Using Capture-the-Flag to Enhance the Effectiveness of Cybersecurity Education

Kees Leune
Salvatore J. Petrilli, Jr.
leune@adelphi.edu
petrilli@adelphi.edu
Adelphi University
Garden City, New York, United States of America

ABSTRACT

Incorporating gamified simulations of cybersecurity breach scenarios in the form of Capture-The-Flag (CTF) sessions increases student engagement and leads to more well-developed skills. Furthermore, it enhances the confidence of students in their own abilities. Our argument is supported by a study in which undergraduate students taking a cybersecurity class were surveyed before and after participating in a CTF.

GENERAL TERMS

cybersecurity; education; capture-the-flag

1 INTRODUCTION

There is ample opportunity to improve the current state of cybersecurity defenses. We argue that many breaches are, at least partially, caused by human error, which, in turn, is regularly the result of insufficient and/or ineffective education [5].

In this paper, we argue that incorporating gamified simulations of cybersecurity breach scenarios in the form of Capture-The-Flag (CTF) sessions strongly support traditional lecture-style instruction, and that it significantly enhances confidence of students in their own abilities.

Three types of CTF games are generally distinguished: quiz-based, in which participants score points by answering questions; scavenger-hunt, or flag-based, in which participants locate and exploit vulnerabilities in systems security in order to gain access to files which contain “flags” in the form of random strings, and king-of-the-hill, or castle-based, in which participants score points by defending a server against attackers. The study presented in this paper combines quiz-based questions and flag-based questions, and does not include a king-of-the-hill component.

We present our statistical findings from a single observation. Although based on a single experiment, we hope that this study begins an on-going discussion and may lead to recommendations for further research regarding the benefits of gamification for students pursuing a career in cybersecurity.

The remainder of this paper is structured as follows: section 2 explains the rationale for our study and provides background. Section 3 described the methods and objectives of our study and section 4 presents our findings. Section 5 summarizes the main conclusions that can be drawn from our work, and section 6 shows possibilities for future research.

2 BACKGROUND

Popular media have been reporting extensively how, despite the availability of many highly advanced technical products and solutions, human decision-making is a key component of data breaches. The point is made very clearly in a survey of approximately 80,000 recent systems breaches that have been summarized in Verizon's annual Data Breach Investigation Reports [15, 16], in which human error, deliberate actions, or insufficient training and awareness are all identified as factors that lead to security breaches.

While these highly visible breaches have led to fragmented legislation [12] and increased awareness of the general public concerning the importance of achieving and maintaining cybersecurity, the methods for teaching cybersecurity at institutes for higher education have not kept up with the high pace of developments. In fact, human error, which may have been minimized through better education and training, has played some role in each of the breaches that were referenced.

Hence, since systems breaches are commonplace, and the root cause of many breaches involves human decision making, engaging in an attempt to improve human decision making skills with regards to cybersecurity problems is a worthy effort. One way by which human decision making skills can be influenced is through education. As such, the goal of this research project is to find new and innovative ways to engage student learning, especially in the arena of cybersecurity.

The problem of lacking education in cybersecurity is a global problem. In “The UK cyber security strategy: Landscape review” [9], the authors observed that it could take up to 20 years to address the skills gap at all levels of education.

Among the many top issues that are facing higher education, a skills gap between the outputs of colleges and universities and the needs of employers [1] is frequently mentioned. Higher Education is not always equipped to respond rapidly to changing demands in labor force skills, and the cybersecurity domain is no exception to that. By embracing traditional teaching methods, which may not align with the needs of potential employers, or with the emerging

SIGITE'17, October 4–7, 2017, Rochester, NY, USA.
© 2017 ACM. 978-1-4503-5100-3/17/10…$15.00
DOI: https://doi.org/10.1145/3125659.3125686

interests of students, colleges and university do not always produce the best outcomes.

Research shows that student motivation is a key predictor of successful educational outcomes. If students are intrinsically motivated to learn something, they may spend more time and effort learning, feel better about what they learn, and use it more in the future [8].

The most notable causes of student disengagement with learning include boredom, alienation, and disconnection between learning activities and real life application of knowledge [11]. Gamification has the ability to address many of these causes: it focuses students on relevant content, provides timely feedback, which improves retention, and supports students' multiple styles of learning [4].

Usually, there are two approaches to using games in education. The first approach seeks the engagement that commercial and widely available games have to foster learning outside the school environment. Games such as *Sid Meier's Civilization* or *World of Warcraft* can provide a challenging and motivating world that requires analyzing, planning, communication skills and others, contributing to improving the problem solving abilities of players. On the other hand, games can be specifically designed to convey traditional content in a different, untraditional form [7].

Intrinsic learning requires the embedding of learning outcomes of a teaching program within the mechanics of a game. It appears crucial that the task learned in the game maps directly onto the challenge faced in the real world [6].

Projecting these findings onto cybersecurity education, we posit that a carefully designed gamified cyberbreach simulation can be incorporated into the traditional university classroom, and used to capture interest and engagement by a variety of audiences. Furthermore, we believe that participation in these simulations can address the root causes identified by [11]. Specifically, we anticipate that students will be more engaged and have an increased understanding how their classroom skills related to realistic real-world scenarios. Consequently, we anticipate that learning outcomes using this approach enhance traditional teaching methods based on lectures followed by written exams.

3 METHODOLOGY

3.1 Research question and Hypotheses

In our study, we set out to answer the following research question:

Research Question. Does the inclusion of realistic simulations of offensive attack scenarios — in the form of capture-the-flag exercises — demonstrably increase the effectiveness of cybersecurity education?

In order to find an answer to this question, we defined a number of hypotheses:

Hypothesis 1. *Self-confidence of students will improve by participating in a capture-the-flag (CTF).*

We are interested in finding out if participating in the CTF helps students gain self-confidence in their skills. Our hypothesis is that participants will generally feel more accomplished and more secure in their abilities after having spent some time in a realistic, but controlled environment.

Hypothesis 2. *Students enjoy participating in a CTF.*

By introducing gamification components, such as the ability to choose 'hacker handles', a shared scoring system to encourage some friendly competition, and the ability to buy hints using points that were scored by solving previous challenges, we anticipate that students will enjoy participation in the CTF. Our expectation is that, by active participation, students will spend more time learning and develop stronger outcomes.

Hypothesis 3. *Students develop stronger practical skills by participating in the CTF.*

Reinforcing theoretical knowledge of cybersecurity methods and techniques by experimenting with them in a controlled environment is expected to reinforce and/or develop strong practical skills in students.

Hypothesis 4. *Participating in the CTF reinforces theoretical concepts.*

Similar to reinforcing practical skills, we expect that participation in the CTF will solidify student's understanding of general concepts and improve learning outcomes.

3.2 Participants

This study was conducted in an elective undergraduate course entitled 'Cybersecurity', which had $N = 24$ students enrolled. Students were asked to fill out a consent form indicating their willingness to participate in this study. A sample of ($n = 10$) undergraduate students agreed to complete a baseline questionnaire after a theoretical block had been completed, but before the block was reinforced using a two-week long capture-the-flag scenario. The questionnaire consisted of seven Likert-scaled questions (scale: 1–5) and 11 free-response questions. In the free-response questions, participants were asked to properly phrase definitions of security-related terms and techniques. After completion of the questionnaires, these questions were subsequently graded out of five points to facilitate analysis.

A large part of our students consisted of male (90%) junior-level (47.6%) and senior-level (33.3%) undergraduate computer science (76.2%) majors. However, since this was a class without prerequisites, other majors (Physics, Mathematics, Business Administration and Information Systems) also participated.

3.3 Materials and Procedures

In the questionnaires[1] we assessed knowledge of the students concerning different attack vectors, and we asked them to identify how confident they were in their skills to recognize, execute and prevent attacks.

Students were provided access to a virtual network, on which a variety of (virtual) servers had been installed. In all, students were tasked to gain access to all servers that they could find, identify flags, elevate privileges to achieve administrative access, and provide a writeup of their activities in which they summarized their case notes and provided recommendations for hardening the environment so that the techniques that they successfully used could not be reproduced.

[1] Questionnaires available via http://cs.adelphi.edu/~leune

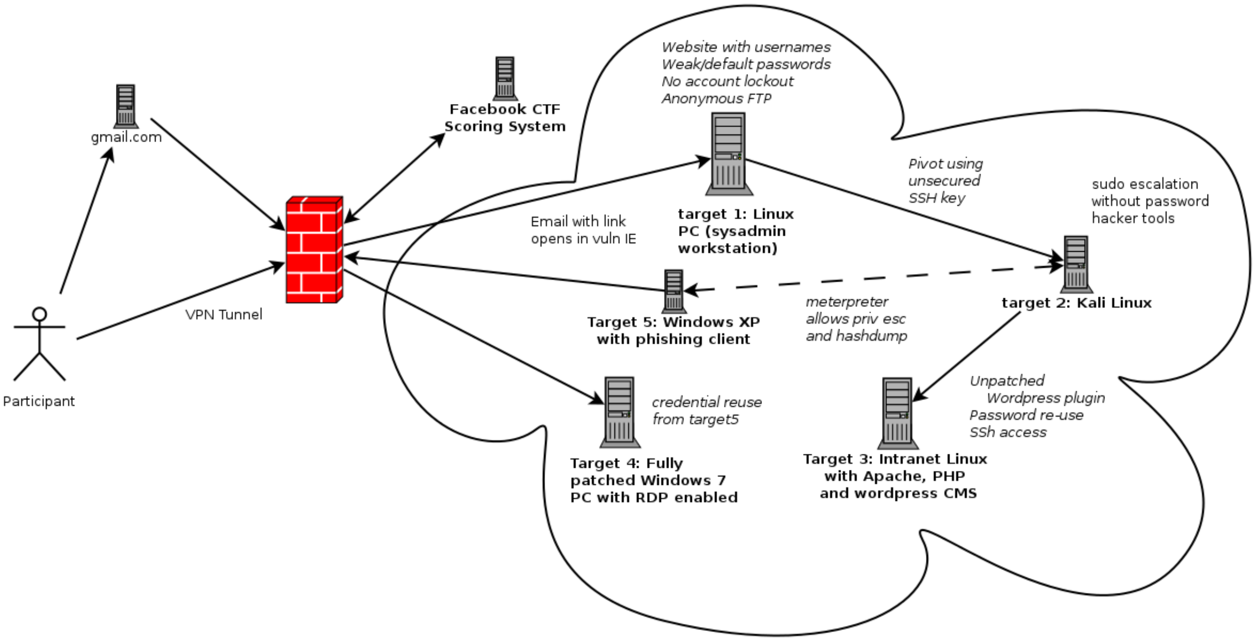

Figure 1: Capture-the-Flag overview

As shown in Figure 1, the virtual network was only accessible via a Virtual Private Network (VPN) tunnel and contained six target virtual machines, running a combination of Ubuntu Linux, Microsoft Windows 7 and Microsoft Windows XP.

With minimal additional instruction, students were asked to play out a variety of scenarios, ranging from simple default passwords (target 1), which allowed access to a server from which students can then pivot their attack by using an unprotected secure shell (SSH) key to pivot their attack to a Kali Linux server running a variety of offensive tools (target 2), to an unpatched Wordpress plugin vulnerabilities hosted on target 4, which re-used the same password for the database access and access to an account capable of elevating privileges using the Linux sudo command.

Furthermore, in an attempt to mimic realistic contemporary attack scenarios, we created a small program that would download email via the POP protocol so that passwords could be intercepted using a packet sniffer (target 5). The server on which the program ran according to a predetermined schedule actively filtered inbound traffic and was offering no externally available services.

As an added twist, the email download program ran on a vulnerable Microsoft Windows XP virtual machine that would open any URLs using Internet Explorer that were contained in the incoming email. Students were able to craft messages and send them from their regular email accounts. Upon receipt, the URL that was opened resulted into a downloaded exploit, which would give students access to the target. On that target, elevating privileges and obtaining a copy of the encrypted password store (SAM) was a logical next step.

Flag category	Amount	Percentage
weak configuration	5	26%
reconnaissance	4	21%
vulnerability	4	21%
password	3	16%
privilege escalation	2	11%
phishing	1	5%
	19	100%

Table 1: Flags per category

The encrypted passwords can be cracked off-line, yielding additional credentials that can were re-used on a fully patched Windows 7 virtual machine, which permitted Remote Desktop Connections.

As shown in Table 1, students were able to collect 19 flags, ranging from easy ones, to ones that required lateral attacks and privilege escalation. The flags were placed on to seven virtual machines, running three different operating systems, and students were able to gain access to and escalate privileges on five of them. Samples of each flag category are described in Table 2.

Vulnerabilities introduced in this simulation mostly focused on insufficiently patched software, password re-use, weak and default passwords, phishing, and weak operational security practices.

In addition to recognizing and exploiting vulnerabilities, students were asked to answer a range of quiz questions. In the CTF design, the flags questions were intended to reinforce skills, while the quiz questions were designed to reinforce knowledge. The quiz questions area broken down by category in Table 3.

Flag category	Sample challenge
weak configuration	Target 2 (Kali Linux) allowed remote login using an unprotected SSH key.
reconnaissance	Target 1 published a staff directory, including usernames, email addresses and job titles to an unprotected website.
vulnerability	Target 3 was running Wordpress with an vulnerable plugin that allowed unauthenticated remote code execution.
password	User on target 1 had easy to guess password.
privilege escalation	User was allowed sudo without password challenge.
phishing	Emailing a special account with trigger opening an arbitrary web page using a vulnerable web browser.

Table 2: Sample questions per flag category. Network architecture depicted in Figure 1

Quiz category	Amount	Percentage
news & general	6	35%
networking	6	35%
definitions	2	12%
attacks	1	6%
cryptography	1	6%
law & ethics	1	6%
	17	100%

Table 3: Quizzes per category

Table 4 includes samples of the quiz questions that students were asked. In scoring, quizzes were generally awarded one point, while flags commonly yielded five points.

Gamification components were introduced by encouraging students to work in competitive teams. A shared score board was kept using Facebook's CTF platform, a publicly available scoring system [2].

4 FINDINGS

Based upon the responses to the pre and post questionnaires, we were able to examine outcomes from participating in the capture-the-flag. Here we share some of our observations. We should note that majority of our results come from an analysis of the post-assessment questionnaire.

We found a direct correlation between the level of enjoyment that participants reported after participation in the CTF activity and an increased confidence in their ability to execute attack methods that were simulated during the exercise. Furthermore, we found that an increased confidence almost directly translates to increased outcomes.

4.1 Strong Confidence in Abilities

[10] found that gamification in the educational setting has the added benefit of increasing student self-esteem with respect to the content area. Our numbers support that statement; we posit that a gamified environment contributes to increased participation, which, in turn, leads to a significant increase in the participant's confidence and abilities to execute cyberattacks.

Our surveys showed that, prior to participating in the activity, students expected to enjoy the exercise. The post-activity assessment demonstrated that students were not disappointed; the data indicates that participants clearly enjoyed taking part in the exercise. This is evident most strongly in the responses to question 7 *Did you enjoy participating in a hands-on capture-the-flag exercise?*, in which students reported a strong affirmative median score of 5 out of 5 in the post-CTF assessment questionnaire.

In addition, our repeated-measures study discovered a significant correlation between a participant's enjoyment in participating and their self-confidence in their ability to perform cybersecurity defense tactics.

The participant's new-found confidence is most strongly illustrated by the answers to question 1 from the pre- and post-questionnaire (*How confident are you in your abilities to execute a typical attack that follows the phases discussed in class?*). A Wilcoxon test found a significant increase in the mean ranks in the repeated-measures study ($Z = 2.719, p < .05$).

All confidence-related questions were significantly correlated, suggesting that confidence is built broadly, and includes confidence in participants' ability to execute, recognize and defend against cyberattacks.

In particular, question 1 was significantly correlated with question 2 *How confident are you in your abilities to recognize a typical attack that follows the phases discussed in class? You may assume that you have sufficient visibility into the infrastructure?*, question 3 *How likely are you to execute certain attack types after reading about them? You may assume that you have permission to do so, and a platform to conduct the attacks on?*, question 4 *Do you feel prepared to defend against real attacks on an actual enterprise network?*, and question 6 *Did participating in a hands-on capture-the-flag exercise enhance your understanding of how attacks are conducted?* ($r = +.716$ and $p < .05$, $r = +.779$ and $p < .05$, $r = +.824$ and $p < .05$, $r = +.857$ and $p < .05$, respectively).

Additionally, we found that students gained an increased appreciation for learning about new cyber-defense techniques. In particular, the answers to question 5 *How prepared are you to keep up with learning about new vulnerability types and attack trends?*, and question 6 *Did participating in a hands-on capture-the-flag exercise enhance your understanding of how attacks are conducted?* were significantly correlated with question 7 *Did you enjoy participating in a hands-on capture-the-flag exercise?* ($r = +.672$ and $p < .05$, and $r = +.701$ and $p < .05$, respectively).

These strong results were matched by students' self-assessments. In particular, question 2 from our survey *How confident are you in your abilities to recognize a typical attack that follows the phases discussed in class? You may assume that you have sufficient visibility*

Quiz category	Sample question	Answer
news & general	What global organization of volunteers aims to advance the state-of-the-art of application security by publishing tools?	OWASP
networking	What protocol is expected to be found in a network flow that involves TCP port 25?	smtp
definitions	In which access control model is the creator of the resource the owner?	discretionary
attacks	What kind of attack will try all possible combinations of a key/passphrase?	bruteforce
cryptography	What hashing algorithm was recently proven to be insecure?	SHA1
law & ethics	Which if the following is a federal law enforcement agency: CIA, NSA, DHS, or US Secret Service?	US Secret Service

Table 4: Sample questions per quiz category

into the infrastructure. was scored with a median score of 4 in the post-CTF assessment, and it was significantly correlated with question 4 *Do you feel prepared to defend against real attacks on an actual enterprise network?* (median: 2.5), and question 6 *Did participating in a hands-on capture-the-flag exercise enhance your understanding of how attacks are conducted?* (median: 4.5) The correlations were $r = +.703$ and $p < .05$ and $r = +.773$ and $p < .05$, respectively.

4.2 Strong Observed Outcomes

Participants spent more time on the CTF than they anticipated: in the pre-assessment, an anticipated median time commitment of ten hours was reported, but in the post-assessment, students reported having taking a median 20 hours to participate.

Their efforts paid off. After completing the post-CTF assessment questionnaire, participants generally scored high. Overwhelmingly, participants were able to define and explain the consequences of password re-use (median: 5), phishing (median: 4) and weak configurations (median: 4). Significant increases in outcomes were observed in participants' ability to describe the risks of using weak passwords (median: 4).

Interestingly, when examined jointly, the quiz and flag components were not contributing to the post-assessment score. However, when examined separately, they were contributors to the post-assessment score. This indicates that quizzes and flags measures two different educational aspects: quizzes measure terminology and flags measure application.

A multiple linear regression model was constructed to predict participants' post assessment score on their quiz score and categorization of the quiz topics. A significant regression equation was found ($F(2, 36) = 7.374$, $p < .05$), with an R^2 of .291. Participants' predicted post score is equal to 31.188 + .968(QUIZSCORE) + .418(CATEGORY). Participants' post score increased by a factor of .968 for every point they scored on a quiz in the CTF. Both QUIZSCORE and CATEGORY significant contributors to the regression model.

Additionally, a multiple linear regression model was constructed to predict participants' post assessment score on their flag score and categorization of the flag topics. A significant regression equation was found ($F(2, 36) = 6.412$, $p < .05$), with an R^2 of .263. Participants' predicted post score is equal to 31.831 + .145(FLAGSCORE) + .332(CATEGORY). Participants' post score increased by a factor of .145 for every point they scored on a flag in the CTF. Both FLAGSCORE and CATEGORY are significant contributors to the regression model.

These findings are consistent with those that were found by [3], who found that gamification in the classroom makes computer science education more interesting and effective for students.

4.3 Reality vs. Perception

One post-assessment question asked students to describe the risk of weak passwords, and to provide suggestions how to mitigate the risk. One students answered that "weak passwords are a vulnerability caused by the use of a generic, default, or otherwise easily guessed password for an account. These can include common passwords (e.g., password, qwerty) or easily accessed personal information (e.g., last name, birth year, pets name). This is dangerous as there are methods where you can guess millions of passwords for a single account and a weak password can be easily guessed/cracked. You can prevent this by educating users on what a secure password should be and also implementing password controls like length on users when they are creating passwords."

This answer is interesting, given that the issue of weak passwords (password=*password*), as well as password re-use were central to successfully completing the simulation exercise. This answer clearly indicates that the student understood this aspect of securing a system.

In another question, students were asked to describe *port scanning*, explain when the technique would be used, and what preventative measures can be taken. One student responded that port scanning happens "when someone scans a computer on a network to determine what ports are open on that computer. This will help determine what protocols are running behind those ports, and in turn what software is running behind that protocol. Nmap is a popular tool. To prevent this, you can monitor incoming traffic and blacklist IPs that try scanning too many nodes. You can also prevent direct access to your network from the outside through a firewall/IPS."

Since the students were provided access to the test network without knowing any additional information, the first step they were expected to perform was to conduct a portscan. This student clearly understood not only how to run a tool (nmap), but also understood its significance.

Other questions were not answered as well as these two examples were. For example, when asked to discuss lateral attacks, which were not prominently present in the CTF, answers ranged from ones that exhibit little or no understanding ("Always keep your eyes on the server. Investigate suspicious IP.") to ones that indicate a more sophisticated view ("Using a vulnerable computer to branch over to another computer on the same network. getting into a 3rd party company to attack the target company through the unprotected connection. Can be defended by implementing controls all along the network.").

Most students were able to successfully execute the attacks that were prominently features in the CTF, which appears to support the hypothesis that participation in CTF-like exercises positively impacts the ability to execute and recognize offensive cyber techniques.

The responses showed clear correlation between the perceived level of understanding (questions 9–18) and enjoyment (question 7) that students obtained from participating in the CTF and the actual understanding of vulnerability types *weak configuration*, *weak passwords* and *unpatched software vulnerabilities*, as evidenced by their ability to define and describe these vulnerability types.

The significant correlation between increased confidence in the ability to recognize and the ability to respond to cyberattacks is relevant, since these vulnerability categories were prominently present in the CTF exercise. Students were asked to recognize and exploit vulnerabilities in Symposium, a Wordpress Plugin [14], as well as in Microsoft Windows XP [13].

5 CONCLUSION

Our first hypothesis was that self-confidence of students will improve by participating in a capture-the-flag (CTF). The study confirmed that this is indeed the case. Having the ability to practice — potentially dangerously — techniques in a controlled environment solidified students' confidence in their own ability to execute, recognize and defend against attacks.

Our second hypothesis was that students enjoy participating in a CTF. First and foremost, one of the main conclusions from this experiment was that students overwhelmingly enjoyed participating in the exercise. Most students were very engaged and spent a significant amount of time in trying to solve the exercises. Research has shown that engaged students generally have improved learning outcomes; this study confirmed that CTF-like events positively contribute to student's enjoyment in the course work and their engagement with it.

While perception and enjoyment are important factors in teaching, the end-result is what really matters. Our third hypothesis was that students develop stronger practical skills by participating in the CTF. This was confirmed as well. Student's understanding, as measured by the post-CTF assessment, of attack scenarios that were played out during the CTF was higher than the scenarios that were not.

Lastly, our fourth hypothesis was that participating in the CTF reinforces theoretical concepts. This hypothesis was less clearly proved. While post-CTF assessment outcomes were generally high, they were not necessarily significantly higher than during the pre-CTF assessment. It appears that a CTF in which flags (skills assessments) and quizzes (knowledge assessment) are combined do not have a strong positive outcome on learning theoretical concepts. However, based on the data presented in section 4.2, we suspect that separating exercise in a knowledge-based CTF and in a skill-based CTF will lead to improved outcomes.

Having tested out hypotheses, we can now answer our research question: "Does the inclusion of realistic simulations of offensive attack scenarios — in the form of capture-the-flag exercises — demonstrably increase the effectiveness of cybersecurity education?" Taking into consideration what we learned, we must answer this positively. CTF exercises increase the effectiveness of cybersecurity education, provided that they are designed appropriately. If the desired learning outcomes include strong practical skills, a CTF that is flag-based is good use of time. Combining flag-based questions and quiz-based questions in a single CTF may be less effective.

6 FUTURE WORK

Future research can and will hopefully entail collecting longitudinal data on the effects of gamification in a cybersecurity major/track. It is still to be determined which of the suggested course content and activities are most effective. One way to approach a broader study might include tracking incoming freshmen and the influence of gamification until the end of their senior year. By means of such a study, researchers might be able to determine if gamification has statistically significant influence in better preparing students to enter the field of cybersecurity. In addition, a larger quantitative and qualitative study across areas in computer science could prove illuminating.

Furthermore, investigating under which circumstances different CTF forms (quiz, scavenger hunt, king-of-the-hill) are most appropriate is a worthwhile endeavor.

REFERENCES

[1] Ebersole, J. (2014). Top issues facing higher education in 2014. *Forbes*.

[2] Facebook (2017). Facebook CTF Platform. Available at https://github.com/facebook/fbctf.

[3] Firdausi, N., Prabawa, H., and Sutarno, H. (2017). Improve student understanding ability through gamification in instructional media based explicit instruction. *IOP Conference Series:* Journal of Physics.

[4] Geelan, B., de Salas, K., Lewis, I., King, C., Edwards, D., and O'Mara, A. (2015). Improving learning experiences through gamification: A case study. *Australian Educational Computing*, 30(1).

[5] Liginlal, D., Sim, I., and Khansa, L. (2009). How significant is human error as a cause of privacy breaches? an empirical study and a framework for error management. *Computers & Security*, 28:215–228.

[6] Linehan, C., Kirman, B., Lawson, S., and Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1979–1888.

[7] Lopes, R. (2014). Gamification as a learning tool. *International Journal of Development and Educational Psychology*, 2:565–574.

[8] Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 4.

[9] National Audit Office (2013). The UK cyber security strategy: Landscape review.

[10] Richter, R., Raban, D., and Rafaeli, S. (2015). Studying gamification: The effect of rewards and incentives on motivation. *Gamification in Education and Business*.

[11] Shernoff, D., Csikszentmihalyi, M., Schneider, B., and Shernoff, E. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, 18(2).

[12] Stevens, G. (2012). Data Security Breach Notification Laws. Technical report, Congressional Research Service.

[13] The MITRE Corporation (2008). CVE-2008-4250. Available from MITRE, CVE-ID CVE-2008-4250.

[14] The MITRE Corporation (2014). CVE-2014-10021. Available from MITRE, CVE-ID CVE-2014-10021.

[15] Verizon Business (2015). 2015 data breach investigations report.

[16] Verizon Business (2016). 2016 data breach investigations report.