

Supporting Marginalized Communities in the Security and Privacy Workforce

A thesis submitted by

Samantha Katcher

In partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science

Tufts University

May 2025

Committee: Dr. Daniel Votipka, Dr. Megumi Ando, Dr. Josephine Wolff

Supporting Marginalized Communities in the Security and Privacy Workforce

Samantha Katcher

Committee: Dr. Daniel Votipka, Dr. Megumi Ando, Dr. Josephine Wolff

0.1 Abstract

Creating more inclusive professional communities and technological interfaces can ultimately foster greater success and diversity in the cybersecurity workforce. The cybersecurity workforce lacks ethnic diversity; the field is predominately men and White or Asian, with only 10% identifying as women, Latine, or Black. Previous studies identified access to supportive communities as a possible disparity between marginalized and non-marginalized cybersecurity professional populations and highlighted this support as a key to career success. In addition to ethnic diversity, the cybersecurity workforce contains many sub-specialized fields. One specifically under-researched sub-specialized sect of the cybersecurity community is network protocol reverse engineering. Protocol reverse engineering (ProtocolREing) consists of taking streams of network data and inferring the communication protocol. ProtocolREing is a critical task in malware and system security analysis.

Several ProtocolREing automation tools have been developed, however, in practice, they are not used because they offer limited interaction. Instead, reverse engineers (ProtocolREs) perform this task manually or use less complex visualization tools. To give ProtocolREs the power of more complex automation, we must first understand ProtocolRE's information and interaction needs to design better interfaces. We interviewed 16 ProtocolREs, presenting a paper prototype ProtocolREing automation interface, and asked them to discuss using the tool and suggest missing information and interactions. We designed our prototype based on existing ProtocolREing tool features and prior reverse engineering research's usability guidelines. We found ProtocolREs follow a flexible, hypothesis-driven process and identified multiple information and interaction needs when validating the automation's inferences.

We also focus on these community experiences by conducting a survey of 342 cybersecurity professionals to identify differences in perceptions and experiences of belonging across demographic groups. Our results show a discrepancy between experiences for different gender identities, with women being more likely than men to report instances of harassment and encountering unsupportive environments because of their gender. Psychological safety was low across all demographic groups, meaning participants did not feel comfortable engaging with or speaking up in the community. Based on these results we provide recommendations to community leaders and provide suggestions for future interaction design

for ProtocolREs.

Creating more inclusive professional communities and technological interfaces can ultimately foster greater success and diversity in the cybersecurity workforce.

Acknowledgments

I would like to thank my advisors Dr. Daniel Votipka and Dr. Diane Souvaine for the help and support they have provided throughout my time as a graduate student at Tufts. Additional thanks to the other members of my thesis committee, Dr. Megumi Ando and Dr. Josephine Wolff, and all the collaborators and co-authors in the Tufts Security and Privacy Lab.

Thanks to the MITRE Corporation for awarding me the Advanced Graduate Degree Program Fellowship which provided the funding to allow me to earn this degree.

Finally, I would like to extend my heartfelt gratitude to my family, friends, and Happu, all of whom have supported me throughout this journey. Please know that your encouragement, guidance, and kindness have made a significant impact.

Samantha Katcher

The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author. MITRE has approved this work for public release and unlimited distribution; public release case number 24-3822. ©2025 The MITRE Corporation. All Rights Reserved.

Contents

0.1	Abstract	i
1	Introduction	1
2	Related Work	2
2.0.1	Defining cybersecurity experiences	2
2.0.2	Defining ProtocolREing	4
2.0.3	Existing tools to support ProtocolREing	5
2.0.4	User Studies with SoftwareREs	6
3	A Survey of Cybersecurity Professionals’ Perceptions and Experiences of Safety and Belonging in the Community	8
3.1	Introduction	8
3.2	Methods	10
3.2.1	Survey Design	10
3.2.2	Recruitment	13
3.2.3	Data Analysis	13
3.2.4	Ethical Considerations	15
3.2.5	Limitations	15
3.3	Participants	16
3.4	Perceptions of Belonging and Social Experiences (RQ1)	17
3.5	Subcommunity Participation (RQ2)	21
3.6	Supportive and Unsupportive Experiences (RQ3)	23
4	An Investigation of Interaction and Information Needs for Protocol Reverse Engineering Automation	25
4.1	Introduction	25
4.2	Method	27
4.2.1	Prototype Interface Design	28

4.2.2	Interview Protocol	33
4.2.3	Recruitment	34
4.2.4	Analysis	35
4.2.5	Limitations	36
4.3	Ethical Considerations	37
4.4	Results	37
4.4.1	Participants	37
4.4.2	ProtocolREing Process (RQ1)	38
4.4.3	Information Needs From Automation (RQ2)	42
4.4.4	Interaction Needs With Automation (RQ3)	45
5	Discussion and Recommendations	47
5.0.1	Recommendations for Researchers	49
5.0.2	Recommendations for ProtocolREing Tool Developers	50
5.0.3	Community Leaders Recommendations	51
	Bibliography	53
A	Survey	73
A.1	General Belonging	73
A.2	Close Relations	75
A.3	Workplace	75
A.4	Organizations	76
A.5	School	76
A.6	Broader Security Community	77
A.7	General	77
A.8	Demographics	79
A.9	Final	80
B	Additional Figures and Tables	81
C	References of Tools Mentioned	91

Chapter 1

Introduction

In today's digital age, security and privacy are critical concerns for individuals, organizations, and governments alike. However, there is a growing recognition that traditional approaches to security and privacy have not adequately addressed issues of equity and inclusivity. In particular, the use of security tools and practices has been shown to disproportionately impact certain populations, such as marginalized communities and individuals from underrepresented backgrounds [167, 66, 65, 136, 71, 40]. Additionally, there is a need to ensure that the workforce responsible for developing and implementing security measures is diverse and representative of the communities they serve [177, 71, 91, 110]. There is very little diversity in the cybersecurity workforce [83, 26] and prior work has found that even individuals who are able to join this community feel marginalized [74]. Further, expected hiring surges across industry [54, 129] and government [144] risk further entrenching these problems if we do not identify the root cause. Therefore, a more equity and inclusivity-minded approach to security and privacy is necessary to ensure these critical concerns are addressed in a way that is fair and inclusive for all individuals and communities.

My work aims to increase considerations of equity and inclusivity in security and create better representation for people from marginalized populations in the security community and in security tool-building. Additionally, this work looks at creating equitable systems to build security and privacy tools and in places that do not yet consider the impact of equity. My research takes a two-pronged approach solution: 1) surveying the computer security community to understand educational and support experiences and 2) interviewing experts to understand their RE workflows to build more usable tools that lower the barrier to use in the future.

Chapter 2

Related Work

Chapter 3’s contribution lies in a focused exploration of belonging in the cybersecurity community, differentiating it from other studies [75, 160, 187, 206]. While little work has focused on understanding the process and needs related to ProtocolREing, there is a dedicated body of research investigating SoftwareRE, various tools developed for ProtocolREing, and Chapter 4 contributes to the growing body of research on automated techniques for ProtocolREing. Here we describe prior work and how the studies fits into the broader research context.

2.0.1 Defining cybersecurity experiences

Marginalized populations’ experiences in computer science and technology. There is a growing body of research considering issues facing marginalized populations in CS and STEM domains. For example, work studying developers has found marginalized populations are paid less [80, 132] and are less likely to have work accepted by colleagues [134, 196, 186, 24]. Similarly, significant research has investigated issues in CS [161, 117, 38, 116, 39, 168, 37, 206, 67] and technology careers more broadly [207, 34, 154]. Margolis and Allen performed an ethnographic study of the gender gap in CS education [117]. They found women had less coding experience than men in undergraduate programs and perceived CS’s “geek” culture negatively. Subsequent work has documented issues of gendered perceptions of CS [139, 36, 48, 131, 162], which are further entrenched by unapproachable early educational activities [107, 41, 15], lack of representation [190, 15], mentorship support [9, 190, 41, 15, 212], and a non-inclusive culture to diverse backgrounds and experiences [9]. Because we expect many of these trends to be mirrored in cybersecurity, we use this prior work as a lens, guiding our survey questions and analysis. However, we expect cybersecurity may present differences as it is more specialized and the inherent focus on privacy and security scrutiny may make cybersecurity communities less welcoming. This has been found, to some extent, demonstrating several differences in interviewee experiences when studying members of the

vulnerability discovery community—a subset of the cybersecurity community [75].

Marginalized populations’ experiences in cybersecurity. Several prior industry surveys have demonstrated the lack of diversity in the cybersecurity community [27, 84, 182, 95]. This includes ICS2’s annual survey of the cybersecurity workforce, which showed that the younger generation (under 30 years old) of cybersecurity professionals are more diverse. However, this diversity remains limited as only 26% of this generation are women [95]. This survey also found the pathways into cybersecurity differ by gender and race/ethnicity. Women and non-white cybersecurity professionals are more likely to come from a traditional education-based pathway (e.g. college) and less likely to come from an IT background.

In addition to these industry surveys, some academic interview-based studies have examined the challenges marginalized cybersecurity professionals face [75, 150, 187, 193, 160]. Fulton et al. conducted semi-structured interviews with members of the vulnerability discovery community from marginalized populations, uncovering challenges specific to members of marginalized populations, such as a difficulty being taken seriously by others, a reluctance from other community members to share information, and explicit discrimination within the community. Additionally, Fulton et al.’s works discussed the important role mentors played in participants’ experiences [75]. In interviews with 21 cybersecurity professionals, Schoenmaker et al. found some participants believed holding a minority status might cause an increase in an individual’s ability to monitor for security anomalies, as these individuals already have significant experience monitoring for threats regarding personal safety. However, they also observed social conventions and lack of access to resources might make it more difficult for these groups to practice vulnerability discovery [160]. Plato et al. interviewed sixteen women C-Suite executives in cybersecurity to learn about their journeys into leadership and experiences with mentorship, sponsorship, and trusted advisors, as well as experiences of biases and discrimination highlighting how networking, mentorship, and observing leadership styles play pivotal roles in shaping individuals’ trajectories, even with a shortage of female mentors and racial bias making this difficult to accomplish in practice [150]. Each of these studies highlights important challenges marginalized cybersecurity professionals face, but have limited generalizability due to their small samples. Our work expands on these findings with a large-scale survey focusing on community belonging, a central challenge observed in prior work.

Students’ cybersecurity experiences. Some work has investigated existing workforce disparities. This work has primarily focused on student experiences in security exercises [152, 59] and college courses [33] as students take the first steps toward cybersecurity careers. It provides some indications of students’ reasons for abandoning the field (e.g., lack of role models and community, gendered stereotypes) and suggests entry-level hands-on exercises can increase interest. While these education-focused questions are important, cybersecurity professionals face challenges throughout their careers [75] and

prior work has found many are not trained through these traditional educational settings [204, 85]. To address this gap, our work takes a holistic view of cybersecurity professionals’ community environments.

2.0.2 Defining ProtocolREing

ProtocolREing is recovering task of recovering an abstract description (i.e., a protocol) from observed messages, software implementations, physical hardware, or source code, depending on the scenario. Here we focus on reverse engineering a format using only the observed messages. Imagine an analyst given the following three hex messages intercepted from air traffic control broadcasts relaying information about in-progress flights as shown in 2.1.

```
4708b89a0000012c41502052
4708b99a0000012a4f47414e
4708b1330000012a52444f20
```

Figure 2.1: Example hex messages.

The analyst may not necessarily know where these messages originated from and must reconstruct the message format and uncover the semantic meaning of the values presented in each message. The analyst must determine the meaning of these messages by identifying which portions of each message belong to discrete fields (e.g., byte boundaries), the data types of the fields (e.g., float, integer, text), and ultimately, the semantic meaning of the field in the context of the application (e.g., speed, altitude, and destination). For example, the analyst might first split the messages into three 32-bit fields based on similarities and differences across the messages as shown in 2.2.

```
4708b89a 0000012c 41502052
4708b99a 0000012a 4f47414e
4708b133 0000012a 52444f20
```

Figure 2.2: Example hex messages split into three fields.

Considering these three fields, the analyst might conclude the fields are (1) a 32-bit IEEE Float, (2) a 32-bit integer, and (3) a four-character ASCII field. They arrive at these types by observing the range of values for each field and identifying patterns associated with different data types. For example, the ASCII field (3) contains only values within the standard human-readable ASCII range, while the Float field (1) contains values mostly outside that range. Similarly, the leading null bytes in the second field (2) indicate an integer field where the most significant bits seldom change.

Finally, the analyst examines the floating point, integer, and ASCII values and infers these values are semantically related to aircraft. The analyst notes the ASCII values are consistent with airport identification codes, the floating point values are consistent with aircraft cruising altitudes, and the integer values are consistent with aircraft speeds.

2.0.3 Existing tools to support ProtocolREing

Such manual reverse engineering is challenging and time-consuming, especially as ProtocolREs have to review larger volumes of data. To improve their efficiency, there have been several tools developed to help ProtocolREs visualize message data and perform common ProtocolREing tasks.

One of the most well-known tools for examining network data is Wireshark [72]. Wireshark is used to analyze packets' sequences and examine their values. Wireshark allows a ProtocolRE to filter packets by format and packet field values for known protocols (e.g., TCP, HTTP). A ProtocolRE can then view both the raw binary packet representation as hexadecimal data, the packet decoded as text, and the individual packet fields interpreted when the packet is known to belong to a specific protocol.

However, Wireshark does not provide interactive support for users attempting to determine the specification of an unknown protocol. Instead, users must determine the protocol and write a parser to be applied by Wireshark.

Additionally, Wireshark is generally limited to viewing only one packet at a time, making between-packet comparisons difficult. Wireshark does allow TCP streams to be extracted and packets viewed in sequence through the *follow conversation* feature; however, packets from different streams or those not encapsulated in TCP cannot be compared side by side.

Another popular online tool that supports features for ProtocolREing is CyberChef [78]. Commonly called the “cyber swiss army knife.” CyberChef is a popular web application supporting several common ProtocolREing features, such as encoding or decoding Base64, Hex conversions, hash and checksum calculations, IPv6 header parsing, and more. While the ease of use for the web application allows technical and non-technical analysts to manipulate data without managing any custom algorithms, CyberChef offers limited automation or guidance as to which of the available tools would be useful given the raw data.

There has also been significant academic work to develop novel ProtocolREing interfaces. Conti et al. presented a broad survey using visualizations to support security goals [47, 45], including ProtocolREing. However, these visualizations data distributions instead of directly presenting messages. Other examples of these visualizations incorporate temporal elements and multiple file streams [8, 46].

An alternative approach comes from methods supporting investigation of binary files. These methods focus only on visualizing a single file of unknown format. Conti's et al.'s BytePlot interface is a robust example of this approach [47]. BytePlot represents the same data with different visualizations, each tuned to different aspects and reverse engineering process functions. For example, one visualization provides an overview of all binary data in the file, while another visually summarizes the byte-value distribution of individual regions.

We leverage the features and design of these tools as inspiration for the prototype we used in our interviews 4.2.1.

Automation for ProtocolREing. There is a growing body of literature and development effort directed at automation support for ProtocolRE. There are various available tools with autonomous features that prior literature has identified [58, 104, 135]. Broadly, ProtocolREing tooling is classified into two categories: tools that analyze and depend on captured network messages exchanged between two components or tools that leverage source code or binary files from the application itself to uncover the protocol structure. For simplicity, we focused on the former for our paper prototype.

Automation in this context attempts to identify three things: message formats, field semantics, and rules governing the sequence of exchanged messages. Several recent advancements in automated field boundary detection leverage intrinsic repeated structures or distributions of values throughout messages. Kleber et al. utilize Sokal and Michener’s bitwise similarity metric to track the change in congruence between bit values [103, 173]. This metric allows their tool NEMESYS to select candidate field boundaries by identifying inflection points at the rising edges of this congruence metric. Luo et al. proposed using Latent Dirichlet Allocation models to characterize message types and n-grams [113, 22]. This information would be used for clustering algorithms (e.g., UPGMA, K-Means, etc.) to group similar fields from messages that likely shared a type. Similarly, Ye et al. utilized multiple-sequence alignment to justify all messages before using probabilistic inference to cluster messages by their most likely candidate type [210]. Chandler et al.’s BinaryInferno, leverages an ensemble of specialized detectors to identify common data types, and serialization patterns [35]. BinaryInferno identifies field boundaries by first looking for all semantic types consistent with the message bytes, and then finding the over-all description with the highest explanatory power.

Although much of the prior work has developed novel automation techniques for ProtocolREing, it is unclear how these tools can best provide inferences to ProtocolREs to best inform them and support task efficiency. This paper investigates how to close this gap between useful interfaces and automation, empowering the next generation of autonomous tooling for ProtocolREing.

2.0.4 User Studies with SoftwareREs

Investigating SoftwareREs’ processes. While, to our knowledge, there are no prior studies investigating ProtocolREs processes, there is a growing body of research investigating the related task of SoftwareRE. Perhaps the first was by Bryant, who conducted interviews with four professional reverse engineers and developed a sense-making model of the reverse engineering process [25]. Votipka et al. built on this work, conducting 16 retrospective observational interviews with expert reverse engineers investigating an unfamiliar program [202]. Votipka et al. refined the reverse engineering process model into a three-phase iterative model beginning with an overview, then scanning through functions or blocks

of interest, and finally performing focused experimentation to test answer specific questions. Reverse engineers iterate between these phases, continually learning more about the program until they can answer their target questions, such as identifying a vulnerability or understanding a malware infection. This work also suggested usability guidelines for reverse engineering tool development. This model has subsequently been supported in large-scale quantitative measures showing how reverse engineering experience influences analyst actions and strategies [115, 29, 70, 19]. Although the specific context differed from ours, we expect ProtocolREs to have similar processes. Therefore, we utilized the process model and design guidelines proposed from this prior body of research for our paper prototype design (see Section 4.2.1) to ensure the presented interfaces fit the users’ expected workflow.

Studies of SoftwareRE tool usability. Most related to our work, recent studies have assessed the usability of specific tools and techniques for reverse engineering. Using the guidelines proposed by Votipka et al. [203], Mattei et al. assessed the expected usability of 288 reverse engineering tools [118]. Ploger et al. studied the usability challenges for beginners when using a fuzzer and static analyzer, showing that setting up these tools was a significant barrier to their use [151]. Shoshitaishvili et al. developed the HaCRS system, which created an interface allowing users to provide input seeds to mutational fuzzers without needing expertise with fuzzers or even programming knowledge [165]. This demonstrated non-experts’ input improved tool outcomes, but did not assess the usability of this interaction. Finally, Yakdan et al. developed DREAM++, a usability-optimized decompiler leveraging heuristic-based transformations to improve decompiled code readability [208]. This resulted in concrete performance improvements by improving decompiled code readability in a user-informed manner. These studies focus solely on designing and implementing visualizations of single binary files. In contrast, we focus on understanding ProtocolREs’ processes when investigating a collection of multiple network packets. We draw inspiration from these studies to design our prototype and assess what information and interactions ProtocolREs need.

Chapter 3

A Survey of Cybersecurity Professionals' Perceptions and Experiences of Safety and Belonging in the Community

3.1 Introduction

With technology's growing ubiquity, and parallel increases in cyberattacks, skilled cybersecurity professionals are in demand. This demand has outpaced the supply of qualified workers, with some estimates suggesting a four million job shortfall in 2023 [95]. Governments and private institutions are campaigning to increase the number of cybersecurity professionals [68, 69, 23, 153, 93, 191, 194] and the US government has prioritized growing the cybersecurity workforce [21].

While there have been many efforts to grow the cybersecurity workforce, this growth has not increased diversity. Prior workforce surveys show the field as predominantly male, white or Asian¹, with women, Latine, and Black participants constituting fewer than 10% in each survey [27, 85]. In a 2020 cybersecurity professionals survey, SynAck, a platform connecting organizations with cybersecurity professionals who provide security reviews, found women (66%) and members of marginalized ethnicities (47%) were less likely, when compared to men (88%), to believe people of the same gender or ethnicity were given the same opportunities [182]. Furthermore, interviews with cybersecurity professionals from marginalized populations revealed regular instances of othering, hate, and harassment in the workforce [75].

¹We use the term Asian broadly here, as this is how it is used in the cited prior surveys, but we recognize Asian Americans and those from other regions may still be marginalized in the community.

Cybersecurity’s deficiency in diversity creates two problems. First, and foremost, is an equity problem. Members of marginalized populations are driven away from well paying, in-demand careers in cybersecurity. Second, cognitive diversity is essential to secure system design. The more eyes reviewing potentially insecure code, the more thorough a review will be completed and attacks thwarted [114]. People from different genders, ethnicities, and backgrounds provide a fresh perspective to solving complex security problems [57, 125]. As cybersecurity hiring increases, we must prevent furthering existing ethnic and gender disparities by identifying and understanding factors underlying lacking diversity to improve recruitment and retention among marginalized populations.

Recent work investigates the career challenges cybersecurity professionals face through a survey broadly with cybersecurity professionals [10] and interviews with marginalized cybersecurity professionals [75]. Both populations indicated that the most significant challenges were the result of the difficulty of getting started, e.g., navigating unstructured resources to develop necessary skills, and the stress and uncertainty of the market, e.g., trying to find work for which they qualified. Non-marginalized cybersecurity professionals found support from their peers crucial, viewing the community as inclusive, while marginalized cybersecurity professionals found it challenging or impossible to join the community, hindering their access to necessary support for success.

In this paper, we take further steps to understand the barriers faced by marginalized cybersecurity professionals by focusing on this point of divergence: their community experiences. We surveyed 342 cybersecurity professionals from varying backgrounds (196 men, 128 women, 10 genderqueer; 215 White, 46 Latine, 38 Black, and 31 other ethnicities). We used multiple validated psychometric scales to measure perceptions of belonging [60, 205] and experiences of supportive and unsupportive social environments [81, 188]. We also asked about participation and experiences in specific subcommunities. We address the following research questions:

- **RQ1:** What differences exist in perception (e.g., belonging, psychological safety) and incidents of unsupportive experiences (e.g., othering, hate, harassment) within the cybersecurity community between marginalized and non-marginalized cybersecurity professionals?
- **RQ2:** Do marginalized and non-marginalized cybersecurity professionals differ in their participation and experiences in specific subcommunities (e.g., work, social organizations, online)?
- **RQ3:** What community interactions are perceived as particularly supportive or unsupportive and how do these differ between marginalized and non-marginalized cybersecurity professionals?

The biggest divide among cybersecurity professionals was across gender identities, with women being more likely to report experiencing harassment and unsupportive environments due to their identity. However, across all demographic groups, cybersecurity professionals reported low psychological safety relative to other professions, indicating the difficulty to engage in the community. Conversely, we did

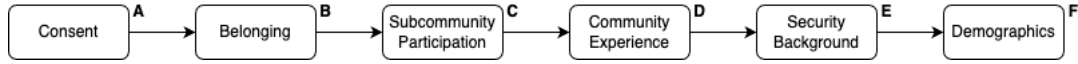


Figure 3.1: Structure of the survey.

not observe low scores on measures of internal belonging (i.e., whether a participant felt qualified and knowledgeable enough to belong in the community). Together these suggest unsupportive forces on cybersecurity professionals are generally external to the individual. Finally, our results suggest early development environments for cybersecurity professionals might be particularly problematic since participants with high school programming experience were less likely to feel psychologically safe. Based on our results, we provide recommendations for cybersecurity community leaders.

3.2 Methods

We seek to understand how practitioners in cybersecurity participate in and perceive belonging within their professional community, and specifically to consider differences and similarities between practitioners from different demographics. We do not place limits on participation (e.g., industry, academia, government), but consider the field of cybersecurity broadly. In this section, we describe the survey design, recruitment methods, data analysis procedures, ethical considerations, and the work’s limitations.

3.2.1 Survey Design

The survey began by requesting participant consent; included three main components aligned with the research questions; and concluded with demographics questions. Figure 3.1 summarizes the survey’s flow. The full survey can be found in Appendix A. Where applicable, we altered validated scales to focus on cybersecurity and the cybersecurity community, and we included attention checks to catch inattentive respondents [124]. The survey was divided into three parts to match our research questions: questions about participants’ belonging within the cybersecurity community generally (RQ1), participation in various subcommunities—listed in Table 3.2—(RQ2), and prototypical community experiences (RQ3). We detail how we asked about each of these topics in turn, then concluded with questions about their security experience and demographics. Figure 3.1 summarizes the survey’s flow. Participants completed the survey in 15 minutes on average.

Perceptions of belonging (RQ1, Figure 3.1.B). We first sought to understand whether participants feel they belong in the cybersecurity community, as prior work found cybersecurity professionals were more successful after finding a community where they could get support and ask questions [75].

Name	Description	Ex. Item	Response Opts.	Items	Agg ¹
Psychological Safety [60]	Whether participants feel safe to express themselves, take risks, and ask questions	"If you make a mistake in the cybersecurity community, it is often held against you"	7-pt; "Very inaccurate" to "Very accurate"	5	Avg
Belonging Uncertainty [205]	Whether participants feel people like them belong in the community	"When something bad happens, I feel that maybe I don't belong in cybersecurity"	7-pt; "Strongly disagree" to "Strongly agree"	3	Avg
Vulnerability Discovery Self-Efficacy [200]	Whether participants believe they have appropriate cybersecurity skills	"I can identify potential security threats to the system"	5-pt; "Not confident at all" to "Absolutely confident"	9	Sum
Online Social Experiences [81]	How often participants experience positive and negative interactions	"Someone in the cybersecurity community has made me feel embarrassed or foolish"	5-pt; "Very Slightly" or "Not at All" to "Extremely"	8	Sum
Hate and Harassment [188, 198]	How often participants experience hate and harassment	"Stereotyping based on perceived demographic characteristics"	4-pt; "Never" to "Frequently"	7	Sum

¹Aggregation function used to combine responses from multiple items to a single score.

Table 3.1: Summary of psychometric measures used in the survey to understand participants' sense of belonging and experiences in the cybersecurity community. The different scales were presented in a randomized order to avoid ordering effects.

We utilized three validated psychometric measures of belonging: *psychological safety* [60], *belonging uncertainty* [205] and *vulnerability discovery self-efficacy* [200]. Table 3.1 provides additional details about each scale.

The psychological safety scale has previously been used to investigate why employees feel comfortable sharing information [43, 166], suggesting organizational improvements [53, 111], and taking initiative [18]. The belonging uncertainty scale has predominately been used investigate feelings of otherness among historically underrepresented groups, for example, among professionals [205] and students [52]. We also ask participants explicitly whether people with similar backgrounds have opportunities to participate in cybersecurity work to assess the question of representation more directly.

Finally, to assess whether participants believed they had the skill to be in the community (i.e., separate from whether they believed others would accept them into the community) we used the vulnerability discovery sub-scale of Votipka et al.'s secure software development self-efficacy scale (SSD-SES), which asks participants to assess their proficiency to identify vulnerabilities [200]. SSD-SES has been used to assess differences in perceived ability between study subgroups [185, 99, 184] and as a measure of learning improvement with educational interventions [201, 76].

(Un)welcoming Community Experiences (RQ1, Figure 3.1.B). To understand concrete experiences that might impact cybersecurity professionals' community participation, we asked a modified version of de Grey et al.'s Online Social Experiences Measure (OSEM), which assesses social support and negativity arising from online social network interactions [81]. OSEM evaluates aspects such as

Name	Description
Close Friends / Mentors	Family, friends and other close mentors who provided either career or other support (e.g., emotional, economical, etc.)
School	A learning community focused on cybersecurity in an academic setting (e.g., class, student-run organization)
Work	Community in participants' workplaces where they are able to discuss cybersecurity topics and receive support
Organizations	Groups outside work and school (e.g., ACM chapters, Women in Security and Privacy)
Online and Conferences	The broader cybersecurity community where participants might meet and talk with other, but not have close or lasting relationships. This includes interactions at cybersecurity conferences or workshops, as well as through online forums (e.g., StackOverflow, Reddit, X (formerly Twitter), public Slack or Discord).

Table 3.2: Types of subcommunities participants were asked about.

emotional, informational, and instrumental support. This measure has been employed in research to understand how online interactions influence mental health and social well-being, particularly in digital communities [138, 42, 192].

To capture hate and harassment, we borrowed from Thomas et al.'s [188] and the Pew Research Center's [198] existing survey questions investigating online hate and harassment. We included four questions about severe negative experiences which OSEM did not include, namely stereotype bias, violence, sexual advances, and doxxing. These questions have been employed in various research contexts to measure experiences of sexual harassment, particularly in professional and educational settings [100, 51, 147].

Subcommunity participation (RQ2, Figure 3.1.C). Community is not a global construct, but instead is specific to the individual [159]. Someone might not feel comfortable communicating with others at a large security conference, but may establish a smaller local community where they feel strong connections and receive support. Therefore, we investigated how participants' experiences varied across subcommunities—specifically, those described in Table 3.2, drawn from prior work [75]. We asked how many of each type of subcommunity participants were members of, how frequently they discussed cybersecurity concepts and how helpful they found each subcommunity, and, for each subcommunity, we asked at least one subcommunity-specific question to allow a better understanding of participant's relationship to the subcommunity. We randomized the order participants were asked about each subcommunity to avoid ordering effects.

Examples of supportive and unsupportive community experiences (RQ3, Figure 3.1.D). Next, we sought to understand what makes participants feel particularly welcome or unwelcome. We asked participants to describe an experience where they felt particularly well supported which could involve explicit assistance, encouragement, or any positive influence that aided the participant's professional growth. We also asked participants to describe a particularly unsupportive experience, which could include instances where the interaction was harmful or hindered their professional progress.

Cybersecurity background and demographics (Figure 3.1.D/E). We finished by asking about participants’ cybersecurity background, i.e., the extent their work focuses on cybersecurity, whether and what kind of cybersecurity training they have received, when they began programming, the age they became interested in cybersecurity, and the age they first received cybersecurity career support. We ended with demographics questions like gender, ethnicity, and education.

3.2.2 Recruitment

Recruiting cybersecurity professionals is a difficult task because they are a small, well paid population with significant demands on their time [102, 90]. Our challenge was compounded by the fact that we weighted our sample toward marginalized cybersecurity professionals, who make up a small fraction of this small workforce [27, 85].

We used several recruit methods, including contacting cybersecurity professional organizations’ leaders; advertising in public (i.e., X (formerly Twitter), LinkedIn, and Reddit) and private (i.e., Slack and Whatsapp) online spaces; recruiting at cybersecurity conferences; and contracting Qualtrics for a curated panel. Participants recruited through organizations, online, and conferences were given a study description and entered into a raffle for one of 25 \$50 Amazon gift cards. For the Qualtrics panel, we instructed Qualtrics to identify panelists working in cybersecurity, with a majority being women and at least 30% non-white. Panelists were paid \$25.

Our recruitment messages indicated that anyone currently working (or having worked in the last two years) in cybersecurity could participate. We did not mention the study’s intent to compare responses between marginalized and non-marginalized cybersecurity professionals to avoid a potential backlash [157, 96] due to increasing antagonism and polarization around diversity, equity, and inclusion efforts from a segment of the population [128, 62].

3.2.3 Data Analysis

Next, we outline our quantitative and qualitative analyses.

Qualitative analysis. To analyze the two free response questions in Part C, we used iterative open coding [178]. Two researchers collaboratively reviewed the first 35 responses to generate the codebook. Then, the same two researchers iteratively coded responses in rounds of twenty. After each round, the coders compared responses, resolved disagreements, and updated the codebook as necessary. After six rounds of independent coding (i.e., 120 responses), the coders achieved a Krippendorff’s α of 0.858 for *what* participants experienced and 0.835 for *who* the experience was with. Both are above the recommended level of agreement [86]. The remaining responses were divided evenly among the coders

Factor	Description	Baseline
Gender	Gender participants identify as	Man
Ethnicity	Ethnicity participants identify as	White
<i>Optional</i>		
Yrs. Exp.	Number of years participants have worked in cybersecurity	–
Yrs. Until Mentor	Age when participant first had a mentor who helped them learn about cybersecurity	–
Helpful Mentor	Whether participant reported having someone close (mentor/family member/friend) who helped them learn about cybersecurity	False
HS Prog.	Whether participant had high school programming experience	False
Job/ Seniority	Current job role (junior, senior non-leadership, senior leadership, or not currently working in cybersecurity)	Junior

Table 3.3: Factors used in regression models. Categorical variables are compared individually to the baseline.

and coded independently. The final codebook is included in the supplemental materials [180].

Quantitative analysis. In our statistical tests, we limited our dataset to participants who identified as men or women and were White, Black, and/or Latine. We did not include other demographics for statistical tests because there was an insufficient number of participants to produce generalizable results. We include 289 participants in the reported statistical analysis.

For the vulnerability discovery self-efficacy, online social experiences, and harassment questions, we used a poisson regression as the scales were scored using a sum of the Likert responses. As the harassment questions from Thomas et al. are not part of a validated scale, we first computed Cronbach’s α over participants’ responses to the four harassment questions to test their internal consistency [123]. These questions had “good” internal consistency ($\alpha = 0.806$), so we chose to treat them as a single measure like the other scales. For the psychological safety and belonging uncertainty scales, we used linear regressions as the outcome variables were a percentage and an average, respectively. To generate our initial models, we included all the factors listed in Table 3.3. Because it is possible some explanatory variables are not independent, which would violate the regressions’ assumptions [31, pg. 67-106], we tested for multicollinearity and found there was no significant correlation between factors. We then conducted model selection on all possible combinations of these factors, only considering models that included gender and ethnicity as they were our key variables of interest and selected the model with minimum Bayesian Information Criteria (BIC) [170, 155].

For each subcommunity, we used Kruskal-Wallis H tests, to compare subcommunity membership, discussion participation rates, and helpfulness Likert responses. We began each comparison with an omnibus test over all demographic groups. If the result was statistically significant, we applied planned

pairwise comparisons between non-marginalized and marginalized groups for gender and ethnicity, i.e., men to women, White to Black, White to Latine.

Finally, we applied Pearson’s χ^2 tests to compare responses between top-level code categories between men and women for themes that were mentioned by at least five participants [73] for our free-response questions. We focused on gender differences as we do not have sufficient data points across races/ethnicities to produce generalizable results. For categories mentioned by five or fewer of a single gender group, we perform a Fisher’s Exact Test instead of a χ^2 test [64].

3.2.4 Ethical Considerations

This study was approved by our institution’s ethical review board. Since this survey asks about multiple sensitive topics including experiences of harassment, psychological safety, and social experiences, participants were informed about our data collection and secure storage practices and that they could stop participating or skip a question at any time.

3.2.5 Limitations

Our study has several limitations that should be considered when interpreting our results.

Self-report responses. As is common in online studies with self-reported data, some participants may not approach the survey seriously, may try to take the survey multiple times, or may fabricate responses to qualify for compensation. To account for these behaviors, we deterred multiple responses by using a browser cookie and followed best practices for removing inattentive responses, e.g., we removed those that failed attention checks, were significantly faster than average, or provided nonsensical responses to open-ended questions (N=385). Also, we received over 500 automated responses where more than 50 identical or nearly identical responses were submitted within a very short period—often within a minute. We removed these responses as they were received as they clearly did not represent a legitimate response.

Inauthentic responses were a challenge in this study. To mitigate this, we primarily recruited from venues with a high likelihood of cybersecurity professionals, leveraging community relationships and in-person recruiting. Qualtrics panel participants were recruited independently of our study, reducing their motivation to lie, and their open-ended responses were consistent with those from professional venues.

Demographic distribution and US-centric population. While we made significant efforts to recruit participants from marginalized populations, many identities have limited representation in our sample (e.g., genderqueer, Middle Eastern, indigenous peoples). Therefore, any results from their responses may not generalize beyond our sample. We give descriptive statistics regarding their responses to provide indications of potential trends for future work to investigate, but avoid conducting statistical

tests relating to these identities or making broader statements about their responses. Similarly, our small sample sizes preclude investigation into the effects of intersectional identities. We expect there are important differences introduced by intersectionality, as prior work has shown in other domains [141, 174], but we refrain from investigating them to avoid overgeneralizing their personal experiences.

Despite attempting to recruit broadly, our participant pool was predominantly US-centric, with 279 out of 342 respondents from the US. We expect experiences of cybersecurity professionals in other regions will be similar, but there are likely critically important differences; we encourage further work focus on other geographic areas.

Survivor and recall biases. Our recruitment was limited to currently or recently employed cybersecurity professionals. It is likely many members of marginalized populations considered becoming cybersecurity professionals or worked in the field, but faced substantial challenges and chose to switch professions. Our results inherently do not account for these individuals, so our findings may skew toward a more positive portrayal of cybersecurity. We attempt to capture some of this adversity by asking participants to consider their experiences throughout their career when answering all questions, but they may not clearly remember events from years ago [158].

Demand effects. Participants might be motivated to report more or less unsupportive experiences based on political or cultural views or other social factors. Some participants from marginalized populations might under-report unsupportive experiences to avoid being seen as “whining” or as not earning their success [79, 175]. Alternatively, non-marginalized participants might over-report unsupportive experiences to counter what they see as “woke” popular perceptions [96, 128, 62]. To identify these biases, we include multiple community inclusivity measures and open-ended questions to capture participants’ experiences from multiple vantage points. However, these effects likely narrow any differences we might identify between non-marginalized and marginalized populations.

3.3 Participants

Table 3.4 provides a summary of our 342 participants’ demographics, divided by gender and race/ethnicity. Most of our participants identified as men (57%) or women (37%). The vast majority of our participants identified as either White or of European descent (63%), Black or of African descent (11%), and/or Hispanic or Latine (13%). Our participants were mostly located in the US (N=279).

The majority of our participants reported having taken at least one programming course in high school (N=234). On average, participants reported 7.5 years of security experience and had job titles including leadership positions, managerial roles, technical positions, and specialized roles related to secu-

	Men	Women	Genderqueer	White	Black	Latine	Total
Men	196 (57%)	-	-	125 (37%)	22 (6%)	28 (8%)	196 (57%)
Women	-	128 (37%)	-	83 (24%)	14 (4%)	17 (5%)	128 (37%)
Genderqueer	-	-	10 (3%)	5 (2%)	2 (1%)	1 (0.3%)	10 (3%)
No Answer	-	-	8 (2%)	-	-	-	8 (2%)
White	125 (37%)	83 (24%)	5 (2%)	215 (63%)	-	-	215 (63%)
Black	22 (64%)	14 (4%)	2 (1%)	-	38 (11%)	-	38 (11%)
Latine	28 (8%)	17 (5%)	1 (0.3%)	-	-	46 (13%)	46 (13%)
Avg. Yrs in Sec.	8.3	6.6	5.1	8.1	6.8	8.1	7.5
Heterosexual	169 (49%)	110 (32%)	2 (1%)	186 (54%)	32 (9%)	40 (12%)	291 (85%)
Gay/Lesbian	4 (1%)	3 (1%)	1 (0.3%)	6 (2%)	1 (0.3%)	1 (0.3%)	8 (2%)
Bisexual	8 (2%)	11 (3%)	3 (1%)	12 (4%)	5 (2%)	4 (1%)	24 (7%)
High school	12 (4%)	7 (2%)	0 (0%)	11 (3%)	3 (1%)	2 (1%)	19 (6%)
Some college	15 (4%)	14 (4%)	1 (0.3%)	17 (5%)	5 (2%)	8 (2%)	30 (9%)
Associate degree	12 (4%)	15 (4%)	0 (0%)	21 (6%)	3 (1%)	2 (1%)	27 (8%)
Bachelor's degree	72 (21%)	45 (13%)	5 (2%)	77 (23%)	12 (4%)	22 (6%)	122 (36%)
Master's degree	63 (18%)	39 (11%)	3 (1%)	71 (20%)	14 (4%)	12 (4%)	105 (31%)
Doctorate	14 (4%)	5 (2%)	0 (0%)	17 (5%)	1 (0.3%)	0 (0%)	19 (6%)
Junior role	79 (23%)	43 (13%)	3 (1%)	79 (23%)	13 (4%)	20 (6%)	131 (38%)
Senior role	73 (21%)	55 (16%)	4 (1%)	90 (26%)	18 (5%)	17 (5%)	132 (39%)
Senior leadership	30 (9%)	18 (5%)	0 (0%)	34 (10%)	7 (2%)	8 (2%)	48 (14%)

Table 3.4: Participant demographics divided by gender identity and race/ethnicity. For each cell we provide the number of participants, as well as the percentage of the total participant pool. Note, we only include the top three most common races/ethnicities and participants could mark multiple races/ethnicities, so those numbers will not sum to 100%. Additionally, two participants self-described their gender identities.

ity analysis and engineering, even holding more senior roles (N=178), such as ‘Senior Security Officer’ or ‘CISO’. Our participants’ had a wide-range of job roles (e.g malware analysis, secure development, and SOC operations).

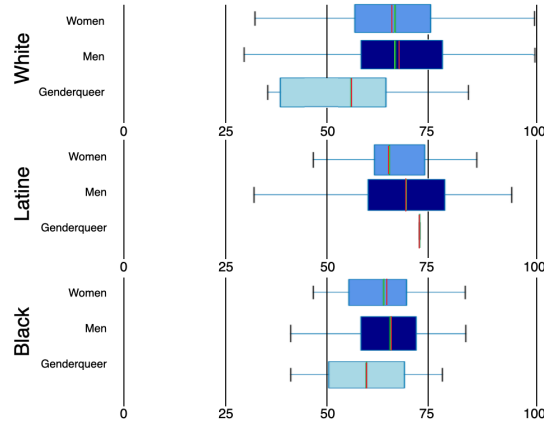
3.4 Perceptions of Belonging and Social Experiences (RQ1)

We found few differences in responses between demographics. However, we observed widespread low perceptions of belonging and that women were more likely to experience more severe forms of hate and harassment.

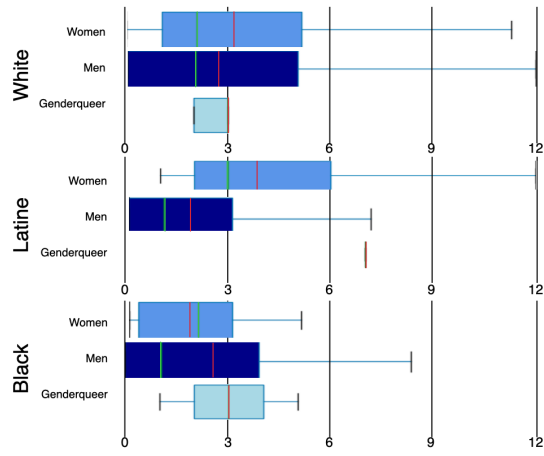
Psychological safety is low for everyone. We observed low psychological safety for all participants: 65.5 on average, which falls into the bottom quartile of scores from a cross-industry benchmark [77]. However, our participants’ average belonging uncertainty was lower (indicating less uncertainty) than samples from prior work, which showed higher uncertainty for both non-marginalized and marginalized groups of professionals [205] and students [52]. This suggests our participants overall feel they belong in cybersecurity, but do not feel comfortable speaking up in the community.

No observed differences in perceptions of belonging between genders or races/ethnicities.

When comparing psychological safety and belonging among genders the average scores for men (psycho-



(a) Psychological safety scores across gender and ethnicity



(b) Severe harassment scores across gender and ethnicity

Figure 3.2: Quantitative results from survey questions about perceptions of belonging and social experiences. The green line indicates the median and the red line indicates the mean for each metric.

logical safety 66.6, belonging uncertainty 14.2), women (64.6, 14.3), and genderqueer (60.0, 13.1) were similar. White participants (65.7, 14.2) reported similar scores as participants who identified as Black (64.5, 13.7) or Latine (67.4, 14.6). The similarities across demographic groups can be seen in Figure 3.2a and Figure B.1, which plot participants' psychological safety and belonging scores, grouped by gender and race/ethnicity.

The psychological safety regression (Table 3.5a) found no statistically significant correlation for gender or race/ethnicity. Psychological safety was negatively correlated with participants who reported taking a programming course in high school. Participants who took high school programming scored 5.9 points lower on average while controlling for other factors ($p < 0.001$), indicating participants who began developing cybersecurity skills earlier feel less safe in the community. The final model for belonging uncertainty did not explain a significant variance ($R^2 < 0.02$), so we do not provide it here or discuss it further.

Severe harassment more common for women and those who enter the field earlier. Focusing

Variable	Value	Coef	CI	<i>p</i>
Gender	Man	—	—	—
	Woman	-2.3	[-5.6, 0.9]	0.170
HS Prog	False	—	—	—
	True	-5.9	[-9.4, -2.4]	0.001*

— Base case (Coef=0, by definition)
*Significant effect

(a) Psychological safety linear regression.

Variable	Value	LE	CI	<i>p</i>
Gender	Man	—	—	—
	Woman	1.2	[1.6, 3.0]	0.015*
Sec Yrs	—	0.9	[0.9, 1.0]	0.016*
HS Prog	False	—	—	—
	True	1.6	[1.2, 2.1]	0.001*

— Base case (Log Estimate(LE)=1, by definition)
*Significant effect

(b) Severe harassment Poisson regression.

Table 3.5: Summary of regression over participant psychological safety (A) and severe harassment (B). R^2 for the psychological safety model was 0.04 and the Pseudo R^2 for the harassment model was 0.06 (corrected Aldrich-Nelson).

specifically on severe instances of negative social experiences, namely violence, stereotyping, doxxing, and sexual advances, Figure 3.2b shows the distribution of severe harassment responses, organized by gender and races/ethnicities. Overall, we note that severe harassment is rare. The average score was 3 out of a possible 12 points and a majority of participants reported "Never" experiencing violence (69.0%), sexual advances (62.0%), or doxxing (60.5%). The exception regarded experiences of stereotype bias, which participants most often reported "Never" experiencing (32.5%), but a non-trivial number reported experiencing it "Rarely" (29.5%), "Occasionally" (25.4%), or Frequently (12.6%).

Women and genderqueer participants reported more frequent occurrences of severe harassment (average frequency scores 3.2 and 3.7 out of 12, respectively) than men (average 2.8). Table 3.5b shows this correlation was statistically significant ($LE = 1.2, p = 0.015$), indicating women were 1.2 $\times$ more likely to report more frequent severe harassment than men. We did not observe a similar statistically significant difference for race/ethnicity.

Again, we observed that high school programming experience correlated with an increase in negative outcomes for participants. Participants with high school programming experience were an estimated 1.6 $\times$ more likely to report severe harassment ($p < 0.001$). We also found participants with more security experience were slightly less likely to report experiences of severe harassment ($LE = 0.9, p = 0.016$).

No observed statistically significant difference in social experiences. We did not observe any statistically significant differences between genders or races/ethnicities on the OSEM scale. Men's (17.9) and women's (17.4) average scores were similar, however genderqueer participants' scores were slightly higher (21.9), indicating a higher rate of unsupportive experiences. Also, White participants

Variable	Value	LE	CI	<i>p</i>
Ethnicity	Black	–	–	–
	Latine	0.9	[0.9, 1.0]	0.364
	White	0.9	[0.8, 0.9]	0.001*
Close Helpful	False	–	–	–
	True	1.1	[1.0, 1.1]	0.011*
HS Prog	False	–	–	–
	True	1.1	[1.0, 1.1]	0.002*
Role	Junior	–	–	–
	Not in Sec	0.9	[0.7, 1.1]	0.020*
	Senior	1.1	[1.1, 1.2]	0.001*
	C-Suite	1.2	[1.1, 1.2]	0.001*

– Base case (Log Estimate (LE)=1, by definition)
 *Significant effect

Table 3.6: Summary of regression over participant vulnerability discovery self-efficacy scores. The Pseudo R^2 (corrected Aldrich-Nelson) for the self-efficacy model was 0.25.

(18.2) reported similar OSEM scores as Black (17.4) or Latine (16.6) participants. The regression over OSEM scores is summarized in the supplemental materials [180]. The only statistically significant correlation was for participants with a helpful close relationship who were expected to have OSEM scores $0.8\times$ participants’ without close relationships ($p < 0.001$). Because this LE is < 1 , this indicates a lower score and less negative experiences, as close relationships likely provide important support.

White security experts have lower vulnerability discovery self-efficacy. On average, White participants reported statistically significantly lower vulnerability discovery self-efficacy (32.8) than Black participants (36.6). White participants’ scores are estimated to be $0.9\times$ Black participants’ scores, holding all other factors equal ($p < 0.001$), as seen in Table 3.6. We did not see a similar difference between Black and Latine participants. We did not see the same stark differences for gender. On average, men’s scores on the vulnerability discovery self-efficacy metric were slightly higher (34.2) than women’s (33.0) and genderqueer participants, but gender does not appear in our final regression model.

Security experts with more experience have higher vulnerability discovery self-efficacy. Participants who have left the field or have yet to enter the workforce had lower scores than those currently working in security ($LE = 0.9$, $p = 0.020$). Participants in more senior ($LE = 1.1$, $p < 0.001$) or C-Suite ($LE = 1.2$, $p < 0.001$) roles reported higher vulnerability discovery self-efficacy than participants in junior roles. Similarly, participants with earlier exposure to programming reported $1.1\times$ higher vulnerability discovery self-efficacy than participants who began programming later ($p = 0.002$). Also, we observed participants who reported having close helpful relationships had statistically significantly higher vulnerability discovery self-efficacy ($LE = 1.1$, $p = 0.012$), echoing the benefits of having a mentor from prior work [75].

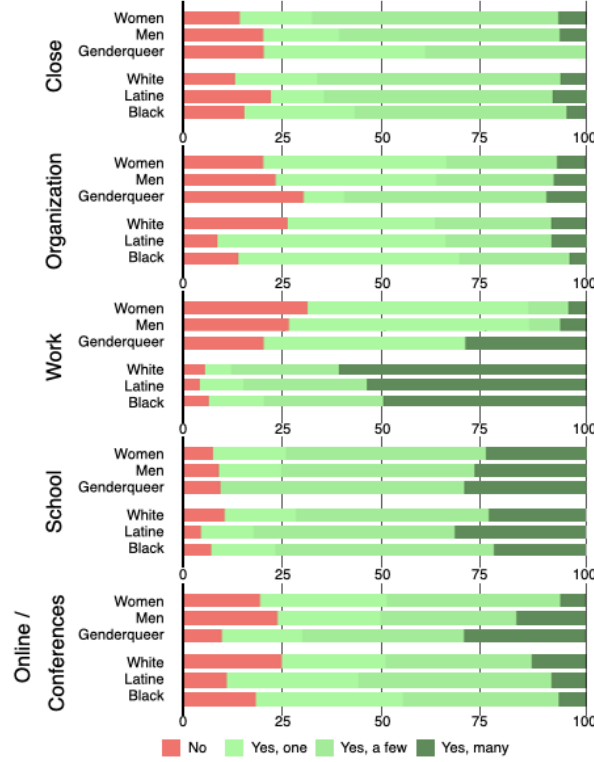


Figure 3.3: Subcommunity membership grouped by gender and race/ethnicity.

3.5 Subcommunity Participation (RQ2)

Next, we turn to participants’ reported subcommunity experiences. Figures 3.3 and 3.4 show participants’ reported membership in and perception of helpfulness, respectively, for each subcommunity, divided by demographic group. For brevity, we show reported rate of discussion in the supplemental materials [180], as there were no clear differences between groups.

No difference in subcommunity membership or rate of discussion. Participants most often reported discussing security at work (57.3%), having close friends/mentors (56.7%), joining learning communities while in school (48.2%), and participating in online forums or conferences (37.1%). We did not observe a statistically significant difference in subcommunity membership or discussion rates for any demographic groups. Participants who reported joining each subcommunity most often reported discussing security occasionally (54.8% - close friend/mentors, 51.1% - online and conferences, 47.3% - organizations) or frequently (51.7% - work, 40.1% - school).

Women prioritized identity-focused organizations. Focusing specifically on the types of organizations participants reported being members of, we observed a significant difference between women and men. Women (39.8%) were statistically more likely than men (14.8%) to join “identity-focused” organizations ($\chi^2 = 18.3, p < 0.001$), such as Women in Cybersecurity. However, these rates flipped for

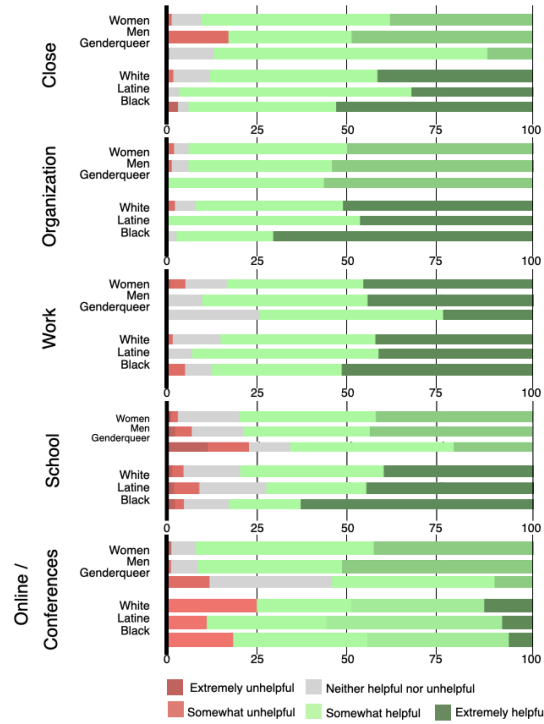


Figure 3.4: Community helpfulness grouped by gender and race/ethnicity.

	Men	Women	Genderqueer	Total
White	106 / 101	72 / 73	3 / 3	183 / 179
Black	23 / 21	18 / 18	2 / 2	43 / 41
Latine	25 / 24	17 / 16	1 / 1	43 / 41
Total	172 / 163	115 / 114	7 / 7	301 / 291

Table 3.7: Participant demographics divided by gender and ethnicity for participants who provided examples of supportive (first number) and unsupportive (second number) community experiences.

general-focus organizations, as men more often joined these groups (59.7%) than women (38.3%), and this difference was statistically significant ($\chi^2 = 6.6, p = 0.010$). We did not observe a similar divide between races/ethnicities for either identity-focused ($\chi^2 = 2.5, p = 0.284$) or general-focus ($\chi^2 = 0.2, p = 0.917$) organizations. We did not observe a similar trend among genderqueer participants as four (of ten) reported membership in identity-focused organizations and four were members of general-focus organizations.

Black participants found community organizations more helpful than White participants.

A majority of both Black and White participants who were members of community organizations perceived them as helpful. However, Black participants skewed significantly ($\chi^2 = 5.5, p = 0.019$) more positive regarding community organizations (74.2% extremely helpful, 22.6% somewhat helpful) than White participants (51.3% extremely helpful, 40.3% somewhat helpful).

3.6 Supportive and Unsupportive Experiences (RQ3)

Finally, we turn to participants' reports of (un)supportive experiences within the cybersecurity community. 307 of the 342 participants responded: 301 described supportive experiences (88.0%) and 291 described unsupportive experiences (85.1%).

We note that a lack of response does not necessarily indicate a lack of supportive or unsupportive experiences, as responding to these questions was optional. Our analysis focuses on trends observed between men and women, as responses from other genders and races/ethnicities were limited. We did not observe clear differences in the percentages of reported supportive and unsupportive experiences across race/ethnicity, so we do not report those numbers for brevity. However, this should not preclude future work from achieving higher recruitment across these demographics.

Women experienced more unsupportive, negative identity-based incidents. Women (N=14, 12.3%) were significantly ($\chi^2 = 5.18, p = 0.023$) more likely to describe encountering an unsupportive environment than men (N=9, 5.5%). For example, one woman shared, "I had a project with a colleague who is not anywhere near as technical as I am, yet he consistently tried to micromanage my technical work, and sometimes told me I was doing things wrong even though he didn't know what he was talking about." When men reported unsupportive environments, these were often due to differing goals or personalities, unlike the devaluation of skills observed with women. For example, one man stated, "[I have been] able to hop projects and or jobs in the past when I felt a workplace was not supportive of the direction I wanted to grow in... Once to avoid policies, and once to avoid a person I did not work with well." Men often reported being able to navigate out of these unsupportive environments relatively easily.

Women (N=16, 14.0%) reported significantly ($\chi^2 = 5.2, p = 0.023$) more negative experiences related to their identity than men (N=10, 6.1%). One woman explained, "I was at a career fair and the person at the booth refused to talk to me. They ignored me ... as I patiently waited and proceeded to talk to people who showed up after me... All while refusing to acknowledge my presence. I figure it had something to do with the fact that I was the only female at the booth."

Men more often reported never having an unsupportive experience. Some participants explicitly said they had no supportive (N=10, 3.3%) or unsupportive experiences (N=97, 33.3%). Men were more likely not to have unsupportive experiences (N=59, 36.2%) than women (N=36, 31.6%), though the difference was not statistically significant ($\chi^2 = 1.4, p = 0.240$). We did not observe a difference between genders for participants reporting no supportive experiences (6 men and 4 women; $p = 1$ using a Fisher's Exact Test).

Multiple mentions of toxic experiences by both genders. Bullying, harassment, fear of retaliation, and doxxing were reported by women (N=9, 7.9%) and men (N=14, 8.6%; $\chi^2 = 27$, $p = 0.600$). One man shared he experienced “years of harassment, doxxing, and impersonation,” including “fake profiles created in white nationalism and hacking forums.” Unwanted attention or sexual advances were mentioned by three women and one man. One participant shared, “A somewhat close friend I had made through a cybersecurity forum had made quite a few uncomfortable sexual remarks which made me question if cybersecurity as a whole was like this or if it was an isolated case.”

Men and women reported some common experiences. In addition to the differences discussed above, we also observed some similarities in unsupportive experiences, though these were typically less frequent. These included feeling unwelcome as newcomers (18 men, 11.0%; 6 women, 5.3%; $\chi^2 = 1.3715$, $p = 0.242$), receiving negative consequences from their own actions (13 men, 8.0%; 10 women, 8.8%; $\chi^2 = 0$, $p = 1$), and difficulty collaborating (22 men, 13.5%; 11 women, 9.6%; $\chi^2 = 0.5$, $p = 0.473$).

There was also general agreement on supportive experiences. Men and women both described receiving career support (41 men, 23.8%; 30 women, 26.1%; $\chi^2 = 0.1$, $p = 0.781$), having their questions answered (26 men, 15.1%; 20 women, 17.4%; $\chi^2 = 0.1$, $p = 0.734$), having positive educational experiences (16 men, 9.3%; 7 women, 6.1%; $\chi^2 = 0.6$, $p = 0.444$), and participating in collaborative problem solving (21 men, 12.2%; 11 women, 9.6%; $\chi^2 = 0.3$, $p = 0.608$).

Women receive support from individuals; men more likely to find groups helpful. Women predominantly cited experiences with specific individuals such as managers, professors, and specific friends or acquaintances (N=47). While the same number of men discussed these individual relationships (N=47), this number is proportionally lower (27.3% of men and 40.9% of women). On the other hand, men discussed supportive experiences with broader groups like online forums or conferences more frequently than women did (34 men, 19.8%; 13 women, 11.3%). One man said, “After sharing a blog post or link to code, someone from the community replied with helpful advice or other areas I could investigate.” This difference was statistically significant ($\chi^2 = 6.4$, $p = 0.011$). While we saw a similar divide between men and women in terms of the people involved in unsupportive experiences, i.e., individuals (27 men, 16.6%; 26 women, 22.8%) and groups (27 men, 16.6%; 14 women, 12.3%), this difference was not statistically significant ($\chi^2 = 2.1$, $p = 0.147$).

Chapter 4

An Investigation of Interaction and Information Needs for Protocol Reverse Engineering Automation

4.1 Introduction

There has been a long history of developing automation for reverse engineering, with significant research and industry investments made [208, 165, 17, 56, 181, 122, 12, 197, 88, 50]. While these efforts have made significant strides, human intelligence is required to supplement automation and will likely remain necessary for the foreseeable future [209, 29, 143, 204]. Therefore, recent research has focused on understanding and supporting reverse engineers' processes [202, 115, 29, 204]. However, there is limited human-automation interaction research for reverse engineering, with existing research focused on existing automation's usability flaws [118], challenges in automation setup [151], or developing novel AI/ML techniques without primarily considering usability [11, 20, 13, 97, 108].

In this paper, we take a first step toward closing this gap. Because reverse engineering is a complex domain, we expect variation in information and interaction needs for specific tasks. Therefore, we first focus on a specific reverse engineering sub-field, *protocol* reverse engineering (ProtocolREing). To our knowledge, prior work investigating reverse engineering's human factors has almost exclusively considered *software* reverse engineering (SoftwareREing), with none studying ProtocolREing. ProtocolREing is the task of inferring a protocol, i.e., structured message sequence where each message segment serves a specific purpose, from a stream of unstructured data [104]. This task is most commonly performed for malware analysis, where an analyst reviews data transmitted between a bot and bot controller to determine what is transferred to write signatures for malware identification and potentially manipulating

these communications. ProtocolREing is also used when investigating the software is challenging. For example, embedded systems often have unique architectures that are not supported by existing disassembly and emulation tools. ProtocolREing can be thought of as attempting to identify patterns in unstructured data, where protocol reverse engineers (ProtocolREs) must consider many possible protocols and determine which fits the data. This differs from SoftwareREing, where software reverse engineers’ (SoftwareREs) task is to interpret a sequence of program instructions with known meaning to determine their function. It is worth noting that ProtocolREing is closer to problems in data science where an analyst investigates data records to infer some relationship. However, in ProtocolREing, there are clear logical and structural relationships between fields in the underlying protocol, restricting the search space of possible protocols. For example, messages are typically divided into fields, multiples of a byte in size.

ProtocolREing’s ambiguity introduces a unique challenge and an opportunity for automation to help ProtocolREs navigate this data. Several tools have been developed to automate ProtocolREing [35, 98]. However, these tools offer limited interactivity only through configuration file changes or modifications to the automation’s codebase. This requires users to be familiar with the automation’s design, creating a high barrier for interaction. Conversely, there are full-featured user interfaces for manual ProtocolREing, most notably Wireshark [72]. However, these tools only support limited automation for matching data to known protocols (e.g., TLS, HTTP). Because of the unique task characteristics and dearth of existing tools, we sought to understand the interaction needs of ProtocolREs to guide the development of the next-generation automation.

Specifically, we seek to answer the following research questions:

- RQ1: What process do ProtocolREs follow?
- RQ2: What information do ProtocolREs need from ProtocolREing automation?
- RQ3: What interactions should ProtocolREing automation support?

Because there is no existing literature on ProtocolREs, we sought to understand ProtocolREs’ processes, e.g., what types of questions they ask about the data, what hypotheses they generate, the specific steps taken to learn more about the data, and what information informs their decisions. We also wanted to understand their unique information and interaction needs throughout.

To answer these questions, we drew inspiration from Votipka et al.’s retrospective observational interviews with expert SoftwareREs to understand their process when investigating unfamiliar programs [203]. As there is no theoretical basis on which to build quantitative assessments, we pursued an exploratory qualitative approach drawing from prior work in expert decision-making [105, 32].

We conducted 16 interviews with ProtocolREing professionals, asking them to interact with a paper prototype interface with automated features. We asked participants to consider sample tool output from

the automation and provide feedback on remaining information needs, missing interactions, and thoughts on how automation can support or hinder their process.

We found that generally, the ProtocolREing process model centers around determining potential field boundaries and datatypes. ProtocolREs cyclically update their hypotheses about field boundaries and datatypes, which are tightly coupled. ProtocolREs incorporate information from various sources, including context research about the application (if available) and any tooling support. ProtocolREs typically operate within field boundaries, occasionally comparing a given field against its neighboring fields to validate or challenge their current working hypothesis. Above all, ProtocolREs need the ability to modify field boundaries, datatypes, and other configuration options quickly and easily as they cycle quickly between hypotheses, testing whether protocols that fit one part continue to fit when looking at other message segments. Any tooling or automation inhibiting this quick, cyclic process is highly disruptive to ProtocolREs. This differs from SoftwareREing, which has been shown to have a natural break between identifying a hypothesis and testing that hypothesis, as the analyst shifts from static to dynamic analysis.

Due to the application-specific nature of ProtocolREing and the inherent ambiguity of the task, automated results should be conservative and easy to reverse, acting as a guide to an iterative search. This follows from the observation that the process was highly iterative as ProtocolREs try multiple possible protocol options before determining the correct ones. This could be seen in our participants' nearly universal belief that automation could not correctly predict an entire protocol, but was instead useful for providing suggestions for parts of the protocol. Instead of taking the suggestion as an end product, they wanted interactions that allowed adjusting the automation's assumptions to refine the analysis and quickly validate any suggestions iteratively. This suggests future work in human-automation interaction is particularly important for ProtocolREing as human review is necessary. Without good interaction design, the automation is unlikely to be useful.

Finally, because the ProtocolREing process is mostly focused on iterating possible datatypes and field boundaries, we observed participants' requests for additional information and interaction were shaped by these two protocol properties. This included providing datatype definitions as a way of restricting automated suggestions or navigating the search space by investigating neighbor field boundaries.

From these results, we offer suggestions for ProtocolREing tool developers and future research.

4.2 Method

We conducted 16 semi-structured interviews with ProtocolREs with significant ProtocolREing experience. We asked them to interact with a paper prototype of an automated ProtocolREing tool to see how they approach ProtocolREing tasks and how they take information from and interact with automa-

tion. Interviews were conducted between February and June 2023. We describe the development of our prototype (referred to as the Prototype Interface), mock use case, and interview protocol, as well as our study recruitment, analysis approach, ethical considerations, and our study’s limitations.

4.2.1 Prototype Interface Design

To shape the interview discussion and provide a reference point, we presented participants with a paper prototype to understand how ProtocolREs interact with automation. A paper prototype is beneficial as it minimizes distractions caused by interface aesthetics and technical issues, ensuring the participants’ focus was on core information needs and user interactions [169].

To produce a realistic interface, we first examined existing ProtocolREing tools and ProtocolREing process literature to identify common features. Then, we defined design guidelines based on prior HCI and SoftwareREing literature. This section describes our feature and guideline development and how we produced the final components of our prototype, which we refer to as prototype views.

4.2.1.1 Tool review

As a preliminary step, we determined the features commonly provided by existing tools to ensure a realistic interface. We leveraged one author’s ten-plus years of experience with ProtocolREing tool development to identify a preliminary list of ProtocolREing tools. Then, we broadly surveyed the web using an online keyword search for other available tools. We searched the terms “reverse engineering” and “Protocol reverse engineering”. We combined them with “frameworks,” “tools,” and “networking.” We searched every resulting link, webpage, or video for downloadable tools until we repeatedly encountered duplicate tools. For each search, we reviewed at least the first page of results.

From our search process, we found a variety of tools broadly categorized as Hex Editors (ImHex, Hexinator, ReHex, Synalysis, and 010 Editor), Interactive analysis tools (Cyberchef), Network analysis tools (BinaryInferno, Nemesys, Netplir, Netzob, and Wireshark), and data description tools (Kaitai). These tools are listed in Table 4.1. Due to our extensive search process, we expect our resulting tools to represent what analysts would be familiar with through their day-to-day functions to assist with ProtocolREing. To confirm this, we asked participants in our interview to state tools they previously used for ProtocolREing. Almost all the tools mentioned had overlapping features with those identified by our search. The exception was tools for packet generation (i.e., Scappy, Fiddler, and FakeNetNG), which went beyond direct message analysis and were used to produce additional message data for testing. We discuss these in detail when describing participants’ ProtocolREing processes (see Section 4.4.2).

	Hex	Decode Type	Set Type	Suggest Type	Suggest Field	Stats	Compare Multiple	Filter /Sort	User Scripting	Use Case
BinaryInferno [35]	✓	-	-	✓	✓	-	✓	-	-	Network
Cyberchef [78]	✓	✓	✓	✓	✓	✓	✓	✓	-	Binary Data
ImHex [92]	✓	✓	✓	✓	✓	✓	✓	✓	✓	Binary Data
Kaitai [98]	✓	✓	✓	✓	-	-	✓	✓	✓	Binary Data
Hexinator [89]	✓	✓	✓	✓	✓	✓	✓	✓	✓	Binary Data
Nemesys [171]	✓	-	-	-	✓	-	✓	-	-	Network
Netplier [210]	✓	-	-	-	✓	-	✓	-	-	Network
Netzob [137]	✓	✓	✓	✓	-	✓	✓	✓	✓	Network
ReHex [44]	✓	✓	✓	✓	✓	-	✓	✓	✓	Binary Data
Synalysis [183]	✓	✓	✓	✓	✓	✓	✓	✓	✓	Binary Data
Wireshark [72]	✓	✓	-	✓	✓	✓	✓	✓	✓*	Network
010 Editor [172]	✓	✓	✓	✓	✓	✓	✓	✓	✓	Binary Data

Table 4.1: Table of features and capabilities of various ProtocolREing tools. The * notation indicates a tool supports the feature via plugins.

4.2.1.2 Feature review

We reviewed each tool for the features they provide for ProtocolREing. This review identified the following features observed in at least one tool. Table 4.1 indicates the features offered by each tool.

F1 **Hex** - All our surveyed tools present raw data first in hex. This offers a view into the raw data in a compressed format, i.e., instead of showing each bit.

F2 **Decode Type** - Many tools allowed users to convert the hex data to various other datatypes, either one datatype or multiple simultaneously.

F3 **Set Type** - Some tooling allowed users to specify and restrict fields to a certain type. This was a common feature of several tools.

F4 **Suggest Type** - When working on ProtocolREing, there is missing information about the structure and type of the messages. The tool can offer users suggested types based on the raw hex values to guide their process.

F5 **Suggest Field** - Some tools try to infer potential field boundaries from the data and present these candidate fields to the user.

F6 **Data Distribution (Stats)** - Another method for ProtocolREs to glean insights about the raw data is through visualizations and summary statistics. Tools can present this information visually with graphs or via text.

F7 **Multiple Data comparisons (Compare Multiple)** - All tools can simultaneously operate on multiple pieces of data, packets, files, or streams, allowing for comprehensive comparisons and analyses. For example, several tools support the incremental development of data parsers to handle various input streams.

F8 **Filter/Sort** - Many tools offered filtering or sorting functionality across hex messages. ProtocolREs could use these to sort by an identified sequence number to preserve inter-message contexts.

F9 User Scripting - Most of the tools could support extended functionality from user-provided scripts.

In addition to the features identified in existing tools, we incorporated additional opportunities for automation interaction. While some tools offer automated support, none allow ProtocolREs to interact with the automation directly through the user interface. Instead, users could only impact the automation by changing a configuration or providing a new template. These are speculative features not provided by prior tools but essential to investigating issues of human-automation interaction.

F10 Split-merge fields - We included the ability to split and merge fields at every byte boundary and automated suggestions for additional field boundary splits beyond the preliminary fields. This additional flexibility allows users to explore the data more meticulously and more easily change their field boundaries as they uncover more information about the data.

F11 Alternative data type and field boundary suggestions - We provide additional possible suggestions besides the preliminary suggested type and field boundaries. This emulates automation, providing the top N options instead of just the option with the most confidence. This allows users to consider several suggestions in parallel and may prompt users to consider potentially overlooked options [214, 213].

4.2.1.3 Design guidelines

Next, we describe the design guidelines that directed the development of our Prototype Interface. We drew on two bodies of literature to determine these guidelines. First, we adopted recommendations from prior work in SoftwareREing tool usability [203, 118]. We also consider guidelines from exploratory visual analysis [164, 163] From these reviews, we established the following guidelines:

D1 Filter, Zoom, and Details on Demand. Prior literature has highlighted the importance of user-defined focus and scope. For our Prototype Interface, this was filtering the data or zooming and drilling down to get a more detailed view [164]. Tools should allow users to get a richer, more detailed view of their data as they need it. These core functionalities enable users to refine their focus and extract only the relevant details to continue investigating their hypotheses.

D2 Input and Output presented in line - To integrate into users' processes more seamlessly, tools should accept inputs and present outputs in the same window or medium with which the user is engaged. For our context, we ensured that users had potential input and output options for each interface view, which were presented in line with the data.

D3 Analysis Tuning - To better suit the needs of their users, tools should allow their functionality to be customized or fine-tuned. This additional flexibility enables users to impart contextual

messages simultaneously, as shown in Figure 4.1.A. Intuitively, each message is shown on a line, with the bytes aligned across messages using a mono-space font in Hex (F1, F7, D2). This format is consistent with ASCII text display, with data ordered left to right, top to bottom.

This interface automatically assigns field boundaries to the raw data and separates the data into candidate fields and suggested types, as shown in Figure 4.1.A (F4, F5). Users can select from alternative data type suggestions the automation has deemed potentially valid (F3, F11, D2, D3). This broad view allows users to get an overview of patterns across all the data or narrow their focus to a single field boundary, type, or message.

Split-Merge View (F3, F4, F1, F7, F5, F10, F11, D2, D3). The second view is similar to the first in how the data is presented (F4, F1, F7, F5, D2), but it gives more control to the user. Users can modify the suggested initial field boundaries, combining fields or splitting to create new boundaries and change the suggested types (F3, F10, D3). Users can adjust field boundaries via a graphical interface presented directly in line with the hex data immediately (D2), as shown in Figure 4.1.B.

In addition to the initially suggested field boundaries, this interface allows the user to choose alternative field splits using a wedge glyph. Similar to the additional suggested data types, this indicator can be a subtle nudge to ProtocolREs to explore other field boundary and data type possibilities when investigating the data (F11). One of the wedge glyphs was highlighted to show a candidate field boundary the automation thought was possible. We used this to ask participants how they preferred incorporating automated tooling.

Histogram View (F1, F6, F7, D1, D4). Another way to visualize the data is through a histogram, as shown in Figure 4.1.C. The visualization is organized by byte position, with each column representing the aligned bytes across the collection of Hex messages (F1, F7). This visualization shows the frequency of different values at a byte offset into a message as a line chart (F6, D1, D4). We order the chart with lower hex values at the top and higher hex values at the bottom. A legend of hex values is shown on the vertical axis. Visualizing the data in this manner allows the frequency of values to be compared within a byte offset or between fields. We removed the suggested types and split-merge capabilities from the prior interfaces to avoid overloading participants with too many new features simultaneously.

Decoding View (F1, F7, F2, F8, D1, D3, D4). Our penultimate view introduces several key features to manage the data: decoding as multiple data types (F2), filtering, and sorting (F8, D1, D3). First, this view showed the data decoded as a variety of standard datatypes, as shown in Figure 4.1.D. This interface allows the analyst to observe the decoded values and judge whether any hex values translate to valid values for multiple common data types (F1, F7, F2, D4). By providing multiple comparisons,

this view facilitates exploration and rapid evaluation of hypotheses for data types where the interpreted values appear random or inconsistent with analyst knowledge.

This view allows analysts to filter or sort messages to refine their focus on specific data (F8, D1, D3), as shown in Figure 4.1.D. We did not restrict how filtering or sorting could be utilized but instead asked participants how they might want to perform these operations. For example, they could modify their focus based on criteria such as timestamps, source and destination addresses, or message content. Because we tried to cover several features but wanted to ensure we were respectful of our participants' time, we could not always cover all views when participants gave more detailed answers earlier in the interview. To avoid missing participants' responses to the combination of all views, we chose to skip the filtering and sorting, which are well-established in UI design practice [163, 126, 142, 63, 189] and already employed in several existing tools in Figure 4.1. The interviewer decided whether to skip filtering and sorting at the time of the interview. These features were skipped in six of the 16 interviews.

Combined View (F1-F8, D1-D4). The final view integrates elements of all prior views as shown in Figure 4.1. The Combined View allows candidate fields to be split and merged (F4, F5, D3, D4), interpreted as various data types (F3, F1, F7, F2, D1, D2, D3, D5), and displays summary visual information (F6, D1, D2, D5). The advantage of this interface is it allows simultaneous comparison across multiple dimensions and gives users multiple options to tweak the data and explore hypotheses. Participants were prompted to explain which interface elements are useful, overbearing, or distracting while performing their tasks.

Due to the nature of our Prototype Interface, we chose not to directly implement a scripting view across our interface. Instead, we prompted participants during the interview to describe how they would like to extend each view with scripting or how they would like to interact with the combined interface programmatically.

4.2.2 Interview Protocol

Interviews were conducted and recorded on Zoom and lasted 60 minutes on average. The same author conducted all interviews to ensure consistency. Participants were asked to review each view and provide feedback. Our study protocol is summarized in Figure 4.2, and the script and additional interview materials are provided in an OSF folder [14].

Participant background and experiences in ProtocolREing (Figure 4.2.A,B). The interviews began with an open-ended question asking participants to describe their ProtocolREing backgrounds and expertise. We concluded this section by asking participants to describe ProtocolREing tools they commonly use.

Interview scenario. Participants were told to determine the protocol of an example set of captured messages. We used the example given in Section 2.0.2 to populate our interface but did not provide context to participants. The interview started with a sample of the message data shown as hex, with one message per line. We instructed participants that each message was transmitted in the same direction (host A → host B), so each line was a different instance of one-directional communication.

View interaction (Figure 4.2.C-G). Next, we showed participants the different views described in Section 4.2.1.4. Participants were progressively walked through the different views in the following order: Hex View, Split-Merge View, Histogram View, Decoding View, and finally, Combined View.

We asked participants to walk through their approach to investigate this data using all of our prototype views. For each view, we asked participants to review the automation output and say if they had any hypotheses about the data, if they thought any output was incorrect, and what other features or information they would need to complete their process. We asked participants to “think aloud” to reveal their ProtocolREing approaches and decision-making criteria. We asked follow-up questions to investigate participants’ thought processes, information and interaction needs, or any sources of confusion. To challenge participants’ perspectives, we introduced an incorrect automated suggestion across all views and added a noisy errant message during the Split-Merge View.

The interview scenario was developed independently of the design of the Prototype Interface. While we consistently used the airline data scenario for testing, the interface was designed to be versatile and applicable to many other protocols beyond just this use case. We aimed to capture genuine interactions and decision-making processes while allowing flexibility for participants to suggest feature improvements that better align the tool’s workflow with their ProtocolREing process.

4.2.3 Recruitment

We posted on reverse engineering-related channels on social media (e.g., LinkedIn and Reddit), posted on Slack channels for various popular reverse engineering tools (e.g., BinaryNinja and Ghidra), shared our recruitment message with the leadership of social organizations focused on reverse engineering, and sent emails to a curated panel of security professionals who made themselves available for user studies. Anyone who responded to these recruitment efforts was asked to fill out a screening survey that asked questions about experience with ProtocolREing. Specifically, we asked participants to rate their skills in reverse engineering, vulnerability discovery, malware analysis, and software development on a Likert scale ranging from “Basic (Limited exposure)” to “Expert (An acknowledged source of knowledge).” Additionally, we asked respondents to submit their resumes detailing their reverse engineering experiences to contextualize their responses and help us understand their background and familiarity with ProtocolREing.

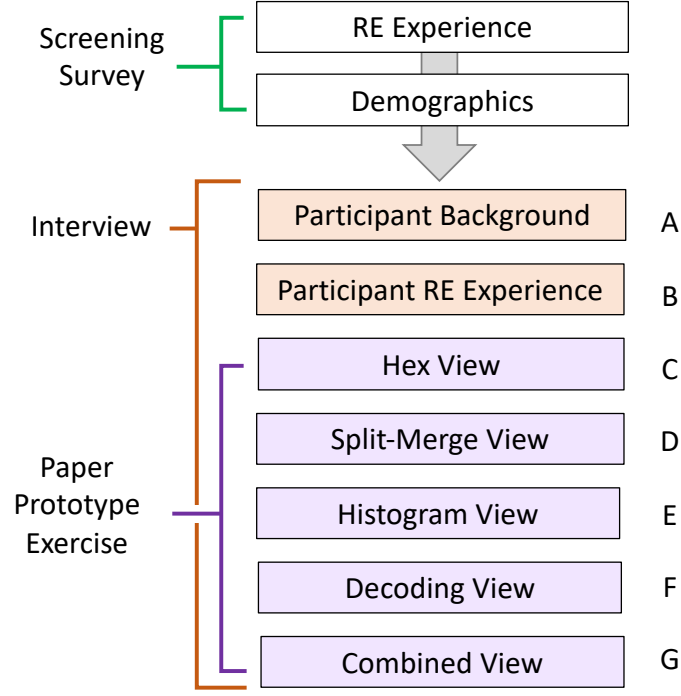


Figure 4.2: Study recruiting and interview structure.

We interviewed all candidates with self-reported “advanced” or “expert” skill levels in at least two categories. P5 did not provide their experience level in the survey, but from their CV, we determined they met our interview criteria. We also required participants to be English-speaking and at least 18 years old. While our recruitment focused on smaller communities focused on reverse engineering, we still received several inauthentic responses. We removed responses with duplicate answers to open-ended questions and very similar responses (e.g., email addresses with a few letter changes) within a short period. While this slowed recruitment efforts, it did not impact our data collection as it was straightforward to determine whether a participant *actually* had ProtocolREing experience once the interview began. Specifically, one participant was removed from our study and not included in our analysis because they did not understand basic ProtocolREing concepts, such as converting hex to ASCII, and could not provide specifics about their process when asked to elaborate. Additionally, this participant reported conflicting information during their pre-screening survey and at the beginning of the interview (i.e., their job title and roles performed), indicating issues with this participant.

All participants were compensated with an Amazon.com gift card valued at \$40 USD.

4.2.4 Analysis

In this study, we employed iterative open coding [178, pg. 101-122] to analyze the interviews and identify key themes related to participants’ approaches to ProtocolREing and their interaction with the automations’ suggestions. We primarily followed an inductive approach, allowing codes to arise from the data. However, we used some predefined codes from previous work as a foundation to guide our framing.

Specifically, we drew initial codes regarding reverse engineers’ processes and usable tool guidelines as described in Section 4.2.1.2 and from prior studies [202, 119]. We expanded on these as necessary to capture unique aspects of ProtocolREing.

The interviewer and one additional researcher collaboratively coded two interviews and discussed the results with the entire research team to develop the initial codebook. The codebook, given in Appendix ??, centered around three high-level categories: their ProtocolREing processes, information needs, and automation interaction needs. We also code other items, such as their doubts about the automation’s suggestions and any additional requested features. Because we presented multiple views, we applied codes uniquely for each view to capture differences between their components. This allowed participants to preempt features provided in future views, suggesting that those features are general needs instead of specific needs in certain contexts.

Using this codebook, two researchers independently coded the remaining 14 interviews, meeting after every two interviews to compare codes and resolve disagreements. After each round, the full research team discussed the codes and updated the codebook accordingly; these changes were reapplied to previously coded interviews. Because the researchers jointly reviewed every coded transcript, we did not calculate inter-rater reliability [121].

Next, we performed axial coding to identify connections within and between our applied codes, allowing us to extract higher-level themes and relationships [178, pg. 123-142]. We specifically wanted to identify the different information and interaction needs and our participants’ reservations regarding automation.

4.2.5 Limitations

This research shares limitations that are common to other exploratory qualitative work. First, our sample size is small, and our results may not generalize. For most findings, we indicate the portion of participants who stated the theme to provide a general indication of prevalence. However, participants who do not mention a theme may still agree but may not have thought to mention the theme. Therefore, we chose not to perform statistical comparisons, and our prevalence results may not generalize beyond our sample; instead, we suggest directions for future work. We attempted to recruit broadly to capture diverse perspectives. Because we reach saturation of themes, we expect our results to represent many approaches to and opinions on ProtocolREing.

Our results may also be limited by the specific ProtocolREing scenario we used as an example and the prototype views we began with. This priming likely introduced some anchoring bias, i.e., participants are more likely to focus on information and interactions presented than they might if shown other views first [112]. We attempt to account for this bias by utilizing multiple views and developing views to represent a range of existing tools through expert feedback and our tool review (see Section 4.2.1). We chose

our scenario as it is representative of a common class of ProtocolREing problems, i.e., interacting with embedded and legacy systems, and required participants to consider various data types and relationships between fields. We also included probing questions to encourage participants to think beyond our views and scenario. We believe we achieved this as several elements of our ProtocolRE process diagram in Figure 4.3 were inspired by our participants even though they fell outside the direct actions they took in our scenario.

Our study was conducted in a lab setting with a paper prototype. When performing ProtocolREing, participants can directly interact with tools and utilize additional resources like running the code with different configurations, reverse engineering associated binaries, or reading device documentation. While our study does not capture their full ProtocolREing process, it does elucidate their interaction needs with automation and their information needs to complete their tasks properly. Furthermore, several participants remarked that the Combined View was familiar and reflected the types of interfaces they encounter in their daily work. Future work is required to investigate other facets of the ProtocolREing process.

4.3 Ethical Considerations

Before the study, participants were informed about the data collection practices, including the nature of the collected data, how it would be used, and the various measures to protect their confidentiality. Each participant provided informed consent, indicating their voluntary agreement to participate in the study. They were made aware participation was entirely optional and they could withdraw at any time without penalty. Additionally, participants could skip questions or tasks they preferred not to answer. This study was reviewed and approved by our institution’s ethics review board.

4.4 Results

This section presents our findings. We start by describing the participant population before addressing our thematic findings on ProtocolREing process, ProtocolREs’ information needs, and ProtocolREs’ interaction needs. We refer to participants using the notation "PX" where X is an anonymized number unique to each participant. See Table 4.2 for more details on each participant.

4.4.1 Participants

Our participants provided diverse demographic and professional backgrounds, as described in Table 4.2. Of our 16 participants who completed the interview, one identified as non-binary, one declined to share their gender, and the remaining identified as men (N=14). While ethnicity was an optional category,

ID	Role	Education	RE Experience
P1	Developer	Doctorate	Intermediate
P2	PhD Student	Master's	Advanced
P3	PhD Student	Bachelor's	Intermediate
P4	Developer	Some college	Advanced
P5	Reverse Engineer	Bachelor's	-
P6	Reverse Engineer	Master's	Expert
P7	Other (Security)	Bachelor's	Expert
P8	Developer	Bachelor's	Expert
P9	Reverse Engineer	Some college	Advanced
P10	Reverse Engineer	Master's	Intermediate
P11	Developer	Some college	Intermediate
P12	Other (Security)	Doctorate	Advanced
P13	Leadership	Bachelor's	Expert
P14	Leadership	Doctorate	Expert
P15	Other (Security)	Master's	Intermediate
P16	Reverse Engineer	Bachelor's	Expert

Table 4.2: Participant ID, functional roles, education, and reverse engineering skills.

13 participants self-identified as White. Most reside in the US (N=11), and five were located outside North America, including Eastern and Western Europe. Participants were between 19 and 44 years old, with the average age being 33.8 years old. Participants were highly educated, almost all holding a Bachelor's degree (N=13), and about half (N=7) had graduate degrees. This is consistent with the demographics from prior large-scale surveys of SoftwareREs [84, 85, 28]—there are no similar surveys of ProtocolREs to our knowledge—, but our population skews slightly older and more educated. This is expected as we targeted experts. Participants reported various roles and titles in software development, security research, reverse engineering, and academic roles, such as PhD students with a focus in reverse engineering.

We did not observe any difference in the processes, information needs, or interaction preferences between participants who reported intermediate versus advanced reverse engineering experience; therefore, we do not make direct comparisons between experience levels.

4.4.2 ProtocolREing Process (RQ1)

Based on participants' interactions with the prototype and the discussion of their ProtocolREing process, we observed their processes could be described as an iterative cycle of hypothesizing parts of the protocol and assessing whether these definitions fit the data. We present the components of this process model in Figure 4.3. Participants' hypotheses centered on asserting potential field boundaries (Figure 4.3.C) and datatypes (Figure 4.3.D), testing how different configurations challenge or support their beliefs. In this section, we discuss how participants interleave identifying and assessing hypotheses across these dimensions, supported by additional external information to produce a final protocol, as shown in Figure 4.3.G.

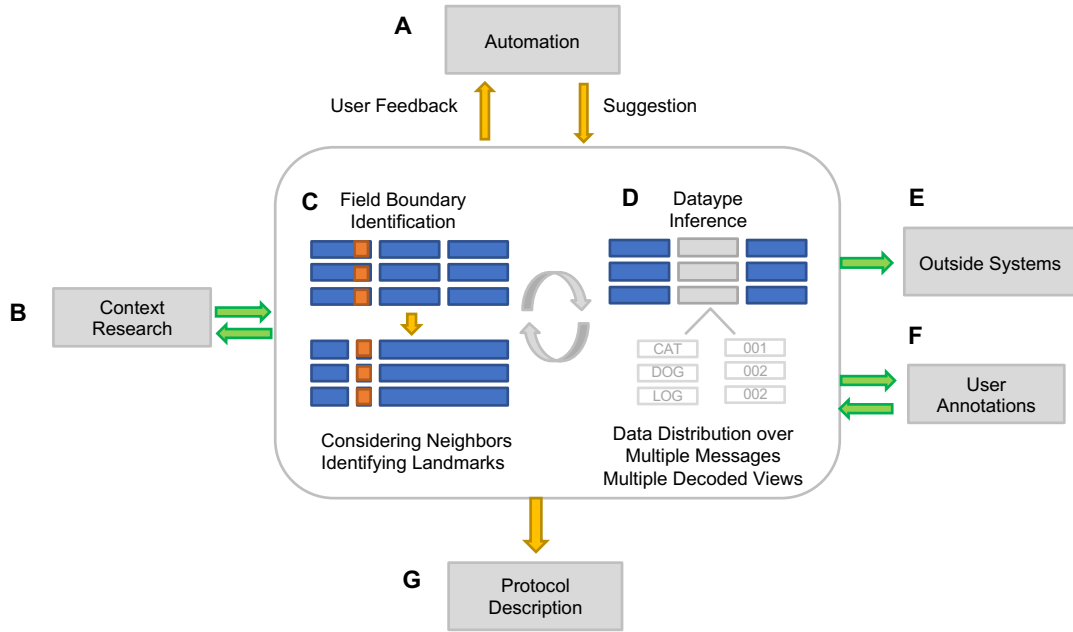


Figure 4.3: Illustration of the Protocol reverse engineering (PRE) process. The core components, Field Boundary Identification (C) and Datatype Inference (D) interactively drive the development of a Protocol Description (G) describing the data, field boundaries, and datatypes. Automation (A) and Contextual Research (B) each inform and are guided by both the current description and user feedback. Users typically start by performing Contextual Research (B) or by identifying fields (C) and types (D). Outside Systems (E) use the description and data for further validation. Throughout the PRE process, User Annotations (F) allow the user to store and recall ideas about the data and their hypotheses.

Participants used landmarks and/or neighboring bytes as indicators to determine field boundaries (Figure 4.3.C). All of our participants expressed the importance of identifying field boundaries during ProtocolREing. Participants primarily described two methods for determining field boundaries: 1) Investigating differences between neighboring bytes or fields. 2) Using easily identifiable patterns as “landmarks” to set field boundaries.

When establishing initial potential field boundaries, most participants described using differences in neighboring bytes to determine natural breaks. For example, P8 chose to segment into fields based on differences in the data distributions, saying they chose that field boundary *“because the repeating parts are grouped together, and the non-repeating parts are grouped in a way that makes sense.”* ProtocolREs can use these subtle indicators between bytes to set initial field boundaries for further investigation.

Additionally, most participants described landmarking using typical networking values like magic numbers or ASCII characters. When evaluating the automation’s suggested field boundaries, P16 described how *“Field one could definitely be like a magic value of some sort. This is the same for every packet, just like a protocol identifier.”* These repeating patterns across multiple messages were a clear indicator to determine field boundaries or details about the data sample. The majority of participants described identifying trends across multiple messages, like P9 said in the Decoding View: *“I’m looking*

at things as columns... and then I see 301 300 298 and I see these values are monotonically decreasing. So now I'm thinking the messages are out of order in the original data."

Participants' assumptions about datatypes in the protocol supported or disproved their hypotheses (Figure 4.3.D). One technique nearly all participants described was using a common, easily recognizable datatype, typically ASCII, as an anchor point to establish initial field boundaries. Most participants recognized ASCII characters from the raw hex values and used bytes inside the valid ASCII range to set initial field boundaries like P8: *"It looks like some kind of ASCII characters because the ranges of the bytes fall in ASCII range."* Similarly, P2 recognized the hex bytes as valid ASCII characters and wanted the automation to take this into account: *"I would like to be able to do something about that to affirm to the algorithm. That's good. Please take this into account."* This idea of incorporating user feedback is described further in Section 4.4.4.

Once the automation decides a type to suggest or the ProtocolRE produces a hypothesized data type, almost all participants wanted the data decoded as that type. By decoding the data as the specific type in-line, ProtocolREs can quickly assess whether the type fits. For example, when trying to determine the type of one of the fields, P7 said, *"I think that field four should be ASCII characters, so if I can tell [the automation] that should be represented as ASCII and see it as that immediately, that would be helpful."*

Several participants found the data distribution presented in the Histogram View helpful when determining datatypes. P2 scanned the Histogram View and was able to determine the final field was ASCII because all the values were contained within the ASCII range, saying, *"the obvious signal amongst this noise is the ASCII characters at the end of the messages."* However, several participants felt the Histogram View was only useful for initially providing an overview of the data but did not provide any additional insights. For example, P15 explained, *"once you realize that it's ASCII text, the histogram is no longer useful."*

Lastly, several participants described how decoding hex values to several datatypes simultaneously benefited their process. P9 described how the tool he typically uses for his job, *"[has] the ability to visualize things side by side and simultaneously visualize more [datatype] interpretations at once."*

Participants regularly switched between identifying field boundaries and datatypes (Figure 4.3.C-D). We found the relationship between field boundaries and datatypes is tightly coupled. ProtocolREs often use their hypotheses about the field boundaries to inform hypotheses about the datatypes, subsequently adjusting their field boundaries and continuously iterating cyclically. Several participants described how a small change to one field can lead to a cascade of other updates. P1 explained, *"For example, you adjusted this boundary now, and the delta in Field 2 became this much more variable across the whole data set also, probably per field."* Because these two aspects are so intercon-

nected, we observed all participants regularly switched between considering datatype and field boundary hypotheses as new patterns and ideas arose.

Participants want automation to integrate with other ProtocolREing systems and information (Figure 4.3.B,E).

Most participants expressed the importance of integrating multiple input and output streams. Early in the interviews, participants often wanted to do context research, like reviewing the binary and source code, comparing findings to established templates, or looking at documentation. P9 said, *“The truth is in the binary. If you can get access to the binary that you can reverse engineer, that is often the quickest way to deal with things.”*

Other participants described turning to documentation related to the application they were investigating that might reveal some information about the protocol. For example, P11 said, *“I would go look at the RFC [Request for Comments] ... to understand how the protocol is designed.”* Similarly, P5 described looking for patents, saying, *“Is there a patent on these people?’ so that this way I can find what they think they’re doing.”*

Several participants described transferring the results of their ProtocolREing efforts to other tools for testing or use in other applications. For example, P15 described their process of switching between multiple tools: *“If this is a lot of traffic, we filter and automate Scapy extraction of data ... Then we can change, and can state in a descriptive language like in Kaitai-struct a definition [of the protocol].”* As another example, P14 described wanting the automation to generate scripts usable by other packet processing tools, saying, *“I would like it to generate decoder code that fits the tool that I am using [Wireshark].”*

Participants emphasized the importance of message ordering by avoiding the provided filtering and sorting features.

We initially thought filtering and sorting features would be beneficial. However, participants had mixed reactions. Of the ten participants who interacted with these features, half felt they were not helpful, as filtering and sorting could disrupt the sequence and ordering of messages. P16 explained, *“Preserving the time of the messages and showing them in time order. It’s extremely important.”*

However, a few participants liked the idea of filters as they would help focus the view only on relevant messages. For example, P9 wanted to utilize filtering visually in the Histogram View to focus on messages within a certain value range, saying *“I’d want to be able to drag a region in the visualization and have it filter the rest of the data dynamically to only that data. It lets you immediately see that the field one values never change; that the field two values are somewhat periodic, and that the field three values never really change.”*

4.4.3 Information Needs From Automation (RQ2)

Given participants’ ProtocolREing process from the prior section, we now discuss themes in the types of additional information participants wanted when working with automation. In particular, we observed two general themes. First, participants often did not expect the automation to be exactly correct due to the ambiguity of the task; therefore, they wanted the automation to take a conservative approach and provide mechanisms for evaluating the outputs. Also, participants wanted ways to inspect and compare multiple suggestions by the automation.

4.4.3.1 ProtocolREs do not trust automation to provide a complete solution, but see value in automation as a support agent

When considering the suggestions, most participants doubted that any automation could be sufficiently accurate to correctly determine a full protocol due to challenges inherent to ProtocolREing. That is, finding samples to train or test ProtocolREing automation on, especially for malware analysis, is difficult due to the rapidly changing and adversarial nature of malware [127, 179]. Several participants shared a similar sentiment to P16, who said, “*Most of the time we’re dealing with new and novel stuff [that] isn’t in the [training] data set.*” Other participants believe current automation does not scale to complex real-world examples. For example, P14 said, “*They don’t work on real software. That’s why I say automated reverse engineering tools don’t really exist because they only really exist on paper.*”

Despite participants’ belief that automation would make mistakes when automatically identifying a protocol, most participants believed the suggestions were still helpful as a starting point for their analysis and to give them ideas for further analysis as an interactive support agent. For example, P12 said, “*the more options that the system gives and says, Hi, we think it is this... the more that the system says here is a possible suggestion the better.*”

Prior work investigating trust in other uses of automation has shown users initially distrust automation, though this reduces with use as they learn its boundaries [106, 156]. Several participants expressed a similar sentiment to P10 when using automation, “*If I’m using the tool and find that the suggestions often are helpful, Then over time my trust in it would develop.*” Similarly, P16 described learning an automated tool’s “*individual quirks, and if it’s suggestions are going to be helpful or not, and what kinds of things it’s better it’s suggesting.*”

Automation should make conservative field boundary and datatype edits. Most participants wanted automation only to edit the presented fields and data types when it was confident and present information with a minimal visual footprint otherwise. P16 said liberal field boundary and datatype edits place a burden on ProtocolREs, creating extra work, saying “*it’s just like 15 other things that you have to go through and like, minimize and click through.*” This is particularly problematic if users cannot

override suggestions or limit the scope of changes made by the automation. For example, P8 described how this would be frustrating, saying, *“If I assign one type, and then it goes and re-assigns types for other fields that I have already reviewed the suggestions for, it would mess me up.”*

Participants want semantic characteristics about the data as suggestion explanations. In addition to wanting the automation to provide conservative suggestions, most participants wanted ways to assess and understand those suggestions, then make refinements (which we discuss further in Section 4.4.4). P16 said, *“With automated tools like this, you kind of have to get some experience with them to figure out their individual quirks. And if it’s suggestions are going to be helpful or not, and what kinds of things it’s better at suggesting.”* To help them make these assessments, participants wanted the automation to justify its suggestions. P1 said, *“I’d want to understand what information is going into that prediction. Why is it picking that?”*

Providing these explanations for automation results is a common request beyond ProtocolREing [133, 82]. However, we further refine this need by observing ProtocolREs wanted these explanations presented as characteristics of the data fitting the suggested datatype’s range. For example, P15 wanted to know if a datatype suggestion was based on a relationship between two fields, explaining the automation should show *“the length of this field is controlled by this field over here.”* Similarly, P9 described a checksum relationship between two field, saying *“these 2 bytes of this data might be a checksum because they are a valid CRC16 with X,Y,Z parameters [fields].”* Additionally, many described looking for these datatype characteristics when inferring the datatypes manually. For example, most participants investigated whether a field’s data was restricted to the ASCII range, indicating it was an ASCII string, or monotonically increasing, likely a counter.

Several participants believed adding some indicator of the automation’s confidence in a result, based on these characteristics, would help users make assessments. P13 described this, saying, *“For any automated tool, I like to show confidence level. [What] does the tool really think percentage wise? This is 90%, it’s probably certain, is it 85? We’re not really confident, but we have a good guess. Is it 50? We’re just guessing out of the blue... That helps not having to go down a wild goose chase.”*

4.4.3.2 ProtocolREs perform analysis by comparing data both within single messages and across multiple messages

Due to the complex nature of ProtocolREing and the large number of analyzed messages, participants appreciated features that allowed different forms of comparison. The Combined View included allowing quick comparisons across all messages using the histogram, viewing multiple messages simultaneously, and seeing the data decoded into multiple datatypes simultaneously. P9 described their information needs saying, *“workflow-wise... [what] I’m used to is much more parallel... viewing the data in multiple*

ways at the same time.”

However, excess visual information can overload ProtocolREs; thus, several participants described wanting to hide parts of views once they were no longer needed. Cluttered visuals are a common challenge in interface design, but our results indicate when different information should be presented. For example, P16 felt the histogram was not needed at all times: *“Having the ability to hide them is useful because it’s something I’ll look at once, and then I never look at them again.”* Similarly, P13 described wanting to *“just ‘x’ [the float and integer windows] out, and they would leave and then say the Hex data or the ASCII field would automatically resize bigger.”* While participants discussed wanting to hide many of the views, no participant indicated they wanted to hide the Split-Merge View, likely because this interaction with fields and data types is central to ProtocolREs’ processes.

Participants mostly wanted to view each field boundary decoded according to its suggested datatype, not multiple datatypes.

To assess the validity of a suggestion, participants often focused on the datatype representation. This included wanting to see the data decoded as the specific datatypes on a per-field basis. For example, P6 said, *“If it could break the message, some of it being ASCII, some of it being the floats, some of it being an integer while still displaying each individual message broken up as a comparison, that would be more helpful.”* This allows ProtocolREs to visualize the parsed protocol data and assess whether the presented data matches the ranges they would expect to see. However, presenting the message decoded as each datatype simultaneously, as we did (See Figure 4.1.D), was viewed as overload for many participants. Instead, they preferred some drop-down option, allowing quick switching between decoding types for comparison instead of showing all at once. P2 explained, *“There is totally enough mental bandwidth to say, ‘click’, and then within a millisecond field four changes [to ASCII]. And then I could just click back to integer if it doesn’t help.”*

Participants wanted to know how datatype suggestions changed when field boundaries were moved.

As the automation provides different field boundary suggestions or the user splits or merges fields, this would cause the automation to suggest additional datatypes for newly formed fields. Some participants indicated they wanted a way to see each of these states of the field boundaries and compare how changes to the fields impact within-field variation and patterns. For example, P1 said, *“I want it to say, for example, you adjusted this boundary, now field 2 became this much more variable.”* This follows from the fact that we observed ProtocolREs regularly switching between assessing field boundaries and data types; thus, information helping them see aspects of this relationship was beneficial.

4.4.4 Interaction Needs With Automation (RQ3)

Finally, we discuss how participants wanted to interact with the automation. These include programmatically extending datatype descriptions, interacting with suggestions and reverting actions, and automatically generating scripts for future ProtocolREing process steps.

Participants wanted extendable datatype descriptions that consider multiple fields. In the Prototype Interface, we only showed a few basic datatypes, but participants discussed wanting the automation to identify a wide range of possible complex non-traditional datatypes like sequence counters or message identifiers. P14 wanted the automation to recognize non-traditional datatypes like hashes saying, *“It would save me time identifying what things are probably integrity checks like series or Md5 or check digits.”* Because some semantic types might be specific to the studied application, participants wanted the ability to write their own type definitions or provide something like C struct definition that could help the automation recognize these relationships in the data. This would also support alternative data encoding—a need expressed by almost all participants—, such as specific string encodings. P9 explained, *“I’d feel limited by ASCII. I know this is probably ASCII data, but what if it’s not a normal code page? what if it’s CJK encoding, which is common in China, or Shift-JIS encoding, which is common in Japan?”*

As discussed in Section 4.4.2, many participants attempted to determine datatypes by looking at relationships between fields, such as message type identifiers or checksums. A few participants reported these trends are typically harder for ProtocolREs to identify by hand; thus, they wanted additional feature support to help identify these relationships, which could be supported through extended datatype representations.

ProtocolREs want to guide automation by constraining field boundaries and datatypes.

As discussed previously, most participants expect the automation to make some mistakes due to the problem’s complexity. Therefore, most participants wanted to incrementally guide the automation by changing a field’s datatype or boundaries and forcing the automation to use those changes when determining future suggestions. P13 explained, *“Most automated testing a lot of times is incorrect...I know patterns because that’s something that I use all the time. Automated systems don’t always know these patterns, or they don’t know something that I’ve seen before, but they haven’t. And so, having that flexibility to change it and then see, does that correct it if I think something is wrong, or does that make it perform better?”*

ProtocolREs need the ability to switch back and forth between possible automation suggestions quickly. As we discussed in Section 4.4.2, ProtocolREs regularly switch back and forth between

field boundaries and datatypes, testing whether different possible protocol specifications fit their data. Therefore, when interacting with the automation, several participants emphasized the importance of quickly and easily reverting actions or configuration settings while working. For instance, they could test how changes in field boundaries or datatypes might impact their understanding of the protocol. Then, they can quickly revert or compare these changes to their original analysis. P7 described using a preview of changes to make a comparison, saying *“if [there’s] some way to split [fields] up and then undo it... preview how it would affect the guesses or the field values, so you split it at this point and then you can see the values of the [fields] and what the guess type for the new fields would look like.”*

Similarly, several participants wanted ways to maintain their mental models and understanding of the data. This can take numerous forms, from leaving in-line comments or sticky notes as P5 described: *“If I’m looking at this in three weeks, when I get pulled to something else; If I come back I have no [idea] when I figured out what this ASCII meant. I have no idea what it meant to me back then.”*

Participants wanted support for generating data to validate automated suggestions and their protocol inferences.

Many participants wanted a mechanism to provide additional input messages or generate output messages that could be validated with an external tool (Figure 4.3.E). Some participants wanted to manually craft new messages to see how specific changes would impact the automated suggestions. For example, P2 said, *“I would like to be able to add a new message and arbitrarily manipulated bytes, and see in real-time how it gets parsed as a float. That would easily allow me to test hypotheses about what this program is showing me.”*

Alternatively, half of the participants requested some form of script generation based on their inferred protocol definitions to use with other tools to assess the validity of their final protocol definition. For example, Wireshark and other packet parsing tools allow users to provide data decoders for new protocols, using these decoders to parse received packets. Multiple participants mentioned they wanted the automation to generate a decoder that could be applied to a packet parser and give it a large volume of packets. If the parser fails, this would indicate the protocol definition is incorrect. For example, P14 said, *“I’d want [the prototype] to be provable... I probably want it to be something that emits C code that then can handle decoding whatever packet capture you like, or maybe it emits a Wireshark dissector.”* Other participants proposed validating their inferred protocols by automatically producing code that would generate packets matching the inferred protocol. Then, these packets could be sent to a server that uses the studied protocol to see if it responds as expected. As P14 explained, *“I would generate packets and send them at real servers and see what happens. The live poking of things can be really useful as a lot of them will drop packets on the floor if they’re wrong or respond with an error.”* A few participants pointed out these scripts would also provide a dual use, as they could be directly used for future security tasks like intrusion detection systems.

Chapter 5

Discussion and Recommendations

Our results provide insights into demographic discrepancies in perceived belonging and community experiences among cybersecurity professionals, alongside overarching trends within the community. Additionally, based on our interviews with professional ProtocolREs, we observed the process model described in Figure 4.3, where ProtocolREs cyclically iterate between adjusting field boundaries and establishing datatypes for fields. These decisions are tightly coupled, and ProtocolREs incorporate information from various external sources, including context research about the application (Figure 4.3.B) and automated suggestions (Figure 4.3.A). Our participants investigated multiple hypotheses in parallel and preferred to view data in multiple formats simultaneously. Our participants were largely distrustful of automation due to low confidence in its performance for novel data; thus, many participants wanted firm control over how automation provided suggestions. Our participants valued the ability to extend the tools they used to support custom-defined datatypes or interface with other ProtocolREing tools they typically use.

We distill common themes from our results and propose actionable recommendations for community and organizational leaders to improve inclusivity and diversity

There is a clear gender disparity in community experiences. Across all research questions, the primary divide observed was between genders. Women are more likely to face harassment and unwelcoming experiences related to their identity. Likely due to these unwelcoming experiences, women were less likely to participate in general-interest security organizations, instead opting for identity-based groups, similar to anecdotes presented by Fulton et al. [75].

While we did not have a large enough sample of genderqueer participants to produce generalizable results, these participants' responses suggest they face an unwelcoming community. Across all our survey questions, genderqueer participants reported lower perceptions of belonging and higher experiences of unsupportive environments and identity-based harassment. Future work should focus on this group to better understand the unique challenges they face.

While these experiences mirrored examples described by interview participants in prior work [75], our work demonstrates a clear gender gap and indicates the scale of the problem.

Black participants’ responses suggest positive outcomes. Turning to differences between races/ethnicities, we only observed significant differences between Black and White participants. Black participants found community organizations more helpful to their development and career success and had higher vulnerability discovery self-efficacy. This is a positive indicator; however, we remain cautious on this finding as the number of Black participants in our sample was small (N=38) and our findings are only indicative of current cybersecurity professionals’ experiences, meaning survivor bias likely plays a role in this result. Further, we stress that while we considered differences in high-level demographic groups, the experiences of members of these groups are not monolithic. Further work is needed to confirm these results with a larger sample and assess the impact of intersectional identities.

Safety perceived as low across participants. Across all demographic groups, we observed low psychological safety scores when compared to results of prior surveys [77]. This lack of perceived safety to engage with others in the community could be internal (e.g., impostor syndrome or perceived lack of knowledge), but our measures of internal belonging and knowledge did not show a similar deficit. This suggests the perceived lack of safety is caused by external forces, which is supported by participants’ multiple instances of reported harassment experiences across demographic groups. These results suggest efforts to improve inclusivity and climate in the cybersecurity community would be universally beneficial.

This finding also points to a larger issue of survivor bias in our results. Prior work has shown people with low belonging uncertainty [205] and high self-efficacy [109] are better equipped to overcome negative external forces (like those we observed) because they have a strong internal view of self. Conversely, individuals without the same strong internal perception of self may not join, or remain in, the cybersecurity community due to these negative forces. While our results cannot make claims regarding the people excluded from the community, they point to a potentially high dropout rate and motivate future work investigating this problem.

More work is needed to ensure early cybersecurity education is inclusive and supportive. There have been significant efforts to increase early development programs for cybersecurity-interested students [94, 145, 30, 130, 120, 140, 148]. These programs are important, but our results suggest more work is necessary to investigate and improve their associated communities’ inclusivity. Our results indicate cybersecurity professionals who begin skill development early are more likely to face unsupportive environments.

We expected early engagement would lead to stronger perceptions of belonging. While higher vulnera-

bility discovery self-efficacy correlated with high-school programming experience, these early experiences also correlated with lower psychological safety and increased reports of severe harassment, particularly among marginalized groups. This likely contributes to higher dropout rates, emphasizing the need for welcoming and inclusive early education.

While we did not find direct evidence about this, we speculate women are more comfortable participating in gender-specific affinity groups than in general support groups, which may relate to the higher rate of severe harassment women reported, as supported by [75], which described participants avoiding certain groups due to negative experiences.

ProtocolREing differs from SoftwareREing in the distinction between generating and testing hypotheses.

While we expected there would be some similarity between ProtocolREing and SoftwareREing as they have similar goals and are often performed by the same people, there are key differences in process due to the inherent difference in the ambiguity of the underlying task, as discussed in Section 2.0.2. Comparing our process model (Figure 4.3) with three-phase SoftwareREing model defined in prior work [203, 25], they are both hypothesis-driven and iterative. However, the clear difference is that the ProtocolREing process is more condensed, meaning there are not clear distinctions between parts of the process as participants discuss performing hypothesis generation and testing using the same interfaces and sometimes simultaneously. Thus, we present a cyclic process centered on the dual problems of field and data-type identification, whereas prior work found the SoftwareREing model can be broken into three distinct phases SoftwareREs progress through and use different tools to support, dividing hypothesis generation and testing. This distinction creates unique needs for ProtocolREing automation and interaction, as ProtocolREs want to use automation as an interactive support agent, and cannot trust it to produce final answers.

5.0.1 Recommendations for Researchers

While this study provides direct recommendations for tool development, we can draw out valuable insights for researchers investigating ProtocolREing. First, we identify a process model for ProtocolREing, which can be used to guide the development of controlled empirical studies of the various tasks. Our study was conducted in an observed setting without a working tool, but future work should build on our foundation to test with a live tool and networking data. This would allow future research to answer questions about the most effective common practices and strategies and what interfaces help save time and streamline the process for ProtocolREs. Similarly, we enumerated the set of ProtocolREs' information and interaction needs, but our study design prevents us from identifying which would be the most useful. Developers must make tradeoffs with limited development time, especially in this domain where many tools and plugins are produced by the open source community in their own time [119], so assessing actual feature

value is necessary.

The information and interaction needs we identified should be assessed with additional populations and settings. First, because we only test with expert ProtocolREs, whether these requirements hold for beginners is not apparent. Repeating our study directly with beginners will likely not be sufficient as beginners likely lack the experience to envision what they might be missing. However, given the additional information and interactions, a future study could assess whether these are helpful for beginners. This would be most useful in defining initial universally usable default views. Finally, future work should investigate how these processes change in alternate settings, specifically when users can access application source code or binary files. Work in this setting will provide a bridge between our work and the prior SoftwareREing literature [202, 204, 115, 29, 70, 19, 25], likely unearthing unique insights at their intersection.

5.0.2 Recommendations for ProtocolREing Tool Developers

While ProtocolREing differs from SoftwareREing due to the ambiguity of the analysis, it is similar to problems in data science where analysts visualize and consider data records in different ways to reveal relationships. We saw this play out with the information and interaction needs identified (Sections 4.4.3 and 4.4.4, respectively) as participants wanted to see the data presented in many different ways, wanted the automation to provide explanations for suggestions, wanted the ability to validate and correct the automation, and wanted to get this information without creating information overload. These are common challenges when using automation to support visual data exploration tasks, thus these similarities suggest existing guidelines from this domain [87, 149, 211, 195, 55] could be leveraged for future automation design. However, our results provide additional insights regarding how these can be best explicitly applied to support ProtocolREs. Specifically, we observed two overarching themes for interface design based on participants' responses. We discuss each in turn.

Organize interactions around datatypes and field boundary characteristics. Throughout our interviews, we observed participants' investigations centered around the unique semantic characteristics of different datatypes and restrictions on possible field boundaries. These components were central to their processes, and shaped their interaction needs. For example, participants specifically liked the Histogram View because it allowed them to determine whether data in a field matched the known bounds of particular datatypes. However, because the view did not allow them to view inter-field relationships, they wanted it extended to capture richer datatype representations. Additionally, the only information we did not hear anyone say they wanted to hide was the field boundaries and datatype labels, further indicating the centrality of this information. Therefore, any ProtocolREing interaction should be designed around datatypes and field boundaries.

Focus interfaces on supporting iterative investigation. The second theme arises from the fact that our participants expected automation to make erroneous suggestions but still saw these tools as helpful in supporting their process of quickly testing variations on potential protocols. They quickly move through the search space, swapping a datatype or field boundary, and want automation to take a similar approach, conservatively suggesting changes to parts of the protocol definition when confident. To support this iterative search, automation should enable easy comparison by neighbor points in the search space by allowing analysts to preview or undo a suggestion and visualize the impact of a change.

5.0.3 Community Leaders Recommendations

Our results indicate a need to improve the culture in cybersecurity to make it more safe and inclusive for everyone, especially women/gender minorities and early career cybersecurity professionals. To this end, we draw on existing best practices from prior work in psychological safety evaluated in other domains [61]. For each best practice, we discuss potential adoption strategies, noting that while these practices were designed for structured workplaces, not all cybersecurity organizations fit this mold (e.g., conference communities, online forums). However, we discuss how the ideas of these practices can be leveraged in less structured environments.

Set the stage. The first step to establish a safer, more inclusive community is for leaders to emphasize safety and inclusivity's value and clearly frame participation in cybersecurity as open to all. This step's goal is to set a shared expectation and vision. For example, #ShareTheMicInCyber promotes the stories and accomplishments of Black cybersecurity professionals, highlighting the impact of their work on the field [176]. Additionally, all major security conferences have established codes of conduct [146, 101, 1, 2, 3, 4] and many have adopted diversity and inclusion statements [49, 5, 6, 7] extolling the importance of a welcoming community, indicating goals for inclusivity, and establishes that hate and harassment that will not be tolerated. While there has been a significant increase in this stage setting recently, it is important that these messages are repeated regularly and within all subcommunities.

Invite participation. While setting the stage is important for creating a shared vision in the community, it is not as meaningful if action is not taken to foster inclusivity. Action is not only the responsibility of leaders, but all members because parts of the cybersecurity community lack clear leaders and structure. Unfortunately, the low psychological safety observed suggests cybersecurity professionals may be less likely to stand up as allies. To counter this issue, community leaders should provide training and support that encourage being an ally and bystander intervention [16], more empathetic responses, and a transition away from a victim-blaming.

Responding productively. Finally, it is paramount that cybersecurity professionals experience actual safety when participating in the community through demonstrated support. The most important practice here is to deter clear instances of hate and harassment, especially targeting women and genderqueer cybersecurity professionals, as those were seen as most prevalent in our results. This response requires transparent and clear guidelines to avoid silencing expression. Our results indicate cybersecurity professionals experience hate and harassment that crosses a clear boundary, according to most existing policies [101, 49, 1, 2, 6, 5, 3, 146, 7, 4], so it is important these actions are sanctioned publicly to demonstrate the community’s commitment to inclusivity. In some subcommunities with less structure, sanctions may be harder to employ, as there is limited central control. These cases demonstrate, again, the need to develop a broad culture of inclusion among community members. For example, it may not be possible for moderators to effectively ban offending users on an anonymous site, as these users can just create new accounts. Allies, instead, might respond with support for the victim and make it clear the offenders’ views are not acceptable or representative of the subcommunity.

Our results also suggest destigmatizing mistakes for beginners, especially in early development phases. Beginners may ask easily searchable questions, which can seem frustrating for overworked security educators, but should not be met with disdain [199]. Instead, using resources like FAQs and detailed walkthroughs can help. Questions that persist should be addressed with care, as having someone reliable to ask is crucial for development. Practicing empathy and patience is vital, as most cybersecurity professionals experience some insecurity.

Bibliography

- [1] Code of conduct. <https://defcon.org/html/links/dc-code-of-conduct.html>. (Accessed 02-13-2024).
- [2] Code of conduct. <https://www.blackhat.com/code-of-conduct.html>. (Accessed 02-13-2024).
- [3] Code of conduct. <https://www.ndss-symposium.org/ndss-code-of-conduct/>. (Accessed 02-13-2024).
- [4] Code of conduct. <https://www.usenix.org/conferences/coc>. (Accessed 02-13-2024).
- [5] Diversity and inclusion. <https://www.ieee.org/about/diversity-index.html>. (Accessed 02-13-2024).
- [6] Diversity and inclusion. <https://www.blackhat.com/html/sustainability.html>. (Accessed 02-13-2024).
- [7] Diversity and inclusion. <https://www.usenix.org/conferences/diversity-and-inclusion>. (Accessed 02-13-2024).
- [8] Kulsoom Abdullah, Christopher P Lee, Gregory J Conti, John A Copeland, and John T Stasko. Ids rainstorm: Visualizing ids alarms. In *VizSEC*, page 1, 2005.
- [9] M K Ahuja. Women in the information technology profession: a literature review, synthesis and research agenda. *European Journal of Information Systems*, 11(1):20–34, 2002. **arXiv:**<https://doi.org/10.1057/palgrave.ejis.3000417>, doi:10.1057/palgrave.ejis.3000417.
- [10] Omer Akgul, Taha Egthesad, Amit Elazari, Omprakash Gnawali, Jens Grossklags, Michelle L. Mazurek, Aron Laszka, and Daniel Votipka. Bug hunters’ perspectives on the challenges and benefits of the bug bounty ecosystem. In *32nd USENIX Security Symposium*, USENIX Sec ’23, Los Angeles, CA, August 2022. USENIX Association.
- [11] Miltiadis Allamanis, Earl T. Barr, Christian Bird, and Charles Sutton. Suggesting accurate method and class names. In *2015 Joint Meeting on Foundations of Software Engineering*, ESEC/FSE ’15,

- page 38–49, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2786805.2786849.
- [12] AllsafeCyberSecurity. Awesome ghidra, 2021. (Accessed 08-11-2021). URL: <https://github.com/AllsafeCyberSecurity/awesome-ghidra>.
 - [13] Uri Alon, Meital Zilberstein, Omer Levy, and Eran Yahav. Code2vec: Learning distributed representations of code. *Proc. ACM Program. Lang.*, 3(POPL), January 2019. doi:10.1145/3290353.
 - [14] Anonymous. An investigation of interaction and information needs for protocol reverse engineering automation. https://osf.io/ernux/?view_only=ba228218aed04a4eb35d1e947f2330de, Sep 2024.
 - [15] Catherine Ashcraft, Elizabeth Eger, and Michelle Friend. Girls in it: The facts. Technical report, National Center for Women & Information Technology, 2012.
 - [16] American Psychological Association. Bystander intervention tip sheet. <https://www.apa.org/pi/health-equity/bystander-intervention>. (Accessed 02-13-2024).
 - [17] Dejan Baca, Bengt Carlsson, Kai Petersen, and Lars Lundberg. Improving software security with static automated code analysis in an industry setting. *Software: Practice and Experience*, 43(3):259–279, 2013. URL: <http://dblp.uni-trier.de/db/journals/spe/spe43.html#BacaCPL13>.
 - [18] Markus Baer and Michael Frese. Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior*, 24(1):45–68, 2003. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/job.179>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.179>, doi:10.1002/job.179.
 - [19] Steffen Becker, Carina Wiesen, Nils Albartus, Nikol Rummel, and Christof Paar. An exploratory study of hardware reverse engineering — technical and cognitive processes. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*, pages 285–300. USENIX Association, August 2020. URL: <https://www.usenix.org/conference/soups2020/presentation/becker>.
 - [20] Benjamin Bichsel, Veselin Raychev, Petar Tsankov, and Martin Vechev. Statistical deobfuscation of android applications. In *2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16*, page 343–355, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2976749.2978422.
 - [21] Joseph Biden. Executive order on improving the nation’s cybersecurity, May 2021. (Accessed 07-21-2021). URL: <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/>.

- [22] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [23] Boeing. Boeing technical apprenticeship. <https://jobs.boeing.com/btap>. (Accessed 01-05-2024).
- [24] Amiangshu Bosu and Kazi Zakia Sultana. Diversity and inclusion in open source software (oss) projects: Where do we stand? In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 1–11, 2019. doi:10.1109/ESEM.2019.8870179.
- [25] Adam Bryant. *Understanding How Reverse Engineers Make Sense of Programs from Assembly Language Representations*. PhD thesis, US Air Force Institute of Technology, 01 2012.
- [26] Bugcrowd. The state of bug bounty. Technical report, Bugcrowd, June 2016. URL: <https://pages.bugcrowd.com/2016-state-of-bug-bounty-report>.
- [27] BugCrowd. Inside the mind of a hacker 2020, 2020. (Accessed 07-21-2020). URL: <https://itmoah.bugcrowd.com>.
- [28] Bugcrowd. Inside the mind of a hacker. <https://www.bugcrowd.com/wp-content/uploads/2023/11/Inside-the-Mind-of-a-Hacker.pdf>, 2023. Accessed: 2024-08-17.
- [29] Kevin Burk, Fabio Pagani, Christopher Kruegel, and Giovanni Vigna. Decomperson: How humans decompile and what we can learn from it. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2765–2782, Boston, MA, August 2022. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity22/presentation/burk>.
- [30] Tanner J. Burns, Samuel C. Rios, Thomas K. Jordan, Qijun Gu, and Trevor 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)*, Vancouver, BC, August 2017. USENIX Association. URL: <https://www.usenix.org/conference/ase17/workshop-program/presentation/burns>.
- [31] A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.
- [32] Janis A Cannon-Bowers and Eduardo Ed Salas. *Making decisions under stress: Implications for individual and team training*. American psychological association, 1998.
- [33] Sang-Mi Chai and Min-Kyun Kim. A road to retain cybersecurity professionals: An examination of career decisions among cybersecurity scholars. *Journal of the Korea Institute of Information Security & Cryptology*, 22(2):295–316, 2012. doi:10.13089/JKIISC.2012.22.2.295.

- [34] Nina Chamblou. Diversity in cybersecurity. <https://www.cyberdegrees.org/resources/diversity-in-cybersecurity/>, 2022.
- [35] Jared Chandler, Adam Wick, and Kathleen Fisher. Binaryinferno: A semantic-driven approach to field inference for binary message formats. In *NDSS*, 2023.
- [36] Sapna Cheryan, Benjamin J. Drury, and Marissa Vichayapai. Enduring influence of stereotypical computer science role models on women’s academic aspirations. *Psychology of Women Quarterly*, 37(1):72–79, 2013. arXiv:<https://doi.org/10.1177/0361684312459328>, doi:10.1177/0361684312459328.
- [37] Sapna Cheryan, Allison Master, and Andrew N Meltzoff. Cultural stereotypes as gatekeepers: Increasing girls’ interest in computer science and engineering by diversifying stereotypes. *Frontiers in psychology*, 6:49, 2015.
- [38] Sapna Cheryan, Victoria C Plaut, Paul G Davies, and Claude M Steele. Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology*, 97(6):1045, 2009.
- [39] Sapna Cheryan, John Oliver Siy, Marissa Vichayapai, Benjamin J Drury, and Saenam Kim. Do female and male role models who embody stem stereotypes hinder women’s anticipated success in stem? *Social Psychological and Personality Science*, 2(6):656–664, 2011.
- [40] CISA. Cisa awards \$2 million to bring cybersecurity training to rural communities and diverse populations. URL: <https://www.cisa.gov/news/2021/10/20/cisa-awards-2-million-bring-cybersecurity-training-rural-communities-and-diverse>.
- [41] Debbie Clayton and Teresa Lynch. Ten years of strategies to increase participation of women in computing programs: The central queensland university experience: 1999–2001. *SIGCSE Bull.*, 34(2):89–93, jun 2002. doi:10.1145/543812.543838.
- [42] David A Cole, Elizabeth A Nick, Rachel L Zelkowitz, Kathryn M Roeder, and Tawny Spinelli. Online social support for young people: does it recapitulate in-person social support; can it help? *Computers in human behavior*, 68:456–464, 2017.
- [43] Christopher J Collins and Ken G Smith. Knowledge exchange and combination: The role of human resource practices in the performance of high-technology firms. *Academy of management journal*, 49(3):544–560, 2006.
- [44] Daniel Collins. Rehcx, 2024. (Accessed 08-11-2024). URL: <https://github.com/solemnwarning/rehex/>.

- [45] Greg Conti. *Security data visualization: graphical techniques for network analysis*. No Starch Press, 2007.
- [46] Gregory Conti and Kulsoom Abdullah. Passive visual fingerprinting of network attack tools. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 45–54, 2004.
- [47] Gregory Conti, Erik Dean, Matthew Sinda, and Benjamin Sangster. Visual reverse engineering of binary and data files. In *International Workshop on Visualization for Computer Security*, pages 1–17. Springer, 2008.
- [48] Joel Cooper. The digital divide: The special case of gender. *Journal of computer assisted learning*, 22(5):320–334, 2006.
- [49] Cas Cremers and Engin Kirda. Diversity and inclusion. <https://www.sigsac.org/ccs/CCS2024/code-of-conduct/diversity-and-inclusion.html>. (Accessed 02-02-2024).
- [50] DARPA. Darpa — cyber grand challenge, 2016. (Accessed 02-26-2017). URL: <http://archive.darpa.mil/cybergrandchallenge/>.
- [51] Robert Gordon Kent de Grey. *Friends in High-Tech Places: The Development and Validation of the Online Social Support Measure*. PhD thesis, The University of Utah, 2018.
- [52] Anne Deiglmayr, Elsbeth Stern, and Renate Schubert. Beliefs in “brilliance” and belonging uncertainty in male and female stem students. *Frontiers in psychology*, 10:442252, 2019.
- [53] James R Detert and Ethan R Burris. Leadership behavior and employee voice: Is the door really open? *Academy of management journal*, 50(4):869–884, 2007.
- [54] Alicia Dietsch. Winner of the at&t diversity and inclusion champion award 2021. URL: <https://cybersecurity.att.com/blogs/security-essentials/winner-of-the-att-diversity-and-inclusion-champion-award-2021>.
- [55] Kyriaki Dimitriadou, Olga Papaemmanouil, and Yanlei Diao. Explore-by-example: An automatic query steering framework for interactive data exploration. In *SIGMOD ’14*, pages 517–528, New York, NY, USA, 2014. ACM. URL: <http://doi.acm.org/10.1145/2588555.2610523>, doi:10.1145/2588555.2610523.
- [56] Adam Doupé, Marco Cova, and Giovanni Vigna. Why johnny can’t pentest: An analysis of black-box web vulnerability scanners. In *Proceedings of the 7th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, DIMVA’10*, pages 111–131, Berlin, Heidelberg, 2010. Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=1884848.1884858>.

- [57] Michelle Drolet. Diversity in cybersecurity: Barriers and opportunities for women and minorities. <https://www.csoononline.com/article/571811/diversity-in-cybersecurity-barriers-and-opportunities-for-women-and-minorities.html>, 2021.
- [58] Julien Duchêne, Colas Le Guernic, Eric Alata, Vincent Nicomette, and Mohamed Kaâniche. State of the art of network protocol reverse engineering tools. *Journal of Computer Virology and Hacking Techniques*, 14:53–68, 2018.
- [59] Michael H. Dunn and Laurence D. Merkle. Assessing the impact of a national cybersecurity competition on students’ career interests. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education, SIGCSE ’18*, page 62–67, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3159450.3159462.
- [60] Amy Edmondson. Psychological safety and learning behavior in work teams. *Administrative science quarterly*, 44(2):350–383, 1999.
- [61] Amy C Edmondson. *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. John Wiley & Sons, 2018.
- [62] Rachel Fairbank. Psychologists persevere in ed work despite growing backlash against racial equity efforts. <https://www.apa.org/monitor/2024/01/trends-anti-equity-diversity-inclusion-laws>, 2024. (Accessed 02-08-2024).
- [63] Daniel Fallman. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 225–232, 2003.
- [64] Ronald A Fisher. On the interpretation of χ^2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.
- [65] Lisa Y Flores, Rachel L Navarro, and S Joseph DeWitz. Mexican american high school students’ postsecondary educational goals: Applying social cognitive career theory. *Journal of Career Assessment*, 16(4):489–501, 2008.
- [66] Lisa Y Flores, Rachel L Navarro, Hang Shim Lee, Dorothy A Addae, Rebecca Gonzalez, Laura L Luna, Ricardo Jacquez, Sonya Cooper, and Martha Mitchell. Academic satisfaction among latino/a and white men and women engineering students. *Journal of Counseling Psychology*, 61(1):81, 2014.
- [67] Cynthia E Foor, Susan E Walden, and Deborah A Trytten. “i wish that i belonged more in this whole engineering group:” achieving individual diversity. *Journal of Engineering Education*, 96(2):103–115, 2007.

- [68] National Initiative for Cybersecurity Careers and Studies. Cybersecurity education and training assistant program. <https://niccs.cisa.gov/cybersecurity-career-resources/cybersecurity-education-and-training-assistance-program>. (Accessed 01-05-2024).
- [69] National Initiative for Cybersecurity Education. National initiative for cybersecurity education cybersecurity workforce framework. <https://www.cisa.gov/national-initiative-cybersecurity-education-nice-cybersecurity-workforce-framework>. (Accessed 01-05-2024).
- [70] Irina Ford, Ananta Soneji, Faris Bugra Kokulu, Jayakrishna Vadayath, Zion Leonahenahe Basque, Gaurav Vipat, Adam Doupé, Ruoyu Wang, Gail-Joon Ahn, Tiffany Bao, and Yan Shoshitaishvili. "Watching over the shoulder of a professional": Why Hackers Make Mistakes and How They Fix Them. In *Proceedings of the IEEE Symposium on Security and Privacy*, May 2024.
- [71] Nadya A Fouad and Mercedes C Santana. Sctt and underrepresented populations in stem fields: Moving the needle. *Journal of Career Assessment*, 25(1):24–39, 2017.
- [72] Wireshark Foundation. Wireshark, the world's most popular network protocol analyzer, 2024. (Accessed 08-11-2024). URL: <https://www.wireshark.org/>.
- [73] Karl Pearson F.R.S. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50(302):157–175, 1900.
- [74] Kelsey R Fulton, Samantha Katcher, Kevin Song, Marshini Chetty, Michelle L Mazurek, Chloé Messdaghi, and Daniel Votipka. Vulnerability discovery for all: Experiences of marginalization in vulnerability discovery. *Proc. of the IEEE*, 2023.
- [75] Kelsey R. Fulton, Samantha Katcher, Kevin Song, Marshini Chetty, Michelle L. Mazurek, Chloé Messdaghi, and Daniel Votipka. Vulnerability discovery for all: Experiences of marginalization in vulnerability discovery. In *Proceedings of the 44th IEEE Symposium on Security and Privacy*, IEEE S&P '23, 2023.
- [76] Kelsey R. Fulton, Daniel Votipka, Desiree Abrokwa, Michelle L. Mazurek, Michael Hicks, and James Parker. Understanding the how and the why: Exploring secure development practices through a course competition. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [77] David A Garvin, Amy C Edmondson, and Francesca Gino. Is yours a learning organization? *Harvard business review*, 86(3):109, 2008.

- [78] gchq. Cyberchef, 2024. (Accessed 08-11-2024). URL: <https://github.com/gchq/CyberChef>.
- [79] Thomas Gilovich, Dacher Keltner, and Richard E Nisbett. Being a member of a stigmatized group: stereotype threat. *Gilovich, Thomas; Keltner, Dacher; Nisbett, Richard E., Social psychology, New York: WW Norton*, pages 467–468, 2006.
- [80] John W. Graham and Steven A. Smith. Gender differences in employment and earnings in science and engineering in the us. *Economics of Education Review*, 24(3):341–354, 2005. URL: <https://www.sciencedirect.com/science/article/pii/S0272775704001104>, doi:10.1016/j.econedurev.2004.06.005.
- [81] Robert G Kent de Grey, Bert N. Uchino, Brian RW Baucom, Timothy W. Smith, Avery E. Holton, and Edward F. Diener. Enemies and friends in high-tech places: the development and validation of the online social experiences measure. *Digital Health*, 5(1), 2019.
- [82] Wenbo Guo, Dongliang Mu, Jun Xu, Purui Su, Gang Wang, and Xinyu Xing. Lemna: Explaining deep learning based security applications. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS '18*, page 364–379, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3243734.3243792.
- [83] Hackerone. 2016 bug bounty hacker report. Technical report, Hackerone, San Francisco, California, September 2016. URL: <https://hackerone.com/blog/bug-bounty-hacker-report-2016>.
- [84] Hackerone. 2019 hacker-powered security report. Technical report, Hackerone, San Francisco, California, December 2019. URL: <https://www.hackerone.com/resources/reporting/the-hacker-powered-security-report-2019>.
- [85] Hackerone. The 2020 hacker report. Technical report, Hackerone, San Francisco, California, December 2020.
- [86] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007. URL: <http://dx.doi.org/10.1080/19312450709336664>.
- [87] Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, April 2012. URL: <http://doi.acm.org/10.1145/2133806.2133821>, doi:10.1145/2133806.2133821.
- [88] Hex-Rays. Plug-in contest 2018: Hall of fame, 2019. (Accessed 05-30-2019). URL: <https://www.hex-rays.com/contests/2018/index.shtml>.
- [89] Hexinator. Hexinator, 2024. (Accessed 08-11-2024). URL: <https://hexinator.com/>.

- [90] Eoin Hinchy. Voice of the soc analyst. <https://www.tines.com/reports/voice-of-the-soc-analyst>, 2022.
- [91] (ICS2)². Cybersecurity workforce study. Technical report, (ICS2)², August 2018. URL: www.isc2.org/-/media/ISC2/Research/2018-ISC2-Cybersecurity-Workforce-Study.ashx.
- [92] ImHex. Imhex, 2024. (Accessed 08-11-2024). URL: <https://imhex.werwolv.net/>.
- [93] Ohio Cyber Range Institute. Ohio cyber range institute: Unlocking potential, securing the future. <https://www.ohiocyberrangeinstitute.org/>. (Accessed 01-06-2024).
- [94] SANS Cybersecurity Institute. Girls go cyberstart. (Accessed 05-27-2020). URL: <https://girlsgocyberstart.org/>.
- [95] ISC2. Isc2 cybersecurity workforce study 2023. https://media.isc2.org/-/media/Project/ISC2/Main/Media/documents/research/ISC2_Cybersecurity_Workforce_Study_2023.pdf?rev=28b46de71ce24e6ab7705f6e3da8637e, 2023.
- [96] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3):405–431, 09 2012. arXiv:<https://academic.oup.com/poq/article-pdf/76/3/405/17120500/nfs038.pdf>, doi:10.1093/poq/nfs038.
- [97] Alan Jaffe, Jeremy Lacomis, Edward J. Schwartz, Claire Le Goues, and Bogdan Vasilescu. Meaningful variable names for decompiled code: A machine translation approach. In *2018 Conference on Program Comprehension, ICPC '18*, page 20–30, New York, NY, USA, 2018. Association for Computing Machinery. doi:10.1145/3196321.3196330.
- [98] Kaitai-Project. Kaitai, 2015. (Accessed 08-11-2021). URL: <https://kaitai.io/>.
- [99] Harjot Kaur, Sabrina Amft, Daniel Votipka, Yasemin Acar, and Sascha Fahl. Where to recruit for security development studies: Comparing six software developer samples. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 4041–4058, 2022.
- [100] Robert G Kent de Grey, Bert N Uchino, Brian RW Baucom, Timothy W Smith, Avery E Holton, and Edward F Diener. Enemies and friends in high-tech places: the development and validation of the online social experiences measure. *Digital Health*, 5:2055207619878351, 2019.
- [101] Engin Kirda and David Lie. Code of conduct. <https://www.sigsac.org/ccs/CCS2024/code-of-conduct/code-of-conduct.html>. (Accessed 02-02-2024).

- [102] Solomon Klappholz. A third of cyber security pros report crumbling work-life balance. <https://www.itpro.com/security/a-third-of-cyber-security-pros-report-crumbling-work-life-balance>, 2023.
- [103] Stephan Kleber, Henning Kopp, and Frank Kargl. NEMESYS: Network message syntax reverse engineering by analysis of the intrinsic structure of individual messages. In *12th USENIX Workshop on Offensive Technologies (WOOT 18)*, Baltimore, MD, August 2018. USENIX Association. URL: <https://www.usenix.org/conference/woot18/presentation/kleber>.
- [104] Stephan Kleber, Lisa Maile, and Frank Kargl. Survey of protocol reverse engineering algorithms: Decomposition of tools for static traffic analysis. *IEEE Communications Surveys & Tutorials*, 21(1):526–561, 2018.
- [105] Gary A Klein. A recognition-primed decision (rpd) model of rapid decision making. *Decision making in action: Models and methods*, 5(4):138–147, 1993.
- [106] Jan H. Klemmer, Stefan Albert Horstmann, and Nikhil Patnaik. Using ai assistants in software development: A qualitative study on security practices and concerns. <https://synthical.com/article/b790c6f0-7738-4c20-b2fe-a6308b3135a6>, 4 2024. arXiv:2405.06371.
- [107] Andrew J Ko. Attitudes and self-efficacy in young adults’ computing autobiographies. In *2009 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 67–74. IEEE, 2009.
- [108] Jeremy Lacomis, Pengcheng Yin, Edward Schwartz, Miltiadis Allamanis, Claire Le Goues, Graham Neubig, and Bogdan Vasilescu. Dire: A neural approach to decompiled identifier naming. In *2019 IEEE/ACM International Conference on Automated Software Engineering, ASE ’19*, pages 628–639, 2019. doi:10.1109/ASE.2019.00064.
- [109] Robert W. Lent, Steven D. Brown, and Gail Hackett. Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45(1):79 – 122, 1994. URL: <http://www.sciencedirect.com/science/article/pii/S000187918471027X>, doi:10.1006/jvbe.1994.1027.
- [110] James Andrew Lewis. The cybersecurity workforce gap, January 2019. (Accessed 08-06-2021). URL: <https://www.csis.org/analysis/cybersecurity-workforce-gap>.
- [111] Jian Liang, Crystal I. C. Farh, and Jiing-Lih Farh. Psychological antecedents of promotive and prohibitive voice: A two-wave examination. *Academy of Management Journal*, 55(1):71–92, 2012. arXiv:<https://doi.org/10.5465/amj.2010.0176>, doi:10.5465/amj.2010.0176.

- [112] Falk Lieder, Thomas L Griffiths, Quentin J M. Huys, and Noah D Goodman. The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review*, 25:322–349, 2018.
- [113] Xin Luo, Dan Chen, Yongjun Wang, and Peidai Xie. A type-aware approach to message clustering for protocol reverse engineering. *Sensors*, 19(3):716, 2019.
- [114] Thomas Maillart, Mingyi Zhao, Jens Grossklags, and John Chuang. Given enough eyeballs, all bugs are shallow? revisiting eric raymond with bug bounty programs. In *2016 Workshop on the Economics of Information Security, WEIS '16*, 2016.
- [115] Alessandro Mantovani, Simone Aonzo, Yanick Fratantonio, and Davide Balzarotti. RE-Mind: a first look inside the mind of a reverse engineer. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2727–2745, Boston, MA, August 2022. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity22/presentation/mantovani>.
- [116] Jane Margolis, Rachel Estrella, Joanna Goode, Jennifer Jellison Holme, and Kim Nao. *Stuck in the shallow end: Education, race, and computing*. MIT press, 2017.
- [117] Jane Margolis and Allan Fisher. *Unlocking the clubhouse: Women in computing*. MIT press, Cambridge, MA, 2002.
- [118] James Mattei, Madeline McLaughlin, Samantha Katcher, and Daniel Votipka. A qualitative evaluation of reverse engineering tool usability. In *Proceedings of the 2022 Annual Computer Security Applications Conference, ACSAC '22*, New York, NY, USA, 2022. Association for Computing Machinery.
- [119] James Mattei, Madeline McLaughlin, Samantha Katcher, and Daniel Votipka. A qualitative evaluation of reverse engineering tool usability. In *Proceedings of the 2022 Annual Computer Security Applications Conference, ACSAC '22*, New York, NY, USA, 2022. Association for Computing Machinery.
- [120] Luke McCormack. Us cyber challenge, September 2022. (Accessed 09-17-2022). URL: <https://www.uscyberchallenge.org/>.
- [121] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019.
- [122] Gary McGraw and John Steven. Software [in]security: Comparing apples, oranges, and aardvarks (or, all static analysis tools are not created equal). Technical report, Cigital, 2011. (Accessed 02-26-2017).

- [123] Robert K McKinley, Terjinder Manku-Scott, Adrian M Hastings, David P French, and Richard Baker. Reliability and validity of a new measure of patient satisfaction with out of hours primary medical care in the united kingdom: development of a patient questionnaire. *Bmj*, 314(7075):193, 1997.
- [124] Adam W Meade and S Bartholomew Craig. Identifying careless responses in survey data. *Psychological methods*, 17(3):437, 2012.
- [125] Uta Menges, Jonas Hielscher, Annalina Buckmann, Annette Kluge, M Angela Sasse, and Imogen Verret. Why it security needs therapy. In *European Symposium on Research in Computer Security*, pages 335–356. Springer, 2021.
- [126] Matthew B Miles and A Michael Huberman. *Qualitative data analysis: An expanded sourcebook*. sage, 1994.
- [127] Jaron Mink, Hadjer Benkraouda, Limin Yang, Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, and Gang Wang. Everybody’s Got ML, Tell Me What Else You Have: Practitioners’ Perception of ML-Based Security Tools and Explanations.
- [128] Rachel Minkin. Personal experiences with online harassment. <https://www.pewresearch.org/social-trends/2023/05/17/diversity-equity-and-inclusion-in-the-workplace/>, 2023. (Accessed 02-08-2024).
- [129] Sam Mintz. Boston cyber training company gets \$1 million from dhs for diversity efforts. URL: <https://www.bizjournals.com/boston/inno/stories/news/2021/11/01/boston-cyber-training-company-gets-1m-from-dhs.html>.
- [130] Jelena Mirkovic and Peter A. H. Peterson. Class capture-the-flag exercises. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, San Diego, CA, August 2014. USENIX Association. URL: <https://www.usenix.org/conference/3gse14/summit-program/presentation/mirkovic>.
- [131] Phoenix Moorman and Elizabeth Johnson. Still a stranger here: Attitudes among secondary school students towards computer science. In *Proceedings of the 8th Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE ’03*, page 193–197, New York, NY, USA, 2003. Association for Computing Machinery. doi:10.1145/961511.961564.
- [132] Laurie A. Morgan. Glass-ceiling effect or cohort effect? a longitudinal study of the gender earnings gap for engineers, 1982 to 1989. *American Sociological Review*, 63(4):479–493, 1998. URL: <http://www.jstor.org/stable/2657263>.

- [133] Azqa Nadeem, Daniël Vos, Clinton Cao, Luca Pajola, Simon Dieck, Robert Baumgartner, and Sico Verwer. Sok: Explainable machine learning for computer security applications, 2023. URL: <https://arxiv.org/abs/2208.10605>, arXiv:2208.10605.
- [134] Dawn Nafus. ‘patches don’t have gender’: What is not open in open source software. *New Media & Society*, 14(4):669–683, 2012. arXiv:<https://doi.org/10.1177/1461444811422887>, doi:10.1177/1461444811422887.
- [135] John Narayan, Sandeep K Shukla, and T Charles Clancy. A survey of automatic protocol reverse engineering tools. *ACM Computing Surveys (CSUR)*, 48(3):1–26, 2015.
- [136] Rachel L Navarro, Lisa Y Flores, Hang-Shim Lee, and Rebecca Gonzalez. Testing a longitudinal social cognitive model of intended persistence with engineering students across gender and race/ethnicity. *Journal of Vocational Behavior*, 85(1):146–155, 2014.
- [137] Netzob. Netzob, 2024. (Accessed 08-11-2024). URL: <https://github.com/netzob/netzob?tab=readme-ov-file>.
- [138] Elizabeth A Nick, David A Cole, Sun-Joo Cho, Darcy K Smith, T Grace Carter, and Rachel L Zelkowitz. The online social support scale: measure development and validation. *Psychological assessment*, 30(9):1127, 2018.
- [139] SH Nielsen, LA von Hellens, and Sharon Wong. The male it domain: You’ve got to be in it to win it. In *Proceedings of the 12th Australasian Conference on Information Systems (ACIS 2001)*, pages 1–12, 2001.
- [140] NIST. Cybersecurity competitions — nist. <https://www.nist.gov/itl/applied-cybersecurity/nice/community/community-coordinating-council/cybersecurity-skills-0>, 2023. (Accessed 02-08-2024).
- [141] Samantha Nix and Lara Perez-Felkner. Difficulty orientations, gender, and race/ethnicity: An intersectional analysis of pathways to stem degrees. *Social Sciences*, 8(2), 2019. URL: <https://www.mdpi.com/2076-0760/8/2/43>, doi:10.3390/socsci8020043.
- [142] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.
- [143] Timothy Nosco, Jared Ziegler, Zechariah Clark, Davy Marrero, Todd Finkler, Andrew Barbarello, and W. Michael Petullo. The industrial age of hacking. In *2020 USENIX Security Symposium*, USENIX Security ’20, pages 1129–1146. USENIX Association, August 2020. URL: <https://www.usenix.org/conference/usenixsecurity20/presentation/nosco>.

- [144] US Department of Homeland Security. Secretary mayorkas announces most successful cybersecurity hiring initiative in dhs history, July 2021. (Accessed 08-06-2021). URL: <https://www.dhs.gov/news/2021/07/01/secretary-mayorkas-announces-most-successful-cybersecurity-hiring-initiative-dhs>.
- [145] Plaid Parliament of Pwning. picoctf. (Accessed 05-27-2020). URL: <https://picoctf.com/>.
- [146] IEEE Symposium on Security and Privacy. Code of conduct. <https://www.ndss-symposium.org/ndss-code-of-conduct/>. (Accessed 02-13-2024).
- [147] Zhen Xin Ong, Liz Dowthwaite, Elvira Perez Vallejos, Mat Rawsthorne, and Yunfei Long. Measuring online wellbeing: a scoping review of subjective wellbeing measures. *Frontiers in psychology*, 12:616637, 2021.
- [148] PACTF. Pactf. (Accessed 05-27-2020). URL: <https://2019.pactf.com/>.
- [149] Adam Perer and Ben Shneiderman. Systematic yet flexible discovery: Guiding domain experts through exploratory data analysis. In *IUI '08*, pages 109–118, New York, NY, USA, 2008. ACM. URL: <http://doi.acm.org/10.1145/1378773.1378788>, doi:10.1145/1378773.1378788.
- [150] Toni C Plato. Women c-suite executives in cybersecurity: Transformational experiences and gender barriers on their leadership journeys. 2021.
- [151] Stephan Plöger, Mischa Meier, and Matthew Smith. A qualitative usability evaluation of the clang static analyzer and libfuzzer with CS students and CTF players. In *2021 Symposium on Usable Privacy and Security*, SOUPS '21, pages 553–572. USENIX Association, August 2021.
- [152] Portia Pusey, Mark Gondree, and Zachary Peterson. The outcomes of cybersecurity competitions and implications for underrepresented populations. *IEEE Security & Privacy*, 14(6):90–95, 2016. doi:10.1109/MSP.2016.119.
- [153] Laura Quintana. Hack your way to a new career in cybersecurity: Cisco networking academy offers new ethical hacker course. <https://blogs.cisco.com/learning/hack-your-way-to-a-new-career-in-cybersecurity>, October 2023. (Accessed 01-05-2024).
- [154] Jessica Rafael. How three stem leaders forged a path for others through fearlessness. *US Black Engineer and Information Technology*, 44(1):22–25, 2020.
- [155] Adrian E Raftery. Bayesian model selection in social research. *Sociological methodology*, pages 111–163, 1995.

- [156] Neele Roch, Hannah Sievers, Lorin Schöni, and Verena Zimmermann. Navigating autonomy: Unveiling security experts' perspectives on augmented intelligence in cybersecurity. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 41–60, Philadelphia, PA, August 2024. USENIX Association. URL: <https://www.usenix.org/conference/soups2024/presentation/roch>.
- [157] Laurie A Rudman. Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, 74(3):629, 1998.
- [158] David L Sackett. Bias in analytic research. In *The case-control study consensus and controversy*, pages 51–63. Elsevier, 1979.
- [159] Robert J Sampson. Local friendship ties and community attachment in mass society: A multilevel systemic model. *American sociological review*, pages 766–779, 1988.
- [160] Koen Schoenmakers, Daniel Greene, Sarah Stutterheim, Herbert Lin, and Megan J Palmer. The security mindset: characteristics, development, and consequences. *Journal of Cybersecurity*, 9(1):tyad010, 2023.
- [161] Elaine Seymour and Nancy M. Hewitt. *Talking About Leaving: Why Undergraduates Leave the Sciences*. Westview Press, 2000.
- [162] Jenessa R. Shapiro and Amy M. Williams. The role of stereotype threats in undermining girls' and women's performance and interest in stem fields. *Journal on Sex Roles*, 66(3):175–183, 2012. doi:10.1007/s11199-011-0051-0.
- [163] Ben Shneiderman. *Designing The User Interface: Strategies for Effective Human-Computer Interaction*, 4/e (New Edition). Pearson Education India, 1987.
- [164] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pages 364–371. Elsevier, 2003.
- [165] Yan Shoshitaishvili, Michael Weissbacher, Lukas Dresel, Christopher Salls, Ruoyu Wang, Christopher Kruegel, and Giovanni Vigna. Rise of the hacrs: Augmenting autonomous cyber reasoning systems with human assistance. In *Proc. of the 24th ACM SIGSAC Conference on Computer and Communications Security, CCS '17*. ACM, 2017.
- [166] Enno Siemsen, Aleda V Roth, Sridhar Balasubramanian, and Gopesh Anand. The influence of psychological safety and confidence in knowledge on employee knowledge sharing. *Manufacturing & service operations management*, 11(3):429–447, 2009.

- [167] Romila Singh, Nadya A Fouad, Mary E Fitzpatrick, Jane P Liu, Kevin J Cappaert, and Catia Figueredo. Stemming the tide: Predicting women engineers' intentions to leave. *Journal of Vocational Behavior*, 83(3):281–294, 2013.
- [168] Daryl G Smith. *Diversity's Promise for Higher Education: Making It Work*. JHU Press, 2020.
- [169] Carolyn Snyder. *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann, 2003.
- [170] Elliott Sober. Instrumentalism, parsimony, and the akaike framework. *Philosophy of Science*, 69(S3):S112–S123, 2002. doi:10.1086/341839.
- [171] Nemesys Software. Nemesys, 2024. (Accessed 08-11-2024). URL: <https://www.nemesys-soft.com/>.
- [172] Sweetscape Software. 010 editor, 2024. (Accessed 08-11-2024). URL: <https://www.sweetscape.com/010editor/>.
- [173] Robert R Sokal and Charles D Michener. A statistical method for evaluating systematic relationships. 1958.
- [174] David M. Sparks, Steve Daniel Przymus, Allison Silveus, Yohanis De La Fuente, and Cassandra Cartmill. Navigating the intersectionality of race/ethnicity, culture, and gender identity as an aspiring latina stem student. *Journal of Latinos and Education*, 22(4):1355–1371, 2023. doi:10.1080/15348431.2021.1958332.
- [175] Claude M. Steele, Steven J. Spencer, and Joshua Aronson. Contending with group image: The psychology of stereotype and social identity threat. volume 34 of *Advances in Experimental Social Psychology*, pages 379–440. Academic Press, 2002. URL: <https://www.sciencedirect.com/science/article/pii/S0065260102800090>, doi:10.1016/S0065-2601(02)80009-0.
- [176] Camille Stewart, Lauren Zabierek, and Katelyn Ringrose. #sharethemicincyber. <https://www.sharethemicincyber.com/>, 2021.
- [177] Margaret-Anne Storey, Alexey Zagalsky, Fernando Figueira Filho, Leif Singer, and Daniel M German. How social and communication channels shape and challenge a participatory culture in software development. *IEEE Transactions on Software Engineering*, 43(2):185–204, 2016.
- [178] Anselm Strauss and Juliet Corbin. *Basics of qualitative research*, volume 15. Newbury Park, CA: Sage, 1990.

- [179] Octavian Suciu, Connor Nelson, Zhuoer Lyu, Tiffany Bao, and Tudor Dumitras. Expected exploitability: Predicting the development of functional vulnerability exploits. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 377–394, Boston, MA, August 2022. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity22/presentation/suciu>.
- [180] Supplemental materials. <https://osf.io/k3m8g>. doi:DOI10.17605/OSF.IO/KUCBA.
- [181] Larry Suto. Analyzing the accuracy and time costs of web application security scanners. *San Francisco, February*, 1, 2010.
- [182] Synack. Synack cybersecurity diversity and inclusion report. <https://www.synack.com/diversity-report/>, 2020.
- [183] Synalysis. Synalysis, 2024. (Accessed 08-11-2024). URL: <https://synalysis.com/>.
- [184] Mohammad Tahaei, Ruba Abu-Salma, and Awais Rashid. Stuck in the permissions with you: Developer & end-user perspectives on app permissions & their privacy ramifications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–24, 2023.
- [185] Mohammad Tahaei and Kami Vaniea. Recruiting participants with programming skills: A comparison of four crowdsourcing platforms and a cs student mailing list. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.
- [186] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Raineart, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science*, 3:e111, 2017.
- [187] Steven Terrell and Kembley Lingelbach. A study of female cybersecurity professionals. *Issues in Information Systems*, 24(3), 2023.
- [188] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, Damon McCoy, Sarah Meiklejohn, Thomas Ristenpart, and Gianluca Stringhini. Sok: Hate, harassment, and the changing landscape of online abuse. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 247–267, 2021. doi:10.1109/SP40001.2021.00028.
- [189] Jenifer Tidwell. *Designing interfaces: Patterns for effective interaction design*. ” O’Reilly Media, Inc.”, 2010.
- [190] Eileen M Trauth. Mapping information-sector work to the work force. *Communications of the ACM*, 44(7):74–75, 2001.

- [191] Thomas Trevethan and Grace Williams. The time is now to secure the future. <https://www.paloaltonetworks.com/blog/2023/10/secure-the-future/>, October 2023. (Accessed 01-06-2024).
- [192] Chih-Hsiung Tu. The measurement of social presence in an online learning environment. In *International Journal on E-learning*, volume 1, pages 34–45. Association for the Advancement of Computing in Education (AACE), 2002.
- [193] Sherry Turkle. *The Second Self: Computers and the Human Spirit*. Mit Press, 1984.
- [194] Visa University. Visa payments cybersecurity certification. <https://www.visauniversity.com/en/certificate-programs/visa-certification/visa-payments-cybersecurity-certification>. (Accessed 01-06-2024).
- [195] Manasi Vartak, Sajjadur Rahman, Samuel Madden, Aditya Parameswaran, and Neoklis Polyzotis. Seedb: Efficient data-driven visualization recommendations to support visual analytics. *VLDB Endowment*, 8(13):2182–2193, September 2015. doi:10.14778/2831360.2831371.
- [196] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark G.J. van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. Gender and tenure diversity in github teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, page 3789–3798, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2702123.2702549.
- [197] Vector35. Vector35/community-plugins, 2019. (Accessed 05-30-2019). URL: <https://github.com/Vector35/community-plugins/tree/master/plugins>.
- [198] Emily A. Vogels. Personal experiences with online harassment. <https://www.pewresearch.org/internet/2021/01/13/personal-experiences-with-online-harassment/>, 2021. (Accessed 01-06-2024).
- [199] D. Votipka, E. Zhang, and M. Mazurek. Hacked: A pedagogical analysis of online vulnerability discovery exercises. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1589–1606, Los Alamitos, CA, USA, may 2021. IEEE Computer Society. URL: <https://doi.ieeecomputersociety.org/10.1109/SP40001.2021.00092>, doi:10.1109/SP40001.2021.00092.
- [200] Daniel Votipka, Desiree Abrokwa, and Michelle L. Mazurek. Building and validating a scale for secure software development self-efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, page 1–20, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3313831.3376754.

- [201] Daniel Votipka, Hongyi Hu, Bryan Eastes, and Michelle L. Mazurek. Toward a field study on the impact of hacking competitions on secure development. In *Proceedings of the 4th Workshop on Security Information Workers*, WSIW '18, Baltimore, MD, 2018. USENIX Association. URL: <https://wsiw2018.l3s.uni-hannover.de/papers/wsiw2018-Votipka.pdf>.
- [202] Daniel Votipka, Seth Rabin, Kristopher Micinski, Jeffrey S. Foster, and Michelle L. Mazurek. An observational investigation of reverse engineers' processes. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 1875–1892. USENIX Association, August 2020. URL: <https://www.usenix.org/conference/usenixsecurity20/presentation/votipka-observational>.
- [203] Daniel Votipka, Seth Rabin, Kristopher Micinski, Jeffrey S. Foster, and Michelle L. Mazurek. An observational investigation of reverse engineers' processes. In *29th USENIX Security Symposium (USENIX Security 20)*, Boston, MA, August 2020. USENIX Association. URL: <https://www.usenix.org/conference/usenixsecurity20/presentation/votipka-observational>.
- [204] Daniel Votipka, Rock Stevens, Elissa M Redmiles, Jeremy Hu, and Michelle L Mazurek. Hackers vs. testers: A comparison of software vulnerability discovery processes. In *Proceedings of the 39th IEEE Symposium on Security and Privacy*, IEEE S&P '18, 2018.
- [205] Gregory M Walton and Geoffrey L Cohen. A question of belonging: race, social fit, and achievement. *Journal of personality and social psychology*, 92(1):82, 2007.
- [206] Miranda Wei, Pardis Emami-Naeini, Franziska Roesner, and Tadayoshi Kohno. Skilled or gullible? gender stereotypes related to computer security and privacy. In *IEEE Symposium on Security and Privacy*, 2023.
- [207] Elizabeth Williams. Actualizing gender and racial diversity inclusion in computing fields. *Issues in Information Systems*, 24(4), 2023.
- [208] K. Yakdan, S. Dechand, E. Gerhards-Padilla, and M. Smith. Helping johnny to analyze malware: A usability-optimized decompiler and malware analysis user study. In *IEEE S&P '16*, pages 158–177, May 2016. doi:10.1109/SP.2016.18.
- [209] Khaled Yakdan, Sergej Dechand, Elmar Gerhards-Padilla, and Matthew Smith. Helping johnny to analyze malware: A usability-optimized decompiler and malware analysis user study. *2016 IEEE Symposium on Security and Privacy (SP)*, 00:158–177, 2016. doi:doi.ieeecomputersociety.org/10.1109/SP.2016.18.
- [210] Yapeng Ye, Zhuo Zhang, Fei Wang, Xiangyu Zhang, and Dongyan Xu. Netplier: Probabilistic network protocol reverse engineering from message traces. 2021.

- [211] Ji Soo Yi, Youn ah Kang, John Stasko, and J.A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007. doi:10.1109/TVCG.2007.70515.
- [212] Danielle M. Young, Laurie A. Rudman, Helen M. Buettner, and Meghan C. McLean. The influence of female role models on women’s implicit science cognitions. *Psychology of Women Quarterly*, 37(3):283–292, 2013. arXiv:<https://doi.org/10.1177/0361684313482109>, doi:10.1177/0361684313482109.
- [213] Tianyi Zhang, Zhiyang Chen, Yuanli Zhu, Priyan Vaithilingam, Xinyu Wang, and Elena L. Glassman. *Interpretable Program Synthesis*. Association for Computing Machinery, New York, NY, USA, 2021.
- [214] Tianyi Zhang, London Lowmanstone, Xinyu Wang, and Elena L. Glassman. Interactive program synthesis by augmented examples. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST ’20, page 627–648, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3379337.3415900.

Appendix A

Survey

Helpfulness Scale. Extremely helpful, Somewhat helpful, Neither helpful nor unhelpful, Somewhat unhelpful, Extremely unhelpful

Agreement Scale. Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree

A.1 General Belonging

This section asks about your experiences in the cybersecurity community and any support in these settings you have been provided toward your security education and career.

For the following statements, please indicate the extent to which they reflect your experience in the cybersecurity community. *1 (Very inaccurate), 2, 3, 4 (Neither accurate nor inaccurate), 5, 6, 7 (Very accurate)*

1. In cybersecurity spaces, it is easy to speak up about what is on your mind.
2. If you make a mistake in the cybersecurity community, it is often held against you.
3. People in cybersecurity are usually comfortable talking about problems and disagreements.
4. People in cybersecurity are eager to share information about what doesn't work as well as share information about what does work.
5. Keeping your cards close to your chest is the best way to get ahead in the cybersecurity community.

Please indicate your level of agreement with the following statements: *Agreement Scale*

1. Sometimes I feel that I belong in cybersecurity, and sometimes I feel that I don't belong.

2. When something bad happens, I feel that maybe I don't belong in cybersecurity.
3. When something good happens, I feel that I really belong in cybersecurity.
4. People from different backgrounds have equal opportunities to participate in the cybersecurity community.

For the following items, think about your interactions with your professors/peers/colleagues in the computer security community. To respond, indicate to what extent you felt this way. *Very Slightly or Not at all, A Little, Moderately, Quite a Bit, Extremely*

1. Interactions with someone in the field prevented me from working on my goals or other important things.
2. Someone in the cybersecurity community has encouraged me when I felt like quitting.
3. I have felt supported by someone in the cybersecurity community who agreed with my point of view.
4. I have been unable to fall asleep while thinking about a negative interaction I had with someone in the cybersecurity community.
5. There are people in the cybersecurity community please ignore the first part of this statement and mark "Extremely".
6. Someone in the cybersecurity community has cheered me up when I was feeling down.
7. Someone in the cybersecurity community has made me feel embarrassed or foolish.
8. There is someone in the cybersecurity community I can turn to for advice about handling problems.
9. There is someone in the cybersecurity community I could turn to for advice about making career plans or about changing my job.

Have you ever experienced any of the following behaviors directed at you in the context of the cybersecurity community? *Never, Rarely, Occasionally, Frequently*

1. Lack of response or rejection of contributions or questions.
2. Conflict or interpersonal tension between you and another community member.
3. Written or spoken language that made you feel unwelcome (e.g. profanity, racist jokes, sexual imagery, hostility, rudeness, name calling).
4. Stereotyping based on perceived demographic characteristics.

5. Threats of violence, stalking.
6. Unsolicited sexual advance or comments.
7. Impersonation or malicious publication of personal information (doxxing).

A.2 Close Relations

This section asks about your experience with family, friends and other close mentors, and any support they have provided toward your security education and career.

1. Do you have an family, friends or other close mentors you go to for help when you're trying to learn difficult security concepts? *Yes, one; Yes, a few; Yes, many; No*
2. How often do you discuss technical topics related to your security education and career with your family/friends/mentors? *Never, Rarely, Occasionally, Frequently*
3. How helpful do you find the security-related guidance that your family/friends/mentors give you?
Helpfulness scale

A.3 Workplace

This section asks about your experiences in your workplace and any support in these settings you have been provided toward your security education and career.

1. Are you a part of any workplaces where security concepts are discussed? *Yes, but security concepts are rarely discussed; Yes, security concepts are sometimes discussed; Yes, security concepts are often discussed; No, but I was previously employed in security; No*
2. What is the primary focus of the company you work for? *Non-technical - critical infrastructure (hospitals, power, etc.), Non-technical - non-critical infrastructure, Security - defense (intrusion detection/response, system defense/hardening), Security - offense (penetration testing, vulnerability analysis), Other technical - software development, Other technical - network/system administration, Other technical - hardware development, Other*
3. How often do you discuss technical topics related to your security education and career in your workplace? *Never, Rarely, Occasionally, Frequently*
4. How helpful do you find the security-related guidance given to you by colleagues in your workplace?
Helpfulness scale

A.4 Organizations

This section asks about your experience with any organizations you participate in and any support you have received toward your security education or career. We consider an organization as any group outside your work/classes where people meet regularly to discuss technical topics of interest (e.g., local ACM chapter, Women in Security and Privacy). Please answer the following questions only considering security-related organizations.

1. Are you a part of any organizations where security concepts are discussed? *Yes, I participate in one organization where security concepts are discussed; Yes, I participate in a few organizations where security concepts are discussed; Yes, I participate in many organizations where security concepts are discussed; No*
2. What kinds of organizations are you a part of? *“Identity-based (e.g. Women in Security, LGBTQ+ in Security, Blacks in Cyber)”, “Topic-based (e.g. malware analysis working group)”, “General security group”, “Other”*
3. How often do you discuss technical topics related to your security education and career with people in these organizations? *Never, Rarely, Occasionally, Frequently*
4. How helpful do you find the security-related guidance given to you by people in organizations you are a part of? *Helpfulness scale*

A.5 School

This section asks about your experiences in your school and any support in these settings you have been provided toward your security education and career. This section only pertains to academic situations, e.g., classes, professors, peers, organizations.

1. Have you taken any classes where security concepts are discussed? *Yes, one; Yes, a few; Yes, many; No*
2. Are you a part of any school organizations where security concepts are discussed? *Yes, I participate in one organization where security concepts are discussed; Yes, I participate in a few organization where security concepts are discussed; Yes, I participate in many organization where security concepts are discussed; No*
3. How often do you discuss technical topics related to your security education in your school? *Never, Rarely, Occasionally, Frequently*
4. How helpful do you find the security-related guidance given to you by professors and peers? *Helpfulness scale*

A.6 Broader Security Community

This section asks about your experience with the broader security community, including conferences/workshops or online when asking questions about or discussing computer security topics. We define online community discussions as any discussion about security concepts in a public online forum (e.g., StackOverflow, Reddit, Twitter, public Slack or Discord).

1. Have you participated in the broader security community (conferences/workshops or online security communities)?

‘Yes, I participate in one public conference or online community; Yes, I participate in a few public conferences or online communities; Yes, I participate in many public conferences or online communities; No

2. Please indicate any security conferences or workshops you have participated in.
3. Please indicate any forums or social media platforms you use for interacting with the public online security community.
4. How often do you discuss technical topics related to your security education and career with people in the broader security community? *Never, Rarely, Occasionally, Frequently*
5. How helpful do you find the security-related guidance given to you by people in the broader security community? *Helpfulness scale*

A.7 General

We have asked several questions pertaining to your experiences with various communities surrounding you—close contacts, school/workplaces, organizations, and the broader security community. For the next section, we’ll ask you to consider all the experiences you’ve had with others in the security community. For both questions, we ask that you do not name specific individuals.

1. Please describe one particularly good experience you had with a community you are part of (and mention which community—close contacts, school/workplaces, organizations, or the broader security community). This could be any experience where you felt the other individual was supportive and helped your development or career in a tangible or intangible way.
2. Please describe one particularly bad experience you had with a community you are part of (and mention which community—close contacts, school/workplaces, organizations, or the broader security community). This could be any experience where you felt the other individual was not

supportive and the interaction was harmful to your development or career in a tangible or intangible way.

Now we will ask some questions pertaining to your background and experience in computer security.

1. Do you work in a role where you are asked to perform security tasks? *Yes, this is the primary focus of my job; Yes, this is a part of my job, but not the primary focus; I previously worked in a role where security was my primary focus; I previously worked in a role where security was part of the job; No*
2. Please indicate the approximate number of years you have worked in security.
3. What is/was your position title?
4. Did you choose to leave? *Yes, No*
5. If you feel comfortable sharing, what were your reasons for leaving?

Please indicate your level of agreement with the following statements: *Agreement Scale*

1. I am interested in continuing my security education.
2. I am interested in pursuing or continuing to pursue a career in security.
3. I am well-prepared for a career in security.

Please indicate:

1. Have you participated in any of the following types of security education? (Select all that apply) *Capture-the-flag, wargames, or other online security competitions (e.g., picoCTF, crackmes, iCTF), Penetration testing lab (e.g., Hack the box) or cyber range exercise, Professional certification course (e.g., GIAC Security Essentials, Certified Ethical Hacker), Conference workshop (e.g., Defcon Village workshops), MOOC security course (e.g., Coursera Cybersecurity Specialization), Academic course, Other, I have not participated in any security education*
2. Where did you typically rank when participating in CTFs or other online security competitions? *Top 25% of participants, 25-50% of participants, 50-75% of participants, Bottom 25% of participants*

Please indicate how confident you are in the following statements: *Not confident at all, Slightly confident, Somewhat confident, Moderately confident, Absolutely confident*

1. I can perform a threat risk analysis (e.g., likelihood of vulnerability, impact of exploitation)
2. I can identify potential security threats to the system

3. I can identify the common attack techniques used by attackers
4. I can identify potential attack vectors in the environment the system interacts with (e.g., hardware, libraries)
5. I can identify common vulnerabilities of a programming language
6. I can identify the common please ignore this question and select "Absolutely confident"
7. I can design software to quarantine an attacker if a vulnerability is exploited
8. I can mimic potential threats to the system
9. I can evaluate security controls on the system's interfaces/interactions with other software systems
10. I can evaluate security controls on the system's interfaces/interactions with hardware systems

Cybersecurity development experiences

1. When was the earliest time you remember first being interested in computer security? Please indicate your approximate age: *(Number)*
2. When was the earliest time you had someone (e.g., friend, family member, colleague) in your life who you could go to for security education support? Please indicate your approximate age: *(Number)*
3. Aside from direct educational support, do you have anyone (e.g., friend, family member, colleague) who support your educational pursuits in security (e.g., encouragement, monetary support)? *Yes, I have one person who has provided non-educational support; Yes, I have a few people who have provided non-educational support; Yes, I have many people who have provided non-educational support; No*
4. When was the earliest time you had someone (e.g., friend, family member, colleague) in your life who supported your pursuit of a security education aside from direct teaching? Please indicate your approximate age: *(Number)*

A.8 Demographics

1. In which country do you currently reside?
2. How old are you? *18-19, 20-24, 25-29, 30-35, 35-39, 40-44 , 45-49, 50-54, 55+, Prefer not to answer*

3. What is your ethnicity? *White or of European descent, South Asian, Hispanic or Latino/a/x, Middle Eastern , East Asian, Black or of African descent, Southeast Asian, Indigenous (such as Native American, Pacific Islander, or Indigenous Australian), Prefer to self-describe, Prefer not to answer*
4. What is your gender? *Woman, Man, Transgender Woman / Trans Feminine, Transgender Man / Trans Masculine, Non-Binary / Genderqueer / Gender Fluid, Two Spirit, Prefer to state, Prefer not to answer*
5. What is your sexual orientation? Do you identify as: *Bisexual, Gay/Lesbian, Heterosexual/Straight, Don't know, Prefer to self-describe, Prefer not to say*
6. What is the highest degree or level of school you have completed? *High school, Some college or currently enrolled, Associate's degree, Bachelor's degree, Master's/ Professional degree, Doctorate degree, Prefer not to say*
7. Did you take any programming classes or training in high school? *Yes, one; Yes, a few; Yes, many; No*
8. Which range matches most closely your total, pre-tax household income over the last fiscal year? *< \$29,999 , \$30,000 - \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 - \$124,999 , \$125,000 - \$149,999, \$150,000 - \$174,999 , \$175,000 - \$199,999 , > \$200,000, Prefer not to answer*
9. Which range matches most closely your total, pre-tax household income when growing up? (before 18 years old)? *Same as Question 8*

A.9 Final

1. If you like, we may contact you for one of the following reasons. Please indicate what you like to be contacted for (you may select multiple): *Follow-up interview (i.e., questions related to this study); Future research (i.e., questions related to other computer security topics); Raffle for one of 25 \$50 Amazon gift cards; None of the above*
2. Please provide your email address so we can contact you for the reasons selected previously. If you chose to only be contacted for the raffle, your email address will be deleted after the raffle has been completed. Your email will not be used for any purpose beyond those you indicated in the previous question.

Appendix B

Additional Figures and Tables

We provide tables and figures excluded from the main text for brevity. This includes a summary of participants responses regarding belonging uncertainty (Figure B.1), social experiences (Figure B.2), vulnerability discovery self-efficacy (Figure B.3), frequency of participating in discussion with different subcommunities (Figure B.4), the regression results for OSEM scores (Table B.1), the qualitative codebook derived from the open responses about where participants described their experiences (Table B.2), as well as the types of positive (Table B.3) and negative experiences (Table B.4).

Variable	Value	LE	CI	<i>p</i>
Gender	Man	–	–	–
	Woman	0.9	[0.9, 1.1]	0.610
Close Hepful	False	–	–	–
	True	0.8	[0.8, 0.9]	0.001*
– Base case (Log Estimate (LE)=1, by definition)				
*Significant effect				

Table B.1: Summary of regression over participant OSEM scores. The Pseudo R^2 (corrected Aldrich-Nelson) for the OSEM model was 0.11.

Code	Definition
Community at large	The broader collective of individuals engaged in the cybersecurity field, representing a diverse range of backgrounds and expertise.
School / Work	Settings where cybersecurity professionals either receive education and training or actively apply their skills within an organizational context.
Specific community member	An individual within the cybersecurity community, such as a mentor, peer, or leader, who plays a distinct role in providing support, guidance, or expertise.
Social media	Online platforms where cybersecurity professionals engage in discussions, share information, and collaborate with peers, expanding their networks beyond traditional face-to-face interactions.
Technical capability	The proficiency and expertise of cybersecurity professionals in understanding and applying technical concepts, tools, and methodologies to address security challenges.
Group meeting or Event	Scheduled gatherings or conferences within the cybersecurity community where professionals convene to exchange knowledge, discuss emerging trends, and network with peers.
Authority figure	Individuals within the cybersecurity community who hold positions of influence or expertise, such as mentors, industry leaders, or organizational executives, and provide guidance and direction to others.

Table B.2: Qualitative codes derived from open responses about where participants had positive experiences.

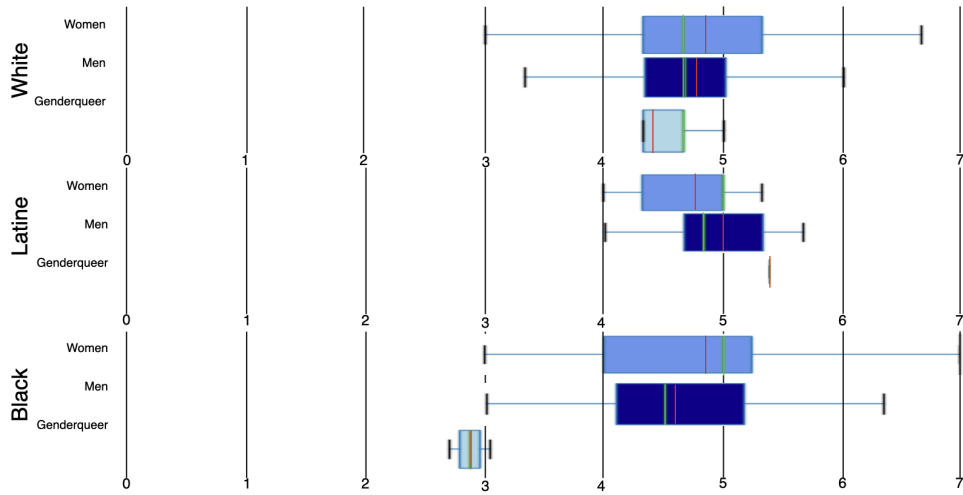


Figure B.1: Belonging and uncertainty scores from participants across gender and ethnicities.

In this section, we give the full definition for the different entries in our codebook.

Code	Definition
Financial support	Monetary assistance provided to cybersecurity professionals for educational, research, or career advancement purposes within the cybersecurity community.
Career support	Guidance, mentorship, and resources offered to cybersecurity professionals to aid in their professional development, job advancement, and career growth within the cybersecurity community.
Support outside security	Assistance, resources, or guidance offered to cybersecurity professionals from sources external to the cybersecurity community, such as friends and family.
Positive interaction	Constructive, collaborative, and supportive engagements between cybersecurity professionals that foster mutual respect, learning, and growth within the community.
Group work impact	The collective effect of collaborative efforts and teamwork among cybersecurity professionals on project outcomes, problem-solving, and innovation within the cybersecurity community.
Learning opportunities	Educational experiences, training programs, workshops, and resources available to cybersecurity professionals to enhance their knowledge, skills, and expertise in the cybersecurity community.
No good experience	Instances where cybersecurity professionals did not report any positive experiences.
COVID	The impact of the COVID-19 pandemic on the cybersecurity community.

Table B.3: Qualitative codes derived from open responses about the type of positive experiences participants described.

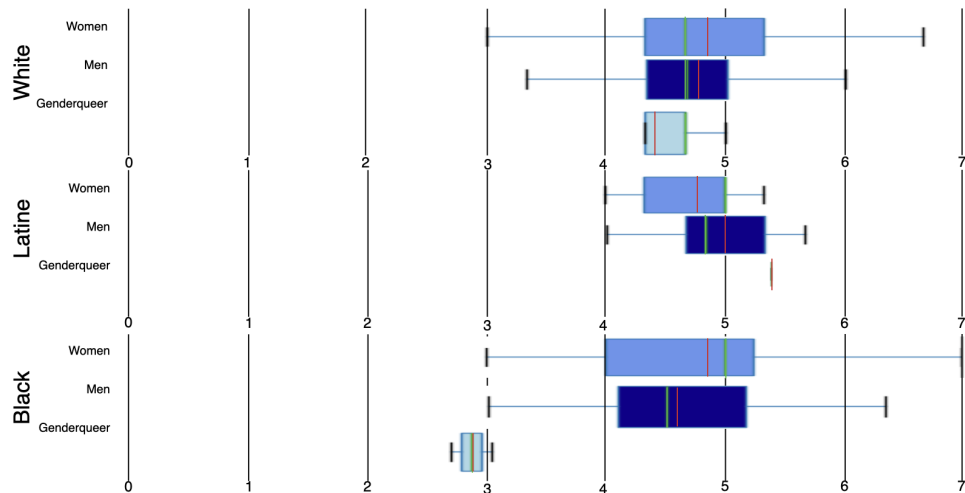


Figure B.2: Modified Online Social Experience Measure scores from participants across gender and ethnicities.

Code	Definition
Negative response from identity	Adverse reactions or treatment received by cybersecurity professionals based on their identity, such as gender, race, or ethnicity, within the cybersecurity community.
Unsupportive environment	A work or social setting within the cybersecurity community that did not provide adequate support, resources, or encouragement for growth and success.
Challenges getting started	Difficulties and obstacles encountered by individuals when initially entering the cybersecurity communities, including barriers to access, lack of mentorship, or limited opportunities.
Unwelcoming actions	Behaviors or actions within the cybersecurity community that convey hostility, exclusion, or discrimination towards certain individuals or groups.
Harassment	Persistent and unwanted behaviors, such as verbal abuse, intimidation, or threats, directed towards cybersecurity professionals, creating a hostile or intimidating environment.
Imposter syndrome	Feelings of self-doubt, inadequacy, or fear of being exposed as a fraud experienced by cybersecurity professionals, despite evidence of their competence and accomplishments.
Lack of recognition	A failure within the cybersecurity community to acknowledge or appreciate the contributions, achievements, or expertise of certain individuals, leading to feelings of undervaluation or invisibility.
Unwelcoming to newcomers	Attitudes, practices, or behaviors within the cybersecurity community that deter or discourage new entrants from fully participating or integrating into the community, hindering their transition and integration.
Difficulty with retention	Challenges faced by organizations or communities in retaining cybersecurity professionals, leading to high turnover rates, talent loss, and instability within the workforce.
No bad experience	Instances where cybersecurity professionals did not report any of negative or adverse experiences.
Social media	Online platforms and networks used by cybersecurity professionals for communication, collaboration, information sharing, and networking purposes, facilitating connections and knowledge exchange within the community.
Frauds and scams	Deceptive or fraudulent activities targeting cybersecurity professionals.
Difficulty collaborating	Challenges encountered by cybersecurity professionals when attempting to work together or collaborate effectively on projects, including communication barriers, conflicting priorities, or resource constraints.

Table B.4: Qualitative codes derived from open responses about the type of negative experiences participants described.

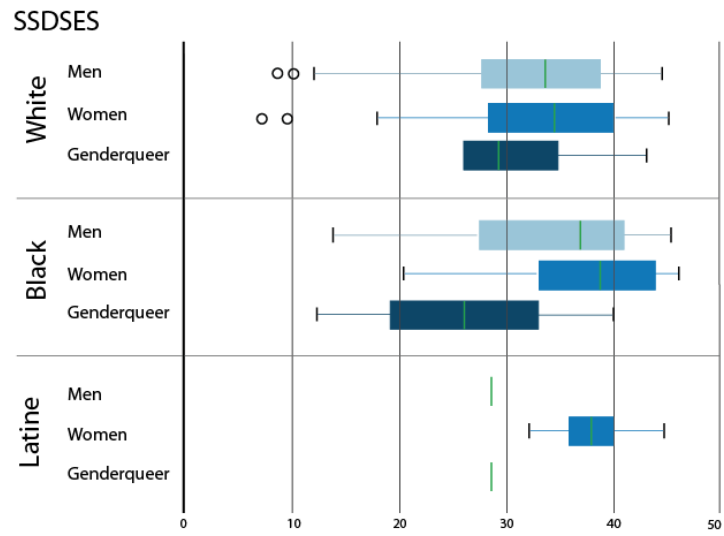


Figure B.3: Vulnerability discovery self-efficacy scores from participants across gender and ethnicities.

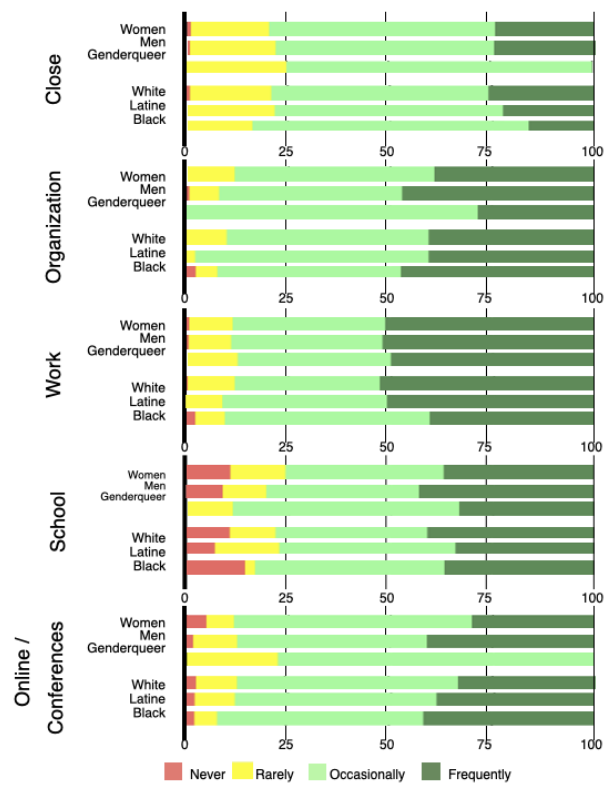


Figure B.4: Rate of discussion by subcommunity grouped by participant gender and race/ethnicity.

Code	Description
Preserving mental model	Participant wants to preserve their mental model of the current state of the protocol RE task .
Context research	Participant performs research into the context for the protocol RE task to inform the process, field choices, structure, etc.
Look for counters / magic numbers	Participant looks for well-known magic numbers / sequential counters.
Investigating and confirming hypotheses	Participant performs investigation and confirmation of hypotheses regarding the data.
User provided tests / data	Participant uses their own test and input data as part of the RE process.
Looking for patterns	Participant looks for patterns in the data.
Landmarking	Participant looks for landmarks in the data.
Preprocessing	Participant performs preprocessing of the data prior to examination.

Table B.5: Codebook definitions for ProtocolREing processes

Code	Definition
<i>Visualization</i>	Participant wants changes to how data was presented or plotted.
Format / presentation of data	Participant wants changes to the format or presentation of the data.
Alternate plotting options	Participant wants alternate plotting options.
Final presentation of data for reporting	Participant wants feature related to output of results for reporting.
<i>Alternative I/O</i>	Participant wants means to move data into or out of the system.
External tool integration / interoperability	Participant wants integration or interoperability with an external tool.
User provided data	Participant wants to input a message or data and see how it is represented in the tool.
Generates Scaffolding	Participant wants tool to automatically generate code / interface for parsing data or IO with another tool.
Programmatic interaction with data structure	Participant wants to interact with the data in a programmatic way.
<i>Focus</i>	Participant wants to control the focus / layout / UI elements of the tool or data.
Searching	Participant wants to search data.
Bit Level View	Participant wants to view data at the bit-level.
Sorting	Participant wants to sort data.
Moving windows in and out of display	Participant wants to be able to use the interface across multiple windows, such as to facilitate use with other tools or to retain focus
Easy interface for drilling down	Participant wants tool to facilitate focusing on a subset of the data.
Hide parts of interface	Participant wants to show / hide parts of tool interface.
Filtering, general	Participant wants to select specific rows within the data to perform filtering actions.
Filtering by value	Participant wants to filter rows by the value of a field or fields.
Filtering by message sequencing	Participant wants to filter rows by a criteria and include messages with a relative offset (before or after) to those rows matching the criteria.
User filters data and reruns analysis	Participant wants to filter data and see how analysis changes.

Table B.6: Tool Features codebook.

Code	Definition
<i>Field Boundaries</i>	Participant wants a feature related to or for manipulating field boundaries.
Automatic field boundary assignment	Participant wants automation to assist with assigning field boundaries.
Identifying connected features	Participant wants a feature identifying which pieces of information are related.
User defines connected features	Participant wants a feature allowing them to specify which pieces of information are related.
Automatic checksum management	Participant wants a feature related to automatic calculation or validation of checksum.
User choice in field boundaries	Participant wants to be able to set the field boundary themselves.
Bit boundaries	Participant wants boundaries at the bit level.
Filter / sorting applies across fields	Participant wants filtering and sorting capabilities that can be performed on one field should be able to apply to either reorder and filter data in other fields
Dotted vertical lines for visualizing breaks	Participant wants a visual indicator for visualizing splits before making the split.
Variable message size	Interface should support non-aligned field boundaries in messages which are variable length.
Variable field size	Participant wants field sizes to vary across messages.
Provide field boundary alternatives	Participants wants a feature which provides alternative field boundaries to one selected.
Automatic field boundary assignment	Participant wants a feature which automatically assigns field boundaries based on a condition.
<i>Data type</i>	Participant wants a feature related to or for manipulating data types.
Automatic data type assignment	Participant wants a feature related to automatic data type assignment.
User restricts data type and rerun analysis	Participant wants a feature related to automatic analysis with data type restrictions.
Additional supported encoding types	Participant wants a feature related to supporting additional text encoding formats.
User choice of endianness	Participant wants to be able to view / choose the endianness of the data types used.
User choice in mixed-type view in-line	Participant wants to select data and see it decoded as a bunch of different types simultaneously.
Provide data type alternatives	Participant wants tool to show alternative datatypes which are consistent with the data.
Decode as datatype	Participant wants to choose data and see it decoded as a particular datatype.
Decode as float	Participant wants to see data decoded as a float.
Decode as hex	Participant wanted to see data decoded as a hex.
Decode as ascii	Participant wanted to see data decoded as ASCII.
Additional supported data types	Participant wants additional data types provided.

Table B.7: Tool Features codebook (continued).

Code	Definition
<i>Meta data</i>	Participant wants summary statistics, or metadata.
Timestamp	Participant wants to see timestamp when data was collected.
Summary statistics	Participant wants tool to calculate statistics such as min, max, median, mode for field values.
Identify data distribution	Tool should identify if data is drawn from a well known or user supplied distribution.
Histogram by field	Participant wants to see a histogram distribution for each field.
Show counts	Tool should display counts of messages / field values.
Confidence Metric	Automated system should provide a confidence measurement for inferred splits / types (OR) Interface should allow participant to record how confident they are in a piece of data.
Size of fields	Tool should give the size of fields based on field boundaries.
<i>State-based modification</i>	Participant wants features related to undo, redo, preview, or storing annotations.
Look back at previous states	Participant wants to be able to look at previous tool states for comparison.
Sticky notes / in-line comments	Participant wants to annotate data / interface with information, recording their mental model into the interface
Preview before change	Participant wants to see a preview of what will happen by making a change, comparing current state with next state
Undo button	Participant wants to undo last action, moving interface to previous state.
Impact of changes in field variance	Participant wants a tool feature allowing them to see the impact of changes on the variance of fields.
<i>Displaying connected data</i>	Participant wants features allowing brushing, or co-ordinated multi-view.
Coordinated Views / Brushing	Participant wants to select data and have related pieces of data highlighted throughout the interface.
Color coding	Participant wants data to be color coded by type or value such as all bytes in ASCII range shown with a fixed color.
<i>Other</i>	
Reason for automated choice	Participant wants information as to how the automated system made a choice.
Solution steps for automated decisions	Participant wants tool to provide specific steps that the automation took when developing a solution.
General customization	Participant wants a feature related to general customization of the tool.

Table B.8: Tool Features codebook (continued).

Code	Definition
Feature exists in our tool and is good	Participant expressed a positive opinion of a specific tool feature.
Don't add to our tool	Participant does not want feature added in a working version of the tool.
Add as improvement	Participant suggested a specific feature that would improve the tool.
Negative opinion of feature	Participant expressed a negative opinion of a specific tool feature.
Talked about generally	Participant generally talked about this feature without a comment on positive or negative attributes.
G5 - Readability improvements	Mentioned usability guideline G5 from Votipka et al. CITATION
G4 - Analysis method selection	Mentioned usability guideline G4 from Votipka et al. CITATION
G3 - Static + Dynamic Analysis method selection	Mentioned usability guideline G3 from Votipka et al. CITATION
G2 - Input + Output Analysis method selection	Mentioned usability guideline G2 from Votipka et al. CITATION
X - Auditability	Mentioned auditability of process.

Table B.9: Codebook for Other Tooling

Code	Description
Reason for automated choice	Participant wants information as to how the automated system made a choice.
Visualization	Participant wants changes to how data was presented or plotted.
Alternative I/O	Participant wants means to move data into or out of the system.
Focus	Participant wants to control the focus / layout / UI elements of the tool or data.
Field Boundaries	Participant wants a feature related to or for manipulating field boundaries.
Data type	Participant wants a feature related to or for manipulating data types.
Meta data	Participant wants summary statistics, or metadata.
State-based modification	Participant wants features related to undo, redo, preview, or storing annotations.
Displaying connected data	Participant wants features allowing brushing, or coordinated multi-view.

Table B.10: High-level tool features proposed by participants.

Appendix C

References of Tools Mentioned

Tool	Reference
FakeNetNG	https://github.com/mandiant/flare-fakenet-ng
Scappy	https://scapy.net/
Ghidra	https://ghidra-sre.org/
Fiddler	https://www.telerik.com/fiddler
Binary Ninja	https://binary.ninja/
Web Dev tools	https://developer.chrome.com/docs/devtools
Netzob	https://github.com/netzob/netzob
Kaitai	https://kaitai.io/
Cyber Chef	https://gchq.github.io/CyberChef/
EtterCap	https://www.ettercap-project.org/
Wireshark	https://www.wireshark.org/
Python	https://www.python.org/
Ida Pro	https://hex-rays.com/ida-pro/
HackRF	https://greatscottgadgets.com/hackrf/
Hexdump	https://man7.org/linux/man-pages/man1/hexdump.1.html
Canterdust	https://github.com/Battelle/cantordust
PwnTools	https://github.com/Gallopsled/pwntools
Make their own tool	N/A
Proprietary tool	N/A

Table C.1: Codebook for Other Codes