How Students Reason about Cybersecurity Concepts

Travis Scheponik, Alan T. Sherman,
David DeLatte, Dhananjay Phatak
Cyber Defense Lab
Dept. of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, MD
{tschep1,sherman,dad,phatak}@umbc.edu

Linda Oliva
Department of Education
University of Maryland, Baltimore County
Baltimore, MD
oliva@umbc.edu

Julia Thompson, Geoffrey L. Herman
Illinois Foundry for Innovation in Engineering Education
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{jdthomp,glherman}@illinois.edu

***Abstract*—Despite the documented need to train and educate more cybersecurity professionals, we have little rigorous evidence to inform educators on effective ways to engage, educate, or retain cybersecurity students. To begin addressing this gap in our knowledge, we are conducting a series of think-aloud interviews with cybersecurity students to study how students reason about core cybersecurity concepts. We have recruited these students from three diverse institutions: University of Maryland, Baltimore County, Prince George's Community College, and Bowie State University. During these interviews, students grapple with security scenarios designed to probe student understanding of cybersecurity, especially adversarial thinking. We are analyzing student statements using a structured qualitative method, novice-led paired thematic analysis, to document student misconceptions and problematic reasonings. We intend to use these findings to develop Cybersecurity Assessment Tools that can help us assess the effectiveness of pedagogies. These findings can also inform the development of curricula, learning exercises, and other educational materials and policies.**

Keywords—cognitive interviews; cybersecurity; Cybersecurity Assessment Tools (CATS); thematic analysis, misconceptions.

I. Introduction

National reports reveal a growing need for cybersecurity professionals [1]. As educators wrestle with this demand, there is a corresponding awareness that we lack a rigorous research base that informs how to meet that demand. This awareness is reflected in the recent creation of cybersecurity education programs by the National Science Foundation and the Department of Defense. Similarly, the NICE framework was developed to articulate a common lexicon for cybersecurity education [2], and the 2013 IEEE/ACM Computing Curriculum has added cybersecurity content to the undergraduate curriculum in computing [3]. Of particular interest to this paper is the need to develop rigorous assessment tools that can measure student learning and identify best practices. We initiated the *Cybersecurity Assessment Tools (CATS)* project[1] to address this need.

This paper presents one step in the process of designing and validating these assessment tools: rigorously documenting how students reason about cybersecurity concepts. To our knowledge, no formal studies have previously explored student cognition and reasoning about cybersecurity. Cybersecurity lies at the confluence of several disciplines, including computer science, engineering, information systems, networks, cryptography, human factors, and policy. Cybersecurity is an evolving field with new concepts and methods invented on an ongoing basis. Therefore, we focus our study on how students develop and use adversarial models to share their reasoning about security scenarios and what misconceptions they reveal about core cybersecurity concepts that were previously identified from our Delphi study [4].

To reveal these misconceptions and problematic reasonings, we are interviewing students at three institutions while they discuss security scenarios: University of Maryland, Baltimore County (UMBC), Prince George's Community College, and Bowie State University (a Historically Black College or University).

Our interviews constitute the second major step in a three-step research plan [5]. In fall 2014, as the first step, we carried out two Delphi processes to identify the core concepts of cybersecurity [4]. We created our interview prompts building on the most important identified concepts. Next year, as the third step, we will use the findings from this study to guide the creation and validation of the assessment tools.

This work was supported in part by the U.S. Department of Defense under CAE-R grants H98230-15-10294 and H98230-15-1-0273 and by the National Science Foundation under SFS grant 1241576.

[1] http://www.cisa.umbc.edu/cats/index.html

In this paper, we present a brief background on student cognition and how they learn. We then describe the design of our study and present preliminary findings to show the promise of the approach. These findings can inform the development of rigorous assessment tools as well as the design of curriculum and instruction for cybersecurity.

II. Background

There is a wealth of research on how students learn complex technical science, technology, engineering, and mathematics (STEM) concepts. This research has revealed that people develop misconceptions about the physical world from their intuitive interactions with the world (e.g., children believe that the world is flat or believe that heavier objects fall faster than lighter objects) [6, 7]. These misconceptions can become so robust that student knowledge can often even appear to be theory-like, providing predictive power and maintaining a degree of consistency across contexts [7]. For example, it has been asserted that students often hold to a naïve theory about physics that resembles the now rejected Impetus Theory [8].

In contrast, research on student understanding of computing concepts reveals that student knowledge is fragile, easily shifting based on contextual cues [9, 10]. For example, in studies of students' understanding of computational state, students would reveal on average four different conceptions of state over a one-hour interview [10]. Critically, these conceptions were often mutually exclusive (e.g., conceiving of state as the inputs and outputs of a system vs. conceiving of state as the stored information of a system) and were frequently revealed in close succession–within minutes or seconds of each other [10]. Similarly, when studying student understanding of Boolean logic statements, students would solve the same problem using different concepts and different problem-solving approaches based on small perceptual cues such as the presence of a truth table or the "cover story" of the logic problem (e.g., students discussing implication (if p then q) in the context of a drinking bar and legal drinking ages vs. in the context of a made-up card game) [11].

It has been argued that student cognition is much more fragmented in computing contexts because students must wrestle with the "science of the artificial" rather than the science of the physical world [12]. While student misconceptions about physical systems are robust from years of observations, student knowledge is unreliable in computing contexts. From this perspective, it is important not only to document misconceptions, which may not be stable, but also to explore the problematic reasonings that students use that may lead to the formation of misconceptions. For example, students may focus too much on specific technologies or overgeneralize from a single case study, leading to poor understanding of new scenarios.

Because we posit that student knowledge will be fragile, we use DiSessa's Knowledge-in-Pieces (KiP) theory [13, 14]. KiP argues that student knowledge is originally a loosely connected collection of knowledge pieces called phenomenological primitives [14]. Students construct their understanding in the moment in response to the perceptual cues available to them. Importantly, what novice students perceive to be relevant in a context is dramatically different from what experts perceive, so student knowledge appears to be even more chaotic or unpredictable to an expert. Expertise is the ability to organize knowledge into cohesive conceptual structures and explanations [8]. In cybersecurity, we posit that a significant aspect of this cohesive structure can be described as "adversarial thinking," including the ability to organize a scenario into an adversarial model.

Adversarial thinking involves reasoning about actions and goals in a context in which there might be bad actors attempting to defeat those goals and carry out their own nefarious actions. Such reasoning requires an understanding of the goal requirements, as well as an understanding of who are the bad actors and what are their objectives, resources, access, capabilities, knowledge, motivations, and risk tolerance. It also requires a technical understanding of the computer systems and their potential vulnerabilities. Our Delphi processes revealed that adversarial thinking, and the associated management of trust and information in computer systems and networks, is the core of cybersecurity [4].

No prior research has documented student misconceptions about cybersecurity concepts nor how they use adversarial models to guide their reasoning. The NICE framework [2] and professional certification tests, such as CISSP [15], provide a basis for identifying standards in terminology, information, and notation, but do not fundamentally tell us about how students learn or reason about cybersecurity concepts. Similarly, the 2013 IEEE/ACM Computing Curriculum articulates some learning goals that institutions may want to adopt for their students, but does not provide any guidance for how students actually learn those topics [3].

III. Specific Aims

This study draws from fundamental theories about student cognition to develop new observations and theories about student reasoning. We can then use these findings to inform theories about how students learn cybersecurity and how we can support more efficient or better learning. Because the secondary goal of this project is to develop assessment tools, and because we are using KiP to guide the design of our study, our research questions focus on documenting the broad range of ways that students misunderstand cybersecurity concepts and struggle to reason using adversarial thinking.

We explore the following research questions.

1) In what ways do students misunderstand core cybersecurity concepts such as authentication, confidentiality, integrity, and availability?

2) Is an adversarial model a significant part of students' thinking regarding security designs?

3) What problematic inferential patterns do students use when reasoning about security scenarios?

4) Do the ways that students identify and address vulnerabilities in practical security scenarios suggest gaps in knowledge that can be mitigated through curricular enhancements?

IV. METHODS

We carried out interviews to engage students as they reason about security scenarios to document and describe their misconceptions and problematic reasonings.

A. Interview Subjects

We recruited students from courses in cybersecurity at three institutions that are ethnically and academically diverse. We selected all students who volunteered to be interviewed as long they had completed or were currently enrolled in at least one course focused on cybersecurity. We interviewed a total of 26 students: twelve from UMBC, ten from Prince George's Community College, and four from Bowie State University. We compensated students $10 for their participation in the interviews.

B. Interview Process

The interviews took place in a classroom or conference room at the subjects' home campus. All interviews lasted approximately one hour. Oliva conducted all interviews, and Sherman and Scheponik were present for most of them, observing and asking follow-ups for each question after Oliva completed her dialog. To begin each interview, Oliva explained the purpose of the study, asked for informed consent to participate, audio record, and video record, and collected some demographic information about the subject's degree program, cybersecurity courses taken, and cybersecurity experience. We transcribed the recordings to facilitate analysis. UMBC's IRB office approved our research protocol.

C. Interview Protocols

We developed three interview protocols (called Alpha, Bravo, Charlie), each comprising four separate questions or interview prompts. The team selected these questions from a larger pool of candidate questions created in Fall 2015. Each protocol includes a diverse set of questions covering a range of ideas, contexts, difficulty, and question types. The goal of each question is to encourage the subjects to reveal how they think about important cybersecurity concepts by having them talk about how they solve specific cybersecurity problems presented to them. We delivered the questions in order of increasing complexity.

The team developed and refined the questions during brainstorming sessions. We developed a template to identify the concepts that were covered in the questions and specified exemplary responses. For each question, we planned ways to respond for "hits" (when the subject gave a reasonable response) and for "strikes" (when the subject struggled to come up with a reasonable response). For example, if a subject suggested a flawed security measure, we might ask them to explain what would happen in a concrete situation chosen to expose the flaw. For some strikes, we might provide a diagram to stimulate further discussion.

The interviewer told the subjects that she was a novice in cybersecurity and that they would be prompted to provide as much detail as possible in their responses. Oliva gave a written copy of each question to the subjects and we encouraged them to sketch diagrams to facilitate their explanations. We encouraged subjects to explain what they meant, the reasons behind answers, and the meanings of any terms used. The prompting continued until subjects reported that they could provide no further explanation.

We focused on familiar, yet conceptually rich and open-ended scenarios that did not require detailed technical knowledge. The following example (Charlie-1) is typical:

> "Bob's manager Alice is traveling to country B and is planning on giving a sales presentation. Bob receives an email with the following message: `Bob, I just arrived in country B and the airline lost my luggage. Would you please send me the technical specifications for our new product? Thanks, Alice.' What should Bob do?''

Another example (Bravo-1):

> "While Mary is traveling she decides to do some shopping online. She is connecting from a computer in a hotel business center. What are some of the cybersecurity issues that might arise? Sketch a figure to illustrate your explanation."

D. Analysis

Thompson and Scheponik are leading the analysis of the interviews using a novice-led paired thematic analysis approach [16]. This approach allows researchers to investigate student conceptual understanding while addressing issues of "expert blind-spot [17]." The lead researcher for the analysis is a novice in terms of cybersecurity knowledge, but with expertise in educational research and metacognitive ability. This researcher is paired with a cybersecurity content expert who also analyzes the interviews and provides insights into how student responses are exemplary or problematic. We center our analysis on the learning of the lead researcher.

In this analysis approach, the content expert first reviews the interviews and codes sections as either "correct" or "incorrect," and provides short clarifying comments as appropriate. For example, labeling a section as "correct, but the student only provides a partial answer." This initial review by the content expert allows for a richer and more targeted analysis process. Following the initial review, the content novice familiarizes herself with the interview and carries out a line-by-line read through to develop a better understanding of students' reasoning. She asks herself a series of questions during the analysis phase, including: "Why do I think this response is 'correct' or 'incorrect?,'" "What is the scope or range of the student's response?," "What is the viewpoint of the student when he or she answered this question?," and "How did the question being proposed prompt the students response?" The lead researcher takes copious notes of the interviews and then talks through the interviews once a week with the content expert. These questions are intended to produce a better understanding of student reasoning and to explicate the expert's tacit knowledge.

We identify themes within the interviews relating to student reasoning and common misconceptions through the conversations and notes. When themes are identified, the two researchers convey topics to the whole group and bring forth excerpts as appropriate. We developed the themes discussed in

the results section from a subset of the interviews. The researchers will expand on these themes through the analysis of the whole data set.

V. Preliminary Findings

Having recently begun the analysis of the interviews, we offer two preliminary examples of themes. See Appendix for an explanation of selected terms and concepts from cybersecurity.

A. Students conflate confidentiality with integrity

The first theme reveals that students incorrectly reason that the use of encryption prevents adversaries from modifying data as it traverses a network. This reasoning reveals a misconception as students are conflating confidentiality with integrity. Encryption prevents adversaries from reading the data and does not necessarily prevent them from modifying the content of the message.

The following excerpt illustrates confusion between authentication, confidentiality, and integrity.

> "**Interviewer**: So man-in-the-middle. What are some of the things you would want to do to mitigate that?
>
> **Subject:** Encryption. You would want to use some sort of public-private key encryption where you can verify through a third party that…ideally through a third party that the person who is sending it is actually them or actually you are actually you, so that when the other person is receiving it, it hasn't been tampered with."

The explanation is incorrect because the subject states that encryption will ensure that the data have not been tampered with during transmission.

B. Students conflate authentication with authorization

The second theme revealed that students incorrectly reason that when a person is authorized to use a resource, they have proven their identity (and vice versa). Some students incorrectly reason that proving that someone is allowed to use a resource implies that the person is who they claim to be. When presented with a scenario that requires both authentication and authorization, this faulty reasoning may permit a mischievous entity to masquerade as an authorized entity.

For example, when a student was asked about placing sensors to comply with Nuclear Test Ban Treaty, the student responded, "Country A only has, let's call it the authentication group. So, they can only see data from the sensors, and do sensor checks making sure the sensors work and to make sure those are still their sensors." While the subject correctly identifies the need for groups to determine authorization to resources, the subject has named the group "authentication group." This indicates that the subject is conflating aspects of authorization and authentication.

VI. Conclusions and Future Work

Our preliminary results suggest that students do not distinguish sufficiently between core concepts in cybersecurity. These conflated concepts suggest that students use a form of "satisificing" in their reasoning [18], becoming too easily satisfied that a system is secure after identifying only one possible source of security for a system rather than seeking to explore the adversarial space more thoroughly.

Analysis of our 26 interviews will continue with the goal of identifying more misconceptions and uncovering or elucidating problematic reasonings that give rise to shortcomings in student understanding. When analysis of the interviews is complete, we will use the findings to inform the prompt questions and distractor responses within the development of assessment tools to measure student learning in cybersecurity. For example, we may target questions and responses on the differences of confidentiality and integrity, correlating to our preliminary findings. The analysis of these misconceptions and problematic reasoning provide rich insights into how cybersecurity education can be measured and improved.

Appendix: Cybersecurity Terms and Concepts

This section defines selected important terms and concepts from cybersecurity. For more information, see Schneier [19].

Cyber refers to computers or computer networks.

Cybersecurity is an interdisciplinary field that concerns the management of information and trust in an adversarial cyber world. It integrates people, policies and procedures, and technology. Contexts of interest include any situation that involves computers or information in electronic form, including computer systems, computer networks, databases, and applications.

Four essential concepts include confidentiality, authentication, integrity, and availability. *Confidentiality* refers to keeping information secret from unauthorized entities. Encryption is a tool for keeping information confidential. An encryption function mixes a plaintext with a secret key in a complicated way to produce ciphertext, with the intention that an eavesdropper seeing only ciphertext cannot decrypt the ciphertext to produce the plaintext (without knowledge of the secret key).

Authentication refers to the task of, say, Alice convincing Bob that a message purporting to have originated from Alice did indeed come from Alice. Digital signatures and message authentication codes are tools for achieving authentication. For example, Alice can sign a message using her private signature key. Using Alice's public verification key, Bob can verify Alice's signature.

By contrast, *authorization* refers to whether an entity is allowed to perform some action, for example, reading some data or gaining access to some computer system.

Integrity refers to the problem of detecting whether data (either at rest or in transit) have been modified. Cryptographic hash functions are useful tools for achieving integrity. A hash function takes an arbitrarily long input and produces a short fingerprint (also called a tag) such that, if any change is made to the input (even just one bit), then with overwhelming probability the tag will change.

Availability refers to systems, services, and networks being up and running.

ACKNOWLEDGMENTS

We would like to thank the students who participated in the interviews and the faculty and staff who facilitated the interviews at the three institutions.

REFERENCES

[1] M. C. Libicki, D. Senty, and J. Pollak, "Hackers Wanted: An Examination of the Cybersecurity Labor Market," RAND Corporation, Santa Monica, CA2014.

[2] NIST, "The National Cybersecurity Workforce Framework," National Institute of Standards and Technology, Washington, DC2013.

[3] M. Sahami, S. Roach, A. Danlyuk, E. Cuadros-Vargas, S. Fincher, R. Dodge, et al., "Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science," IEEE/ACM2013.

[4] G. Parekh, T. Scheponik, D. DeLatte, G. L. Herman, L. Oliva, D. Phatak, et al., "Identifying core concepts of cybersecurity: Results of two Delphi processes," in prep.

[5] J. Pellegrino, N. Chedowsky, and R. Glaser, Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press, 2001.

[6] S. Vosniadou, "The conceptual change approach and its reframing," in Reframing the conceptual change approach in learning and instruction, S. Vosniadou, A. Baltas, and Z. Vamvakoussi, Eds., ed Amsterdam: Elsevier, 2007, pp. 1-15.

[7] M. T. H. Chi, "Commonsense conceptions of emergent processes: Why some misconceptions are robust," The Journal of the Learning Sciences, vol. 14, pp. 161-199, 2005.

[8] G. Ozdemir and D. B. Clark, "An overview of conceptual change theories," Eurasia Journal of Mathematics, Science, & Technology Education, vol. 3, pp. 351-361, 2007.

[9] D. N. Perkins and F. Martin, "Fragile knowledge and neglected strategies in novice programmers," presented at the First Workshop on Empirical Studies of Programmers, 1986.

[10] G. L. Herman, C. Zilles, and M. C. Loui, "Flip-flops in students' conceptions of state," IEEE Transactions on Education, vol. 55, pp. 88-98, 2012.

[11] G. L. Herman, M. C. Loui, L. Kaczmarczyk, and C. Zilles, "Describing the what and why of students' difficulties in Boolean logic," ACM Transactions on Computing Education, vol. 12, pp. 3:1-28, 2012.

[12] H. A. Simon, The Sciences of the Artificial. Cambridge, MA: MIT Press, 1996.

[13] A. diSessa, N. Gillespie, and J. Esterly, "Coherence versus fragmentation in the development of the concept of force," Cognitive Science, vol. 28, pp. 843-900, 2004.

[14] A. A. diSessa, "A bird's eye view of 'pieces' and 'coherence' controversy," in Handbook of Conceptual Change Research, S. Vosniadou, Ed., ed Mahawha, NJ: Lawrence Erlbaum, 2008, pp. 35-60.

[15] CISSP. (2016). CISSP. Available: https://www.isc2.org/CISSP/Default.aspx

[16] D. B. Montfort, G. L. Herman, S. A. Brown, H. M. Matusovich, and R. A. Streveler, "Novice-led paired thematic analysis: A method for conceptual change in engineering," presented at the 2013 American Society for Engineering Education Annual Conference and Exposition, Atlanta, GA, 2013.

[17] M. J. Nathan, M. W. Alibali, and K.R. Koedinger, "Expert blind spot: When content knowledge & pedagogical content knowledge collide," Institute of Cognitive Science, 2005.

[18] R. Brown, "Consideration of the orgin of Herbert Simon's theory of 'Satisficing' (1933-1947)," Management Decision, vol. 42, pp. 1240-1256, 2004.

[19] B. Schneier, Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2 ed. New York: Wiley, 1996.