Mental Models of Privacy and Security

L. JEAN CAMP

Security research has produced a rich array of tools and techniques [1], [3], [10]. Anonymity, confidentiality audits, and hardening tools are widely available. Individuals regularly report very high degrees of concern about privacy and security, and these concerns have been steadily increasing [20], [21]. Yet security tools are not widely used, and even freely available privacy-enhancing applications have not been widely adopted. The lack of adoption of security and privacy technologies is in part due to a failure of effective risk communication. In fact, the concept of risk management as a valuable element of computer security has only been recently accepted [4], [13]. Security communication as risk communication has not been seriously engaged.

In terms of risk communication, a large body of literature addresses how to communicate complex risks in simple terms. Risk communication has long been concerned with medical, financial, environmental, and lifestyle risks as indicated by Blackwell's *Risk Analysis* journal and the other publications of the Society for Risk Analysis. Beyond

Digital Object Identifier 10.1109/MTS.2009.934142

 1932-4529/09/$26.00©2009IEEE

the academy, risk communications are embedded in our daily lives, from package inserts to the warnings on every package of cigarettes. Yet risk communication methods have not been applied to computer security.

There is no single uniform method for risk communication. One method is to simplify, and sometimes overstate, risks to change behavior. Examples include efforts to increase awareness of fetal alcohol syndrome (e.g., messages to consume exactly zero alcohol while pregnant), and widely contested public drug education efforts (e.g., "this is your brain on drugs," showing a cooking egg).

Another method is to utilize the existence of human risk heuristics and to communicate somewhat more information using simple mental models of risks. A third method is to provide detailed information and demand rational responses from individuals in their risk behavior, as with the paper inserts into prescription medication.

In computer security, simplification is common, while detailed information is available to anyone with network access. In contrast, there is no documented case of the use of mental models in computer security risk communication.

The mental models approach could significantly improve risk communication in the case of computer security. The particular mental models that will be discussed here are: physical, medical, criminal, warfare, and market models. Our strongest conclusion is that mental models can be used to improve risk communication. The second, untested, conclusion is that the best model may be the medical model.

Risk Perception

The same individuals who buy curtains for privacy and lock doors for security often will not similarly invest in electronic security [27]. Physical security is understood to be a result of physical, tactile risk-avoiding behavior. In contrast, network security is less well understood, and thus there is greater uncertainty in decision-making. Decision-making under uncertainty has been characterized as follows [16]:

> "...[P]eople do not appear to follow the calculus of chance or the statistical theory of prediction. Instead they rely on a set of heuristics which sometimes lead to reasonable judgments and sometimes lead to severe and systematic errors."

Individuals systematically fail to respond to detailed risk information in a manner that would be predicted by strict rationality. These heuristics and biases are well-documented and widely understood in risk communication. The systematic biases indicate that straightforward data-based-user communication, ironically, flies in the face of data on human risk perception.

Consider the quote above in light of a leading ethnographic study of how users act in order to secure their machines. The 2004 study [11] illustrates that the users' practices are little more than token behaviors designed to prevent some unknowable harm that is not understood. For example, the most effective person in the study tilts her screen a certain way in order to protect information while she is in her own physically enclosed office. The location of the screen does not in any way affect the security of a machine with respect to online security threats. The user might as well bury a chicken while facing east at midnight, in terms of computer security efficacy. And the user in this profile is a thoughtful, responsible person who has a position that indicates significant intelligence. It is not a "stupid user" or "careless user" problem. However, her models of network security fail to provide her with a set of effective and rational decision-making tools. Effective mental models could assist her in her judgments.

Individuals use heuristics and biases in evaluating risks. These are widespread and consistent: anchoring, the law of small numbers, irrelevant data, and scariness. (Note the references that follow are drawn from those that resulted in the 2002 Bank of Sweden Prize in Economic Sciences for the authors of [32].)

Anchoring is one consistently ineffective mechanism for evaluating risk. Because unusual risks receive more newspaper and media coverage, individuals consider more newsworthy risks to be more prevalent than non-covered risks. Anchoring is not only awareness. Anchoring also refers to quantitative estimates, and describes peoples' tendency to refer to a recent number and use that number. For example, indicate to a user that x% of people will experience road rage while driving. Then ask how many will experience identity theft. Most users will respond with the irrelevant number of x%.

Anchoring not only reflects grounding of a risk but also response to a risk in three ways: irretrievability, imagination, and insufficient adjustment [32].

Irretrievability refers to the memory of the most recent rather than the most relevant information. The most relevant may not be easily retrievable. Humans can keep at most about fifty mental elements readily available at any time [19]. Thus highly specific and detailed computer security information may be less retrievable than more general but less informative mental models.

The ability to imagine an error or outcome alters risk assessment. Explanations of highly technical failures may cause eyes to glaze over. Starting a description with "assume a generator on the well-defined space" is not likely to create an immediately retrievable or

easily imaginable response. For an effective response to be adequate, the threat and response must be imaginable. "Create passwords that are hard to remember, do not reuse passwords, and don't write them down," does not create an easy to imagine response. In contrast, "passwords are like toothbrushes—they should be used for one purpose, kept in a clean place, and never shared" communicates a requirement for unique passwords and secure storage.

When people are clearly informed of an error in risk estimation, they exhibit insufficient adjustment. For example, if a person guesses a number that is incorrect and is told it is incorrect, they will change only a little. If the number is an order of magnitude off, their initial anchoring is not corrected. So if individuals underestimate risks initially then future estimates are unlikely to be adequately updated. In terms of computer security, if a person initially visits well-known web sites without encountering malware, the person may choose to lower the security settings on the machine (or fail to improve them) as they continue to browse.

Another example of a failed but perennially and consistently used heuristic is the "law of small numbers." There is no law of small numbers. The only real law is the "law of large numbers." There is no law of small numbers. In order to converge to a normal distribution there must be a large number of events.

The classic experiment is to present individuals with the following thought experiment:

Imagine a state with a large hospital and a small hospital. The large hospital expects roughly 400 births a month. The small hospital has forty births a month. Which hospital is more likely to have half boys?

Of course, the large hospital is far more likely than the small hospital to have a mean that represents the overall mean of the population of babies born. In every experiment, individuals suspect that the smaller hospital is more likely to have half. In every version of this experiment individuals expect small samples to be more representative than larger samples [31].

The law of small numbers appears in the most unusual places. In Indiana, the *Bloomington Herald Tribune* ran an article saying that military recruits from Indiana were more likely to die than any other recruits [14]. The staff writer had interviewed multiple military leaders from Indiana. There was no explanation of "why Indiana soldiers had a greater probability of dying in Iraq."

> Risk communication methods have not been applied to computer security.

In fact, any soldier from Indiana sent to Iraq is no more likely to die in terms of distribution of random chance than any other soldier. It was only the case that (at that time), with 1000 soldiers from fifty states suffering violent and untimely deaths, the law of large numbers could not apply. There was simply a distribution of mean rates of fatality and one state would be at the high end of that distribution. Indiana happened to be that state, for which it had less than a 1 in 50 chance.[1] In contrast, the *Herald Tribune* noted that every state had seen roughly the same percentage of wounded soldiers. With more than ten thousand wounded soldiers (at that time) the tragic number was large enough that the law of averages could and continues to apply to the wounded.

The law of small numbers is alive, widely believed, and completely unfounded. Flipping one coin and getting heads does not indicate that the next coin will be tails. Flipping one hundred coins suggests that roughly fifty will be heads, and fifty tails. Yet instead of expecting regression towards the mean, individuals take past experiences as more representative than any rational observation would suggest.

The end result of the law of small numbers in cybersecurity is arguably the opposite of the end result in the brutally incorrect newspaper article, which increased fear. People who have had some small number of risky interactions, for example, downloading illegal content from insecure file-sharing networks, believe these small numbers of events to fully illustrate their own unique risk profile. Having played the odds and won, individuals will believe that their risk perception is rational and fully formed. In contrast, a more rational calculation is that they have had unlikely success in avoiding security risks in a dangerous cyber world. A heuristic that identified behaviors as risky, rather than sharing risk statistics, would arguably be more useful, given human faith that each of us has his or her own unique distribution.

An unrelated heuristic, but one that is consistently experimentally validated, is irrelevant data [30]. For example, a computer on a tidy desk or in a more friendly color may appear more trustworthy, but neither desk nor hue determines in any way the machine's vulnerability to attack. Individuals estimate risks and include all data in their estimates, including the wildly irrelevant. Each individual has a level of certainty about his or her own estimate. For example, suppose we ask someone, "How likely are

[1]States with large numbers of soldiers due to population or location of large bases, e.g., North Carolina, New York, and California, had death rates closer to the mean. Indiana therefore had a less than obvious 1 in 50 chance of being the furthest from the mean.

you to be struck by lightening?" Provision of irrelevant information clearly does not change the statistical certainty of the estimate, by definition, but irrelevant data greatly increases people's confidence in their initial estimates. Provision of irrelevant information is prevalent in information security today. For example, the issuer of an SSL certificate is, for the vast majority of users, completely irrelevant. Either it is a "known" issuer or a "not known" issuer. The website has nice colors (and is more likely to be perceived as trusted) or the website is a loud lime green (thus possibly perceived as less trustworthy). However, more information will validate the user's belief regardless of how correct the initial guess was or was not. Confirmation of beliefs based on irrelevant information suggests that minimal mental models would be preferable over detailed information that cannot be evaluated by the end user.

Finally, information security is systematically underrated because of the sheer lack of dread. People respond to fear and respond even more strongly to terror: wolves, snakes, fangs, and things that go bump in the night interact with risk perception in a systematic and predictable manner. People systematically over-invest in avoiding horrifying deaths (burning, drowning) and under-invest to avoid deaths that are not as frightening. Making security less virtual and more frightening can increase risk awareness. Associating virtual risks with more tactile risks—wild animals, disease, crime and war—can increase sensitivity to and awareness of risks [32].

In summary, exact expert information on computer security is unlikely to result in humans responding with the calculus of risk. Numerical and statistical models of risk will not result in corresponding user investments to reduce the expected value of loss. In fact, communicating the exact risk will instead be itself irrational. Humans have consistent practices in estimation of individual risks. Ignoring human behavior in risk communication is, ironically, itself irrational.

In other domains where there is a high degree of uncertainty, the use of mental models has been found to be more effective, and to result in more effective behavior. Several inchoate mental models that are clearly present, but not systematically utilized, in computer security.

Security and Mental Models

The mental models approach has been effectively used to better communicate environmental risks as well as drug risks [18]. Mental models can be powerful; for example they can reduce misinterpretation when communicating complex and interdependent medical risks [15].

Mental models can also have perverse effects. They are not a panacea. When experts communicate to naive users using models, then these users can take more than is intended by the metaphor. When experts offer mental models by using metaphors, the metaphors are used to explain a particular element of a system based on an expert understanding. Counterproductive implications about risk, control, and best behaviors may be drawn from implications of a metaphor. Mental models are a powerful tool, but not without flaws or complexity. Yet use of mental models in risk communication is a tool that computer security experts have not yet embraced [18].

In order to understand security using mental models, five possible mental models for computer security failures are examined here: physical security, medical infections, criminal behavior, economics failure, and warfare. Each model has different implications for the end user, according to the implied role for the individual in the model, as well as different implications for "safe" behavior. For each case there are also examples in computer security that illustrate that the model can be applicable. Which models of security are best and under what conditions? Also, how can these models effectively be communicated to end users? Clear risk communication is critical for the digital autonomy required for full on-line citizenship. If naïve users are unable to effectively evaluate their own risks and decisions, no amount of technology will be empowering for those users.

Physical Security Model

Processes for physical security are fairly well understood and often tacked onto digital systems. Any attacks require physical presence. Perimeters can be defined in a straightforward manner, and policed. Safes, doors, and other physical security devices have ratings based on time to break. Yet the security process must fit both the threat and the asset. Physical threats do not address the core condition of communications technology—that it crosses space and makes boundaries permeable. For example, consider the canonical case of an extremely secure room with a computer that would be almost physically impossible to reach—with an Ethernet cable running through the thick wall. Of course, that physical security matters is clear from the value of searching others' trash—both in corporate dumpster diving and by suburban identity thieves. Yet on more than one security evaluation, the existence of the wall has been shown as an indicator of network security practice.

Control mechanisms for the management of physical papers fails in the case of digital networked information. For example, paper ballots have rules of custody. Paper ballots are governed by being in a sealed container, and being under the control of at least two people at all times. However, the use of physical review and examination has notoriously failed in the case of electronic voting in the United States where paper ballot processes

have been haphazardously applied to electronic systems [12].

Using a simple physical model from automobiles, is LoJack worth the investment? (Lojack is a system based on radio frequency location identification of an automobile. When a car is stolen, it can be tracked and recovered. Lojack is not visible, so care thieves are unable to determine by examination if they will be tracked in a given car.)

Yes, Lojack prevents theft and solves a real problem. In addition it creates a positive externality. That is, when Lojack is in use by some people, car thefts go down for an entire neighborhood. What about car alarms? They do make noise. However, they are easily disabled and have not been shown to prevent auto theft. Car alarms also create a negative externality; they make neighborhoods less pleasant and decrease social capital. Digital security mechanisms are similarly complex.

Are we living surrounded by car alarms? Is there a digital LoJack that each person can invest in, and begin positive network effects? How can end users distinguish between those technologies that are annoyances (but may make us feel better) and those that offer true progress? To the extent that the model of physical security can be used to communicate implications of virtual security, it can be leveraged.

The physical mode has positive mechanisms because the user is required for physical security. Physical security implies individual and localized control. Physical security tends to imply a capacity to isolate failure that does not apply to computer security. Physical security does require both investment by the end user (e.g., locks) and a community investment in protection (e.g., law enforcement). In many neighborhoods, physical security requires a community investment. In the case of neighborhood watch, the group of neighbors coordinates to improve their joint security.

To the degree that physical security can be used a model, it can motivate users to secure their computers. The dread of a nighttime intruder can be used to illustrate the dread of a computer intruder, who can likewise steal valuables. Yet the differences between the virtual and physical make this a most difficult metaphor to leverage as a mental model.

Medical Model

In August of 2003, the computer security world, already weary after a summer of headline-grabbing security problems, rallied to defend systems against yet another Internet worm, the blaster variants. This time there was a difference. In addition to the worm itself, there was an opponent of the worm, worm_blastMS. Although similar to previous blaster strains that exploited Windows RPC vulnerability, this "good worm" gained permissions through the security hole, then patched the infected system, preventing further malicious code from attacking. The "good" worm was spreading immunity using the same weaknesses and mechanisms used by the malevolent worm [3].

Although this worm still managed to bring down several networks due to a denial of service as the worm enforced patching (including Air Canada and Lockheed-Martin), the security industry's response was mixed. Some systems were undeniable saved, suffering only the requirement of instant patching, in an environment where far too many users failed to protect themselves. The best systematic defense to the worm removed autonomy and control from those who had failed to patch their systems. The worm_blast was dynamic and followed the patterns of waves of expansion common to any who study infectious disease.

The medical model for malicious code is grounded in the diffusion patterns of malicious code and infectious diseases, in the importance of heterogeneity in the larger network, and in the importance of the identification and response to a virus [9], [17]. The studies of network security have recently stressed the ecosystem of security.

Blaster is an example of a virus with a pseudo-immune response. The only viable "immune response" for infected users was to download the software that removed the virus and install that code from windowsupdate.com. However, by blocking windowsupdate.com, Blaster prevented that response. After some time, Microsoft moved the server. This addressed the issue, as it was the IP address and not the domain name that was blocked.

> The mental models approach could significantly improve risk communication for computer security.

The public health metaphor communicates important elements of networked security. First, it illustrates that the person most harmed may not be the one initially infected. Public health also communicates to each person that they are likely to be targets. The individual risk model and the need for individual hygiene communicate individual responsibility as well as individual risk. No person believes they are not at risk for illness, unlike crime or warfare where there must be specific targeting.

The policy implications of the public health model are also worthy of consideration. A significant implication of public health is that there is a need for coordinated public response. However, the need for coordinated public response

does not override the rights of individuals to make choices about their own health, even if that implies some risk to others.

In this case the threats are the viability and reliability of the overall system. The threat is a generic one, of malicious code. Threat migration requires responsible action for all users, just as good health requires that we all wash our hands. The requirement for responsible user action creates few new risks. One risk is that users will see a fake email "from Microsoft" containing a "patch." However, this is not a new risk and end users who regularly patch their machines will realize that no Microsoft patch is ever sent over email.

The public health model best communicates the risk and ability for end users to protect themselves, not through heroic action but rather through consistent hygiene. First, the public health model correctly implies that everyone is at risk. Second, the public health model conveys the importance and continued autonomy in the face of risk. Third, the public health model communicates shared responsibility for community health.

> A mental models approach addresses risk perception behaviors, and recognizes that a strictly rational approach is, ironically, irrational.

Criminal Model

As suggested in the summary above, a critical element of the physical mental model is that of crime. Given that computer crime certainly exists, the criminal model is worthy of consideration. The prosecutor in the Melissa virus case argued that this is "simply crime," arguing further that, "Law enforcement can employ technology, too, and track down virus writers and hackers through the electronic fingerprints they invariably leave behind" [5]. However, as of January of 2004 Norton antivirus offered 30 084 discrete virus descriptions and Symantec has recorded 4397 additional descriptions on their public web sites (not double-counting links to each other's sites, but counting variants. This number includes viral code or worms for which either has online descriptions including date of release and payload.) There have been relatively few prosecutions of authors of malicious code. Some of these are described in this section, illustrating that there is little correlation between the severity of the violation and resulting prosecution. Severity may be measured in inherent damage by the payload or by the extent of the distribution; in any case there is no consistency in the 1.74×10^{-4} prosecution rate.

There have been remarkable successes. Robert Morris was identified almost immediately as the creator of the first Internet worm, and was sentenced to three years probation. Morris is now on the faculty at M.I.T., so the prosecutorial message appears mixed. The prosecution took into account the stated (if not widely believed) intent of Morris as research and the role of Morris in helping to resolve the situation once it had escalated. Nonetheless, the model of hacker-made-professor remains. The worm also led to the passage of the Computer Fraud and Abuse Act in the United States, a tool that was not available to the prosecution in Morris' case.

Christopher Pile, the initiator of the SMEG family of viruses in 1995 received less generous treatment, arguably because there was no feasible possibility that he was engaged in research. Pile was the first person sentenced under the Computer Misuse Act in the United Kingdom and received an 18-month prison term.

Chernobyl has one of the most destructive payloads yet recorded. Chernobyl destroys data by beginning at the disk sector zero and writing randomly generated data until the computer crashes. Before overwriting the data, Chernobyl corrupts the machines' BIOS. The Taiwanese author of that virus, Chen Ing-hau, escaped prosecution because there had been no Taiwanese complaints against him and his actions against foreign nationals were not illegal. The virus was most destructive in South Korea and Hong Kong.

In contrast the author of the relatively benign Melissa virus was sentenced to 20 months in prison, committed to community service, and fined. Melissa spread rapidly and randomly entered an obscure Simpsons' joke in MS Word documents.

Jan De Wit authored the Anna Kournikova virus, and turned himself in. (The worm was named after the image file to which it was attached). He received 150 hours of community service. The worm's payload was self-reproduction via MS Outlook.

Onel de Guzman of the Philippines authored the "I love you" virus. The so-called Love Bug altered image files and template files, thus severely damaging the document on Windows-based web servers. It altered MP3 files. It added itself to all visual basic scripts. However, authoring such a virus was not a crime in the Philippines and thus the author remains free.

Four Israeli teenagers were arrested for developing The Goner in 2001. Because of their age the final judgments against them are not public. The SoBig virus is believed to have originated in China. Similarly other malicious rapid code could have been developed by any of a millions of people.

These cases are only a few of the recorded prosecutions for the release of a malicious code. Even these few relatively successful prosecutions bring forward the problems of jurisdiction, proof, and legal expertise that are required for successful criminal prosecution.

The prosecution of creators of malicious code has not proven successful. Yet this does not imply that the mental model of virus as crime cannot be used to better inform the population. In the last decade, malicious code has been recognized as criminal and malicious behavior. There has been increased discussion of when malicious code can be socially acceptable (so-called hacktivism) with an understanding that it is criminal by default. This remarkable cultural change has been primarily in the computing population, and is reflected in computer science and engineering schools, which have evolved from treating malicious hacks as technical feats to addressing such action as ethical failure. Consider the change from Verton in 2002 [33] to 2004 [26].

If the sole problem is crime then the narrow response is aggressive law enforcement, even in universities. Finally, when there is a crime the victim suffers a significant loss. Because of this, some organizations are hesitant to report computer crime. However, the Computer Emergency Response Team/Coordinating Center (CERT/CC) offers anonymity and help for organizations that have suffered digital networked assault.

If computer crime is the problem, then increased surveillance is an obvious part of any solution. Yet the interaction between privacy and security on the network is subtle and decreases in privacy often decrease security as well. Data surveillance requires, perversely, the assured availability of that which one would protect [23]. Therefore, while malicious code on the network is criminal, the criminal mental model may not be the most useful. Yet enforcement against phishers, identity thieves, and network intruders is undeniably important [25].

> The public health metaphor communicates important elements of networked security.

The computer crime model indicates that the end user is responsible for taking particular actions in order to avoid making him or herself a specific target of opportunity. In crime there are experts (e.g., police) but each individual citizen has responsibilities to avoid risk. The ability of experts to protect end user in this model is limited, and requires cooperation of the experts. By linking computer security to larger crimes, the user may better experience himself or herself as potential victim. When computer crime is the issue, then the end user has to experience him or her self as a potentially vulnerable target.

The crime model is tightly linked to the physical model. In the case that the crimes are modeled as intrusions, as noted above, individual incentives may be effectively communicated. Yet the criminal model may call for actions for which end users have no tools. There is no mechanism to enable neighborhood watch—I cannot as easily watch my neighbor's Microsoft Windows as his Anderson windows. While computer crime is undeniably a reality, the use of crime as a mental model for computer security is problematic, not only because it is itself based in the physical model.

Warfare Model

That the warfare metaphor has been internalized is clear from the choice of terms used in network security, as well as underlying design concepts. Firewalls are perimeter defense technologies. Similarly, intrusion detection is based on defense of a trusted interior and a trusted exterior. DMZs are another metaphor that embeds the concept of computer security as warfare [8], [9], [34].

The slammer virus represents both the potential and the problems with the warfare metaphor. Most individuals had no knowledge that they were at risk. Education was made more difficult as the initial response was, "I don't have a database—that is something for web servers."

Slammer illustrates the need for coordinated action, which is clearly an element of war. The spread of Slammer could be easily detected at the network layer and thus could be completely stopped. (It sent a distinctive 376-byte UDP packet to port 1434.)

In the case of Slammer, law enforcement failed. In market terms, there was little economic incentive for home users to be concerned about their own machines and less information about the need to do so.

The SoBig argues for malicious code as terror as well as theft. SoBig takes over systems to be used to send spam and assault other systems. 9/11 illustrated that the most effective attack against an advanced system is hijacking the system to leverage its destructive power. Network security must address the reality that networks can be hijacked and used against their owners and the network community at large.

Warfare as a metaphor is applicable in the critical need for speed in the response, the potential catastrophic results of a loss, and the focus on the control of resources.

In communicating the need for citizens to be alert, and the importance of individual action for collective security the warfare metaphor is powerful. However, warfare implies a temporary state and an identified state actor as initiator of an assault.

The warfare model rightly indicates that perimeter security and constant diligence are necessary.

The warfare metaphor offers valuable insights. It is possible to construct a perimeter on a network. Once the perimeter is secure then this does not end the potential for loss. There must be layers of security, intelligence gathering on one's own network, and there should be an eternal state of readiness. Firewalls cannot be shut down while the network is live.

In this case the assets to be protected included the integrity of information on the computers as well as the availability of the network. Risk responses could include detection and prevention of certain types of content from entering the network. Slammer could have been stopped at the edge of any network.

The warfare metaphor rightly indicates that perimeter security and constant diligence are necessary. The warfare model communicates the existence of a determined implacable enemy. The warfare model has the potential to leverage horror to the extent that the horrors of war, rather than abstractions of heroism or glory, are well understood. The efficacy of the terror element of the warfare model may be bounded by limits of imagination in the U.S., where there has been no active military conflict on the nation's soil since 1865. The warfare model risks the continued user behavior of complete dependence on experts by removing end user responsibility. Indeed, the warfare model in terms of national security and the model of security in computer security are ideologically opposed. The national security model advocates top down control of information and resources. The network security model advocates and even requires a free flow of information. These differences are most clearly spelled out in [22].

Market Model

Security and network vulnerabilities can be seen as an economic failure. Vulnerabilities, in particular, can be seen as a market failure, an externality [7], [28]. Computer security failures cause downtime and costs.

Three common ways in which security failures from one system harm another, better secured, system are through shared trust, increased resources, and the ability for the attacker to confuse the trail [7]. Shared trust is a problem when a system is trusted by another, so the subversion of one machine allows the subversion of another (for example, when passwords for one machine are kept on another).

The second issue, increased resources, refers to the fact that attackers can increase resources for attacks by subverting multiple machines. This is most obviously useful in brute force attacks, for example in decryption or in a denial of service attack. Using multiple machines makes a denial of service attack easier to implement, since such attacks may depend on overwhelming the target machine.

Third, subverting multiple machines makes it difficult to trace an attack from its source. When taking a circuitous route an attacker can hide his or her tracks in the adulterated log files of multiple machines. Clearly this allows the attacker to remain hidden from law enforcement and continue to launch attacks. The last two points suggest that costs to hackers fall with the number of machines (and so the difference between the benefits of hacking and the costs increases), similar to the way in which benefits to phone users increase with the number of other phones on the network. Both the decreased cost and the difficulty of tracking are costs that are not borne by the owner of the passive, otherwise unharmed, but subverted machine.

A fourth point is the indirect effect security breaches have on users' willingness to transact over the network. For instance, consumers may be less willing to use the Internet for e-commerce if they hear of incidents of credit card theft. This is a rational response if there is no way for consumers to distinguish security levels of different sites.

Because security is an externality, the pricing of software and hardware does not reflect the possibility of and the extent of the damages from security failures associated with the item. Pricing has, therefore, not proven to be an effective mechanism for addressing failures of end users to secure their own machines [5].

Externalities and public goods are often discussed together—these are both forms of market failures. Yet despite the similarities, the range of mechanisms to deal with an externality is, traditionally, broader than the range of options available to address a public good. A common example of a public good is national security, and it might be tempting to think of the analogies between national security and computer security. National security, and public goods in general, are generally single, indivisible goods. A pure public good is something that is both non-rival (in that my use of it doesn't affect yours) and non-excludable (once the good is produced, it is hard to exclude people from using it). The aggregate security of the network, computer security, by comparison, is the sum of a number of individual firms' or peoples' decisions. It is important to distinguish computer

security from national security (i.e., externalities from public goods) because the solutions to the public goods problem and to externalities differ. The government necessarily directly addresses the production of public goods, whereas there are a number of examples where simple interventions by the government have created a more efficient private market such that trades between private economic parties better reflect the presence of externalities.

SoBig is an exemplar of security as an externality. SoBig was motivated by the ability to subvert the computers of naive end users in order to implement fraud through phishing and spam. The creator of SoBig has not been detected by law enforcement. In fact, the lack of consideration of agency in computer crime laws creates criminal liability for those with computers subverted by SoBig as they are, in fact, spamming, phishing or implementing DoS attacks from their own home machines. SoBig is unique in that it was the first widespread attack that was clearly identified as having an economic incentive. At the First Workshop on the Economics of Computer Security, such an attack was identified as a theoretical possibility the year before it occurred.

The model of computer attacks as infection does not apply because the large financial motivation for creation of a computer virus is not addressed. Biological viruses, except in extreme use in terror or warfare, are not driven by intelligence or by the search for profit. Biological viruses and medical failures share evolution with economically motivated and coordinate actors. Yet while biological viruses and infections may be called opportunistic, they do not search and optimize in the manner of coordinated and funded computer attackers.

The model of computer crime as warfare fails in the SoBig example because the computer virus subverts but does not destroy assets. In this case the assets are the availability of the network. There are providers who assist in preventing denial of service attacks and targeted assaults. For example, when Microsoft came under attack from the MyDoom worm, the company had an agreement with the Linux-based Akamai to provide content in the case of such an attack. The reliance on Akamai and the resulting distribution of content allowed Microsoft to mitigate any effects of this DDoS attack. Of course, this particular attack was simplistic, and based on IP address, making avoidance itself feasible. However, few organizations will face a denial of service attack. Not all of those that do can call on content servers for assistance.

In the case of a DOS attack, the threats are downtime and loss of customers as well as near term loss of transactions.

Threat mitigation includes making internal billing for security prevention such that the externalities are addressed. For example, if every department is charged for time spent patching their computers, then departments will not want to implement the constant stream of patches for Microsoft products. If departments are provided free patching support, and charged based on the vulnerabilities of the network, then the pricing reflects the externality. The new risk for changing this accounting system is that it creates incentives for decision makers to spend too much time looking for vulnerabilities and not enough time on other sources of internal risk. Thus recognizing the role of economics in security, either as mental model or a practical tool, does not mitigate all risks.

The market model can be used to provide accurate information about costs in an organization. For example, there may be credits for security expenditures and costs to not patching. This is the opposite of most organizational rules. In most organizations there is a cost to information technology services providing support for addressing vulnerabilities or bad security practices. The units in a business that do not address security are shifting costs to others. Having security failures priced to reflect true market costs is not a solution to the problem of risk communication. Using security to communicate to users would require users to accept costs to help others if the concept of security as an externality were used. If computer security as economics can indicate to end users that they have generically valuable assets (processing power and connectivity) that any thief might target, as opposed to having them believe that they must be visible targets particularly worthy of assault, then economics can assist naive end users in thinking about their own investments in computer security.

The model of computer crime as warfare fails in the SoBig example because the computer virus subverts but does not destroy assets.

What is the Problem?

Enumeration of the heuristics of risk perception indicates that not only is the goal of educating users about the statistical and technical details of security risks difficult, it is also likely to be unproductive. There is no reason to assume that these systematic biases do not apply in the case of computer security. The use of mental models addresses risk perception behaviors, recognizing that a strictly rational approach is, ironically, irrational.

> The units in a business that do not address security are shifting costs to others.

The different examples and metaphors currently used as inchoate mental models all indicate different responses by the user. Crime suggests investigation of every virus and worm by central authority, with the possibility of a neighborhood watch. The criminal metaphor requires tracking and prosecution. Physical security indicates training in lock-down, and perhaps the integration of security and tactile devices. The concept of warfare requires tight constraints on the network, with limited connectivity. The warfare metaphor further mitigates individual user responsibility. Economic metaphors suggest pricing and the increased use of contracts as well as civil torts for irresponsible individuals who harbor zombie machines. The public health or illness metaphor implies coordinated public action with an appreciation of an individual's right to refuse treatment with virtual or physical patching. The medical model shows particular potential, as it combines horror with appeals to personal responsibility, and indicates that everyone should be concerned with the risk.

Each metaphor offers a different solution to a different facet of the security problem. Each model communicates to end users particular images and activities if properly used.

Mental models are not perfect as tools, nor are they perfectly understood in terms of user response. Of course, nothing is a panacea. Using mental models in communications may create perverse implications, particularly in the atheoretical manner that is current practice. End users may take all the implications of the metaphors. Therefore when communicating with policy makers, media, and non-technical users the computer security expert should consider which metaphor correctly communicates user expectations. Mental models should more widely used, and even more widely researched, in communicating computer security risks.

References

[1] R. Anderson, "Unsettling parallels between security and the environment," in *Proc. Inaugural Workshop on Economics and Information Security*, Berkeley, CA, 2002.

[2] R. Anderson, *Security Engineering: A Guide to Building Dependable Distributed Systems*. New York, NY: Wiley, 2001.

[3] M. Bailey, E. Cooke, F. Jahanian, D. Watson, and J. Nazario, "The blaster worm: Then and now." *IEEE Security and Privacy*, vol. 3, no. 4, pp. 26-31, Jul. 2005; http://dx.doi.org/10.1109/MSP.2005.106.

[3] M. Bishop, *Computer Security: Art and Science*. Pearson, 2003.

[4] B. Blakley, E. McDermott, and D. Geer, "Information security is information risk management," in *Proc. 2001 Workshop on New Security Paradigms*. New York, NY: ACM, 2001, pp. 97-104.

[5] L.J. Camp and C. Wolfram, "Pricing security," in *Proc. CERT Information Survivability Workshop* (Boston, MA), Oct. 24-26, 2000, pp. 31-39.

[6] L.J. Camp, *Trust and Risk in Internet Commerce*. Cambridge, MA: M.I.T. Press, 2000.

[7] L.J. Camp and S. Lewis, Eds., *The Economics of Information Security*. Boston, MA: Kluwer, 2004.

[8] S. Cass, "Listening in," *IEEE Spectrum (Special Report on Intelligence and Technology)*, vol. 40, no. 4, pp. 32-37, 1995.

[9] D. Denning, "Cyberplagues," *in Information Warfare and Security*. Boston, MA: Addison-Wesley, 1998, ch 10, pp. 269-282.

[10] D. Denning, *Cryptography and Data Security*. Boston, MA: Addison-Wesley, 1982.

[11] P. Dourish, R. Grinter, J. Delgado de la Flor, and M. Joseph, "Security in the wild," *Personal and Ubiquitous Computing*, vol. 8, no. 6, pp. 391-401, 2004.

[12] S. Freeman and J. Bleifuss, *Was the 2004 Presidential Election Stolen?: Exit Polls, Election Fraud, and the Official Count*. St. Paul, MN: Stories Press, 2006.

[13] S.J. Greenwald, K.G. Olthoff, V. Raskin, and W. Ruch, "The user non-acceptance paradigm: INFOSEC's dirty little secret," in *Proc. 2004 Workshop on New Security Paradigms* (Nova Scotia, Canada), Sept. 20-23, 2004.

[14] Staff writer, "Mothers of fallen NW Indiana soldiers mourn as toll passes 1000," *Bloomington Herald-Tribune*, Sept. 8, 2004, pp. A1.

[15] H. Jungermann, H. Schutz, and M. Thuring, "Mental models in risk assessment: Informing people about drugs," *Risk Analysis*, vol. 8, no. 1, pp. 147-155, 1998.

[16] D. Kahneman, P. Slovic, and A. Tversky, *Judgment Under Uncertainty*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[17] R. Kephart and W. Chess, "Computer networks as biological systems," *IEEE Spectrum*, May 1993.

[18] M.G. Morgan, B. Fischhoff, A. Bostrom, and C.J. Atman, *Risk Communication: A Mental Models Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[19] A. Newell and H. Simon, *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972.

[20] Pew Internet & American Life Project, "Fear of Online Crime," 2001; http://www.pewinternet.org/PPF/r/32/report_display.asp, accessed Nov. 2005.

[21] Pew Internet & American Life Project, "America's Online Pursuits," 2003; http://www.pewinternet.org/PPF/r/106/report_display.asp, accessed Nov. 2005.

[22] H. Nissenbaum, "Where computer security meets national security," *Ethics and Information Technology*, vol. 7, no. 2, pp. 61-73, June 2005.

[23] B. Schneier, *Beyond Fear*. New York, NY: Copernicus, 2003.

[24] M. Smith, "Melissa was 'a colossal mistake' says author," *Sophos: Anti Virus for Business*, May 2, 2002; http://www.sophos.com/virusinfo/articles/melissa2.html.

[25] C. Stoll, *Cuckoo's Egg*. New York, NY: Pocket Books, 1990.

[26] N.B. Sukhai, "Hacking and cybercrime," in *Proc. 1st Ann. Conf. Information Security Curriculum Development* (Kennesaw, GA), Oct. 08, 2004. New York, NY: ACM, pp. 128-132.

[27] P. Syverson and A. Shostack, "The paradoxical value of privacy," in *The Economics of Information Security*, J.L. Camp and S. Lewis, Eds. Boston, MA: Kluwer, 2004.

[28] H. Varian, "System reliability and free riding," in *The Economics of Information Security*, J.L. Camp and S. Lewis, Eds. Boston, MA: Kluwer, 2002.

[29] V. Tuesday, "Anatomy of an attack: A race against time," *Computerworld*, vol. 35, no. 12, p. 57, May 19, 2001.

[30] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124-1131, 1974.

[31] A. Tversky and D. Kahneman, "Belief in the law of small numbers," *Psychological Bulletin*, vol. 76, pp. 105-110, 1971.

[32] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, no. 4481, pp. 453-8, 1981.

[33] D. Verton, *The Hacker Diaries: Confessions of Teenage Hackers*. Osborne/McGraw-Hill. 2002.

[34] P. Wallich, "Getting the message," *IEEE Spectrum (Special Report on Intelligence and Technology)*, vol. 40, no. 4, pp. 38-43. 1995.