An Examination of Industry Standards of Success within Penetration Testing Groups

Mollie Ducoste, Rachel Bleiman, Trinh Nguyen, and Aunshul Rege
Temple University, mollie.ducoste, rachel.bleiman, thuy-trinh.nguyen, rege@temple.edu

***Abstract* - Penetration testing groups can be used as an ethical proxy to study cybercrime groups, as both parties share the common goal of identifying and exploiting weaknesses in their targets' systems. Pentesters often use existing industry standards to guide their performance and practices, but little research has investigated how these standards operate in simulated cybersecurity exercises. Using the experiences of college students in the 2018 and 2019 National Collegiate Penetration Testing Competition (CPTC), a simulation of a professional real-world penetration test, this study seeks to further examine pentesting metrics. Metrics from industry standards of pentesting practices are compared to the metrics identified by the CPTC participants, revealed through semi-structured group interviews. Industry metrics include standards, such as methods, information gathering, attack generation, quantity of findings, quality of findings, and reporting of findings. Other additional metrics identified by the CPTC participants include skills of the team, the environment, expectations, and the relationships among group members. This study uses a qualitative methodological approach to examine the metrics of success identified by pentesters as they reflect on their decisions, actions, and performance.**

Index Terms - Cybersecurity competitions, Penetration testing, Pentesting metrics

INTRODUCTION

Penetration testing involves cybersecurity industry professionals replicating the actions and decisions of cybercriminals by operating on the offensive side of a simulated cyberattack. Penetration testers, or pentesters, often engage in real-time cybersecurity exercises, where they play the role of cybercriminals, and thus serve as good proxies for the latter. The industry has developed several standards for measuring the efficacy of penetration testing (pentesting) exercises. However, there is a dearth of information on the application of these standards to simulated cybersecurity exercises. Specifically, this paper addresses two questions. First, do these standards capture all possible measures of decisions, actions, and performance? Second, how can a qualitative social science methodological approach be used to capture these standards?

This paper shares the results from a qualitative study on students that participated in two pentesting competitions. The researchers specifically sought to compare industry standards of success in pentesting with those success standards identified by the students in the pentesting competition. Because students who participate in pentesting competitions often seek a future career in the cybersecurity industry, the researchers were curious to observe whether the standards the students used to measure their own success in pentesting competitions would reflect the industry standards of success. To the best of the authors' knowledge, existing literature on pentesting has not come to a consensus on an exhaustive list of success measures for pentesting competitions, nor has existing literature used self-reported data from competitors to inform these success measures.

The metrics of success reported by the pentesters in the competition were analyzed and then compared to existing industry standards of success. While the industry standards of success contain technical aspects of pentesting, the present study also discovered non-technical success measures that relate to the stages of pentesting, the overall testing environment, and the strength of findings. While the technical readiness of pentesters certainly contributes to an aspect of success, this paper looks at the competitors' perceptions of their preparedness and performance, which is inherently non-technical in nature.

The next section will review the existing literature on pentesting and its function in cybersecurity, existing standards of success within the pentesting industry, and the prevalence of cybersecurity and pentesting competitions. Subsequently, information on the primary data source for this study is provided, followed by an outline of the methods used to collect and analyze the data. Next, the results and analysis section describes how the existing standards of success, along with newly identified metrics, were found in the pentesting competition. The discussion section details the implications of these findings and notes limitations of this study. Finally, the paper concludes with the authors' takeaway thoughts, lessons learned, and directions for future research.

LITERATURE REVIEW

Over the years, research on hackers have expanded beyond a criminal perspective, encompassing the study of both legal hacking (white hats) and illegal hacking (black hats) behaviors [1]. Pentesters are one example of legal hackers that work within the cybersecurity industry [2].

I. Penetration Testing and Industry Standards

Pentesting is one avenue for ethical hackers to use their skills to discover vulnerabilities and protect computer systems [2]. Pentesters are tasked with simulating targeted attacks on a company's system to determine any weaknesses in their infrastructure [2,3]. They perform a range of duties that are comprised of information gathering, attack generation, and response analysis [2,3]. The widespread recognition of pentesting techniques warranted the development of standardized practices by government agencies and other related industries [2,4]. Several trade groups, such as the Open Web Application Security Project (OWASP), Communications and Electronic Security Group from the United Kingdom, and the Institute for Security and Open Methodologies (ISECOM) are well known organizations that have outlined industry standards for the best pentesting practices [2]. ISECOM created the Open Source Security Testing Methodology Manual (OSSTMM) that has been peer-reviewed across multiple disciplines, which provides evidence-based procedures for accurate penetration and security testing [4]. Additionally, the Payment Card Industry Data Security stands (PCI-DSS) was created for industries related to debit, credit, e-wallet, POS, and ATM cards, and the International Organization for Standardization (ISO/IEC-27001) outlines how to establish information security for private, non-profit, and profit organizations [4].

II. Industry Metrics for Penetration Testing

A few existing standardized practices for measuring successful pentests have been developed and acknowledged by the pentesting community [2,4]. The different guidelines indicate varying components that contribute to a successful pentest in the industry. Several of the measures tend to fall under the following six categories: methods, information gathering, attack generation, quantity, quality, and reporting. Maximizing all six categories helps contribute to a more successful pentest, which ensures the process is accurate, useful, and ethical.

IIa. Methods

A primary component of a successful pentest is grounded in its methodology [3,5,6]. The methodologies employed during a pentest refer to whether the individual is using the most appropriate and realistic means to obtain secure information [3]. Therefore, the art of pentesting consists of different specializations, such as network, host, application, and social engineering [3,4,5,6,7]. The information that is accessible to the tester will determine if black box (no information on test target), white box (all necessary information on test target), or gray box (partial information on test target) strategies are used [3,6,7]. Open communication with the client will create transparency and provide pentesters with the client's rules of engagement, which will prevent particular methods from resulting in harm to the company [4,5,7]. These methods should closely mirror real-life hacking scenarios and not involve information that would be unknown to cyber adversaries [4,5,7]. Therefore, pentesters will carefully gather information on the client's company, tailor attack methods accordingly, and analyze the results of attempted attacks to generate a successful pentest [2].

IIb. Information gathering

Selecting the appropriate procedures for pentests require a thorough analysis of the system and collection of all available information on the target [2,3,6]. Information gathering, or reconnaissance, is often the first and critical step in performing pentests, because the tester must understand all the different input vectors in an application that can be attacked [2,4]. Passive reconnaissance refers to gathering information without accessing the network [4]. Active reconnaissance refers to connecting to the network to examine preliminary responses from the target [4]. The more information that is obtained during this process creates a better opportunity for a successful pentest [3]. This step can be completed with various existing tools [2,3], but successful pentesters have the creative ability to customize their own tools.

IIc. Attack generation.

Using the vulnerabilities discovered from the information gathering phase, pentesters create or identify exploits to use in the system attack [2,4]. Pentesters iterate through each input vector (vulnerability point in an application), and for each input vector, they develop possible attack strings [2]. These attack strings are used to generate attacks which are attempted against the target's system and are necessary for its exploitation [2,3]. Successful target exploitations will allow pentesters to gain access to the resources contained in the system [4].

IId. Quantity

The thoroughness of pentesters is correlated to the number of vulnerabilities they find. The most objective measure of pentesting outcomes is the quantity of vulnerabilities identified in a system. Determining the quantity of these findings refers not only to the sheer number of identified security concerns left susceptible to potential cybercriminals but also to the number of unique components of those identified vulnerabilities [2]. Additionally, successful pentesters can not only obtain access but should gain access multiple times without detection [6].

IIe. Quality

Moreover, the outcome of pentests can be equally dependent on the quality of the discovered vulnerability, and the extent that it poses a severe risk to the organization [3]. A pentester could conceivably have explored more components of the system without finding high-risk vulnerabilities [2]. Take, for instance, a pentester who has found only a few vulnerabilities, yet those few vulnerabilities could lead to more harmful security breaches or theft of information if they were to be exploited by a cybercriminal. Pentesters who gain access to

administrative accounts or privilege escalation may have a more fruitful conquest [6]. Thus, the quality of the vulnerability discovered can be a more valid measure of success.

IIf. Reporting

Efficient pentests include the reporting of findings and recommendations to the organization, which is a constant process throughout the assessment. In the response analysis or validation phase, testers determine if each attempted attack on a system's vulnerability was successful and document their results [2,4]. Specifically, at the culmination of the pentest, the information gathered from testers during all phases of the assessment are used to provide the client with a summary of attack methods, vulnerabilities, and mitigation plans [4,6]. Exceptional pentesters will prioritize reporting on vulnerabilities dependent on their probability of exploitation, challenge and cost to mitigate, and effect on the business [6]. The OSSTMM indicates several procedures for reporting, such as transparent reporting and different ways to classify discovered limitations in security systems. Pentesters should provide accurate reports along with a comprehensive post-debriefing so that client companies can best understand the insecurities in their network [5,6]. Clear and detailed reports allow the company to repeat the attack patterns and remedy the security concerns [3]. Clients appreciate pentesters that can offer reports on technical analyses with the goals of the company at the forefront [6].

III. Cybersecurity Competitions

The high number of security breaches each year has led to a high demand for people with a background in information technology to fill the cybersecurity workforce [8]. Therefore, academic institutions have incorporated pentesting in their curricula and have recognized the importance of sharing resources to increase the advancement of pentesting curricula in higher education [8]. In addition, educating students on pentesting skills requires hands-on activities inside and outside of the classroom [8]. Beyond the classroom, cybersecurity and pentesting competitions take place at a regional and national level to provide students with opportunities to learn and practice their skillsets [8]. Unlike cybersecurity competitions, such as the National Collegiate Cyber Defense Competition (CCDC) and CyberPatriot [8], contestants in pentesting competitions are not focused on defending a network but on performing specific duties to simulate a real-world pentest [9]. The goal of pentesting competitions is to engage students with experiences in their future work environments.

CURRENT STUDY

The National Collegiate Pentesting Competition (CPTC) provides an opportunity for undergraduate students to experience a simulation of a real-world pentest. During the CPTC, teams of students from all over the world in the field of cybersecurity are tasked with the goal of discovering, triaging, and mitigating critical security vulnerabilities of a fictional company [9]. The students serve on the offensive team as pentesters and must use their technological, communication, and collaboration skills to effectively take part in the competition. This three-day competition is organized and located at the Rochester Institute of Technology in Rochester, New York. Similar to a real-world pentest, the teams have an objective to find weaknesses in the company's security using their technical prowess, communicating these technical concepts to both technical and non-technical audiences, and working together with team members to achieve success [9]. At the end of the competition, the teams present their project deliverables, which are detailed findings compiled by the teams, to the judges. Based on the presentations of the teams' deliverables, the judges score and rank each team.

METHODS

Past studies have specified the advantages of using qualitative methods to uncover a more detailed understanding of the complex processes involved in hacking [10,11,12]. Particularly, interviews create a dialogue between researchers and their participants, where researchers can probe participants with tailored questions to ascribe additional meaning to social contexts [13].

As such, the researchers conducted semi-structured interviews with 18 pentesting teams from the 2018 and 2019 National CPTC event.

Participants shared their individual experiences about strategies employed, challenges faced during the competition, and overall experiences and performances during the competition. The researchers adopted a holistic approach, allowing the pentesters in the competition to identify their own measures of success, regardless of if they turned out to be technical or non-technical metrics.

The interviews lasted no longer than an hour for each team. All the interviews were manually transcribed and coded. This study employed a grounded-theory approach that adopted inductive coding strategies [13]. This allowed for themes to emerge as the researchers transitioned between initial codes to more focused codes and related the themes to theoretical frameworks [14]. This study was approved by the ethics board at the authors' home institution.

RESULTS AND ANALYSIS

Results from the 18 team interviews revealed that the industry standards of methods, information gathering, quantity, quality, and reporting were indeed relevant. Interestingly, these teams also identified four additional contributing success factors that they identified over the course of the competition: skills, environment, expectations, and relationships. The following section details the ways that the pentesting teams found the combined nine factors to contribute to their overall levels of success throughout the course of the competition.

I. Industry Metrics

Ia. Methods

All 18 teams identified their methods over the course of the competition, describing how these methods contributed both to their success and failures. One of the largest contributing factors of success as related to a team's methods had to do with their ability to remain flexible and adjust to changing external factors that were outside their control. One team member stated: "We did have semi-strategies going in [to the competition, but] because the environment changes a lot … [it] also changed how we all approached the [penetration] test. It was, I would say, changing over time." Another team discussed how they changed roles based on the changing environment: "We each have our loosely defined roles. And we have to be very flexible, creative, and have the ability to adjust on the fly. So we can't restrict ourselves to any kind of attack plan." When describing their ability to react quickly to changing factors, one team stated: "there's some...a few things that we react [to]-as soon as we see it, we'll react to it immediately. If we see an outdated operating system, we know exactly what exploits to try first. Or if we find a list of usernames, then we would try those lists of usernames and then do a password spray attack immediately. So there's certain actions where, once we see it, we automatically just do it."

The pentesting teams also identified instances in which their methods of attack were flawed, thus contributing to incremental failures in the competition. For example, one team found that their methods of communication and information sharing were not efficient: "We had information … in different areas -some of it was in the Google Drive, some of it's on the white board, some of it's just like -'did anyone hit that computer,' and we're all sitting there like 'uhh, we scanned it, yeah.' So for me that was, I felt like, sometimes we kind of lost focus there, because we were like, 'did anyone actually ever get on this?' So I wish... Like a more formal method of how to share findings." Similarly, another team found that their process of prioritization was flawed: "During the competition, we were so focused on trying to privilege escalate our access. We landed on a Linux server, but we weren't able to get anything fruitful from the Linux server, so we were trying to escalate our privileges for the longest time. And [then we] realized, this is not the attack path that we should be going down."

Ib. Information Gathering

A commonly reported method was the information gathering, or reconnaissance, phase. All 18 teams discussed their use of information gathering as an important and near-constant step in further generating their attacks. For example, one group reported that, "we'll find something, we'll try to attack it, we'll give up, then recon something else, recon, attack it, give up", showing the importance of this phase in being able to generate attacks. Another group explained that they did "constant recon throughout the whole day…we had to constantly look around, constantly enumerate" and another group said that "most of our reconnaissance we did was more passive in the background. So while that was going on, we could do other stuff", which demonstrates how information gathering is a continuous process.

Ic. Quantity

Teams named the importance of finding a high volume of vulnerabilities in a pentest to be associated with success. 12 of the 18 teams mentioned the quantity of vulnerabilities in the course of the interviews. One pentesting team described their approach to finding high quantities of vulnerabilities early on in the competition: "We try to find the low hanging fruit. So whatever services or things we might be able to find that are vulnerable, and then exploit and then from there, we'll do post exploitation stuff where we either tried to gain persistence or pivot through whatever we exploit to try to get to somewhere else." Another team member reflected that finding more quantities of vulnerabilities may have been helpful to their overall success: "Finding more vulnerabilities, even if we didn't personally need them, in retrospect, might have been helpful as well."

Id. Quality

All 18 teams identified the quality of their vulnerability findings as a factor of their overall success. The overwhelming majority of vulnerability findings identified as high in quality were deemed as such because they led to further findings. One team member described their intent to find high-quality findings to help with future findings: "we tried to target things we felt were juicy in nature. So that it would be – let's say, something with credentials, something with user information, something that we can leverage later on in the competition." Another team described a high-quality finding as "a breakthrough that changed into multiple breakthroughs."

Many teams identified high-quality vulnerability findings as a major turning point in the competition. For example, one team member stated: "One of the key turning points was when we were able to actually craft an exploit that we didn't think we could craft and that allowed us to escalate our privileges up to the maximum in the environment. And we were able to get into most of the boxes through that. And that was a pretty advanced manual piece of work to identify that as a vulnerability, to figure out how it could be exploited, and then craft that exploit that actually did the deed for us. So that was it, that was a very exciting moment." The emergence of high-quality findings also appeared to direct the entire team's ability to find more vulnerabilities. Another team member described how their elevated access helped them direct their attention to other important findings: "We had discovered a number of vulnerabilities on those that would allow us to gain elevated access. So, [we] worked on exploiting those and then from there started to explore different aspects of the new services."

Results from team interviews also pointed to the idea that high-quality and high-quantity findings often go hand in hand. One team member described how after finding a high-

quality vulnerability, they then proceeded to look for high quantities of vulnerabilities with the new elevated access they had just obtained: "After that, once we were able to obtain that, we used what information we got and used it to spread across the network and get as much stuff as we can." Contrarily, another team decided to refrain from continuing to find higher quantities of vulnerabilities, because they found that an early high-quality finding was more beneficial to their overall goals in the competition: " [we were] looking into doing privilege escalation but, I guess we put that to the side a little bit, because we had a domain admin so early [which] was not ideal." The pentesting teams did not state whether they thought a high-quality or high-quantity find was better than the other; rather, they seemed to place equal value on both kinds of findings in relation to their overall success.

Ie. Reporting

16 of the 18 teams discussed the reporting process as it related to their overall ability to achieve success during the competition. Teams were constantly in the process of debriefing the client organization when representatives would enter their rooms with questions or concerns. One team noted that these constant interruptions served as a source of stress for the competitors. Similarly, another team described the stress of interacting with clients, as well as their resulting response: "every time someone walked in the room, I'm like, 'alright, who is it, who do I need to talk to, how do I need to interact with them?"

Another major instance of reporting to clients occurred in the form of the final report to be presented to the client company at the end of the competition. Many teams described the need to remain organized and take detailed records of their findings in-time. For example, one team member stated: "One thing we do when we are [penetration] testing is, once we find something, we immediately take the appropriate screenshots for evidence, and we put them in an organized way in our Google Drive, in terms of folders and things like that. There's always an organized approach to [penetration] testing; in that case, later on, we don't need to worry about where we're going to find specific evidence and things like that." Similarly, another team member found that "maintaining organization is important because when we were writing the report, I kept on asking every single person, 'hey, where's that screenshot at, did you have this, or do you know where it is?' And, like, it's because... all we did was we just piled everything into one folder and we had to look for it, it's like looking for a needle in a haystack."

For the most part, the authors found that industry metrics applied to the CPTC, and, interestingly, they found four additional metrics that are discussed in the next subsection.

II. Team-Identified Metrics

IIa. Skills

Each pentesting team was comprised of members with varying technical backgrounds and skillsets. All 18 teams that were interviewed discussed how their skillsets contributed to their successes and failures during the course of the competition. One team mentioned how the diversity of their members' skillsets was a major advantage during the competition: "We know everyone has different backgrounds, everyone has different skillsets, so, [we would ask each other] 'hey is this a good idea?', '[are] you familiar [with] this, what would you do here?' … some people bring different perspectives that you might not have thought of." Some teams had an assortment of expertise that they brought to the competition. For instance, one team had members with skills that included web application expertise, binary reverse engineering, Linux, and scripting, while another team had members with different skills, including technical writing, forensics, active directory security, and Windows.

While skillsets were instrumental in achieving success for the teams, they also mentioned a number of instances where their skillsets were not suited for certain tasks or environments. One team member stated: "We do lot of training on single machines. So, there's a lot of focus on going in-depth on a single machine and that doesn't really take into account some of the more dynamic interactions between networks and machines. And competitions such as this, I think that we underprepared a little bit for analyzing how we can use different aspects of the more interconnected network to leverage and gain additional privileges to other machines." Another team member stated: "We should know windows better. A lot of us are focused on Linux and so we're good at that." It is clear that in the cases that skillsets were lacking, a solution would be to obtain more skills to overcome similar obstacles in the future. In other words, these teams felt that an increase in skills is highly correlated to an increase in success.

IIb. Environment

17 of the 18 teams discussed their environment as a contributing success factor. Many teams described the technical characteristics of the competition when discussing environmental factors. For example, one team discussed how the simulated pentesting environment was helpful in their overall training: "They did a good job of mimicking a real-world environment and adding that human aspect to it which is critical for getting the job done as a professional…it was a great learning experience ... being able to deal with the people and the things you have to do along with that while also being able to be technically competent and get the job done." Similarly, another team member stated, "I learned a whole lot about what it's like to be in that kind of business environment even if it was just a simulated business."

In addition to experiencing the simulated systems that tested their technical prowess, students also had the opportunity to gain experience with social interactions that would simulate real world engagement with clients. For instance, during the competition there were interactions between the student teams and the competition organizers, who served as the client company seeking updates. There were also interactions with the judges during the formal presentations, which simulated teams presenting findings to

the people who hired them. The participants received social interactions which are just as important as the technical ones and can impact one's overall experience in the competition. These interactions are also particularly important because pentesters need to know how to speak to clients and formally present their findings.

While the technical environment facilitated teaching about systems that pentesters must engage with, the social environment enabled teams to learn how to formally engage with clients in stressful situations. Thus, a properly structured technical and social environment is crucial for teams to feel comfortable and perform well. Furthermore, it is evident that some of the opinions related to the environment are specific to the competition setting, rather than a general pentesting environment. However, the difficulties that many teams have with juggling tasks, particularly those related to client interactions, are likely relevant within a real-world pentesting team.

IIc. Expectations

13 of the 18 teams discussed how their expectations prior to starting the competition affected the extent of their successes. For example, one team discussed their difficulties adapting to their unmet expectations: "What the competition expects can be very different from what the reality happens to be for a lot of teams. And also because there is so much chaos … sometimes things don't go as planned. I think everyone was aware in this competition, that there were some aspects … that did not go as planned for everyone. And it can be hard to deal with that and adapt." Another team discussed how their expectations about what they should prioritize were not the same as what the competition expected: "I'd say as a whole, if we had potentially tried to focus more preparation efforts on some of those [competition expectations], it might have been a bit more helpful."

Similar to the concept of environment, it is important to note that the teams' expectations are likely specifically related to the competition setting. However, it is also highly plausible that real-world pentesters could also find a disconnect between prior expectations and reality.

IId. Relationships

16 of the 18 teams discussed how their relationships within their teams contributed to their successes. Many teams described their strong bonds prior to the competition. One team member stated: "We're all friends. We joke around a lot and work well. Make each other feel comfortable working around each other. It doesn't feel like too tense an atmosphere of work." Similarly, another team contributed their success to their ability to work well together: "Since we have such great camaraderie, everything pretty much went according to plan."

Relationships also played a role in a team's ability to support each other and maintain active roles within the group. For example, one team member stated: "So there was definitely a segregation of roles where we basically gave people specific roles as things were coming up … There was still fluidity, we could still work on different things, but we definitely assigned roles." Another team member found that they were stronger as a team than separately, stating: "The team supporting each other made this much better than it would be if anyone was going to try this … like a one-man army style; that's not going to work."

DISCUSSION

Competitors identified the existing industry metrics as indications of their performance, showing that the standards of success in pentesting competitions reflect the industry standards of success. However, competitors also identified several other metrics as standards for their success or failure in addition to the industry standard metrics of methods, information gathering, quantity, quality, and reporting. A key finding in this study is the juxtaposition between the more technical existing industry standards of success, and the more non-technical measures as identified by the student competitors. As such, the skill range of the team members, their environment and ability to adapt to their environment, their expectations, and the relationships among the members of the group are all also important in measuring pentesters' successes.

With any study, there are limitations. In this study, the most prominent limitation is the single data source of the national CPTC. The findings may differ across other pentesting events, which may be structured differently. Additionally, industry pentesters are more seasoned than the student participants interviewed at the CPTC. Collectively, the differences across pentesting environments and skill level of participants may impact the generalizability of these newfound metrics. However, a social science lens, such as the one used in this study, helped unearth additional factors that would otherwise have remained invisible in discussing the efficacy of pentesting exercises. The authors hope that these additional metrics may be examined in the future and possibly incorporated into the existing industry standards.

Another limitation is that the authors were unable to find evidence of the "attack generation" industry metric. There are several potential reasons that could explain why this metric was absent. First, it could be because of the nature of the competition and how the event is structured. Second, it may relate to the skillset and level of expertise of the students who were competing in the competition. A third possible explanation is that the researchers did not ask the necessary questions during the interviews to extract this information from the participants. A final explanation is that while evidence of the "attack generation" phase did not appear in the interviews, it may have been present in the technical logs of the competition, which were not analyzed for this study.

CONCLUSION

This paper sought to address two issues: whether existing industry standards applied to pentesting exercises and whether a qualitative social science methodological approach could be used to capture these standards. The results showed support for existing industry standards and also identified new

metrics, which collectively shape operations of pentesters. Both researchers and industry professionals might want to further explore the following ideas:

1. While the industry has developed a set of metrics, it should consider integrating these newly identified metrics into their frameworks and researching these collectively.
2. Researchers and industry professionals should examine the entire set of metrics to identify any connections between specific metrics, such as skillsets and quantity, in which certain skills could be correlated with a greater number of findings.
3. Research should test the generalizability of these collective metrics by studying whether they apply to other competitions, as in, examining whether these same types of metrics can be applied elsewhere.
4. Both parties can compare student experiences in pentesting to professional ones to determine if professional pentesters also prescribe to the metrics identified by the student participants in this study.
5. While this study used qualitative group interviews as its data source, analysis of the technical logs from the pentesting portion of the competition might reveal other technical metrics that could supplement both the industry and team-identified metrics.
6. These metrics can be used to study cybercriminal behaviors, organizational dynamics, and decision making in real-time cyberattacks.

The authors hope that these collective metrics will benefit cybersecurity competition organizers, cybersecurity researchers, and the pentesting community.

ACKNOWLEDGEMENTS

This research was supported by the NSF EAGER Award #1742747 and partially supported by the NSF CAREER Award #1453040. The authors thank the National CPTC for allowing data collection at their 2018 and 2019 events. Additionally, the authors thank all of the competing teams who agreed to be interviewed. Lastly, the authors thank their university's Institutional Review Board for ensuring that this study met ethical standards.

REFERENCES

[1] Xu, Z., Hu, Q., and Zhang, C. (2013). Why computer talents become computer hackers. *Commun. ACM*, 56(4), 64–74.

[2] Halfond, W. G. J., Choudhary, S. R., & Orso, A. (2011). Improving pentesting through static and dynamic analysis. *Software Testing, Verification and Reliability*, *21*(3), 195–214.

[3] Bacudio, A., Yuan, X., Chu, B., and Jones, M. (2011). An overview of pentesting. *International Journal of Network Security & Its Applications, 3(6).*

[4] Shah, S. & Mehtre, B.M. (2015). An overview of vulnerability assessment and pentesting techniques. *J Comput Virol Hack Tech,* 11, 27-49.

[5] Dimkov, T., van Cleeff, A., Pieters, W., & Hartel, P. (2010). Two methodologies for physical pentesting using social engineering. Proceedings of the 26th Annual Computer Security Applications Conference, 399–408.

[6] Geer, D., & Harthorne, J. (2002). Pentesting: A duet. 18th Annual Computer Security Applications Conference. Proceedings., 185–195.

[7] Herzog, P. (2010). OSSTMM 3. *Institute for Security and Open Methodologies.*

[8] Li, C. (2015). Pentesting curriculum development in practice. *Journal of Information Technology Education: Innovations in Practice, 14*, 85-99.

[9] GlobalCPTC. (n.d.). Retrieved January, from https://globalcptc.org/

[10] Holt, T.J. (2007). Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant behavior, 28(2),* 171-198.

[11] Hutchings, A. (2013). Hacking and fraud qualitative analysis of online offending and victimization. *Global Criminology: Crime and Victimization in a Globalized Era,* 93-112.

[12] Rege, A. (2016). Not biting the dust: using a tripartite model of organized crime to examine India's Sand Mafia. *International Journal of Comparative and Applied Criminal Justice, 40(2),* 101-121.

[13] Bachman, R., & Schutt, R. (2017). Fundamentals of Research in Criminology and Criminal Justice (4th ed.). Sage Publications.

[14] Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory. Chicago, IL: Aldine Publishing Company.