Editors: Mohamed Kaâniche, mohamed.kaaniche@laas.fr | Aad van Moorsel, aad.vanmoorsel@ncl.ac.uk

Quantitative Security Metrics: Unattainable Holy Grail or a Vital Breakthrough within Our Reach?

William H. Sanders | University of Illinois at Urbana-Champaign

If systems could be made perfectly secure, security metrics would be simple: a system would be either secure or not. Unfortunately, for most systems, perfect security is elusive (and likely unattainable) because of their scale and complexity as well as real-world cost considerations. Such systems include almost all critical infrastructure systems, including those instrumental to power, transportation, aerospace, telecommunications, healthcare, emergency response, banking, and e-commerce systems.

The unattainability of perfect security implies that we should endeavor to build systems that can tolerate partially successful intrusions while still providing the service expected of them, possibly through techniques that provide resiliency. With that approach, security can no longer be thought of as a binary property. Thus, measuring a system's resilience to attacks is important, letting us make rational design decisions to increase system security.

It's long been well understood that you can calculate useful estimations of systems' reliability against accidental failure. It's also well understood that trying to calculate systems' level of security against possibly intelligent, determined, well-funded, and creative adversaries is a far greater challenge. Nevertheless, even a less-than-perfect predictive capacity, if its limitations are respected, is clearly better than none at all. Without promising perfection, such a capacity would offer crucial support to decision making that impacts system security.

Everybody's Talking about Them

Realization of the need for security metrics isn't new. An important goal in almost every research road map for security is the ability to quantify a system's degree of security. For example, in 2003, the Computing Research Association (CRA) named information-systems risk management as one of four grand challenges in trustworthy computing.[1] In 2005, the InfoSec Research Council named enterprise-level security metrics a hard problem,[2] and the President's Information Technology Advisory Committee placed "metrics, benchmarks, and best practices" ninth on its list of cybersecurity research priorities.[3]

Likewise, there's no shortage of security metrics. Broadly speaking, current metrics fall into the following three categories.

First, *organizational security metrics* are those used to describe and track how effectively organizational programs and processes achieve cybersecurity. They can be used to help plan investment in IT architectures or technologies as well as aid in the creation, sustainment, and termination of security programs and program elements.

Organizational metrics are either program metrics or process metrics. An example program metric is the *Guide for Assessing the Security Controls in Federal Information Systems and Organizations*;[4] an example process metric is the Systems Security Engineering Capability Maturity Model.[5] Both approaches specify maturity levels and define areas requiring assessment.

Second, *technical security metrics* indicate the level of security a specific system exhibits. Most famously, the Common Methodology for Information Technology Security Evaluation (CEM)[6] defines the evaluation process for a set of evaluation assurance levels,

1540-7993/14/$31.00 © 2014 IEEE Copublished by the IEEE Computer and Reliability Societies March/April 2014

each of which has a set of assurance requirements. Also in widespread use is the Common Vulnerabilities and Exposures list (http://cve.mitre.org), which can be used to score a system.

Finally, *operational security metrics* are used to describe, and hence manage, the risks to an operational environment. They include

- measures of operational readiness or security postures (for example, how well a system can be expected to perform, given an assumed threat environment),
- measures used in risk management (for example, security compliance),
- metrics describing threat environments, and
- metrics supporting incident response and vulnerability management.

Operational metrics could also include metrics produced as part of normal operations that can serve as input to other security metrics. The estimation of operational metrics generally requires experimental or empirical measurements. An example operational metric is SCORE (Security Consensus Operational Readiness Evaluation; www.sans.org/score).

Broadly speaking, most metrics approaches either focus on specifying procedures to follow during system design or are backward-looking, in that they document known vulnerabilities instead of predicting future behavior. Mathematical approaches typically have been based either on formal methods, aiming to prove that certain security properties hold given a specified set of assumptions, or on statistical methods applied to specific system components (for example, intrusion detection systems). An example of informal approaches is red teaming, in which a team of experts tries to compromise a system.

However, all those approaches have limitations. Process guidelines can improve security but can't quantify the security that has been obtained. Formal methods aim either to prove absolute security (which usually isn't possible) or to find problems (which is useful but offers no quantification). Red teams can find problems but again provide no predictive quantification of security. Existing measurement-based metrics are lagging indicators of observed security and hence aren't predictive.

The Security Metrics Challenge

The ultimate challenge, then, is to create a scientific foundation, methods, and tools for quantitative predictive assessment of security metrics that are applicable to large-scale systems throughout their life cycle. Such a capability would let organizations answer these questions (from presentations of the CRA's four grand challenges):

- How much risk am I carrying?
- Am I better off now than I was this time last year?
- Am I spending the right amount of money on the right things?
- How do I compare to my peers?
- What are my risk transfer options?

Furthermore, a scientifically founded security metrics methodology would let design engineers answer these questions:

- Is design A or B more secure?
- Have I made the appropriate design-time tradeoffs among timeliness, security, and cost?
- How will the system, as implemented, respond to a specific attack scenario?
- What's the most critical part of the system to test, from a security viewpoint?

Such a methodology would also let us understand how technical security metrics impact organizational security metrics.

Although the answers to the previous questions have their basis in quantitative security metrics, the goal is to obtain the insights needed to answer the questions, not to obtain the absolute numeric values of the metrics themselves. That is an important point to remember; numbers in themselves aren't insights but can form the basis of insights if interpreted correctly.

The Path Forward

Broadly speaking, creating the capability to answer questions such as those I just discussed will require progress in four areas.

Appropriate Security Metrics

We must find ways to integrate the multiple metrics I've outlined, and others not yet defined, such that they can help answer those questions. Attaining that goal will require an understanding of the relationships among security metrics and their overall relationship to different stakeholders' questions. We should design and integrate the metrics such that we can use them to gain insights throughout the system life cycle, including design, implementation, configuration, operation, and upgrade or modification.

Methods for Estimating Metrics

Methods must be developed to quantify whatever set of security metrics is chosen for a system. Although much work toward that goal has been done, much more remains. Clearly, we need multiple approaches, including formal methods, probabilistic methods, benchmarking and experimentation, classic risk assessment, threat and vulnerability assessment, whiteboarding and red teaming, and informal and semiformal methods. Significant thought must be given to the nature (for example, mean, variance, and percentile) and accuracy of

the metric values needed to gain the desired insights.

A Comprehensive Security Argument Methodology

We need an overall security argument that can relate business and technical security metrics to one another and provide a convincing, overarching predictive assessment of system-level, end-to-end security. Toward that end, we need methods that

- promote understanding of how to combine seemingly disparate types of evidence into a convincing overall argument,
- define a calculus for decomposing requirements into subrequirements that can be validated independently, and
- specify the relationship among different pieces of evidence gathered during assessment.

Effective Security Metric Evaluation Tools

Ultimately, for the developed metrics and approaches to be useful, we need ways to put them into practitioners' hands. This challenge might seem mundane compared to the ones I just discussed. However, in reality, if security metric evaluation tools are to be widely used, as much thought must go into their usability as into the analysis algorithms they implement.

The goal of developing a scientific foundation, methods, and tools for quantitative assessment of security metrics might seem daunting. However, the challenge is tempered by the realization that the purpose of metrics is almost always to gain insight and make decisions, rather than to quantify absolutely a particular numeric metric. This suggests that we develop methods that can produce relative metrics, which might be easier to create than ones that quantify security absolutely.

The challenge should also be tempered by the realization that progress in the directions outlined here, even without a complete and final solution, would represent a useful step forward for practitioners. Today, security design decisions in industry are often either made with the aid of a trusted analyst who examines the situation and gives advice based on personal experience, or made collectively, on the basis of informal discussions among diverse stakeholders. Although both approaches can result in good design or configuration decisions that increase system security, there's no way to audit the decision process and no quantifiable way to rank alternative decisions. Progress in the four areas I described earlier in this article will provide

- a way for diverse stakeholders to express security objectives and concerns in a common terminology,
- a quantifiable ranking of alternative security policies and architectures, and
- an auditable decision process.

Indeed, this has been an active research area over the last decade, as evidenced by high interest in topical conferences such as the Workshop on Security Metrics (MetriCon), the International Workshop on Security Measurements and Metrics (MetriSec), and the International Workshop on Quantitative Aspects in Security Assurance. Researchers have made important foundational contributions, and there's strong reason to believe that greater success will follow.

In short, although the holy grail of predictive, absolute, and quantitative metrics might not be attainable soon, we mustn't let the perfect become the enemy of the good. To twist a saying of George E.P. Box, "Essentially, all security metrics are wrong, but some are useful." Those that are useful will help us gain insights and thus make appropriate design, configuration, and operation decisions to increase system security. ■

References

1. *Four Grand Challenges in Trustworthy Computing*, Computing Research Assoc., 2006; http://archive.cra.org/reports/trustworthy.computing.pdf.
2. *Hard Problem List*, InfoSec Research Council, Nov. 2005; www.infosec-research.org/docs_public/20051130-IRC-HPL-FINAL.pdf.
3. President's Information Technology Advisory Committee, *Cyber Security: A Crisis of Prioritization*, US Nat'l Coordination Office for Information Technology Research and Development, Feb. 2005; www.nitrd.gov/pitac/reports/20050301_cybersecurity/cybersecurity.pdf.
4. *Guide for Assessing the Security Controls in Federal Information Systems and Organizations*, NIST Special Publication 800-53A, rev. 1, US Nat'l Inst. of Standards and Technology, June 2010; http://csrc.nist.gov/publications/nistpubs/800-53A-rev1/sp800-53A-rev1-final.pdf.
5. *ISO/IEC 21827:2008: Information Technology—Security Techniques—Systems Security Engineering—Capability Maturity Model (SSE-CMM)*, Int'l Org. for Standardization and Int'l Electrotechnical Commission, 16 Oct. 2008.
6. *Common Methodology for Information Technology Security Evaluation*, ver. 2.3, CCMB-2005-08-004, Common Criteria, Aug. 2005; www.commoncriteriaportal.org/files/ccfiles/cemv2.3.pdf.

William H. Sanders is a Donald Biggar Willett Professor of Engineering, the interim head of the Department of Electrical and Computer Engineering, and the director of the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. Contact him at whs@illinois.edu.