Quantifying the Strength of Security Systems

David Lie
Department of Electrical and Computer Engineering
University of Toronto

M. Satyanarayanan
School of Computer Science
Carnegie Mellon University

Abstract

Security researchers and practitioners lack techniques to quantitatively evaluate the strength of security systems against a determined attacker. Currently, evaluation is either qualitative, such as through security certification standards, or ad-hoc, such as through penetration testing and auditing. In this paper, we propose a framework that if applied to security systems, would produce quantitative measures that can be used to compare and appraise security systems relative to each other. Our framework utilizes public challenges in conjunction with an independent organization that mounts the challenges, regulates their implementation and certifies the results in an attempt to provide normalized measures. Unlike various ad-hoc challenges that have been run in the past, we believe our framework can create a quantitative, challenge-based security evaluation infrastructure that is fair, sustainable and flexible.

1 Introduction

Security has become of great importance in our computerized world today, motivating a myriad of security products that claim to protect individuals and organizations from computer security threats. However, it is often difficult for lay individual users, and even professional security practitioners alike to evaluate which security products truly enhance the security of their systems. Those who wish to secure their computing systems are reduced to browsing through marketing literature that qualitatively describes the various features, capabilities, design process and philosophy of the system. For example, security certification standards such as TCSEC/Orange Book [3], ITSEC [5] and Common Criteria [1] specify product capabilities and the design process taken to increase the product's level of assurance

Many other disciplines benefit from the ability of practitioners to assign quantitative measures to the quality of their products. In civil engineering, practitioners can make engineering decisions that make systems arbitrarily robust by designing their product with a quantitative safety margin because they have accurate measurements of both the strength of the materials they use, as well as the required load they must bear. Note that it is not just designs but also *implementations* that are subject to quantification and evaluation in these disciplines. One can, for example, stress test a building component to discover flaws in workmanship or installation. This is in addition to analytically evaluating its architecture and design process for flaws. Closer to home, other facets of computing, such as networking, storage and processor architecture have well-developed benchmarking techniques. These enable engineers to measure the quality of their designs and implementations against their competition and relative to their design goals.

We believe that the ability to quantify important security properties will advance security engineering just as it has advanced other disciplines. The most ardent advocate of quantification in science and engineering was Lord Kelvin (1824-1907), who stated in 1883:

> *"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be."*

While Kelvin's viewpoint is more extreme than our own, we do believe that this quote captures the essence of our thinking. Our hope is that this paper will stimulate constructive debate and controversy, leading eventually to action.

2 Difficulties Unique to Security

Security attributes differ in two fundamental ways from the many other attributes that have been quantified in science and engineering. First, *malicious intent* lies at the very heart of security. Without malicious intent there would be no need for security solutions. The depth of knowledge, tenacity of purpose, sophistication of tools and potential for damage of an attacker are all factors that must be taken into account when defining the security strength of a component or system. It is hard to imagine any meaningful characterization of security properties that fails to take these factors into consideration. These inherently human factors contrast with the objective and impersonal factors that are involved in quantifying other attributes. Although nature is sometimes viewed as an adversary in the natural sciences, the adversarial nature of human attacks is on a fundamentally different plane.

Second, *time dependence* is inherent to security properties because of continuing advances in computing technology. A cryptographic technique that is considered safe today may be vulnerable a decade hence. Similarly, attack techniques progress with time and the discovery of a certain technique may facilitate the compromise of systems that were previously secure. For example, one just has to consider the increasing sophistication of viruses, root kits and spam to see that any system that was considered secure several years ago may have questionable security today. In other words, the threat environment is not stagnant but changes (almost always for the worse) over time. This is in contrast to attributes studied in the natural sciences, where the laws of nature are invariant across space and time.

To address these unique characteristics of security, we propose the use of open challenges or contests. The idea of mounting a challenge to establish security is not new, and skepticism has been expressed about this approach. For example, both Schneier [6] and Spafford [7] have criticized contests for being unfair. This is because previous contests, which have been usually run by the security product vendor, have inconsistent and arbitrary rules, making it impossible to compare the challenge used on one product with that of another. They have also questioned the sustainability of contests as a way for evaluating systems since the monetary incentives are usually not large enough to attract skilled attackers. While we concur with their concerns, we believe that they apply to the way challenges are currently being implemented. Fairness, sustainability, along with flexibility are not incompatible with challenges in general. Therefore, we have three requirements for a successful challenge-based security framework:

- *Fairness:* the challenges must be constructed such that the results are comparable across products – in other words the challenge framework must treat all submissions fairly.
- *Sustainability:* the scheme must be economically viable and self-sustaining – all parties involved must believe they stand to make substantial financial gains depending on their abilities to develop and analyze secure systems.
- *Flexibility:* the scheme must flexible enough to accommodate evaluation for a variety of security system designers, from security product vendors to academic security researchers.

We advocate this approach in addition to (not instead of) conventional techniques such as analysis and verification. Systems and components that claim security properties have always benefited from open and public evaluation: the AES competition organized by NIST and the RSA factoring challenge have been some examples of this. Open challenges in other disciplines, such as the X Prize for aerospace [11] and the DARPA Grand Challenge in autonomous vehicles [2] have spurred innovation and motivated organizations to invest millions in beating the challenge. Thus, we believe that given the correct framework, public challenges can be valuable in adding a quantitative dimension to a security evaluation.

We will describe our challenge-based evaluation framework in the next section, and then follow by analyzing how it meets the three criteria for success in Sections 4, 5 and 6. We then discuss some open questions and future directions in Section 7, and draw our conclusions in Section 8.

3 Evaluation Framework

In this section we describe an evaluation framework that we believe meets the criteria set out in Section 2. The goal of our framework is to assign accurate quantitative measures of security that are widely trusted and are comparable across many different systems. One of the difficulties is that malicious intent is difficult to simulate in a controlled environment, motivating the use of public challenges in our framework.

We propose that the entire process of evaluation and certification (including the conducting of challenges) be the responsibility of a widely-known and trusted organization. This may be a reputable company (such as Verisign, Inc.), an independent nonprofit agency (such as Underwriters Laboratories), or a government-mandated entity (such as the FAA, FDA, or FCC).

Our process of certification via public challenges is illustrated in Figure 1. To obtain certification for a system, a party must submit it to the organization and specify three things:

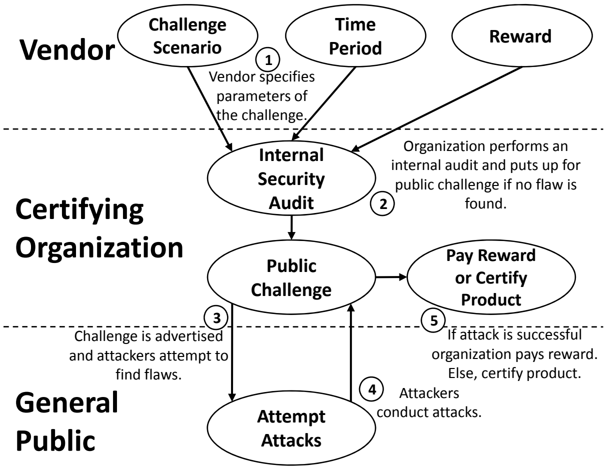
Vendor
Challenge Scenario
Time Period
Reward
1
Vendor specifies parameters of the challenge.
Certifying Organization
Internal Security Audit
2
Organization performs an internal audit and puts up for public challenge if no flaw is found.
Public Challenge
Pay Reward or Certify Product
3
Challenge is advertised and attackers attempt to find flaws.
5
If attack is successful organization pays reward. Else, certify product.
General Public
4
Attackers conduct attacks.
Attempt Attacks

Figure 1: Certification Process. The process of certification is illustrated along with the party taking the action.

- a well-defined *challenge scenario*, which includes a goal for the attacker to achieve and a set of constraints on the attacker (for example, no physical access to the system). The scenario should reflect one that a real attacker might reasonably face should the system be deployed.
- a *time period* for which the challenge will be available.
- a *reward* to be paid to the first attacker who completes the challenge and reports success within the time period.

The organization first performs a thorough internal security audit of the submitted component or system using in-house experts and state-of-the-art verification and validation techniques. If any security flaw is found, the certification process is terminated without a challenge phase. Only those submissions that pass this screening step are presented to the public for challenge. If the challenge time period expires without a successful attack, the organization certifies this fact and includes the three parameters above as part of the certification. The time-dependent nature of the security environment, as discussed in Section 2, is addressed by including an expiry date with the certification, which we typically expect to be a few years into the future. In order to extend certification beyond the expiry date, the component or system must be resubmitted for audit and challenge.

If a system survives the challenge, there is one qualitative metric and two quantitative metrics that can be derived. The qualitative metric is the challenge scenario, which sets out the goals and rules of the challenge. However, for systems that would be direct competitors because they perform the same function, this component of the challenge could likely be made to be very similar if not completely identical. For example, a set of username/password authentication systems could all be subjected to challenges where attackers were allowed to try active attacks on a running system, as well as eavesdrop on successful authentications of users with the same username and password combinations across all systems. It is the role of the certifying organization to try and maintain the value of its certifications by strongly encouraging submissions to specify comparable challenge scenarios whenever possible. Such decisions could be made by an impartial appointed committee, or a council composed of representatives from stakeholders, as is done for SPEC [8] and TPC [9].

When challenge scenarios are similar, customers can make direct quantitative comparisons with the remaining two parameters: the reward size and the challenge time period. It is clear that larger rewards and longer time periods are indicative of stronger security, although the details of this relationship are unknown at this time. Intuitively, a larger reward size will attract a larger number of would-be attackers, while a longer period allows the attackers to try a larger number of attacks and to analyze the system in greater depth.

4 Fairness

This framework is fair (our first criteria) because all challenges are administered by the same organization. The organization must have a high profile, be well-known, and be widely trusted so as to attract the attention of a large number of potential attackers and security experts. The designation of a centralized testing and certification organization is common in many mature industries. For example, Underwriters Laboratories which administers the familiar “UL” symbol seen on many electrical products, certifies 19,000 new products and has its symbol appear on 21 billion products a year [10].

As long as all attackers do not collude, they are motivated to come forward with a successful attack as soon as possible so as to avoid being scooped by other attackers. Naturally, both the reward and challenge duration must be large enough to attract a reasonable number of attackers with adequate skills, and to give them enough time to mount a successful attack. A submission that uses a small reward and a short challenge duration would, at best, receive certification that is implicitly weak. Strong certification is the incentive that motivates submissions to use large rewards and long time periods. Systems that will be used to protect extremely valuable assets would, by their nature, need to have large rewards to have any credibility. It is natural to expect that customers will demand strong certification for components and systems that are used to protect high-value assets.

5 Sustainability

Our second requirement is that the scheme must be self-sustaining. This means that ultimately the rewards must be large to convince professional security analysts (or the companies that employ them) that it is worth their time to continually try to beat these challenges. On the other hand, the organization must find a way to fund the rewards for challenges that are successfully broken. Finally, the certifications handed out must be reputable and valuable so as to convince vendors of future security products to approach the organization with submissions of their own.

To fund these contests, we propose that each vendor that desires certification must pay a fee that is proportional (though not larger or equal) to the size of the reward they select for the challenge. This forces the vendor to share a portion of the risk in posting a large reward. On the other hand, because the certification fee is smaller than the reward for the challenge, vendors will be highly motivated to use this leverage to help certify their products.

The organization collects the full certification fee even if the challenge phase is skipped due to flaws revealed by the internal security audit phase. In each such case, the organization is able to collect its fee while assuming no risk of losing the reward money. This results in profit for the organization, and provides a strong incentive for it to attract and retain in-house security expertise to conduct rigorous audits. The certification fee is bounded on the lower end by the cost of the security audit the organization must perform, but this is a cost that any reputable security vendor would already have to pay today.

If the organization is diligent in its job, it should be able to build up a fairly large "float" of money that it can use to pay rewards for successful attackers. This float is funded by instances where the organization is able to find a weakness without having to make the challenge public, and instances where the challenge is made public, but no successful attack is found. This model is very similar to that of insurance companies: they do their own analysis of each customer to determine whether the risk of insuring them is acceptable. If their analysis is done well, then statistically the number of customers on which they make a profit will outweigh the number of customers on which they lose money. This enables large rewards (possibly many hundreds of thousands of dollars), making it feasible to attract very skilled security analysts. If potential attackers are presented with a substantial reward – perhaps several times a typical yearly salary, they will be motivated to devote more of their time to attacking systems set out in challenges.

Because the fee paid by the vendor is smaller than the reward, it is very important that the organization have a way of identifying vendors who propose systems. Otherwise, a colluding vendor and attacker could take advantage of the organization. For example, Alice, a small security product vendor could place a hidden trap-door in her system that eludes the scrutiny of the organization. She then tells Bob, an attacker, about the trap-door, allowing Bob to claim the reward once the challenge is made public. Alice and Bob both split the profit since the reward is larger than the fee. By identifying vendors who propose challenges, the organization is able to attach a reputation to the vendor, only allowing them to propose challenges for larger amounts once they have shown that they can consistently produce strong systems that are not broken. Naturally, as in the insurance industry, the potential for fraud exists. However, this should not preclude a vigilant organization from sustaining a healthy business.

6 Flexibility

The final criterion for our evaluation framework is that it must flexible enough to be applicable to a broad variety of stakeholders. For example, it must be able to accommodate a variety of security systems, as well as accommodate systems that are designed to protect extremely valuable assets as well as those designed with economy of cost or ease of use in mind. In addition, some parties submitting systems for evaluation may not have financial profit as a goal. For example, open source developers and academics may also wish to have their work benefit from certification.

The evaluation framework provides two avenues of flexibility to encompass both a broader variety of systems and a broader range of users. First, the challenge scenario can be varied to fit the requirements of a wide variety of systems. As mentioned before, the only requirement is that systems claiming to fulfill similar purposes should be evaluated with under a similar scenarios. Second, the requirement that the fee be proportional to the reward size may prevent challengers with limited financial means such as universities or open source developers from participating. However, the ability to choose both the size of the reward and the challenge duration allows vendors with limited financial means to participate in the certification process by selecting a long challenge period. Software that survives long challenge periods is recognized as highly secure. For example, open source university software such as $\TeX$ and Qmail [4] are widely regarded as being of high quality and secure mainly because of the long period of time that they have survived with no bugs or security flaws being found. The main disadvantage with reducing the reward and lengthening the challenge period is that it delays the time for a product to come to market. This is, of course, less of a concern for those not seeking to make a profit.

There is also an opportunity to leverage educational efforts in the certification process. Security courses offered worldwide could use current challenges as the basis of class projects. This would add a large pool of smart and knowledgeable attackers working under expert guidance to the challenge framework. It would also give students experience in playing the roles of attackers, thus giving them a valuable perspective for broadening and deepening the knowledge that they gain from security courses.

7 Open Questions and Future Work

While we believe such a framework is feasible, there are several open questions that need to be answered should its implementation be attempted. In general, we find that these problems require some technical innovations, but also will likely require collaboration with other fields.

How do we bootstrap the framework? Initial funds are required to create a pool of funds, out of which rewards could be paid. In addition, the organization running the challenges also needs to develop and acquire expertise in eliminating insecure proposals before mounting them as a public challenge. During this initial stage, it is likely that insecure systems may successfully pass the organization's internal review resulting in a flaw found by the public and a reward being paid out. It is also possible the larger than usual rewards must initially be used to attract attention and establish the organization's reputation. Thus, it is likely that some initial funds must be sunk into the organization without any hope of a near-term return. We believe that an organization that is at least partially, if not entirely publicly funded would be the most likely initial embodiment of such an organization. Answering this question will likely require collaboration with experts in public policy.

What is the relationship between reward size and challenge period? How long should the challenge period and reward be to attract a significant number of attackers? It is likely in both cases that the relationships are not linear. For example, it seems intuitively false that a challenge that survives 10 months is 10 times as secure as one that only lasts 1 month. Similarly, it seems unlikely that a reward that is $1 million dollars is only going to attract 10 times as many attackers as one that is $100,000. We believe that game theory holds the greatest promise for answering these questions. For example, game theory has been successfully applied to model many problems in economics and open markets.

How do we detect and avoid fraud in the framework? The organization must be able to identify collusion between attackers and challengers to avoid being defrauded of its money. This not only requires ways of identifying parties across several challenges, but also in detecting patterns of behavior. We believe that data-mining and behavioral pattern recognition techniques may hold the answer to this question.

8 Conclusions

As Lord Kelvin has stated and history has shown, quantitative metrics are an invaluable tool for improving technology. Security is an aspect of our computing infrastructure that is of critical importance, yet lacks comprehensive techniques for quantitative analysis. In this paper, we have argued for a framework that involves public challenges mounted by a well-recognized certifying organization. We believe that our proposed framework can be successful because it has mechanisms that support fairness, sustainability and flexibility in the long run.

Acknowledgements

We would especially like to thank Ashvin Goel and the rest of the Security Reading Group at the University of Toronto for their suggestions and insightful discussions. We would also thank the anonymous reviewers for their helpful comments. David Lie is supported by an NSERC Discovery Grant. Satya is supported by the National Science Foundation (NSF) under grant number CNS-0509004. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSERC, the NSF, Carnegie Mellon University, or the University of Toronto.

References

[1] The common criteria evaluation and validation scheme. `http://www.niap-ccevs.org/cc-scheme/cc_docs/` Last Accessed: 07/26/2007.

[2] The DARPA grand challenge competition. `http://www.darpa.mil/grandchallenge` Last Accessed: 07/26/2007.

[3] DoD 5200.28-STD. Trusted computer system evaluation criteria, Dec. 1985.

[4] qmail. `http://cr.yp.to/qmail.html` Last Accessed: 07/26/2007.

[5] K. Rihaczek. The harmonized ITSEC evaluation criteria. *Computer Security*, 10(2):101–110, 1991.

[6] B. Schneier. The fallacy of cracking contests. `http://www.schneier.com/crypto-gram-9812.html` Last Accessed: 07/26/2007.

[7] E. Spafford. Hacker challenges: Boon or bane. http://www.ieee-security.org/Cipher/PastIssues/1996/issue9602/issue9602.txt Last Accessed: 07/26/2007.

[8] SPEC – standard performance evaluation corporation. http://www.spec.org/ Last Accessed: 07/26/2007.

[9] Transaction processing performance council. http://www.tpc.org Last Accessed: 07/26/2007.

[10] Underwriters Laboratories. http://www.ul.com/ Last Accessed: 07/26/2007.

[11] The X prize foundation. http://www.xprize.org Last Accessed: 07/26/2007.