4th Workshop on Cyber Security Experimentation and Test (CSET '11)

August 8, 2011
San Francisco, CA

Opening Remarks

Sean Peisert and Stephen Schwab, CSET '11 Program Co-Chairs

Summarized by Sean Peisart (peisart@cs.ucdavis.edu)

The 4th Workshop on Cyber Security Experimentation and Test (CSET) was held on August 8, 2011. In its first three years, CSET's focus was largely on testbeds and experimentation relating to testbeds, reflecting its origins as the DETER Community Workshop. In its fourth year, the focus was broadened to equally emphasize the nascent science of cybersecurity, i.e., measurement, metrics, data, simulations,

and models, as those subjects will also strongly influence the theory and practice of experimental security research. Additionally, the chairs sought to make CSET more of a workshop in the traditional sense. Depending on subject matter, some talks were set up to be highly interactive, 45-minute discussions between presenter and audience. In some cases, similarly themed papers were presented in sessions in which the three talks were presented in 20 minutes each without questions and then all three papers were discussed for 30 minutes together.

Overall, based on reactions from both presenters and audience members, the new scope and format for CSET was a success. Out of 30 submissions, 12 papers were accepted and were well received by the audience.

Security Experimentation and the Real World

Summarized by Peter A.H. Peterson (pahp@cs.ucla.edu)

Should Security Researchers Experiment More and Draw More Inferences?

Kevin S. Killourhy and Roy A. Maxion, Carnegie Mellon University

The rhetorical title of this title was answered immediately by Killourhy with a resounding "Yes!" Explaining, Killourhy stressed that not all empirical work should be considered an experiment, per se, and that experimental practice in security research could be improved. Only 54% of studies in keystroke dynamics were comparative (evaluating a matrix of tools and data sets for comparison on the same grounds), and only 7.5% were inferential (drawing statistical inferences from data, rather than simply reporting results).

A particularly troubling trend in security is the "one-off" evaluation, where a new technology is evaluated against a home-grown dataset. The researcher performs the evaluation, finds a benefit, and declares victory. Unfortunately, in these cases it is impossible to know how well the technology compares to others, because no comparative evaluation was performed (and often neither the dataset nor the tool is public). Comparative experiments—standard in other sciences—show the differences between pairs of techniques and workloads, so as to show their differences.

In addition to comparative studies, Killourhy stressed the importance of statistical inference for experiments, such as comparative experiments. Reporting only which tool performed the best on which data set is not enough, because "security technologies don't have an error rate, they have many error rates, depending on the factors [in the experiment]." In contrast, statistical inference can show not only which tool is best for which data set, but by how much, what the confidence of that result is, what relationships exist between tools and data sets, and more.

A lively discussion ensued, focusing on how to improve security experimentation. Availability and functionality of "research code" hinders good experimentation. Roy Maxion (CMU) suggested adopting "structured abstracts" as used in medicine, where specific language about methodology and result are included, allowing papers to be quickly surveyed. Terry Benzel (ISI) noted that repeatability is often difficult to achieve, even for the original investigator, because of the changing software and hardware environment (and suggested use of testbeds to help mitigate this problem). Cynthia Irvine (NPGS) followed up by pointing out that, often, a sponsor is looking to solve a problem (not to perform comparative studies). The room generally discussed challenges with performing comparative studies with older tools and workloads in the face of rapidly changing threats. Others voiced a need for sponsors to prioritize enabling testing by others.

From there, the discussion turned to facilitating comparative studies, asking whether it would be better to share detectors, or data sets, as well as the difficulties of doing either, which range from proprietary concerns to sensitivity of workload information and more. Killourhy stressed that "what we're doing isn't working," and that almost anything we could do would be an improvement, saying that "the problem doesn't go away because it is inconvenient." The room agreed—but no silver bullet was evident.

No Plan Survives Contact: Experience with Cybercrime Measurement

Chris Kanich, Neha Chachra, and Damon McCoy, University of California, San Diego; Chris Grier, University of California, Berkeley; David Wang, Marti Motoyama, Kirill Levchenko, Stefan Savage, and Geoffrey M. Voelker, University of California, San Diego

Testbeds enable research that could not easily be performed in the real world. However, some kinds of research must necessarily be performed in the real world. This talk described the goals, procedures, and results from some real-world research involving cybercrime.

Two necessary things for engaging with and observing cybercriminals are verisimilitude and scale. Verisimilitude is the quality of being authentic—an important quality when performing research of this kind. If one is pretending to be a customer, it is important to appear to be an authentic customer, regardless of whether you are purchasing end-user goods offered through spam or purchasing computational or other resources offered on the underground market. This can have challenging repercussions; the researchers found

it necessary to use a native speaker of Russian in order to navigate the forums and communicate in a natural way. Scale is another important issue; the underground market is a large organization, and it is difficult to see the big picture without a significant and broad effort.

This inspired a long discussion on basically two points. First, people considered the work from an experimental perspective, wondering how researchers could best identify how representative their data was. Geoff Voelker said that when possible, researchers would try to measure similar things from various vantage points in order to try to determine how well the data matched. Related challenges arise due to being blocked or having data "poisoned" by criminals who "got wise" to the investigation. Kanich underscored that they try to be upfront about claims relating to the data and state that the observations are limited by many practical concerns.

The second major topic was about the ethics of this kind of research. One participant said, "Your papers are usually great. How the hell do you get the ethical backing to do this stuff?" Kanich responded that first, funding did not come from government sources, and that they tried to consider whether they truly defrauded the parties involved. They considered that those parties who purchased goods did not need to purchase through them, and they did not keep their money. Furthermore, rather than creating new spam from scratch, the researchers used "double-agent" machines to modify instructions for downstream spam bots that would already have sent spam to potential customers.

A number of people asked about IRB oversight and posited that IRBs are currently, by and large, medically oriented and are not sensitive to cybercrime issues. The researchers described their relationship with IRBs, lawyers, and funding, and stated again that they take a consequential approach, asking whether they would do harm in the course of the research. Additionally, the papers for their major studies each include a section on the ethics of their methodology and actions. During this discussion, Doug Maughan (DHS) highlighted the forthcoming Menlo Report on ethical principles for ICT research and suggested that ICT researchers should, like Dave Dittrich, find their way onto IRBs for the future.

Ultimately, important, timely, and fascinating data resulted from—and will continue to come from—this ongoing research. At the same time, the inevitable debates about representivity, ethics, legality, and funding will continue alongside them.

Experimental Methodology

Summarized by Peter A.H. Peterson (pahp@cs.ucla.edu)

Salting Public Traces with Attack Traffic to Test Flow Classifiers

Z. Berkay Celik, Jayaram Raghuram, George Kesidis, and David J. Miller, Pennsylvania State University

George Kesidis argued for greater "statistical hygiene" in security experimentation. Their work focused on identifying, discussing, and attempting to mitigate the way in which the timing characteristics of sample botnet traces used for evaluating flow-classifiers can inadvertently affect the results.

Due to the lack of publicly available traces of attack traffic, many researchers construct synthetic attack traffic as training and testing targets for their flow classifiers. One technique is to combine benign background traffic from a corporate network and traffic from a defanged botnet running in a testbed. In this way, the background traffic is salted into the botnet traffic, providing an ostensibly realistic trace.

However, when flow characteristics of the botnet traffic are statistically distinct from the background traffic, they may be identified by machine learning algorithms as meaningfully identifiable characteristics of botnet traffic, even though they may be artifacts of the trace synthesis process. In turn, this can make the flow classifiers appear to be more successful in evaluations than they may be in real life, because the synthetic trace can be artificially straightforward to classify. In the paper, the authors worked to investigate and overcome these issues, including comparing how various scenarios and machine-learning algorithms combined to produce various results. Researchers in this area would do well to consult the paper for more details.

The focus of the talk and discussion was more about these kinds of issues in general, and expanded beyond the paper itself into issues of experimentation and statistical forensics. For example, another classic issue affecting research results is "double dipping," such as when training sets (or sets used to derive ground truth) are used as targets for testing. This is a specific problem for machine learning, but also affects any research where the evaluation phase can be unintentionally (or intentionally) biased toward the solution.

Discussion for this session was combined with the next two papers.

Beyond Simulation: Large-Scale Distributed Emulation of P2P Protocols

Nathan S. Evans and Christian Grothoff, Technische Universität München

This paper was presented by Matthias Wachs and Bart Polot.

When test requirements grow larger than even significant testbeds can handle, researchers often turn to simulation. However, the fidelity of the simulation can be poor, because of inadvertent mistakes when the simulation is constructed from the real-world counterpart. And, in any case, the process of building a simulation can have a large cost in terms of time and human resources.

While simulation allows great scale, it has a high translation cost. On the other hand, the scale of emulation solutions may be limited, but allows the experimenter to acquire data that directly reflects the original implementation of the tool in question. Accordingly, Wachs and Polot presented the GNUnet framework, which is a scalable emulation framework for peer-to-peer protocols, capable of accurately supporting many emulated hosts through judicious sharing of local resources. They described the resource-sharing design of GNUnet, which includes the use of shared memory and fast messaging techniques as well as centralized management of peers. They also described their experience testing an 80,000-peer emulation of a Kademlia DHT on a small cluster as a test case and example of the frameworks.

GNUnet makes some tradeoffs in order to achieve its goals. For example, it does not support timing control, so it may not be suitable for latency-sensitive tests. There are also other characteristics of the design that may affect its suitability for particular purposes, such as interference from the underlying OS, scheduling, similarity of peers, etc. And, of course, test code must be written using the GNUnet framework, which has its own cost. Interested readers should see the paper or presentation audio for more information.

Automating Network Monitoring on Experimental Testbeds

Michael Golightly, Princeton University; Jack Brassil, HP Laboratories

Jack Brassil presented this work on Netflowize, a prototype tool for Emulab-type topologies that is able to automatically add experiment-wide instrumentation nodes to the topology. This ability, along with a set of tools, allows for flexible monitoring of resource consumption across the test environment. Netflowize uses existing NetFlow probes and collectors available on Emulab and DETER; while these tools are available on these testbeds, they are typically used by testbed operators. Netflowize allows researchers to leverage these tools in a straightforward and accurate way.

Netflowize automatically determines where in the experimental topology to place the probes. Where a naive approach could add too many probes, Netflowize minimizes this. Netflowize has two modes: "lightweight" mode uses existing infrastructure to perform monitoring in a transparent way; "heavyweight" is also transparent, but deploys additional hardware resources in the experiment to serve as the probes. This has the benefit of not creating load on experimental nodes, but requires more hardware.

Netflowize is under active development. Future work includes better error and redundancy handling, more user-accessible "knobs," and efficiency improvements. Other developments may include extending beyond NetFlow to use other tools, and more. Brassil pointed the participants to a URL where prototype code was available (contact him if you are interested).

Discussion

Following this presentation, the floor was opened to questions from the workshop participants to the authors of all three papers.

Stephen Schwab (ISI) asked George Kesidis about best practices for the application of machine learning in experimentation. Kesidis responded that you don't need an expert, but "good statistical hygiene" is a must. I took that to mean that new users of ML techniques may not recognize the necessity of separating testing and training sets, even though they might not make those kind of "double dipping" mistakes in other experimental areas.

Roy Maxion (CMU) asked all three presenters what they felt made rigorous experimentation with high-confidence results hard or easy. Jack Brassil suggested that if our community (like others) would centralize to common, shared tools, it would be easier to verify results and insist on more rigorous experimentation. Kesidis suggested that reproducibility is achievable, but that it is too hard to adequately specify experimental conditions within the confines of a conference paper. Kesidis did say that we can point readers to online resources. Kesidis also said that if you're doing research, your goal should be to prove your results and enable them to be accepted. This might include open sourcing the work and making sure that others are able to recreate it (subject to confidentiality concerns, etc.). He said, "It's not that other fields are that much better than we are," but we should still be doing it.

Jelena Mirkovic (ISI) suggested that recreating experiments isn't always very straightforward—even if the code is open source and available. Matthias Wachs suggested that this could be solved with better communication between

researchers, and accordingly invited the participants to email them directly with any questions regarding GNUnet.

Bots and Overlays

Summarized by Kevin Killourhy (ksk@cs.cmu.edu)

Challenges in Experimenting with Botnet Detection Systems

Adam J. Aviv and Andreas Haeberlen, University of Pennsylvania

Adam Aviv described a hypothetical situation in which a researcher evaluates a new botnet detector. Ideally, the researcher conducts a series of representative tests, both on her own network and also on others'. She compares the results to prior work, and she makes the detector and data available so that other researchers can reproduce her results. A survey of 20 research papers on botnets suggested that this ideal is often sacrificed to practicalities. Challenges include establishing ground truth, creating a production-quality detector, and obtaining permission to deploy it.

Behind these challenges, Aviv said, the big concern is privacy. His statement prompted a lively discussion about whether privacy is the biggest problem—compared to lack of ground truth, standard methodology, and statistical analysis—and whether those other problems can be solved without tackling the privacy issue. Aviv explained that if privacy were not a problem, researchers could share data and do apples-to-apples comparisons of different detectors. It was suggested that, if nothing else, privacy is a good excuse for not sharing data.

Returning to the survey of botnet-research papers, Aviv showed that the majority of papers used an overlay methodology for their evaluation. For this method, two separate network traces are needed: botnet traffic from a simulation or sandbox, and standard traffic from the researchers' network or another source. Then the two traces are integrated into a data set for detector evaluation. A participant observed that one cannot know that the standard traffic is clean (i.e., untainted by bot traffic). Aviv agreed and raised a host of other concerns, including the introduction of artifacts by overlaying traffic from different networks.

Inspired by how PlanetLab helped distributed-systems researchers, Aviv asked, "Can we do better together?" To start the discussion, he sketched out a straw-man solution: operators of various networks give researchers access to a box on their network. The boxes would be fed NetFlow data, and researchers could install their detectors on boxes across many networks. Detector output would be vetted by the network operators; when free of sensitive data, the results would be returned to the researcher for further analysis.

Participants discussed whether incentives such as access to bleeding-edge detectors would encourage network operators to participate. In the end, Aviv offered this solution as a place to start; improvements are welcome.

ExperimenTor: A Testbed for Safe and Realistic Tor Experimentation

Kevin Bauer, University of Waterloo; Micah Sherr, Georgetown University; Damon McCoy, University of California, San Diego; Dirk Grunwald, University of Colorado

Kevin Bauer introduced Tor as being, simultaneously, a production-quality public service and an ongoing research project. Tor is a low-latency overlay network on which users can send and receive TCP traffic anonymously. At the same time, researchers are constantly modeling, testing, and improving the network. These two aspects of Tor sometimes come into conflict, causing researchers to adopt a range of research methods, with mathematical models at one end and worrisome use of the production Tor network at the other. Simulators, like Emulab, and PlanetLab make up the middle of this range.

A good experimental testbed for Tor research would scale to Tor-like size, ensure reproducibility, have realistic traffic properties, and be otherwise ethically sound and easy to use. With PlanetLab, one quickly runs into scalability and reproducibility issues; resources are limited, and node assignments change from one allocation to the next. With the Tor network itself, scalability is not a problem but reproducibility and ethical issues are; Tor users expect privacy—-the very reason they use Tor.

Bauer explained the design of the ExperimenTor testbed (http://crysp.uwaterloo.ca/software/exptor). Realistic routers were modeled using publicly available data. Clients were modeled by studying aggregate real-world traffic (e.g., amount of traffic per service and number of clients per country). The ModelNet emulation framework was chosen since many applications can be run unmodified on the testbed just by linking with the ModelNet libraries. Early experience with the testbed is promising. An emulator has been deployed across four institutions and used in two ongoing research projects: effects of link-based router selection, and a redesign of Tor's congestion control.

Discussion focused primarily on two issues: the general utility of the testbed and the reproducibility of Tor research. One participant noted that this approach seemed like a procedure for network-based research in general, not just Tor research. Bauer agreed that the ModelNet emulation framework is very general, but his present efforts are focused on Tor. Participants also reacted to the issue of reproducibility with experi-

ments on real networks and other testbeds. One participant wondered whether results that cannot be reproduced are worth reporting.

Methodology and Getting Real Data

Summarized by Kevin Killourhy (ksk@cs.cmu.edu)

On the Design and Execution of Cyber-Security User Studies: Methodology, Challenges, and Lessons Learned

Malek Ben Salem and Salvatore J. Stolfo, Columbia University

Malek Ben Salem explained how work on masquerade detection has been hindered by a lack of masquerade data. For instance, she wanted to test the conjecture that an attempt to steal information often manifests as extensive and abnormal search behavior. To test such a claim, one needs data not only from legitimate computer usage but also from attempts to steal information. Observing attack-like behavior under laboratory conditions can be a challenge, and this talk concerns her experiences trying to add rigor to the process of gathering such data.

Ben Salem enumerated steps for designing and conducting a user study: state a hypothesis, identify experimental variables, establish a control group, choose a sampling procedure, and estimate the necessary sample size. She emphasized the need to control variability and reduce bias among users. In practice, these steps require some careful thought. For instance, to ensure that subjects acted as they might if they were participating in a real attack, she provided them with scenario narratives. Subjects were told to imagine themselves at a co-worker's unattended computer, trying to find sensitive financial information.

For discussion, she offered several recommendations and lessons learned. Get IRB approval early, since the process can be slow. For data that is released publicly, subjects should sign waivers regardless of the planned data sanitization; subjects do not always have the understanding or foresight to thoroughly sanitize their data. Conduct pilot tests and post-experiment questionnaires to identify any unforeseen issues and provide additional insight. As an example, the preliminary narrative did not provide a name for the imaginary co-worker. Talking with pilot subjects revealed that they would have used the name when searching for information, and so a name was added.

Experimental Challenges in Cyber Security: A Story of Provenance and Lineage for Malware

Tudor Dumitras, Symantec Research Labs; Iulian Neamtiu, University of California, Riverside

Tudor Dumitras led with an anecdote about what happens when researchers ignore issues of experimental validity: the IROP Keyboard,a piece of hardware satirically proposed by Zeller, Zimmermann, and Bird, does not have the I, R, O, or P key because studies have shown that most coding errors involve those keys. More serious issues of validity arise when tracing the lineage and establishing the provenance of a malware artifact.

When analyzing malware samples to determine which one came first and how they evolved, several methodological questions arise. Is the reconstructed lineage correct (i.e., what is ground truth)? What am I really measuring, and are my data representative? How well does this work now, and, more important, how well will it work tomorrow? These questions assess the validity of an experiment. Threats to validity come in several forms—*construct validity*, whether the metrics measure the right concept; *internal validity*, whether causal inferences can be drawn; *content validity*, whether all data relevant to the concept are used; and *external validity*, whether the results generalize beyond the experiment.

Within the domain of tracing lineage and establishing provenance of malware artifacts, Dumitras explores these threats to validity and offers some solutions. For instance, ground truth is somewhere between hard and impossible to establish for malware lineages, but the same tools can be used to reconstruct the lineage of open-source software (e.g., the Linux kernel). Dumitras also offered Symantec's WINE (Worldwide Intelligence Network Environment) as a helpful service for promoting experimental validity (http://www.symantec.com/WINE).

WINE provides researchers with real-world malware data, gathered as part of Symantec's own security and anti-virus efforts. The data enables research on both static and dynamic properties of malware. Metadata and contextual information are provided for the artifacts (e.g., collection times and infection reports). WINE makes possible reproducible experiments with representative malware samples. Dumitras fielded several questions on the logistics of using the service and whether it might be abused by malware authors or competitors. Users must be on-site at a Symantec Research Lab and under NDA. Dumitras and his colleagues are exploring what IRB-related issues arise for researchers using the data. According to NSF sources, the cost of using this service could be included in a grant budget.

Education

Summarized by Kevin Killourhy (ksk@cs.cmu.edu)

Active Learning with the CyberCIEGE Video Game

Michael Thompson and Dr. Cynthia Irvine, Naval Postgraduate School

Michael Thompson describes CyberCIEGE as a cyber-security game for a broad audience. The game enables instructors to cover a wide variety of security concepts without requiring a lot of prior knowledge of the students. Players are able explore a scenario, approach it in different ways, and even fail as part of the learning process. A scenario-definition language can be used to add new scenarios to the game. Quantitative assessments can also be included, leading one participant to wonder if the game might be used to test the competence of his organization's IT department.

Scenarios involve a simplified network simulation including assets, users trying to access the assets, and attackers trying to compromise them. Players can add computers, routers, and firewalls to the network. They can change ACLs, apply patches, and even adjust aspects of physical security. However, as Thompson explained, students get the experience of configuring a VPN without first going through a CISCO training course. As a demonstration, Thompson walked through one game scenario. The player helps a user access the Web, installs a network filter, and protects trade secrets by isolating another computer from the network.

While there have been no formal assessments, game scenarios are used in Intro to Computer Security at the Naval Postgraduate School, and the game is being piloted at other institutions. Informal feedback suggests that students enjoy the hands-on aspect of exploring networking concepts through the game. One of the lessons learned is that different students approach a scenario in vastly different ways, and the game must provide a lot of feedback to students as they explore a scenario.

Investigating Energy and Security Trade-offs in the Classroom with the Atom LEAP Testbed

Peter A.H. Peterson, Digvijay Singh, William J. Kaiser, and Peter L. Reiher, University of California, Los Angeles

Peter Peterson presented Atom LEAP, an energy-measurement platform, and described his experience using it in an undergraduate research seminar. LEAP is open source, inexpensive, and easy to build. The "energy calipers" provide overall and component-level energy usage (e.g., CPU, RAM, and USB) at very fine granularity. User scripts make it easy to start and stop monitoring.

In the seminar, students used LEAP to investigate security/energy tradeoffs. Instructors presented students with topic areas; the students organized into groups within each area and developed research plans. After the first two weeks, class time was used for group meetings rather than lectures. Through the projects, the instructors intended to have students learn about performance measurement and analysis.

One project measured energy consumption of an AV product scanning a home directory. Another project compared the energy consumption of various compression/encryption schemes, finding that gzip was the most energy efficient. Peterson observed that the projects and the style of the course resembled what a student would encounter in grad school, and that this experience was beneficial for undergrads.

In retrospect, supporting many different topic areas was a lot of work for the instructors. Future iterations of the class might have multiple groups tackle one topic, so groups could red-team other groups' plans. Because the quality of the final reports was uneven, additional coverage of experimental design and statistical analysis is also planned. Nevertheless, students got the message that evaluation is tricky. Peterson encouraged interested instructors to pilot the LEAP technology in their own courses.

Experiences in Cyber Security Education: The MIT Lincoln Laboratory Capture-the-Flag Exercise

Joseph Werther, Michael Zhivich, and Tim Leek, MIT Lincoln Laboratory; Nickolai Zeldovich, MIT CSAIL

Joseph Werther posed the question: How do we get more smart students involved in cybersecurity? He described an effort by MIT Lincoln Laboratory to conduct a capture-the-flag (CTF) event intended to educate and promote interest in security; 53 students from six universities participated. The Wordpress content-management system was chosen as the target; it is a realistic target, and its extensible nature allows students to become comfortable with the base system. Additional components can be tackled incrementally.

The two-day CTF event was preceded by a week of lectures and laboratory exercises. Underlying the effort was the belief that education in "offensive security" enables students to understand the mechanics of a vulnerability and how a system can be exploited. Werther identified three components of learning: reading, building, and experience. All three were incorporated into the lessons. The five classes included coverage of Web applications, Wordpress, server security, and Web-application security. The final class was a work-through of the Google Gruyere hacking exercises. For the CTF event, teams had to defend a Wordpress instance

running in a virtual machine and to attack the other teams' instances. A team's score incorporated offensive success, based on how many other teams' flags were captured, and defensive success, based on how few of their own flags were compromised.

After the exercise, participants filled out a voluntary survey. While acknowledging that the results were not scientific, Werther noted that respondents reported increased interest as well as skill in security work. The organizers plan to conduct more CTF exercises in the future. They hope to expand to more colleges and to improve the robustness and extent of their game monitoring. In the meantime, they are investigating the best way to encourage learning and assess knowledge-gain.

Discussion Panel

Michael Thompson, Peter A.H. Peterson, and Joseph Werther

The panelists were asked what each of their educational efforts replaced. Werther noted a dearth of capture-the-flag exercises, especially with education, not competition, as the principal goal. Thompson explained that CyberCIEGE is used in introductory labs but did not know what exercises they replaced. Peterson explained that, prior to the seminar using Atom LEAP, no such course was offered.

When asked how they measured their learning objectives, all three panelists acknowledged that they were reporting on pilot-stage experiences. They were thinking about how to improve assessments the next time around. One participant suggested that they look to the scientific literature on education and learning.

The panelists were asked whether their experience showed that anyone can learn cybersecurity. Thompson answered that what one needed more than anything was interest. Without an interest in the subject, learning about cybersecurity would be very difficult. For those with interest, Thompson's experience suggested that one could improve success rates by accommodating different learning styles.

In the end, the discussion turned to *Star Trek* and the Kobayashi Maru. Should cyber-security education include an unwinnable scenario? Would facing such a situation be a valuable lesson to students of security? Perhaps.