Capturing Capture the Flag

Further Discussions

MARK GONDREE

Mark Gondree is a security researcher with an interest in games for education and outreach. With Zachary Peterson, he released [d0x3d!], a board game about network security to promote interest and literacy in security topics among young audiences. Gondree is a research professor at the Naval Postgraduate School in Monterey, CA. gondree@gmail.com

Andy Davis is a member of the Cyber Systems Assessment Group at MIT Lincoln Labs. He has helped organize the MIT/LL CTF competition for the last two years and several mini-CTF events at universities in the northeast. Andrew.Davis@ll.mit.edu

Chris Eagle is faculty at the Naval Postgraduate School. He led teams winning DEFCON's CTF competition in 2004 and 2008, and then organized DEFCON's CTF for the next four years, 2009–2012. He is currently designing and organizing DARPA's Cyber Grand Challenge competition. cseagle@nps.edu

Peter Chapman is a graduate student at Carnegie Mellon. He was the first technical lead for picoCTF, an online competition started in 2013 for high school students. He also worked on an attack-defense CTF for the US service academies earlier this year, called IOCTF. peter@cmu.edu

This year, the first USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE) was held, co-located with USENIX Security '14. The summit challenged designers, organizers, gamers, and educators to consider how we assess and improve the current state of security games, both in and out of the classroom.

3GSE featured a panel devoted to capture the flag (CTF) competitions and their use in education, bringing together a diverse group of stakeholders interested in how we both run and evaluate those games. The discussion expressed a fascinating mix of hacker values, student-centric learning approaches, and technical issues inherent to running these complex competitions. I had the opportunity to follow up with our panelists—Peter Chapman, Andrew Davis, Chris Eagle, Portia Pusey, and Giovanni Vigna—to reflect on the highlights of that discussion.

MG: The term "capture-the-flag" has expanded through use. Some might say it has been diluted. Is this confusing? What terminology to distinguish between games seems most useful?

AD: At our CTF, we've had problems with people expecting a type of weekend-long hack-a-thon, where everyone has a project to work on. Some of those types of competitions are advertising themselves as CTFs, so the term is certainly becoming diluted. But we borrowed the term from another game. We've had people show up to our CTF in shorts and a t-shirt, and expect to run around a field stealing flags. So it's partially our own fault.

PC: Within the community, there are some recognized categories but there is a lot of diversity. They'll describe the CTF as: attack-defense, where multiple teams attack each other; *Jeopardy*-style, which is not a great name but refers to challenge-based competitions; and there are war games, which are basically *Jeopardy*-style games that persist, so students can go through challenges and educate themselves at any time. Smash the Stack (http://smashthestack.org/) is one such war game. While these categorizations help people know what to expect when they participate, trying to find new terms to recognize "hidden gems" that fall outside these categories will be useful, going forward.

PP: I would add to that list "inherit and defend" competitions. We also need to distinguish CTF from some specialized games such as *Jeopardy*-style, forensics, and cryptography challenges. Working groups at the Cybersecurity Competition Federation (http://nationalcsf.org/) have begun to brainstorm and define the diverse competition formats.

CE: What I think is lacking is a categorization that communicates the goals of the organizers. Pure competition play, like DEFCON CTF, appeals to a kind of audience, where the goal is to crown or rank the competitors. But there are other CTFs whose goals are more aligned to education. For these, the value may be in post-exercise debriefings and walkthroughs, which you don't necessarily get from a purely competitive league.

Portia Pusey is the project manager and education chair for the National Cyber League (NCL). She is part of the research team at the National CyberWatch Center (NCC) who published and presented on the research conducted for the National Cyber League. She is also leading a study of National Collegiate Cyber Defense Competition (CCDC) coaches and players for the NCC Research Team. edrportia@gmail.com

Giovanni Vigna is a professor at UC Santa Barbara, and the director of its Center for CyberSecurity. Since 2003, he has organized the annual International Capture the Flag competition. iCTF is, today, the largest regularly recurring CTF in existence. Play is open to teams from two- and four-year universities. vigna@cs.ucsb.edu

GV: I agree. My gut feeling is that there are many students who would be excited to compete in CTFs but, because we lack advertised goals and expectations, they are scared of participating. Categories that communicate the "roughness" of the game and level of support provided to the players would help. In general, I would call CTF only competitions in which the teams both attack and defend an asset. I would use other terms for other types of competition (such as "hacking competition" for a challenge-based competition).

MG: During the panel, it became clear that some game designers got negative feedback from the community, for failing to be inclusive of professionals or non-US students. What are the limits of inclusivity in CTFs?

PC: When the people running and supporting the game only know English, that becomes a real barrier to supporting schools in some countries.

PP: Each competition can't be everything to everyone; there is a very important reason for that. We need to give young and novice learners a safe legal place to practice. We can't leave them in the "Wild West." Many high-schoolers that I work with want to build their reputation, and they're getting into trouble. Furthermore, it's not appropriate for minors to interact in the same game environment as adults. And beginners may become disengaged when they have to compete against experts. And, most importantly, games designed to be used in K12 school settings need to protect students' identities and control the types of interactions they have with other players to keep them safe. This comes at the risk of excluding people from games.

MG: How do we build classes around competitions?

CE: For me, I've always found it easier to run a CTF extracurricular activity year-round. Students come and go and may not be able to play year-round, but the timing of these things all over the world is unpredictable and not in harmony with the academic calendar.

AD: We had a professor at a local university use a CTF for their final. I think if you force your students to do a 48-hour, non-stop final where they get beat on, continuously, by more experienced players, that seems pretty cruel.

GV: I disagree with that! I started iCTF in 2001 as an in-class "attacker versus defenders" game, but the defenders claimed it wasn't fun enough. In 2002, we changed the format to "attack-defend," and in 2003 we opened it up to other universities. It has always been in December, as the "final" for my Fall class. I like the idea of taking a student who knows nothing about security, and building up to running them through a competition, and they enjoy it.

PP: I think it's valuable to separate the idea of competitions that are *educational* and competitions that are designed to be used as *education*. For example, the CCDC is a competition that can frustrate the players to tears. And at the same time, the players say it's the best learning experience they've ever had. So, competitions can be educational. But if a competition is going to be used as formal education there are different requirements. Educators need measurable objectives, and the scoring system needs to provide evidence about the learner's progress towards achieving those objectives. Educators need to know what prerequisite skills are required, and what evidence demonstrates that a learner is ready for the competition challenges. Finally, generally speaking, our nation does not have the capacity among educators in the K12 space to teach cybersecurity. Therefore, we need to provide support in the form of background materials and training for the educators to be able to effectively integrate competitions into their classroom teaching.

MG: At 3GSE, we heard that some designers are facing pressure to remove scores, make games non-competitive, or turn competition into a series of tutorials. The thinking, in particular, is that competitive play may be uncompelling to women. Can we do better outreach by changing design?

PP: If I may speak for all of womankind, we are all different; and some of us are highly competitive. Changing the rules or structure of current competitions is not going to work. It seems to me that evidence from the gaming industry and the current landscape of competitions demonstrates that the games/activities/challenges/tasks do not interest most women. The lesson we learn from the gaming industry is that when a game provides challenges that women want to do, they play. So, the trick to engaging more women in cybersecurity competitions is finding competition tasks that women want to do.

GV: Making CTFs non-competitive is not a good solution to inclusivity. The allure of many CTFs is in their competitive and underground atmosphere. We should seek to draw women to CTFs by finding ways to include more women in computer science. I'm not sure this is a CTF problem but, rather, a larger and more systemic issue facing our field.

CE: We can't expect to have substantially higher percentage of women participating in CTFs than are present in computer science as a whole. The right answer is that we need to address that problem, and then we can start to see participation from women in every aspect of the field and not just CTFs.

PC: There is a difference between building a competition that has been "toned down" to make everyone happy and making a competition where we've removed various barriers to entry. There are ways to build challenges so that they don't hamper the competition but are accommodating to players with less experience. In PicoCTF, for example, some simple challenges have tutorials accompanying them. Experienced players were able to skip the tutorials and solve the challenges quickly. To people with no experience, this was some of their favorite sections. They raved about them, and how they had the support to participate in something they really found interesting. Those teams still didn't score very high in the competition, but they didn't care: they walked away with a very positive experience rather than a feeling that this was too hard for them. We hope to see those players again next year, where they may be able to solve more challenges independently. Removing barriers and letting novices participate in some form is one way to increase our diversity.

PP: Efficacy research among the underrepresented in STEM indicates that, if learning or competition experiences provide a developmental sequence of successes, achievement and interest in STEM majors and career paths increases. One study documented that a four-week intervention designed to build efficacy helped girls to overcome societal messages and similar pre-existing notions that "women don't or can't do that"; whatever *that* may be. The Cybersecurity Competition Federation (http://nationalcsf.org/) is an umbrella organization for competitions; they are building a "pathway" of competitions so that students can identify their point of entry in a continuum of competitions, based on their skill level and interest. It will be interesting to see if this supports greater diversity among players.

MG: If CTFs should be competitive and scored, is there value in tying together team performance across games? Or do we need to measure something in addition to the score?

GV: I think that competitions need to have a ranking, but the value is not in the ranking. The value is in the preparation and the active engagement in the game. One thing we are bad at is evaluating the effort leading up to a competition. We measure the moments in the competition when we are there, but it would be great as educators to find a way to assess their progress.

CE: CTF Time (https://ctftime.org/) tries to aggregate information about competitions and weight the games. The organizers have a pretty lofty goal to fill the gap as a team ranking body, but it's hard to do right. CTF teams change and there is no real way to compare competition A to competition B in the absence of a standards body.

PP: The Collegiate Cyber Cup (http://collegiatecybercup.org/) is a national award that uses an algorithm to calculate an individual's score based on their performance in multiple competitions (team or individual). It is similar to NASCAR in that points from different competitions are aggregated to identify a national winner. This idea has merit because it provides formative feedback to the player and quantifies performance for future employers. The algorithm is a clever approach to building a national score from the siloed system of competitions we have now. However, I would like to see a common metric that can be used across all competitions based on tasks from a rigorous job performance model such as the one created for the Department of Energy (https://www.controlsystemsroadmap.net/efforts/Pages/SPSP.aspx).

MG: What factors are most important when designing scoring for a game?

GV: It's a competition and players want to win, so they are very invested in scoring. Most fundamentally, scoring needs to be clear. Most of the times players have been hurt due to scoring, the culprit has been lack of clarity. The scoring rules need to be transparent, and the scoring mechanism needs to be automated, requiring no human or qualitative judgment.

CE: If you are going to award a prize, your scoring mechanism better be pretty stinking good.

AD: Our CTF group has likely spent more time talking about scoring than anything else, and it's a very hard topic. You have to establish what you want to evaluate. It also needs to be simple enough so that there are no surprises. Anyone should be able to look at the scoring algorithm and see that, if I'm doing what I should be doing in the game and I'm doing better at it than anyone else, then my score should be higher.

CE: When you communicate a scoring mechanism, it should be clear to a player what they should try to optimize to win the game. You may think that good defense could win a game, but when you study the scoring metric then you may find that really offense was what you needed to be doing. Whatever you use as a scoring metric, you better be able to measure it reliably and accurately.

PC: From an educational perspective, we would rather avoid the scenario where someone spends two days trying to hack some binary and doesn't get any points out of it because he couldn't get the last few bits. That's a very frustrating experience and, if we want people to keep learning, we don't want to frustrate them to the point where they stop and leave.

MG: How do we use design or scoring to scaffold challenges, to draw players into developing skills?

PC: In PicoCTF, we had leveled challenges. Our buffer overflow challenges were all discrete problems, but they built on each other, in terms of complexity. There was no downside to the simpler challenges that provided scaffolding: Someone who was very experienced breezed through the simpler challenges and, if anything, it made them feel great for solving five challenges back-to-back. For people who are learning, this disentangles complex problems into more isolated skills.

GV: There may be opportunities for giving partial credit for, say, crashing a program in a predictable way rather than demonstrating a full ROP attack; however, creating alternative goals for partial credit on challenges would needlessly complicate scoring and the game. Rather, the solution for scaffolding is making a variety of simpler challenges.

AD: Each of those smaller challenges will be pretty binary in how they can be scored. Either you've achieved the goal by demonstrating the skill, or you haven't.

MG: It sounds like the types of challenges and the algorithm for scoring communicates a set of values, and has the ability to guide novice players to learn or exercise one set of skills over another. As designers, what do you hope players walk away from the game having achieved or learned?

PC: For PicoCTF, we tried very early to develop a list of skills we hoped to build, but we eventually decided no single set of skills was more valuable than the goal of instilling a curious mindset and a sense of empowerment. We wanted students to question everything, as in the mindset of a computer security expert: You don't trust what people tell you; you don't trust the implementation; instead, you test it and you explore. Additionally, we wanted to empower students to tackle new challenges. Instead of seeing a problem and thinking, "I've never learned this before and that's the teacher's fault and I'm not going to do this anymore," we wanted to instill the sense that everything in the competition will be new to you, and you are going to teach yourself all of it, and you are going to be able to do it. Our game's first challenge was an obscure boot error that essentially required you to find the answer on the Web.

CE: As in teaching my class, my primary goal is demystification: demystify the hardware, demystify the software, remove whatever misconceptions students may have, and empower them to delve more deeply into problems independently.

AD: One thing we did at our CTF, after the competition but before we announced a winner, we had each team spend 30 minutes to make a five-slide presentation. They summarized their offensive strategy, their defensive strategy and gave a rough overview. Each team presented five slides in five minutes; they really got to see, for example, that half the teams had firewall rules that just dropped any packet with five As in row. So, if you wanted your attacks to work, you could have just switched your As to Bs. That reflective period and information sharing has been valuable.

MG: What's next? What is the most pressing need in terms of getting better CTFs?

PP: One of the first problems we need to solve for competitions to be used in education is to make them less time-consuming to create. Once a competition has been played, the solutions are known. All new challenges need to be created for the next game. Until we can solve that problem, it's hard to really tackle problems like scaffolding and scoring.

PC: We thought it would be valuable to release the tools we used to host PicoCTF as an open source project. It has increased the diversity of the competitions. For example, one of the teams playing in PicoCTF took our code and hosted their own nation-wide CTF for high school students called HSCTF (http://hsctf.com/). Their twist was to expand the game beyond computer security challenges, to include problems from Project Euler (https://projecteuler.net/).

GV: Ideally, you should be able to just go to a Web site hosting challenges, select the challenges you want, and get VMs you can just spin out to assemble your CTF. In fact, this is exactly what we've recently released in beta, as the iCTF Framework (https://ictf.cs.ucsb.edu/#/framework).

AD: One of the problems is that CTF infrastructure developers are not software engineers. Every CTF we've run has incorporated new features that have required pretty experimental software. Our latest CTF heavily employed a new Android emulator that we had just built. There wasn't an existing, mature product to do these experimental CTF challenges. That may be, in part, because there's no real financial support for building and running these games.

CE: Well, there is a market for internal CTFs, where a company will invite an organizer to run a CTF, for training or team building. In the open, however, most people don't receive compensation for running a CTF. It's not pay to play, and the compensation for organizers is rare.

GV: In releasing our framework, our dream is, eventually, to be able to crowd-source the development of vulnerable services, which is the part of design that requires the most human resources. Once you have the infrastructure more or less right, the services are still the things that take time.

usenix

THE ADVANCED COMPUTING SYSTEMS ASSOCIATION

Publish and Present Your Work at USENIX Conferences

The program committees of the following conferences are seeking submissions. CiteSeer ranks the USENIX Conference Proceedings among the the top ten highest-impact publication venues for computer science.

Get more details about each of these Calls for Papers and Participation at **www.usenix.org/cfp.**

SREcon15
March 16–17, 2015, Santa Clara, CA
Submissions due: January 5, 2015

The second SREcon will take place on March 16–17, 2015, in Santa Clara, CA. Save the date and come join us for two days of highly technical subjects around site reliability and production at scale.

HotOS XV: 15th Workshop on Hot Topics in Operating Systems
May 18–20, 2015, Kartause Ittingen, Switzerland
Submissions due: January 9, 2015

HotOS XV will bring together researchers and practitioners in computer systems, broadly construed. Continuing the HotOS tradition, participants will present and discuss new ideas about systems research and how technological advances and new applications are shaping our computational infrastructure.

2015 USENIX Annual Technical Conference
July 8–10, 2015, Santa Clara, CA
Submissions due: February 3, 2015

USENIX ATC '15 will again bring together leading systems researchers for cutting-edge systems research and unlimited opportunities to gain insight into a variety of must-know topics, including virtualization, system administration, cloud computing, security, and networking.

HotCloud '15: 7th USENIX Workshop on Hot Topics in Cloud Computing
July 6–7, 2015, Santa Clara, CA
Submissions due: March 10, 2015

HotCloud brings together researchers and practitioners from academia and industry working on cloud computing technologies. Cloud computing has gained traction over the past few years, becoming a viable alternative to dedicated data centers and enabling the launch of many prominent companies. However, many challenges remain in the design, implementation, and deployment of cloud computing. HotCloud provides a forum for both academics and practitioners to share their experience, leverage each other's perspectives, and identify new/emerging "hot" trends in this important area.

HotStorage '15: 7th USENIX Workshop on Hot Topics in Storage and File Systems
July 6–7, 2015, Santa Clara, CA
Submissions due: March 17, 2015

The purpose of the HotStorage workshop is to provide a forum for the cutting edge in storage research, where researchers can exchange ideas and engage in discussions with their colleagues. The workshop seeks submissions that explore long term challenges and opportunities for the storage research community. Submissions should propose new research directions, advocate non-traditional approaches, or report on noteworthy actual experience in an emerging area. We particularly value submissions that effectively advocate fresh, unorthodox, unexpected, controversial, or counterintuitive ideas for advancing the state of the art.

LISA15
November 8–13, 2015, Washington, D.C.
Submissions due: April 17, 2015

USENIX's Large Installation System Administration (LISA) conference—now in its 29th year—is the premier conference for IT operations, where systems engineers, operations professionals, and academic researchers share real-world knowledge about designing, building, and maintaining the critical systems of our interconnected world. LISA invited submissions of proposals from industry leaders for talks, mini-tutorials, tutorials, panels and workshops. LISA is also interested in research related to the fields of system administration and engineering. We welcome submission for both papers and posters.