SECURITY METRICS: MEASUREMENTS TO SUPPORT THE CONTINUED DEVELOPMENT OF INFORMATION SECURITY TECHNOLOGY

Shirley Radack, Editor
Computer Security Division
Information Technology Laboratory
National Institute of Standards and Technology

More than 100 years ago, Lord Kelvin (William Thomson, lst Baron Kelvin), the distinguished British mathematical physicist and engineer, observed that measurement is vital to knowledge and to continued progress in physical science. Lord Kelvin stated that: "To measure is to know," and "If you can not measure it, you can not improve it."

These observations on measurements are relevant to our use of information technology (IT). Organizations rely on IT to carry out their daily operations and to deliver products and services to the public. Managers are challenged to use IT effectively and to protect their systems and information from security threats and risks. There have been many past efforts to develop security measurements that could help organizations make informed decisions about the design of systems, the selection of controls, and the efficiency of security operations. But the development of standardized measurements for IT has been a difficult challenge, and past efforts have been only partly successful.

Security metrics are needed to provide a quantitative and objective basis for security operations. Metrics support decision making, quality assurance of software, and the reliable maintenance of security operations. To address this need for more precise measurement of security technology, the Information Technology Laboratory of the National Institute of Standards and Technology (NIST) recently published a report that examines past efforts to develop security metrics and points to possible areas of future research that could lead to improved metrics.

National Institute of Standards and Technology Interagency Report (NISTIR) 7564, *Directions in Security Metrics Research*

Written by Wayne Jansen of NIST, *Directions in Security Metrics Research* provides background information on the various meanings and interpretations that have been applied to the term "security metrics." The report examines critical aspects of security measurement as identified by past efforts and highlights the factors that are relevant to security metrics research. It then focuses on research efforts that are needed to advance the development of effective security metrics. An extensive reference list includes books, papers, and publications on security metrics.

NISTIR 7564, which is summarized in this bulletin, is available at the NIST Web page http://csrc.nist.gov/publications/PubsNISTIRs.html.

What are Security Metrics

In general, a metric implies a system of measurement that is based on quantifiable measures. A method of measurement used to determine the unit of a quantity could be a measuring instrument, a reference material, or a measuring system. The measurement of an information system for security involves the application of a method of measurement to one or more parts of the system that have an assessable security property in order to obtain a measured value. The goal is to enable an organization to evaluate how well it is meeting its security objectives.

The method of measurement that is employed should be reproducible, and should achieve the same result when performed independently by different competent evaluators. Also, the result should be repeatable, so that a second assessment by the original team of evaluators produces the same result. All results of measurements should be timely and relevant to the organization.

Many of the traditional concepts in metrology that are used in the physical sciences, such as the use of fundamental units, scales, and uncertainty, either have not been applied to IT or have been applied less rigorously than in the physical sciences. Available quantitative metrics for IT system security generally reflect an evaluator's reasoned estimates of security. These measures of information system security properties, which are often based on the evaluator's expertise, intuition, and insight, may be subjective and non-repeatable.

Issues in Developing Security Metrics

Past efforts to develop security metrics include the Trusted Computer System Evaluation Criteria of the Department of Defense; the Information Technology Security Evaluation Criteria of the European Communities; the Systems Security Engineering Capability Maturity Model of the International Systems Security Engineering Association; and the international Common Criteria. These arrangements have had only limited success. A review of them suggests some essential factors that need to be addressed by researchers.

- System security is dependent on measurement of both **correctness and effectiveness.** Correctness is the assurance that the security components of a system have been implemented correctly and that they do what they are intended to do. Effectiveness is the assurance that the security components meet their stated security objectives, and that they do not do anything other than the intended tasks.

Correctness is evaluated by examining the ability of the security-enforcing mechanisms to carry out their tasks precisely to the specifications. Correctness can be assessed during the development and operations processes by determining how well the system meets it stated objectives. Effectiveness is evaluated by assessing the strength of the security-enforcing mechanisms to withstand attacks in carrying out their function. This assessment determines how well the security-enforcing components are integrated and work together,

the consequences of any known or discovered vulnerabilities, and the usability of the system.

Security evaluations of correctness and effectiveness are done largely through reasoning rather than direct measurement of actual hardware and software components. Evaluators may make assumptions, and results may not be timely and reproducible. Organizations frequently require the use of standardized procedures and criteria, and conduct evaluator training classes to help eliminate some of the subjective practices. However, more automated methods for evaluating correctness and effectiveness would be useful.

• Security metrics could lead to better assessments of the **leading**, **coincident**, or **lagging indicators** of the actual security state of the system. Leading and lagging indicators reflect security conditions that exist before or after a shift in security. Coincident indicators reflect security conditions that are happening concurrently with a shift in security. If a lagging indicator is treated as a leading or coincident indicator, the consequences due to misinterpretation and reaction can lead to serious problems.

Simple counts, when used as a security measure, can be especially hard to classify and interpret. An increase in the number of viruses detected by antivirus software could be a leading indicator, because the increased activity indicates an elevated threat level; but the count could also be a lagging indicator, because an efficient antivirus mechanism has been implemented. Also, decreased activity could indicate that the antivirus mechanism is losing its effectiveness, other security-enforcing mechanisms are increasingly successful, or the system is simply not being subjected to many attacks.

Many security measures can be viewed as lagging indicators. Over time, better understanding of a system and its weaknesses may lead to system security assessments that reflect a lower security standing and higher associated risk. This is often based on successful attacks on the system or other similar systems that reveal unexpected avenues of attack. Frequent repairs to systems make them more complicated to track. No metrics are available to measure the total state of security of a system.

• **Organizational security objectives** vary because organizations have different purposes, hold different assets, have different exposure to the public, face different threats, and have different tolerances to risk. Also, most organizations do not have sufficient funds to protect all computational resources and assets at the highest degree possible and must prioritize based on criticality and sensitivity.

Security metrics, which organizations use to determine how well they are meeting their security objectives, must meet the needs of different organizations. Since risks and policies are different, it is difficult to establish security metrics that could be used for system comparisons between organizations. There are similarities in high-level security objectives of organizations performing similar work. Security profiles of organizational security requirements and criteria can be used to standardize common sets of core requirements of such organizations for use in comparisons. However, these solutions have limitations insofar as only a portion of needed processes may be covered.

• Measurements of the **qualitative and quantitative properties** of software have been difficult to achieve. Many desired properties such as complexity, usability, and scalability are qualities that can be expressed in general terms, but are difficult to define in objective, useful terms.

Quantitative measures of security properties can be represented by terms such as low, medium, and high. Often, numeric values are used to represent rankings that are qualitative, such as, 1, 2, and 3, instead of low, medium, and high. The numeric difference between ranked values may be significant for some metrics, but may not be significant for security metrics. Quantitative valuations of several security properties may also be weighted and combined to derive a composite value, but these values can be misleading.

Qualitative properties may be intangible and cannot be captured via direct measurement. In cases where no quality can be clearly identified, such as the taste of wine, either a panel of experts rates various qualities using a blind rating or some measurable characteristics that are believed to correlate well with the quality in question are assessed. Developing techniques such as these could improve software security assessments.

• The security measurements of **small components of a system** do not necessarily indicate the **security of the larger system**. Security measurements have been more successful when the target of assessment is small and simple rather than large and complex. An evaluation, which focuses exclusively on cryptographic modules, generally requires less cost and time than an evaluation of a product that incorporates such modules. Larger systems generally have greater complexity and functionality, and the number of possible interactions increases as the number of components in a system increases, requiring more scrutiny and greater cost to evaluate.

Two systems, both of which are considered to be secure, can be connected together resulting in a composite system that is not secure. Composability is a property that would lead to better security measurements; composability would allow the security measurements of small systems to contribute directly to the measurement of the larger systems of which they are a part.

Areas of Research to Improve Security Metrics

Research efforts are needed to address these aspects of security measurements:

• Determine good estimators of system security.
• Reduce reliance on the human element in measurement and inherent subjectivity.
• Offer a more systematic and speedy means to obtain meaningful measurements.
• Provide understanding and insight into the composition of security mechanisms.

NISTIR 7564 identifies the following areas of research, which pose difficult and multifaceted problems for researchers. While these problems may not be solved

completely and quickly, work toward the goals stated above could lead to the development of improved security metrics.

• **Formal Models of Security Measurement and Metrics.** Security measurements that are conceived at a high level of abstraction and formalism are often difficult to interpret and apply in practice, such as when software patches, version updates, and configuration setting changes take place in operational environments. Formal models that depict security properties of operational IT systems and incorporate relevant objects of significance to system security measurement are needed.

The research goal is to establish formal models with a level of detail that is sufficient to enable realistic predictions of operational system behavior and portray security measurements accurately. Attack surface metrics, which uses a formal model defined from an intuitive notion of a system's attack surface (i.e., the ways in which the system can be entered and successfully attacked), is an example of the type of work envisioned. The formal model is characterized in terms of certain system resources—those methods, channels, and data items that an attacker can use to cause damage to the system. The surface measurement model can then be applied to compare attack surface measurements of systems along each of the three dimensions.

Research into formal models could also benefit the design of decision support systems that manage security infrastructure risks by using security metrics to determine security investments. Decision support models that incorporate technical and organizational aspects of a system and also quantify the utility of a security investment based on established principles could be valuable.

• **Historical Data Collection and Analysis.** Predictive estimates of the security of software components and applications under consideration should be extractable from historical data collected about the characteristics of other similar types of software and their vulnerabilities. Organizations could gain insight into security measurements by analyzing historical data collections to identify trends and correlations, and to discover unexpected relationships and interactions.

The research goal is to identify characteristics of software components and applications that can be extracted and used to predict the security condition of other software. Available open source software repositories could serve as a starting point for the data collection, but this approach will require additional effort to incorporate vulnerability information and to identify the points at which the known vulnerabilities first appeared in the code set.

A historical data collection could also be the basis for confirming the validity of independently proposed security measurements and methods of measurement, identifying whether measures are leading, lagging, or coincident indicators, and establishing estimates of latency and uncertainty for identified indicators. The data collection could also help in investigating new methods of detecting expected and unexpected relationships for use as estimators, and in developing mathematical and computational

methodologies to improve analysis of the data collection. A subset of the historical data collection could also be used as reference materials for training or rating the proficiency of security evaluators.

• **Artificial Intelligence Assessment Techniques.** Artificial Intelligence (AI) involves the design and implementation of systems that exhibit capabilities of the human mind, such as reasoning, knowledge, perception, planning, learning, and communication. AI encompasses a number of subdisciplines including machine learning, constraint satisfaction, search, agents and multi-agent systems, reasoning, and natural language engineering and processing. The application of AI to security metrics could lead to ways to reduce subjectivity and human involvement in performing security assessments.

The research goal is to identify areas of security evaluations that could be performed using AI or AI-assisted techniques and to demonstrate their use. Dealing with uncertainty and inconsistency has been a part of AI from its origins. Recently, AI systems have been used to independently formulate, refine, and test hypotheses from observed data to uncover fundamental properties, and to manage uncertainty and inconsistencies. The expectation is that AI technologies can play a similar role in the context of security assessments.

• **Practicable Concrete Measurement Methods.** The current practice of security assessment puts more emphasis on the soundness of the evaluation evidence of the design and the process used in developing a product than on the soundness of the product implementation. The rationale is that without a correct and effective design and development process, a correct and effective implementation is not possible. The emphasis on design and process evidence versus actual product software largely overshadows practical security concerns involving the implementation and deployment of operational systems.

The research goal is to devise methods of measurement that address vulnerabilities occurring in implementation and deployment, and complementing existing security assessment practices that emphasize design and development process evidence. Various forms of black box security testing offer an example of a possible type of concrete measurement method. For example, fuzzing is a type of fault injection technique that involves sending various types of pseudorandom data to available interfaces to discover unknown flaws present in programs and systems. Fuzzing techniques have been shown to be an effective means for detecting security vulnerabilities that otherwise might escape detection.

• **Intrinsically Measurable Components.** Development of computing components that are inherently attuned to measurement and that clearly exhibit security properties would be a significant improvement in the state of the art of security metrics. The research goal is to identify issues of mechanism and component design that facilitate or promote security measurement. Some potential methods include preparing strength of mechanism arguments in conjunction with the design and development of a security-enforcing component; establishing lower and upper bounds on mechanism strength, similar to the

way performance bounds are calculated for sorting, matching, and other essential algorithms used in computing; and applying evaluation criteria during the system design process to establish component properties.

Research results are available for cryptographic mechanisms that would allow bounds on the effort required to breach components to be determined, similar to metrics used to evaluate and identify weaknesses leading to failure in the physical security of storage safes and vaults. Extending this type of analysis to trust mechanisms is a more challenging problem, but not without promise. For example, components that rely on certain surety mechanisms, such as authentication modules designed for passwords or biometric modules for fingerprints, lend themselves to certain types of strength analysis.

Information on NIST Security-Related Publications

For information about NIST standards, guidelines, and other security-related publications, see http://csrc.nist.gov/publications/index.html.

Disclaimer

Any mention of commercial products or reference to commercial organizations is for information only; it does not imply recommendation or endorsement by NIST nor does it imply that the products mentioned are necessarily the best available for the purpose.