic models, Commun. ACM 55 (2012) 77–84.

[32] T. Hofmann, Probabilistic Latent Semantic Indexing, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99, Association for Computing Machinery, New York, NY, USA, 1999, p. 50–57.

[33] I. Marçal, R. E. Garcia, D. Eler, R. C. M. Correia, A Strategy to Enhance Computer Science Teaching Material Using Topic Modelling: Towards Overcoming The Gap Between College And Workplace Skills, in: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 366–371.

[34] J. M. Rouly, H. Rangwala, A. Johri, What Are We Teaching? Automated Evaluation of CS Curricula Content Using Topic Modeling, in: Proceedings of the Eleventh Annual International Conference on International Computing Education Research, ICER '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 189–197.

[35] M. Nadeem, E. B. Allen, B. J. Williams, A Method for Recommending Computer-Security Training for Software Developers: Leveraging the Power of Static Analysis Techniques and Vulnerability Repositories, in: 2015 12th International Conference on Information Technology – New Generations, pp. 534–539.

[36] C. Romero, S. Ventura, M. Pechenizkiy, R. S. Baker (Eds.), Handbook of educational data mining, CRC Press, Boca Raton, FL, USA, 2010.

[37] C. Severance, Discovering JavaScript Object Notation, Computer 45 (2012) 6–8.

[38] M. Droettboom, et al., Understanding JSON Schema, 2019.

[39] Kenneth Reitz, Requests: HTTP for Humans, 2020. https://requests.readthedocs.io/en/master/ [Last accessed on October 23, 2020].

[40] C. Lang, G. Siemens, A. Wise, D. Gašević (Eds.), Handbook of Learning Analytics, Society for Learning Analytics Research (SoLAR), 1st edition, 2017.

[41] J. B. Lovins, Development of a stemming algorithm, Mech. Translat. & Comp. Linguistics 11 (1968) 22–31.

[42] V. Švábenský, P. Čeleda, J. Vykopal, B. Silvia, Dataset: Cybersecurity Knowledge and Skills Taught in Capture the Flag Challenges, 2020. https://doi.org/10.5281/zenodo.4160585 [Last accessed on October 30, 2020].

[43] S. Brišáková, Analyzing Written Solutions of Tasks in Cybersecurity Capture the Flag Games, Bachelor's thesis, Masaryk University, Faculty of Informatics, 2020. Available from https://is.muni.cz/th/j9ru6/.

[44] Scipy.org, scipy.stats.anderson_ksamp, 2020. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.anderson_ksamp.html [Last accessed on October 23, 2020].

[45] Scipy.org, scipy.stats.mannwhitneyu, 2020. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html [Last accessed on October 23, 2020].

[46] Scipy.org, scipy.stats.kruskal, 2020. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html [Last accessed on October 23, 2020].

[47] ENISA, Threat Landscape Report, Technical Report, 15 Top Cyberthreats and Trends, 2018.

[48] Carnegie Mellon University, Plaid Parliament of Pwning, 2020. http://pwning.net/ [Last accessed on October 23, 2020].

[49] UC Santa Barbara, Shellphish, 2020. https://shellphish.net/ [Last accessed on October 23, 2020].

[50] K. Chung, J. Cohen, Learning Obstacles in the Capture The Flag Model, in: 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14), USENIX Association, San Diego, CA, 2014, pp. 1–7.

[51] D. H. Tobey, P. Pusey, D. L. Burley, Engaging Learners in Cybersecurity Careers: Lessons from the Launch of the National Cyber League, ACM Inroads 5 (2014) 53–56.

[52] U.S. Copyright Office, Copyright Law of the United States, 2016. https://www.copyright.gov/title17/92chap1.html#107 [Last accessed on October 23, 2020].

[53] European Union, Copyright in the EU, 2020. https://europa.eu/youreurope/business/running-business/intellectual-property/copyright/index_en.htm [Last accessed on October 23, 2020].