Embracing the Kobayashi Maru: Why You Should Teach Your Students to Cheat

Gregory Conti and James Caroland

Adversaries cheat. We don't. In academic institutions around the world, students understand that they will be expelled if they violate their college's honor code or otherwise fail to play by the institutional rules. The dissonance between how our adversaries operate and how we teach our students puts our students at a distinct disadvantage when faced with real world adversaries who inevitably do not play by the rules. Breaking through the paradigm where students self-censor their ways of thinking to a new paradigm that cultivates an effective adversary mindset is both necessary and possible.

An adversary examines systems and finds weaknesses in trust relationships, human behavior, communications protocols, physical security, and system logic to find exploitable vulnerabilities. By anticipating adversary actions and reactions, ethical actors are far better prepared to build secure systems and perform both defensive and offensive activities successfully. For both the attacker and the defender a devious mind is equally as important as a beautiful mind.

This article describes our experiences in helping students develop an adversary mindset by adopting the Kobayashi Maru training exercise employed in the fictional Star Trek universe. In the Kobayashi Maru exercise, Starfleet cadets were faced with a no-win scenario -- attempt to rescue the crew of a disabled civilian vessel, and be destroyed in the effort, or avoid confrontation and leave the disabled ship and its crew to be captured or destroyed. Famously, Captain Kirk won the scenario by, and this is important, stepping outside the game and altering its rules to his benefit. By deciding to cheat and altering the programming of the Artificial Intelligence driving the exercise, he won the contest.

Lest there be any misunderstanding, our purpose with this article is not to encourage or teach students to cheat *in general*, but to learn to think creatively when considering adversary behavior.

The Challenge

Our variation of the Kobayashi Maru utilized a deliberately unfair exam - write the first 100 digits of *pi* (3.14159...) from memory and took place in the pilot offering of a governmental cyber warfare course. The topic of the test itself was somewhat arbitrary; we only sought a scenario that would be too challenging to meet through traditional studying. By design, students were given little advance warning for the exam. Insurrection immediately followed. Why were we giving them such an unfair exam? What conceivable purpose would it serve? Now that we had their attention, we informed the class that we had no expectation that they would actually memorize the digits of *pi*, we expected them to cheat. How they chose to cheat was entirely up to the student. Collaborative cheating was also encouraged, but importantly, students would fail the exam if caught. To provide additional incentive, we offered a prize to the student who exhibited the most creative and effective cheating technique.

The Techniques

Students took diverse approaches to cheating, and of the 20 students in the course, none were caught. One student used his Mandarin Chinese skills to hide the answers. Another built a small PowerPoint presentation consisting of three slides (all black slide, digits of *pi slide*, all black slide). The idea being that the student could flip to the answer when the proctor wasn't looking and easily flip forwards or backward to a blank screen to hide the answer. Several students chose to hide answers on a slip of paper under the keyboards on their desks. One student hand wrote the answers on a blank sheet of paper (in advance) and simply turned it in, exploiting the fact that we didn't pass out a formal exam sheet. Another just memorized the first ten digits of *pi* and randomly filled in the rest, assuming the instructors would be too lazy to check every digit. His assumption was correct.

The finalists were particularly innovative. The runner-up used two different techniques, a primary and a backup. In his first approach, he remade his desktop nameplate to look legitimate, but included the answers, in fine print, on the side facing him. For his backup plan, he included the answers on a soda can which he concealed with his hand when the proctor walked by, see Figure 1 (right). The winner of the competition created a false book cover for a course text and replaced portions of the text with the answer, matching both color, font, and text size, see Figure 1 (left). He then used hair spray to lightly tack the false cover into place. The result was all but indistinguishable from the original book.

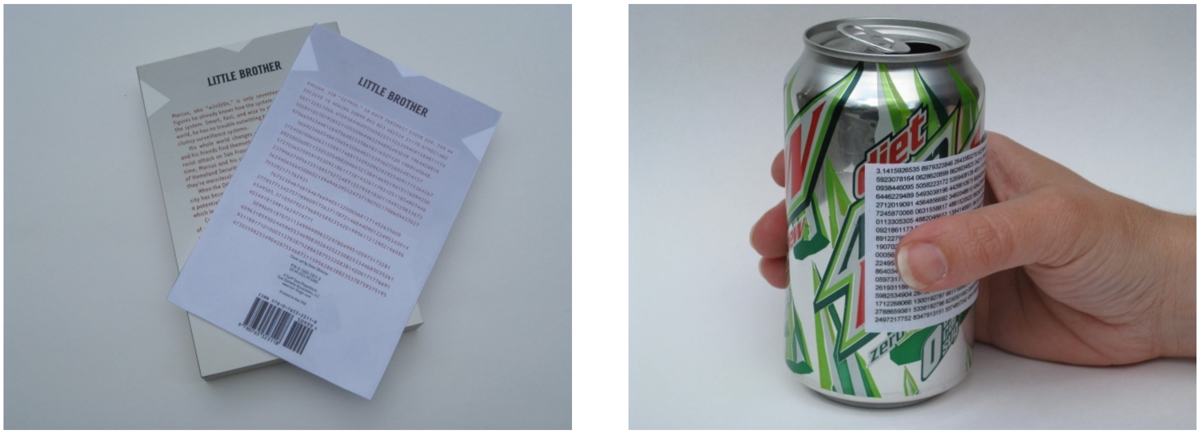

Figure 1: Examples of student work. False book cover containing answers (left) and soda can with answers that could be concealed when the test proctor was nearby (right).

Learning Security Principles from the Cheaters

We learned much from the students during the course of this exercise. Students embraced the test, proved far more devious than their day to day personas let on, and impressed us with their ability to analyze and defeat the inherently flawed classroom system. We drew the following conclusions from observing the techniques students used and through an interactive group discussion where students described their cheating, what they learned, and other techniques they might employ in the future.

Exploit the Environment - Students instinctively analyzed their environment and found weaknesses they could use to their advantage. The presence of computers on the desktop and the fact that they didn't have to clear their desks during the exam provided opportunity to exploit the system. Because students were seated side by side and were partially hidden behind monitors, some students used these characteristics to facilitate their cheating activities.

Exploit Trust - Explicit or implicit trust models are exploitable opportunities. Despite our awareness that the students were cheating, we still inadvertently let our guard down. For example, we wouldn't have stopped a student from using the restroom during the exam. During our group discussion, students suggested that going to the bathroom to cheat would have been an easy-to-implement approach. It is because of our inherent and unconscious trust that we leave ourselves open to exploitation in the physical world and online. As security professionals we must learn to think like the jaded police officer or prison guard who never takes statements and actions at face value.

Exploit Personal Skillsets - Students each possessed diverse skillsets that they could apply to the challenge of cheating. Adversaries do the same. For example, the student who used Mandarin Chinese to write the answers and placed them in plain sight used his uncommon skill to become a formidable adversary.

Exploit the Human - Being lazy, trusting, and predictable, humans are often the weakest link in any security system and students intuitively exploited this fact. One student observed that we rarely handed out worksheets and frequently asked students to provide their own paper. This provided a security gap where they could sneak in an already completed exam and turn it in. Another student suggested instructor predictability and misplaced trust as a potential attack vector. Because we frequently took extra paper from the printer tray to provide to the class, the student said he would preposition answer keys in the printer and then ask us for a sheet of paper. We would then hand them the answers without knowing it, despite coming from a "trusted" source.

Develop Backup Plans - Adversaries rarely seek to accomplish their objectives through a single, all or nothing plan. Several students demonstrated this principle by developing backup plans in case their primary cheating tactic was compromised.

Tips for Teaching Your Students to Cheat

The key to teaching students to cheat is to provide context. Explain to them the objectives of the exercise - learn how an adversary thinks and operates by deliberately loosening traditional rules and tapping their personal creativity. While we advocate teaching students to cheat, instructors must still provide clear boundaries, lest there be misunderstandings. In our case, we made it clear that we expected students to cheat and that getting caught would result in a failing grade, but that this exception to traditional rules of behavior only extended to this exercise and not for other graded events in the course.

We deliberately provided minimal warning for the exercise to increase stress levels and material that couldn't be readily learned through traditional studying. During the exam, we sought to further increase the stress and realism by walking occasionally among the student desks. We didn't try all that hard to catch students, but that wasn't the point. We sought merely to increase pressure by acting as realistic exam proctors. We considered, but chose not to go as far as forcing students into a position where they must cheat *on their own initiative*, but without being told to do so. We believed this would place students into an unfair ethical dilemma, send the wrong message, and that most, if not all, students would simply fail the exam rather than cheat illicitly.

Towards a Larger Adversary Mindset Curriculum

Our Kobayashi Maru exercise was part of a larger set of lessons designed to cultivate an adversary mindset. There isn't space in this article to describe them in similar depth, but highlights are provided below to assist educators in considering more comprehensive approaches.

Early in the course we included the Hackers Are People Too documentary to help students understand the hacker mindset, which is sometimes playful and sometimes adversarial [1]. We also included a "divergence" exercise which was inspired by hacker Dan Kaminsky in the documentary, where he posed the question "What are the alternative uses of a fork?" This seemingly simple question contains significant depth. Typical students frequently encounter "convergence" questions, questions that seek only a single correct answer. Divergence questions, on the other hand, are open-ended and compel students to creatively consider a broad range of answers. We chose this exercise to warm students up to new ways to think about problem solving.

Also early in the course, we held a lock picking lab and taught students how to pick small padlocks. The point here, in addition to a fun, hands-on exercise, was to challenge students' assumptions about physical security and derive commonalities between system security and approaches to understanding and defeating locks.

Our course included Joe Grand, Jake Appelbaum, and Chris Tarnovsky's case study of insecurities in the San Francisco parking meter system which taught students how an adversary might attack critical infrastructure [2]. For future work we are considering including a hands-on hardware hacking exercise to teach students how an adversary might develop or modify hardware, by building a TV-B-Gone (http://www.ladyada.net/make/tvbgone/) universal remote control.

We also included a video by Johnny Long on No Tech Hacking to illustrate how an adversary might use social engineering attacks to compromise humans and human-centric security systems [3]. In the future, we plan to add a phishing email writing contest to allow students hands on exploration of social engineering.

Weekly throughout the course, students read books to explore various aspects of the adversary mindset including: *Ender's Game* by Orson Scott Card which illustrated the need to adapt to intelligent adversaries, *Little Brother X* by Cory Doctorow to teach students the importance of electronic civil liberties and the potential for an adversarial relationship between a government and its citizens, and *Critical System Error* by Joseph Menn to examine the real world actions and reactions between network defenders and online criminals.

Conclusions

Teach yourself and your students to cheat. We've always been taught to color inside the lines, stick to the rules, and never, ever, cheat. In seeking cyber security, we must drop that mindset. It is difficult to defeat a creative and determined adversary who must find only a single flaw among myriad defensive measures to be successful. We must not tie our hands, and our intellects, at the same time. If we truly wish to create the best possible information security professionals, being able to think like an adversary is an essential skill. Cheating exercises provide long term remembrance, teach students how to effectively evaluate a system, and motivate them to think imaginatively. Cheating will challenge students' assumptions about security and the trust models they envision. Some will find the process uncomfortable. That is OK and by design. For it is only by learning the thought processes of our adversaries that we can hope to unleash the creative thinking needed to build the best secure systems, become effective at red teaming and penetration testing, defend against attacks, and conduct ethical hacking activities.

Acknowledgments

We would like to thank Captain TJ White and Peiter "Mudge" Zatko for their feedback and ideas in support of this work. The views in this article are the authors' and don't reflect the official policy or position of the United States Military Academy, the Department of the Army, the Department of the Navy, United States Cyber Command, the Department of Defense, or the United States Government.

References

[1] Ashley Schwartau. Hackers Are People Too. Documentary. Managed Mischief, 2008.

[2] Joe Grand, Jake Appelbaum, and Chris Tarnovsky. "'Smart' Parking Meter Implementations, Globalism, and You." Defcon 17, July 2009. Available online at https://www.defcon.org/html/links/dc-archives/dc-17-archive.html

[3] Johnny Long. "No-Tech Hacking." Defcon 15, August 2007.
Available online at https://www.defcon.org/html/links/dc-archives/dc-15-archive.html

Gregory Conti is an associate professor in the US Military Academy's Department of Electrical Engineering and Computer Science and is responsible for the academy's information security education program. Contact him at gregory.conti@usma.edu.

James Caroland is a member of the US Cyber Command Commander's Action Group and an adjunct associate professor in the University of Maryland University College's Cybersecurity Program. Contact him at jlcarol@cybercom.mil.