Elevation of Privilege: Drawing Developers into Threat Modeling

Adam Shostack
adam.shostack@microsoft.com

Abstract

This paper presents Elevation of Privilege, a game designed to draw people who are not security practitioners into the craft of threat modeling. The game uses a variety of techniques to do so in an enticing, supportive and non-threatening way. The subject of security tools for software engineering has not generally been studied carefully. This paper shares the objectives and design of the game, as well as tradeoffs made and lessons learned while building it. It concludes with discussion of other areas where games may help information security professionals reach important goals.

1 Background

Software rarely becomes secure by luck. Software security usually requires focused engineering activities. Taking action to threat model a system under development is unusual, and usually happens only if there's a security enthusiast on the team, or if an organization has adopted some set of security practices. Both situations are rare, with lamentable consequences for the security of software, systems and critical infrastructures. When threat modeling happens, it is usually done by experts who have learned the art in an apprenticeship model, often an informal one. At its best, threat modeling leads to both the cataloging of threats which are known to experts, as well as the discovery of new threats against a system, all before it is built, allowing the design to be modified to produce a more resilient system. This paper reports on work done at Microsoft as part of ongoing work in developing more secure software. This work to improve the security properties of software (as contrasted with the security features) has been going on in structured ways throughout the company for over a decade. [9, 19] One element of that work has been to bring threat modeling to developers in various forms. The threat modeling work is documented in [27, 32, 13, 14], but one essential tradeoff that underlies this paper is that of the area of expertise of the practitioners. That is, what are the results we can expect from threat modeling done by security experts versus software developers? (Obviously, there is some overlap, and as obviously, there are many whose expertise falls squarely into one camp or the other.)

1.1 Tradeoffs in Threat Modeling

Threat modeling done by security experts has many obvious advantages including domain expertise, implicit knowledge and the apparently intuitive decision making that experts often bring to bear [17]. It is easy to assume that experts must always lead threat modeling activities. However, expert-centered approaches have downsides, and developer-centric approaches have non-obvious advantages. The developers are most aware of what the code really does, and the costs and tradeoffs involved in suggested modifications. Some developers are always present as architecture or design decisions are being made, and do not need to come up to speed on the project. Training them to identify security issues means that the skills are likely to come into play as they do "whiteboard" level designs. In contrast, security experts are rare and expensive. If it were possible to have developers do basic threat modeling, then experts could be used more effectively to find the really unusual problems with a design. Threat modeling is often done late, or not at all unless it is mandated. It is with full awareness of these and other tradeoffs and one other that Microsoft asks all of its software engineers to have a basic familiarity with threat modeling. The remaining tradeoff is how much room a process has for creativity. Those who are new to threat modeling or those who threat model occasionally require a more procedural approach, and procedural approaches are generally at odds with creativity. Elevation of Privilege was created in this constraint space in part to expose non-security experts to the enjoyment that security experts bring to threat modeling.

Elevation of Privilege is a card game for developers which entices them to learn and execute software-centric threat modeling. A sample card is shown in figure 1.

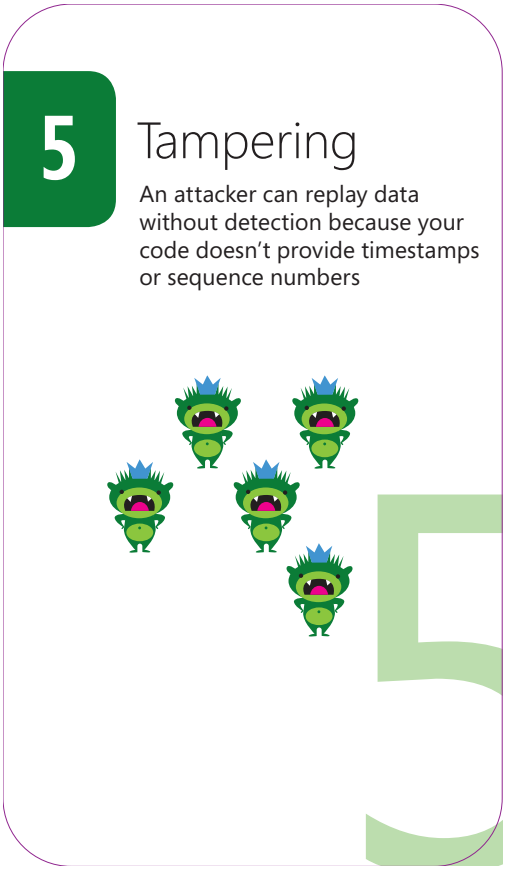

Figure 1: The '5 of Tampering' card from Elevation of Privilege

1.2 Contribution

This work addresses the important problem of bringing security knowledge to engineers, postulates games as a solution, documents the creation of a game to teach threat modeling, shares lessons learned in that process, and provides a framework for considering games in security and how to construct them. The problem of enticing software engineers to allocate scarce attention to information security problems is presented in a new light. This lack of attention is so visible as to be often overlooked and underlies what might be seen as the class of problems typoified by "we brought modern security analysis to 'field' and had a field day."' Works in this class include Hanna et al on medical devices [11], Koscher et al on automobiles [6], the continuing stream of SCADA issues, or the recent announcement of a fixed, symmetric cryptographic key in the product of a major security vendor [16]. The work also presents a model which can be used to transform many checklists into a game format.

1.3 Related Work

There are five groups of related work: they are (1) software security, (2) serious games, (3) security games, (4) game theory games, and (5) those elements of psychology and sociology which can help us understand games.

1.3.1 Applied Software Security

Applied software security is the applied end of how to produce secure software. It is exemplified by Microsoft's Security Development Lifecycle [14]. There are others, including OWASP, and a variety of similar programs. Those similar programs are typically run with a central software security group that exhorts and drives security awareness, processes, tooling, etc. A typical complaint is that these groups have trouble getting "traction" with the development organizations. This makes sense in light of the number of so-called 'ilities' that developers must work on: accessab*ility*. reliab*ility*. managab*ility*, maintainab*ility*. and the like. Performance, security and other non-feature aspects of a product are often lumped in with these. [28]. One large development group studied by the author had 20 sections in their specification template, only one of which was the actual feature to be built. This has two effects which are critical to understand for purposes of this paper, and which motivate the development of Elevation of Privilege. The first effect is that developers are overwhelmed with demands that interfere with shipping features. This effect has been documented as far back as 1975 [5] but as far as I know, no one has attempted to address the issue of overwhelmed developers with games. The second effect is that developers are asked to address perhaps ten to twenty *ilities* outside their non-*ility* core programming competency (such as databases, networks or user interfaces). As such, developers are likely to be (at best) dilettantes in their non-core skills. Thus, a way to train developers which is enticing and potentially even fun is a powerful tool which can overcome real problems in starting conversations around software security. [22] However, a mandate to play a game may not always be welcomed, as discussed in the Results section.

1.3.2 Serious Games

The use of play and games to teach skills or lessons has a long, complex history. Games with structured and explicit learning goals are discussed at least as early as Locke in 1693 [20], and probably far earlier. Perhaps the most useful definition comes from Abt [1], who defined serious games as "games [that] have an explicit and

carefully thought-out educational purpose and are not intended to be played primarily for amusement." Abt's definition is broad, useful, and allows this community to focus attention on the application of games to security issues. Recently, there has been a surge of interest in persuasive games, games for business, games for self-improvement, and 'gamification,' the insertion of game-like mechanics in non-game activities. The security community can take lessons from any of these. Gamification is increasingly popular as a way to help motivate people [29]. Gamification should not be viewed as a panacea or the only approach, as it often fails for a variety of reasons. Gamification may have reached a pinnacle in the area of frequent flyer programs, where people will pay good money to fly back and forth in 'milage runs' whose only purpose is to reach various point levels.

Serious games relate closely to training simulations, but many training simulations have no game-like elements, and should not be classified as (serious) games. A useful distinguisher from [31] is that games involve "the voluntary effort to overcome unnecessary obstacles." For example, the flight simulators used to train pilots include no unnecessary obstacles, and thus, they are not games.

1.4 Security Games

Work in serious games with a security purpose most directly includes Protection Poker. Protection Poker was inspired by Planning Poker[35], an agile development 'game' designed to elicit accurate estimates of the cost of developing software features.[10] Planning Poker consists of effort cards with small integers on them (for example, 1, 2, 3, 5, 10, 20, 50). Each round consists of a developer discussing a feature, and each player plans an effort card face down in front of them. The estimate cards are all revealed simultaneously, and the players who played the high and low cards each discuss their estimate. The game is designed to ensure that different voices are heard in software estimation. Protection Poker is intended to structure discussion about security risk. A customer representative explains each requirement. This is followed by a discussion of threats, each of which is assessed on the basis of ease and asset value. Ease and asset values are then combined into an assessment of security risk. Elevation of Privilege was inspired by Protection Poker. It differs in that the EoP cards are ordered more like a traditional card deck, and the rules provide a very different structure for play, described below. [1] and a no-bid variant of 'Spades.'[2]

Work in security games also includes the 'Capture the Flag' games which have been played at Defcon since 1996 or so[3]. It also includes "Exploit!," a card game developed by Ariel Futoransky of Core Security of Argentina. More recently, other games have been created, including the VOME project's Privacy[2].

1.5 Games for Game Theory

There are a set of games which come from the world of mathematical game theory. The best known of these is probably the Prisoner's Dilemma, but the history includes a set of dueling games proposed by Rand researchers as a way to model nuclear war-fighting options [23, 3]. More recently, it includes FlipIt designed by Rivest et al, to model computer security [24]. These games are distinguished by the fact that they are designed to facilitate mathematical analysis, rather than played for fun, or even a pedagogical purpose.

1.6 Psychology and Sociology

The Elevation of Privilege work draws on a number of areas of psychology and sociology which are worth discussing. A full discussion is beyond the scope of this paper, and references have been selected for accessibility to a security audience. The commonalities here are cognitive biases, as well as the different relationships that experts and beginners have with tasks, and with each other as they execute those tasks.

The first set of related work is Csíkszentmihályi's concept of flow [7]. Flow is a state of undistracted concentration on a task at hand, and is associated with effective performance by experts in many fields. Csíkszentmihályi describes "the person is fully immersed in what he or she is doing, characterized by a feeling of energized focus, full involvement, and success." Many structured approaches to threat modeling actively inhibit flow in both beginners and experts, and few allow it to emerge. The apparent lack of flow in threat modeling by developers was one of the motivators for this work. Other elements of flow include:

1. The activity is intrinsically rewarding
2. People become absorbed in the activity*
3. A loss of the feeling of self-consciousness*
4. Distorted sense of time
5. A sense of personal control over the situation or activity*
6. Clear goals*
7. Concentrating and focusing
8. Direct and immediate feedback*

[1] I am grateful to Gary McGraw for drawing my attention to his Silver Bullet podcast with Laurie Williams.

[2] I am grateful to many friends for teaching me Spades.

[3] There was an official contest at Defcon 4 and then later they became more formalized, with substantial effort devoted to challenges and scoring systems organized by the Ghetto Hackers and others.

9. Balance between ability level and challenge*

I have added trailing * to those where I have personally witnessed regular or systematic failures of threat modeling systems to achieve this property. A full discussion of how threat modeling approaches intersect with flow is beyond the scope of this paper. (I am not attempting to argue that flow is either a necessary or sufficient criteria for assessing security processes, but it does offer a framework for addressing a class of issue.) As an example, many approaches to threat modeling have no clearly stated and achievable goal. For example, a goal might be to "find all possible security problems." This is an exceptionally broad goal and one whose achievement is subject to extended argument. Similarly, many processes require diagrams, but offer no criteria for what constitute sufficient diagramming.

A second set is Klein's work on the performance of experts in high stress jobs such as airline pilots and firefighters. In [17] Klein documents how experts use a technique of recognition-primed decision to rapidly come to a plan. For example, a fireman responding to an alarm might see evidence of fire in a stairwell, and little smoke. This is thus a vertical fire, and should be fought by spraying water down. Other vertical fires, burning longer, may make it difficult to get above, and thus different strategies would be called for. [page 16] These techniques often appear (even to the experts executing them) to be "intuitive," and the result of difficult-to-explain experience. My observations of and interviews with expert threat modelers indicate that many appear to operate with a recognition-primed decision approach and have difficulty explaining it to beginners. When pressed, many security experts fall back on phrases such as "think like an attacker," or "the security mindset." These phrases are as useful to most beginners as saying "think like a professional chef." Elevation of Privilege works, in part, because it provides structure and constraints to the successive evaluation activity. This aids in balancing ability and challenge in those learning to threat model.

A third area is Star and Griesemer's boundary objects[4]. A boundary object is one that is used in different ways by different communities, and allows members of different communities to refer to some object at the boundary of their intersection. The classic example is preparation for a museum exhibit on ornithology. The museum curators thought in terms of exhibited objects and constructing a learning experience around them. The ornithologists thought about the zoological aspects of the specimens. Both are able to communicate by pointing at the actual birds, the object at the boundary of their interactions. In threat modeling, where security and software engineering communities come together, there are several important boundary objects, including software diagrams, bugs, and (potentially) Elevation of Privilege cards. The cards were designed to work as one of these boundaries, with a need to provide affordances[4] to members of both communities.

Finally, the rich literature on cognitive biases is important to security tasks. Well documented biases like availability and priming are at play. The availability bias is that generally only a small subset of things will come to mind because of experience, drama, or recency. For example, people asked about common causes of death might include sharks, and forget obesity. The priming bias is that people's memories and estimates can be primed. For example, when people asked to write down the last 2 digits of their SSN, and then estimate a numerical quantity, their estimates are higher for those with a higher last two digits. Those primed with a set of words starting with Z are more likely to list 'zebra' when asked for a list of animals. It is easy to see how a threat modeling approach centered around 'brainstorming' would fall victim to these and other biases. Kahneman is an excellent resources on these sorts of issues[15].

2 The Game

Elevation of Privilege is a card game designed for 3 - 5 players. The game is given away by Microsoft at trade shows or other events and for download as a PDF or as the Adobe Illustrator source files[5]. Some sample cards are shown in Figure 1 and 2 and we encourage readers to download and print a copy and to play with it.

The game consists of 84 cards, including 2 instruction cards, 1 play and strategy flowchart card, 74 playing cards, 6 reference cards, and an 'about' card. The cards are in six suits: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. This is based on the STRIDE mnemonic introduced by Kornfelder and Garg [18][6]. Each suit consists of cards numbered much like normal playing cards, 2-10, Jack, Queen, King, Ace. (There are 74, not 78 cards because the Elevation of Privilege suit starts at 5, and Tampering starts at 3 because we didn't have sufficient hints to fill out the suit and wanted to avoid repetition).

Each playing card shows a suit, a number, and a threat of the type exemplified by the suit. An example of the threat would be the 3 of Tampering, which reads "An attacker can take advantage of your custom key exchange

[4] An affordance is a term of art from usability, meaning something which advertises its function, such as a push bar on a door, or an arrow on a scroll bar.

[5] `http://www.microsoft.com/security/sdl/adopt/eop.aspx`

[6] STRIDE was designed as a mnemonic, and is often incorrectly referred to as a taxonomy. It works well as a mnemonic, and poorly as a taxonomy.

or integrity control which you built instead of using standard crypto." These threats, or hints, help non-security-experts find problems. Aces are slightly different: each reads "You have invented a new (Suit) attack." This is designed to reward creativity and (quite literally) thinking outside the box. To make it easier to decide if something is covered elsewhere, the deck contains 6 reference cards which list all the threat hints.

Elevation of Privilege is licensed under a Creative Commons Attribution license.[7] We encourage readers to play the game and to play with the game, modifying it and customizing it. Enjoy yourself.

2.1 Preparing to play

An Elevation of Privilege game is usually initiated for one of a few reasons. Those include because a group of developers has a system or feature to threat model, because someone wants to learn or teach the skill, or because someone has picked up a copy of the game and wants to explore. This is a super-set of all non-game motivations to threat model. In any case, it is important to start with a system to be threat modeled, and an architectural diagram of that system should be available. A whiteboard diagram is ideal if participants agree it is reasonably accurate and it shows programs, data flows and data stores. If no such diagram exists, it needs to be created before play starts (see [27].) Players need a way to track bugs. Pen and paper is fast and easy. Last but not least, players need a deck of cards; one created on a black and white printer suffices, but the game is more enticing with a deck printed on glossy card stock.

2.2 Playing the game

Play starts by dealing out the entire deck, and ensuring players are familiar with the rules. Shuffling the deck is optional but encouraged because it puts players in mind of a game. Players should be encouraged to put their cards on the table, and arrange them by suit. Players are encouraged to help each other, unless they're a particularly cut-throat bunch. The rules are as follows.

Play starts with the player with the 2 of Tampering, and then proceeds clockwise around the table in tricks[8]. Starting with the 2 of Tampering is a design choice, see below. Each trick is played 'in' the suit that was led. That is, each player must play a card of that suit if they have one. Playing a card consists of reading it aloud, and explaining how it applies to the system being threat modeled, and putting it in the center of the table. Playing a card where a player knows of a compensating control is less exciting, but still valid, because it allows for discussion of compensating controls, and helps newcomers to threat modeling understand the cycle of discovery and mitigation. If the player has no cards left in the suit that was led, then they may play a card from any suit. After each player has played a card, the trick is won by the player who has played the highest card in either the suit that was led or in the 'trump' suit, Elevation of Privilege.[9] The highest card is the highest value card played in the suit led, unless there was one or more trump card played. If a trump card has been played, the highest value trump card is the winning card. A scorekeeper takes note of the threat and who found it. A point is allocated for each threat played, and optionally each brainstormed threat (see design tradeoffs.) The written rule says to only count the highest card which was actually connected to the system being developed, but in practice this is sometimes discarded to give a deeper involvement to beginners.

Aces are played by the invention of a new threat. This generally requires one of the reference cards and some discussion of 'are these threats equivalent?' This discussion of equivalency between threats distracts somewhat from game play, but offers an opportunity for a deeper discussion of the suits and threat equivalency.

Play then proceeds with the player who won the previous trick. That player leads the next trick by selecting a card from his hand, and playing it as above. The play then proceeds through the trick, and further tricks until players are out of time, cards, or ways to connect the threats on their cards to the system they're threat modeling. Playing through generally takes 60-90 minutes, a time which is highly dependent on the amount of conversation between players. Andy Ellis of Akamai has reported a game which ran 30 minutes per trick, involving deep discussion of each vulnerability and its variants. [8]

After the game, the scorekeeper should create bugs, one bug per threat identified, in whatever system a development team uses to track bugs. Those bugs should be triaged as any other security bugs, and appropriate mitigations or test cases created. (I'm familiar with several teams who ensure that they have a working test case per threat, so even if they're aware of the issue and have already addressed it, test case creation may be a useful out-

[7] Specifically, CC-BY-3.0, as specified at `http://creativecommons.org/licenses/by/3.0/us/`

[8] For those not familiar with American card game terminology, a trick is one time around the table, and is sometimes also called a 'hand.' I'll endeavor to call it a trick to distinguish it from a hand, which is the set of cards held by one player. A trick starts when a player selects a card and 'leads' with it. The trick ends when each player has played one card. The winner of a trick can also be said to have 'taken the trick' (or 'taken the hand.') A trump card is one of a particular suit that "outranks" the suit that was led. As a game player, I was surprised to discover that not everyone is familiar with these terms.

[9] This play mechanic is taken directly from Spades, and I am deeply grateful to Douglas MacIver for pointing out that the name "Elevation of Privilege" makes more sense than my original "Threat Spades."

come.)

3 Designing the Game

3.1 Initial design

Once I was inspired to create a game to help people learn threat modeling, I decided that the key areas for me to improve on prior work would be to better encourage competition and flow (in the sense of Csíkszentmihályi). But how to do so without being a game designer? I borrowed, intentionally and without remorse, because experience teaches that game mechanics matter enormously, and are incredibly hard to get right. Rather than try to create something new, I iterated through games that I had enjoyed, testing each for a possibility of converting it to a threat modeling game. Existing games which come to mind are a good starting point because such games have mechanics that work. I had available as building blocks STRIDE and the hints which we had created for the SDL Threat Modeling Tool. I explored a number of ideas. I looked at board games[10], dice games[11] and card games, looking at War, BlackJack, poker and Hearts before seeing how "Threat Spades" might be made to work with suits based on STRIDE and a question per card. Additionally, Spades is a sufficiently traditional game that there was a much lower risk of patent or trademark issues.

The initial design incorporated space on the cards to write notes about a threat, who discovered it, how many points they earned for it, and a bug number. The bug number space was intended to allow players to use the deck as a tracking mechanism as bugs were entered. (When all cards had bug numbers, you could discard that deck.) The first prototype cards were printed on perforated business card stock, and are shown in figure 2.

The game is intended to promote flow and a sense of accomplishment by starting and ending with something developers know: in particular, whiteboard diagrams and bugs. The whiteboard diagram is a good first step because it's known and understood, and bugs are a good ending step because it will remind people that threat modeling was fun and productive, as well as increasing the likelihood that the threats discovered will be addressed.

[10]Trivial Pursuit appears ideal, but requires more questions than would be generally applicable to a broad range of systems.

[11]I think there's probably a good threat modeling game to be created with dice, for example, with a STRIDE die and perhaps a table of interesting threats, but see discussion of beginner/expert under Design Tradeoffs.

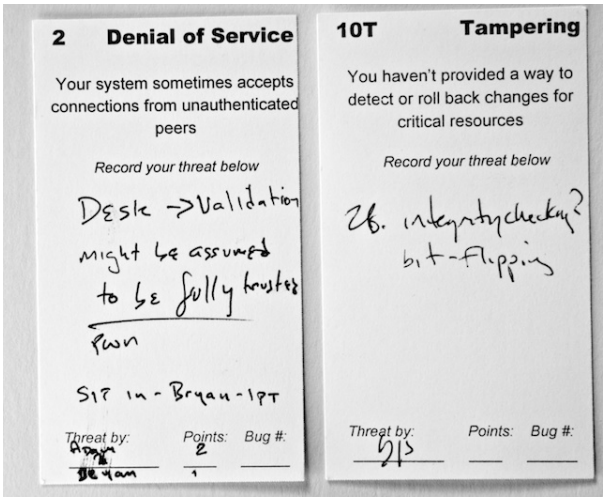

Figure 2: The first Elevation of Privilege card prototypes

3.2 Playtesting and Serendipity

Even having taken the basic playing mechanics from Spades, we made changes, and those changes required playtesting to see how well they would work in practice.

A good deal of how the game came together was a balancing of careful attention to players' reactions and the goals of the game. For example, the final card size was motivated by observations that players had difficulty holding their cards and reading the threats on them. One put their cards face up in front of them, and players started helping each other. Since this seemed a useful feature, I adjusted the final card size to encourage putting one's cards on the table.

Another element that I really wanted to have work was the complex scoring. The initial approaches to scoring were quite complex, and eventually I realized that players (even the dedicated repeat playtesters) were looking to me at the end of every round to ascertain who scored. Scoring was replaced with a much simpler system.

3.3 Design Tradeoffs

Card size The final card size was a tradeoff between not being able to hold the cards in hand (to encourage putting ones cards on the table) and being able to shuffle. Prototype large card decks were large enough that the cards were tough to shuffle or manage. The cards are also large to make it difficult to put laptops in the center of the game table. The cards are 12 cm by 7, in contrast to standard playing cards of 8.75 cm by 5.5.

Game vs Serious Game The game aspects of Elevation of Privilege are important. A serious game with enough 'game' to pull people in can encourage aspects of game play better than 'gamification' of existing tasks.

Play is important to Elevation of Privilege for a number of reasons. Initially, that was because security experts really do have fun when threat modeling. They bring creative and playful elements to the task in a way that beginners cannot. The element of enjoyment, and the differing ways in which experts and beginners engage with the tasks was a key motivator for creating the game. There's an additional issue, which is that for those not steeped in security knowledge, "traditional' threat modeling can be threatening and scary. It's important and no one wants to get it wrong. A number of design elements, including the name, the graphical design, the repeated use of the words 'play' and 'enjoy', and the key message that "Elevation of Privilege is the fun way to get started threat modeling" are all designed to entice developers and reinforce the message that Elevation of Privilege is the fun way to get started threat modeling.

Serve Experts and Beginners Because security experts are familiar with how to threat model, they spend most of their time threat modeling, and relatively little time worrying if they are doing it right. For example, consider a novice chess player, who has to think through '2 spaces on one axis and one on the perpendicular.' That player will take longer per knight move than someone more experienced. The beginner is also likely to explore moves with other pieces first, and thus less likely to make good use of their knights. To continue the analogy, imagine the experience of that newcomer playing speed chess against a master. The experience is likely to be unpleasant and avoided in the future.[12] The mechanics by which Elevation of Privilege focuses threat modeling on the hints on the cards helps both beginners and experts. Beginners are given a set of hints on their cards that they can consider. This allows players to avoid staring blankly, unsure of what to say, and offers them a chance to say "how about this card," or "since this is just a game, I'll play..." (Both happened regularly in playtest). Meanwhile, experts are forced to look for threats that are both on their cards and in the system, or to creatively express threats not on their cards in ways which align with the hint, or to use a (single) Ace. I've also observed experts reading other players' cards and offering them advice on what card to play. This helps the expert be engaged and helpful. Simultaneously, by constraining play to the suit that was led, the set of hints to consider is both constrained, giving players a limited subset to look at even before their turn arrives, and varies through the course of play. Alternate mechanics, such as use of a die roll to select a suit, would not give this "think-ahead" capability to beginners. Game designers should carefully consider if both experts and beginners need to be engaged and thus entertained, and if so, how to balance them.[13]

How many players Elevation of Privilege takes its suggested participant counts from Spades and experience. We set a minimum of three because two player games often require very different mechanics from multiplayer games. We playtested with 5 players, and observed playtesters getting a little bored between their turns, but not leaving. The game might work for 6, but we considered it risky, and since we expected the experiment to fail, we didn't bother with a playtest. One early adopter with a single physical deck attempted to run a game where 20 or so players were put onto 4 teams. That failed to hold everyone's attention, and I'm grateful for the report that it failed.[33]

The hints There are many possible ways to form the hint sentences. We considered at least three forms: the highly personal "your code," the impersonal "there's a problem," and the evocative "an attacker can." Each had reasons and advocates, and this seems like an ideal place to experiment. Because of the difficulty gathering testers, we did not test to see which produced the best results.

Which hints on which cards? We had a pre-existing set of threats from the SDL Threat Modeling Tool, and needed to determine what threats would go with each card. We wanted a mechanic that associated 'better' threats with higher cards, as higher cards are more likely to take a trick, but what is 'better'? Is most impactful better? If so, the high-value cards may represent unusual issues and thus be hard to play, which seems bad for a casual game. If we ordered by prevalence, seemingly trivial threats like cross-site scripting might be dominant. We ended up collapsing a combination of our perceived frequency of encounter, impact and ease of exploiting the threat into an imprecisely ordered list.

Depth versus breadth There were a number of mechanisms in early versions that focused on aspects of expert threat modeling, such as building on threats found by other participants. One such mechanism was the ability to "riff" on a threat suggested by a card, getting points for the riff. That mechanism was exciting to threat modeling experts, and difficult or baffling to newcomers to the craft. We moved these mechanisms to a set of "optional variants."

Electronic or cards As an organization, Microsoft is biased towards shipping software. Shipping the game online would have allowed us to reach more people at lower cost, to track play of the game, and to update it as we learned things. So why a box of cards? There are a number of reasons. As a physical item, the game draws attention[14], and allows people to point at it in ways

[12] Newcomers to threat modeling may well take a similar approach.

[13] Alternately, they can just luck out and not even be able to express such a thing until they're writing up a paper.

[14] As '@SecurityNinja' tweeted "Everyone that walks past my desk has to play with the Elevation of Privilege cards, eye catching!"

that are potentially awkward with a screen. The physical implementation also forces people to sit around a table to play the game in a way which reinforces the game message, reduces distraction and encourages discussion. These advantages to a physical game have made us reluctant to build an online version.

How to start play There are a number of common patterns for starting play of card games, including the player to one side of the dealer, tricks are led moving around the table, and a defined card. The selection of the 2 of Tampering was grounded in several elements. First, it improves flow and the speed of the game by leading to a quick decision. A new player doesn't have to worry about which of their cards to lead. Second, Tampering issues are common–a playtest went poorly when a player led the first trick in Repudiation. Finally, it mirrors the design of Spades, where play starts with the 2 of Clubs.

Graphics All playtesting was done with black and white, graphics-free cards. There was a tradeoff of investing cash in more card production, using it for something else, or spending more or less on graphic design. The graphic design has contributed tremendously to the success of Elevation of Privilege in several ways. First and foremost, it stands out, and invites question and discussion. This is important because people who might otherwise walk by a desk or trade show booth will stop to see what it is. Without that initial contact, they will not learn that there's a game for threat modeling. Many people leave a copy on their desk to invite discussion. (I'm told there's a copy on a desk in the White House for that purpose.) Second, the playful graphic design reinforces the game message when people take it out to play. While the cost of a design project will obviously vary greatly based on project size, the local market and a number of other factors, it may be useful as a scaling aid to understand that we invested less than $10,000 with a local design firm to design and manage production of the boxed card sets. To help those designing games, I've included additional discussion of the design and production process in an appendix.

Sample systems It may be that including a sample system with an 'answer key' would enhance learnability. There is an interesting tradeoff here. On one side is Elevation of Privilege results in real issues in real systems about which the players know. On the other is the ease of getting started, and the flow issue of balance between ability and challenge. I have not observed or heard about substantial trouble in getting started, but that's a sort of trouble which is unlikely to be reported. The frequency of difficulty and the benefit of a sample could be experimentally established.

4 Reception

In discussing results, there are a variety of criteria which might be applied. The question of what the best ones would be is challenging. One obvious and nominally attractive one is to compare threat models produced while playing the game to those produced by experts. If a game could allow beginners to compete with experts, it would be quite a feat. It is also an unrealistically high bar. There are few (if any) fields where expert performance can be achieved without many hours of deliberative practice. In fact, endeavors where mastery is so easily achieved are likely not viewed as having 'experts.' Better measurements may include inducement to begin threat modeling, or to compare threat models produced by the game to threat models produced by novices with other aids or training. The comparison of approaches that can be used by low-engagement practitioners is valuable, as explained in the background section. These attractive comparisons may distract from the fact that a very small portion of the software development in a given year has any threat modeling activity at all. (See, for example, automobiles, medical devices, power grid and other critical infrastructure, et cetera ad nauseum.)

Thus, one useful evaluative criterion is that people are induced to obtain a copy and learn about threat modeling. This is an unusual criteria for a paper, but given regular complaints about the gap between theory and practice, or the lack of security activity in commercial development, it should be considered. Another evaluative criteria might be the establishment of 'flow' in participants. For that, we are pleased to report on one late group of testers, Microsoft Security Development MVPs, all of whom have a strong professional interest in security. After 30 minutes of playing, the group didn't disperse to get lunch, and had to be repeatedly interrupted and reminded that food was only available from 12 to 1.

Outside of Microsoft, Elevation of Privilege appears to be a success at helping security experts introduce developers to threat modeling. By enticing people to threat model, it makes it easier to fit into busy schedules. Results are somewhat tricky to document, which could indicate a problem, or simply that security experts prefer not to admit that their previous efforts to drive threat modeling adoption were not very good. Two years after its introduction, we still receive regular requests for cards. One local company with strong ties to Microsoft recently requested 90 copies for internal use. Martin McKeay has discussed how his six-year-old son attempted to hack his eight-year-old brother's Minecraft account, and that he got the idea from playing with Elevation of Privilege cards. A state agency has commented that when they showed up game in hand, doors previously slammed in their face were opened. The game has also led to im-

proved information sharing. A national government documented an attack under the name "Revelation of Privilege," complete with eight custom cards incorporating their national symbols and details of the attack. (Unfortunately, that means we can't share the delightful slides they created to lead up to the actual details without compromising their privacy.)

An important sub-area of introducing people to the idea of threat modeling is that of inducing students to study information security. One professor reported "[My students] all wanted to keep the cards." [34]. Serious games may be an important way to show students that the STEM subjects of Science, Technology, Engineering and Math can be both enjoyable and creative. Games have also been linked to making STEM subjects attractive to girls[12].

None of this is to say that more formal testing of the stated educational purposes of the game would be a waste. (Those goals are to teach threat modeling, and to do so in a way that brings its participants into a flow state.) Such testing could provide sample systems to a population randomly divided into groups provided with different training or tools, and the results evaluated for some set of values, such as the number of threats discovered, valid threats discovered, time taken, and self-assessed likelihood to bring the training or tool to their professional activity. If the population being studied does not have exposure to security issues, they are unlikely to do very well without any training or tooling. For example, group A could be given the Microsoft SDL Threat Modeling Tool [21], which has been designed to help developers, and group B could be given Elevation of Privilege. (Testing with several Microsoft systems reduces the number of uncontrolled variables, such as STRIDE-centered vs other means of eliciting ideas.) It would also be possible to test from a population of experts brainstorming versus using Elevation of Privilege.

Inside of Microsoft, Elevation of Privilege has been more popular as a training tool than as a threat modeling tool, despite the fact that players produce a threat model. It is worth discussing two reasons why that might be so. One reason is that Microsoft employees are required to threat model and to document their threat models. Microsoft has an SDL Threat Modeling Tool which is both more thorough than Elevation of Privilege and the tool is integrated into SDL process tracking. Related to this, work and play are in many senses opposites. Making a game out of a work requirement doesn't free those burdened by the requirement to play in any full sense of play. Yet it simultaneously risks trivializing the requirement, and engendering a feeling of resentment. The game has been used internally to teach threat modeling, and that turns out to be in accordance with predictions documented in [30]. In that, Smith categorizes "skill expansion" as one of the areas in which games at work can be successful.

5 Lessons Learned

Not everyone plays card games. While obvious in hindsight, it turned out that some players (especially those from non-western cultures), had little or no experience with card games, and terms like "trick" or "led" were not familiar to them. Writing rules to be accessible without being tedious was surprisingly difficult. It turned out presenting the rules as flowcharts was useful here.

Playtesting Getting playtesters for a serious game is serious work. Getting the game tested by the target audience was essential, but getting a set of 3-5 people who didn't care much about threat modeling to commit an hour was a challenge. Games appear to be the easiest work to cancel in a busy day. Over the course of several months with several other commitments, we ran 3 or 4 successful playtests with Microsoft software engineers from a variety of educational and cultural backgrounds. As a result, we did not test areas where testing was warranted. For example, there was no comparison of ways of phrasing the hints, as that's a key mechanic for tying the game to the threat modeling goal. Getting participants to test all three variants was a sufficient hurdle that I didn't do it. One might budget to pay, reward, or otherwise entice testers.

Prototyping As mentioned, testing was done with business card stock and cards laid out in Microsoft Word. I was making cards aligned like playing cards, but the card stock was horizontal. This led to odd text flow issues as cards were edited, and aligning text was challenging. Doing playing card-oriented prototypes on landscape card stock turned out to be a huge pain, and inhibited me from iterating.

Scoring systems Early approaches to scoring were motivated by the desire to tie the game closely to many of the goals of threat modeling. This resulted in a system far simpler than those used in most tabletop role playing games. But that's not really the right comparative point. The system was far too complex and had to be simplified. The scoring system for an early version was as follows:

> Points: 4 for a threat on your card.
> 3 for the first threat on someone else's card, 2 for the next, 1 for the next*
> +1 for taking the trick
> +2 for a face card (J/Q/K)
> +3 for an ace
> ✻ The 4,3,2,1 is the "go broad" variant of the game. If you'd like a "go deep" variant, you can either assign a static or increasing number of points for each additional threat on a card.

In the future, we would select a single goal and align scoring around that, or ensure that there was a support system for scoring.

A game forces participation One interesting lesson, pointed out by an early tester, was that the game forced everyone who showed up to the room to participate. A lack of understanding which could be concealed by nodding along and deferring to others was made quite clear when it was your turn to play a card. If handled well (in an understanding manner, rather than punishment) this is a substantial advantage over other forms of training which allow deficiencies in understanding to be concealed.

Organizational Constraints Moving from a 'day to day' space to a 'game-playing' space is a change that will carry with it complex and nuanced meanings that are highly dependent on the backgrounds and experiences of the players, as well as the culture of the organization. I'll touch on two. The first is that a game gives permission for creative, non-linear exploration, and perhaps as a structure which allows incomplete thoughts. Another issue which was revealed by the game structure was leaders who didn't want 'their' systems attacked. Some of these leaders were managers, others were senior technical staff. A playtester said that they could avoid confrontation about the problem by focusing on getting a point in the game, and then filing a bug. Once something is tracked as a bug, the argument for closing it has to be in writing, and the social elements and pressures were changed. This was an unexpected side benefit.

Jokers Early versions of the game included 6 joker cards. They were intended as jokes, with no impact on game play. For example, the Joker of Tampering reads "You've slipped a new tampering threat into the deck undetected." The Joker of Denial of Service reads "You spilled scotch on the deck, making it unusable." Unfortunately, the similarity of the Jokers to real cards led to confusion, and one tester said simply "What the hell do I do with this?" What we did was remove the jokers.

Physical production Version 1 boxes were too large, and the cards rattled. Combined with not all suits starting at 2, this led to worry that cards were missing. We shrunk the boxes to fit snugly, and also added a parts inventory to the instructions. Production in the US was more expensive than production in China. We also encountered trouble with international shipping because the first boxes had no statement of where they were produced.

Checklist to game framework Elevation of Privilege can be viewed as a game structure for a checklist of items that may reward discussion. For those problems where the elements of the checklist are there to prompt thinking or discussion, the Elevation of Privilege model may be re-used. This would be easiest where the items can be divided into approximately 4-6 suits of 10-14 items.

There is a science and craft of game design Much of the work documented here might have gone faster with more awareness of works on game design, ranging from the applied to the scholarly. In particular, a book by Schell [26] offers solutions for many problems. Much of the work on the science of game design and evaluation has helped ensure that games are not excluded from the academy, while not engaging with the craft or art of design. (For an example, see *Rules of Play* and in particular the debate in the Amazon reviews [25].) Those seeking to design new games should be aware of the distinction.

Graphic design was cheap One of the ways that Elevation of Privilege succeeds at attracting or enticing people is because it just looks cool. It's easy to underestimate the value of that, and hard to discuss in a scientific paper. But tester response was transformed when we brought in colorful cards, and people continue to snatch them up at conferences.

5.1 Future work

There is interesting and valuable work that could be done with Elevation of Privilege, and more broadly with security games.

Possibilities around Elevation of Privilege include comparative testing of the game against other approaches to threat modeling, and we are interested in helping such an effort. There's a possibility of enhancing the game to contain 13 threats per suit. It would be worthwhile to test if a sample system and answer key enhance the value of the game. Translations are underway into Portuguese and Mandarin Chinese which opens fascinating possibilities for work comparing the impact of culture on secure development activity.

Elevation of Privilege (and I hope, this paper) act as proofs that there is interesting work to be done in helping non-experts approach security. Important elements of that include applying games and Csíkszentmihályi's concept of flow to problems in information security. There are a great many subjects in security education, secure development and security operations where one or both of these concepts might help address problems which currently appear intractable.

In reviewing this paper, Cormac Herley commented that threat modeling has received little attention from academics and we lack theory which would usefully underlie a more structured pedagogy, system of measurement or comparison between systems or approaches. Ari Jeuls and John Benninghoff drew attention to cognitive biases and threat modeling, and I believe that studying security methodologies to uncover the impact of cognitive biases and how to overcome them will be fruitful. Additionally, work on the intersection of flow and security tooling may be worthwhile.

6 Acknowledgements

Thanks to Laurie Williams and Austin Hill for inspiration. Jacqueline Beauchere provided funding. Eugene Bobukh spent several hours helping to rank the cards. Thanks also to Rob Reeder for playtesting and advising throughout. Mark Cartwright, Douglas MacIver, Deepak Manohar, Katie Mousouris, Ross Smith, Bryan Sullivan and others playtested. Donald Brinkman, Angela Gunn, Cormac Herley, Ari Juels, Wendy Nather and Dominic White and anonymous referees provided helpful feedback and suggestions on the paper. Jon-Paul Dyson was an invaluable guide to the literature of serious games.

7 Disclaimer

References

[1] ABT, C. *Serious Games*. Viking Press, 1970. (University Press of America, 2002 edition).

[2] BIRCH, D. Playtesting a privacy education game. blog post, April 2011. http://www.chyp.com/media/blog-entry/playtesting-a-privacy-education-game.

[3] BLACKWELL, D. The Noisy Duel, One Bullet Each, Arbitrary Non-Monotone Accuracy. Tech. rep., The Rand Corporation, 1949.

[4] BOWKER, G. C., AND STAR, S. L. *Sorting Things Out: Classification and Its Consequences*. The MIT Press, August 2000.

[5] BROOKS, F. *The Mythical Man Month*. Addison-Wesley, 1975.

[6] CHECKOWAY, S., MCCOY, D., KANTOR, B., ANDERSON, D., SHACHAM, H., SAVAGE, S., KOSCHER, K., CZESKIS, A., ROESNER, F., AND KOHNO, T. Comprehensive experimental analyses of automotive attack surfaces. In *Proceedings of the 20th USENIX conference on Security* (Berkeley, CA, USA, 2011), SEC'11, USENIX Association, pp. 6–6.

[7] CSÍKSZENTMIHÁLYI, M. *Flow: The Psychology of Optimal Experience*. Harper Perennial Modern Classics, 2008.

[8] ELLIS, A. "Ahh, one hour, second trick of Elevation of Privilege". Tweet, January 2012.

[9] GATES, B. Trustworthy Computing. Internal company memo, Jan 2002.

[10] GRENNING, J. Planning Poker or How to avoid analysis paralysis while release planning, April 2002. Short Note, http://renaissancesoftware.net/papers/14-papers/44-planing-poker.html.

[11] HANNA, S., ROLLES, R., MOLINA-MARKHAM, A., POOSANKAM, P., FU, K., AND SONG, D. Take two software updates and see me in the morning: The Case for Software Security Evaluations of Medical Devices. In *Proceedings of 2nd USENIX Workshop on Health Security and Privacy (HealthSec)* (August 2011).

[12] HAYES, E. Using Games and Digital Media to Engage Girls in Computing. White House Champions of Change blog, December 2011. http://www.whitehouse.gov/blog/2011/12/12/using-games-and-digital-media-engage-girls-computing.

[13] HOWARD, M., AND LEBLANC, D. *Writing Secure Code*, 2nd ed. Microsoft Press, Jan 2003.

[14] HOWARD, M., AND LIPNER, S. *The Security Development Lifecycle*. Microsoft Press, June 2006.

[15] KAHNEMAN, D. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, October 2011.

[16] KEIZER, G. Symantec says stop using pcAnywhere after Anonymous threats. Web site, January 2012. http://www.computerworlduk.com/news/security/3332888/symantec-says-stop-using-pcanywhere-after-anonymous-threat

[17] KLEIN, G. *Sources of Power: How People Make Decisions*. The MIT Press, Feb 1999.

[18] KOHNFELDER, L., AND GARG, P. The Threats to Our Products. *Interface* (1999). Interface is an internal Microsoft journal, the article is available at http://blogs.msdn.com/b/sdl/archive/2009/08/27/the-threats-to-our-products.aspx.

[19] LIPNER, S. Evolving Secure Code at Microsoft and Beyond. Blog post at http://blogs.msdn.com/b/sdl/archive/2012/02/01/evolving-secure-code-at-microsoft-and-beyond.aspx, Feb 2012.

[20] LOCKE, J. *Some Thoughts Concerning Education*, (Kindle edition "with active table of contents") ed. Black Swan, Paternoster Row, London, 1693.

[21] MICROSOFT. SDL Threat Modeling Tool. Software Package, October 2008.

[22] NATHER, W. Personal Communication. email, March 2010.

[23] QUADE, E. The Duel with Time of Flight Not Zero. Tech. rep., The Rand Corporation, 1949.

[24] RIVEST, R. L. Illegitimi non carborundum. Invited keynote talk given at CRYPTO 2011.

[25] SALEN, K., AND ZIMMERMAN, E. *Rules of Play*. The MIT Press, 2003.

[26] SCHELL, J. *The Art of Game Design: A book of lenses*. Morgan Kaufmann, August 2008.

[27] SHOSTACK, A. Experiences Threat Modeling at Microsoft. In *Modeling Security Workshop In Association with MODELS '08* (2008), J. Whittle, Ed.

[28] SHOSTACK, A. Engineers are People, Too. Keynote address at Symposium on Usable Privacy and Security http://www.homeport.org/~adam/EngineersarePeopleTooSOUPS2010Shostack.pptx, July 2010.

[29] SINGER, N. Employers and Brands Use Gaming to Guage Engagement. web site, February 2012. http://www.nytimes.com/2012/02/05/business/employers-and-brands-use-gaming-to-gauge-engagement.html.

[30] SMITH, R. The Future of Work Is Play. In *International Games Innovation Conference* (November 2011), IEEE.

[31] SUITS, B. *The Grasshopper: Games, Life and Utopia*, november ed. Broadview Press, 2005.

[32] SWIDERSKI, F., AND SNYDER, W. *Threat Modeling*. Microsoft Press, July 2004.

[33] WHITE, D. Personal Communication, April 2010.

[34] WILLIAMS, L. Personal Communication. email, October 2011.

[35] WILLIAMS, L., GEGICK, M., AND MENEELY, A. Protection Poker: Structuring Software Security Risk Assessment and Knowledge Transfer. In *Engineering Secure Software and Systems*, F. Massacci, S. Redwine, and N. Zannone, Eds., vol. 5429 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2009, pp. 122–134.

Appendices

A Graphic Design

This appendix covers some aspects of the graphic design process with the hope of helping those developing games.

Vendor Selection

Myself and a colleague with more experience getting physical items produced interviewed several firms, and picked one that seemed enthused about the project, offered a good price, and brought a portfolio that went beyond brochures, flyers and websites. The vendor we picked also had things to say about the design and play of games. They were also able to handle project management of physical production. We sat down to discuss the playful nature, and I offered up the ideas I'd had, while being clear that I wasn't a designer, and they didn't need to use my ideas.

Vendor Management

Because large organizations often tend towards conservative designs, I explicitly encouraged the designers to be creative and to push the boundaries of what a large organization would normally be willing to accept. With that guidance, they produced a set of approximately 15 design explorations. Each of these was a collection of images, words and colors that they felt held possibility for further exploration. They used some of my ideas, and those explorations weren't inspiring. We didn't include my initial ideas in the set of ideas to develop further. From the sketches, they produced 4 mock-ups of cards.

I showed those mock-ups to various people on my team, and asked for their opinions, making very clear that I was going to filter their feedback and make decisions. I took copious notes in these design sessions, and aggressively filtered feedback into three categories: must fix, opinions, and not shared. An example of must fix was that the initial designs included hand grenades on one card. Microsoft has a number of policies in place designed to avoid offending our customers, and depictions of weapons requires solid business justification, which we didn't have. Similarly, a reviewer identified that one character looked remarkably similar to a trademarked one, and we replaced that character. We set very high bars for "must fix", or "worth sharing with design firm." Despite those high bars, the feedback process raised confidence that we were finding 'must fix' sorts of items. Filtering the feedback was an essential part of avoiding design by committee.[15] Another essential element was to ensure that we not only budget but also that the designers had time and freedom for design explorations. Much like Brooks' law of "plan to throw one away," [5] designers need time to perform their equivalents of prototyping and evaluation. Part of that freedom for the designers was probably recognition that their design concepts were better than mine.

Timelines

Selecting the firm took roughly two weeks. Initial design explorations took several weeks over the winter holidays. Going from design to mock-ups took another week or two. With a design selected, producing a first full version of the design took roughly two weeks. This was spent laying out cards and text, ensuring that the right number of symbols were on each card, and illustrating face cards. We made a number of adjustments and the designers produced a physical deck of the 84 cards (without printing) and a box mockup on styrofoam. This allowed us to discover that the planned sizes for the cards was too large to hold and shuffle, and we shrunk them. From there, we went into a process of proofing, with the test print runs being physically delivered and checked, and then first production.

[15] This may sound easy, but is an important part of good design.