Information Assurance Measures and Metrics
- State of Practice and Proposed Taxonomy –

Rayford B. Vaughn, Jr.
Associate Professor
Department of Computer Science
Mississippi State University
Mississippi State, MS 39762
(662) 325-7450 (o)
(662) 325-8997 (f)
vaughn@cs.msstate.edu

Ronda Henning
Harris Corporation
Government Communications Systems Division
MS W3/9704
PO Box 9800
Melbourne, FL 32902
(321) 984-6009 (o)
(321) 674-1108(f)
henning@harris.com

Ambareen Siraj
Department of Computer Science
Mississippi State University
Mississippi State, MS 39762
(662) 325-2756 (o)
(662) 325-8997 (f)
ambareen@cs.msstate.edu

Information Assurance Measures and Metrics
- State of Practice and Proposed Taxonomy –

Abstract

The term " assurance" has been used for decades in trusted system development as an expression of confidence that one has in the strength of mechanisms or countermeasures. One of the unsolved problems of security engineering is the adoption of measures or metrics that can reliably depict the assurance associated with a specific hardware and software system. This paper reports on a recent attempt to focus requirements in this area by examining those currently in use. It then suggests a categorization of Information Assurance (IA) metrics that may be helpful to an organization defining a set of measures or metrics for a specific application. This issue of rating and ranking systems in terms of their assurance characteristics was at least partially addressed at a recent workshop on information security system rating and ranking (hereafter simply referred to as "the workshop") held during the period May 21-23, 2001 in Williamsburg, Virginia. Two of the authors were members of the program committee and participated in the workshop discussions. We believe that the provision of security mechanisms in systems is a subset of the systems engineering discipline and that it has a large software-engineering correlation. As software engineers, we understand that the determination and application of measures and metrics is not an exact science, nor is it easily accomplished. As system engineers we understand that these measures must include hardware, processes, and personnel components. IA metrics are essential for measuring the "goodness" of IA as a system element and we believe that overall useful IA metrics are possible. There is general agreement that no single system metric or any "one-prefect" set of IA metrics applies across all systems or audiences. The set most useful for an organization largely depends on their IA goals, their technical, organizational and operational needs, and the financial, personnel, and technical resources that are available. To assist an organization investigating options for IA metrics it is useful to examine various general categories of IA metrics. For example, an IA metric can be categorized as objective/subjective, quantitative/qualitative, static/ dynamic, absolute/relative and direct/indirect. We briefly describe these and other categorizations in this paper, which leads to a proposed taxonomy for IA metrics that can be tailored to an organization's needs.[1]

[1] The categories outlined here are from research at Mississippi State University's Center for Computer Security Research (http://www.cs.msstate.edu/~security) in a larger effort to create taxonomy for IA metrics and measures.

I. Introduction

In today's competitive and dynamic information technology (IT) environment of networks, portals, and software component application servers, enterprises no longer question the need for IT security as an integral component of their enterprise IT architecture. The available security technologies for any one application suite are complex, costly and can be inconvenient to the end user. The convergence of several such application suites into an integrated environment is not only common but may be mandated within the enterprise, and these composite suites are often difficult to evaluate against information security requirements.

The concept of "security metrics", including product evaluation criteria identification, Information Assurance (IA) strength quantification, risk assessment/analysis methodology development, and other related activities have led to the widespread desire for a comprehensible, simple IA measurement technique. This technique or measure has a variety of purposes: e.g., rating security goodness, purchasing a given countermeasure, operating or retiring a given system component. To date, computer science has frustrated these activities by providing neither generally accepted nor reliable measures for rating IT security or requisite security assurance. Furthermore, inconsistent terminology has complicated the development of IT metrics, often confusing single measurements with accepted metrics, such as rating, ranking, quantifying, or scoring measurements. To at least partially address this shortfall in the information assurance science, a workshop was held in Williamsburg, Virginia during the period May 21 through 23, 2001. This paper summarizes the findings of this workshop, identifies important shortfalls, and suggests a proposed taxonomy for IA measures/metrics.

II. A Report on the Workshop and Results

A. The Workshop[2]

The issue of rating and ranking systems in terms of their assurance characteristics was partially addressed at the three day workshop on information security system ratings and ranking (hereafter simply referred to as "the workshop") held during the period May 21-23, 2001 in Williamsburg, Virginia.[3] The goals of this workshop were as follows (taken from the original call for position papers):

- To clarify what researchers and practitioners mean when they refer to IA metrics.
- To debunk the pseudo-science associated with assurance metrics.
- To discover some indirect indicators of security.
- To precisely define the research problems in developing IA metrics methodologies.
- To recap the latest thinking on current IA metrics activities.
- To identify efforts that are successful in some sense, if they exist, and if none exist, reduce expectations on what might be achieved through IA metrics.
- To explore the unintended side effects of ratings/measures (e.g., inflating the numbers to ensure promotion, delay review by higher authority).
- To clarify what's measurable and what's not.
- To scope and characterize the measures to be addressed (e.g., EJB Security, CORBA Security, and/or Microsoft DNA Security) and to explain what happens when several of these measures or applications co-exist in the same enterprise: do they augment each other or cancel each other out?
- To describe how measures should be used in the context of IA, especially to influence purchases and for general resource allocations.
- To identify misapplications of measures, including their description as "metrics" .

[2] Sponsored by the MITRE Corporation and the Applied Computer Security Associates (ACSAC).
[3] While open to the public, the workshop required all participants to submit a short position statement on some aspect of information system security rating and ranking. Thirty seven members of the IA community submitted papers and participated.

Specific outcomes of the workshop were publicly discussed at the DOD Software Technology Conference 2002, held in Salt Lake City Utah April 29 – May 2, 2002 [4], and the Canadian Information Technology Security Symposium held in Ottawa Canada May 13, 2002 [25]. The workshop was also discussed at a closed meeting of the National Infosec Research Council, a U.S. body of funding organizations that set the U.S. national research programs. There is not sufficient space in this paper to present the results completely. The full proceedings can be located at http://www.acsac.org/measurement/. A summary discussion is provided in [1].

B. Workshop Findings and Observations

There is often confusion with the words we use when discussing measurement - metrics, measures, indicators, and predictors are frequently used interchangeably. We are also often confused about what the measurement or metric characterizes (process or product), how to interpret it, and how to validate it. Measurements are generally always possible - they simply tell us the extent, dimensions, capacity, size, amount, or some other quantitative characteristic of the software or system. They are discrete, objective values. Measures are normally not too useful without interpretation, except in direct comparison with other measures to determine is one value is more or less desirable than the other. It's difficult to draw conclusions on measures alone. Only when we relate individual measures to some common terms or framework do they become metrics. Examples might include defects per 1000 lines of code or the number of vulnerabilities found in a particular system scan or penetration attempts per month. Once we establish the metric - we face the problem of interpretation, is the metric is useful, predictive of future behavior, an indicator of aspect of assurance, and the granularity of scale. In civil or mechanical engineering for example, there is a degree of rigor in the proof that certain metrics are true and accurate predictors of a characteristic. Empirical data, trended over time, provides a correlation function. The physical world complies with the laws of physics and many of those laws are well known to engineers. The systems engineering world (to include the software engineering discipline) is not as rigorous as the physical sciences to a large extent, and presents more of a challenge in "proving" the correctness of a measurement

technique. Over the years, we have often proven this lacks of rigor when our systems fail, are unreliable, and are fraught with user complaint. How then can we claim to have metrics that quantify assurance when we do not seem to be able to prove correctness, maintainability, reliability, and other such non-quantifiable system requirements? Prediction relies to some extent on history being an indicator of future behavior. In software and systems engineering this may not be true. Knowing that a particular defensive strategy has worked well in the past for an organization really says very little about its strength or ability to protect the organization in the future. Examples of difficulties that we face in predicting strength (i.e., assurance) include the following.

- Software does not comply with the laws of physics. In most cases, we cannot apply mathematics to code to affect proof of correctness in the way a bridge builder can apply formulae to prove structural strength characteristics. Formal methods in software development have a very useful function and most certainly add to assurance expectations- but they cannot today, or in the near future, realistically prove total system correctness and guarantee assurance. They provide evidence only.

- People, who are by nature error prone, build software. We can measure certain characteristics of our software construction process and the people who labor at it, but in the end - any one of them can intentionally or unintentionally corrupt the system and greatly diminish its assurance. It remains questionable whether or not open systems development is a helpful countermeasure or a version control nightmare.

- Compositions of mechanisms to construct a security perimeter comply with no known algebra. Aggregation of various countermeasures may result in an inherently less secure system. We simply do not know what we have once we put a security perimeter in place. Nor do we have any guaranteed assurance that we implemented the composition properly and resulted in a stronger system if we deployed additional countermeasures. Anyone who has attempted to correctly

configure a firewall will attest to the false sense of security that can occur due to the high likelihood of a single misapplication of a rule or the omission of a single rule, coupled with the propagation of configuration data across an enterprise, and we compound the possibility of an assurance compromise. We remain reliant on the expertise of our systems administrators or security engineers and their specific knowledge to guarantee the correctness of a system.

- It is easier to attack a system today (an assurance issue) than it was 5 years ago due to the pervasive communications and shared knowledge of the Internet. This trend is likely to continue as attack tools are further automated, shared, and explored on a global basis. Whereas once it was reasonably labor intensive to run a password attack on a system - today, one can load up readily available scripts, launch them, go away for a good nights sleep, and collect the results in the morning.

The workshop attempted to address (at least partially) these questions and others. Although many specific techniques and suggestions were proffered to the group, it was apparent to all that some combination of measures was essential. It was also evident that this combination could not generically be applied across all domains of interest. It was clear that measures or metrics adopted by an organization to determine assurance need to be frequently reassessed to determine the applicability and relevance. Attempts to apply a single rating to a system have been attempted in the past and have failed miserably [2,3]. There was also some agreement among the workshop organizers that the problem domain might be best viewed using a non-disjoint partitioning into technical, organizational, and operational categories (i.e., there is some inevitable overlap among these domains that must be accepted). At the workshop, the following categorizations were defined.

- The technical category includes measures/metrics that are used to describe and/or compare technical objects (e.g., algorithms, products, or designs).

- Organizational measures are best applied with respect to processes and programs.
- Operational measures are thought to describe, "as is" systems, operating practices, and specific environments.

An interesting characterization of information security metrics was captured by Deborah Bodeau of the MITRE Corporation [9] who stated that a proper view of these metrics might best be viewed as a cross-product involving what needs to be measured, why you need to measure it, and who you are measuring for. Her characterization of this view in Figure 1 is enlightening.

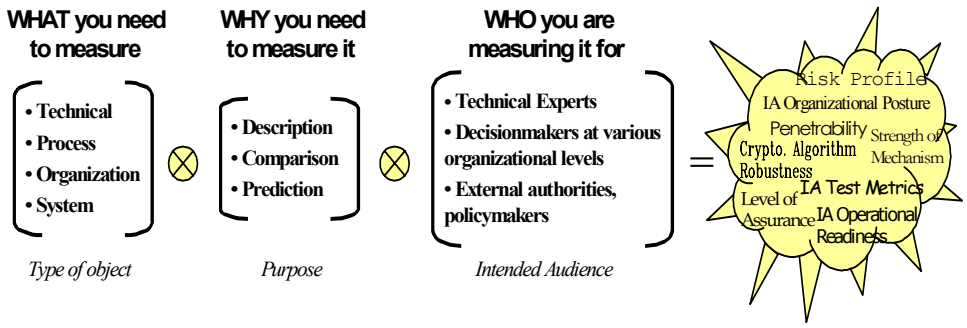

Figure 1: Characterization of IS Metrics (Bodeau)

Another interesting observation made by many of the attendees was that the desired purpose for such measures and metrics seemed to vary between the government and commercial sectors. Government applications seem much more likely to use metrics and measures for upward reporting and organizational reporting. Answering such questions as "what is our current assurance posture", "how are we doing this month compared to last", and "are we compliant with applicable regulations and directives" seemed to be the drivers for the metrics needed. The representatives at the workshop from industry seemed less interested in answering these questions and more inclined to look for answers to the questions like: "how strong is my security perimeter", "what is the return on my security investment", "what is my level of risk or exposure", and "product measures for comparison". The authors of this article also observe that the

commercial sector seemed to have far more interest in technical and operational measures than in process or organizational measures.

The workshop attendees had hoped to find a number of objective, quantitative metrics that could be applied. Although unanimous agreement was not reached, it was apparent to most that such metrics were in short supply, had to be combined with other measures or metrics in a particular context, and were generally not very useful on their own. Many more measures that would be considered subjective and/or qualitative appeared more useful. Examples of such a measure might include adversary work factor – a form of penetration testing. An excellent discussion of this topic is found in [5]. Although penetration techniques are not truly repeatable and consistent, the workshop found great agreement that their results were meaningful and useful. In fact, there was significant agreement at the workshop that penetration

Table of Example Metric Types

Type of metric	Use	Issues
Technical IA metrics, e.g., number of vulnerabilities detectable by scanner, EAL	Differentiate among technical alternatives	Other factors (e.g., interoperability with enterprise management software) may be more relevant to product selection.
Product development process metrics, e.g., ISO 9000, SSE-CMM	Differentiate among product suppliers (surrogate indicator of product quality)	Other factors (e.g., preferred supplier agreements) may be more relevant to product selection.
Acquisition process metrics, e.g., level of information systems (IS) expertise in procurement office	Allocate resources to provide IS expertise, determine level of effort for certification	Process metrics may fail to indicate constraints on acquisition process or procurement office.
Certification level (NIACAP, DITSCAP)	Determine requirements for certification activities, documentation	Relevant factors (e.g., system exposure) may fail to be addressed in definition of certification levels. Identification of activities does not directly indicate required level of effort.

testing was one of the most useful measures of system assurance that exists today. Risk assessments, in their various forms, were also found to be useful measures of assurance. Such assessments are accomplished in a variety of ways, but give a good indication of how one is positioned to withstand attacks on a system. Such assessments also tend to be very dependent on specific organizational objectives and needs, and are therefore very focused to a given environment or user community. The above table was taken from [9] and provides examples of types of IA metrics relevant to IT modernization processes:

III. Workshop Summary[4]

The workshop proceedings [9] list characteristics for "good" IA metrics. Conflicts do exist among these criteria that were not addressed in this first effort due to lack of time. Examples of proposed criteria for IA metrics include:

- Scope. The portion of the IS problem domain the IA metric describes should be clearly characterized.
- Sound foundation. The metric should be based on a well-defined model of the portion of the IS problem domain it describes.[5]
- Process. The metric assessment process should be well defined. The process definition should include qualifications of evaluators, identification of required information, instructions on how specific factors are to be measured or assessed, algorithms for combining factor values into final values, and explanations of sources of uncertainty.
- Repeatable, i.e., a second assessment by the same evaluators produces the same result.
- Reproducible, i.e., a second assessment by a different set of evaluators produces the same result.

[4] These conclusions were taken from the executive summary of the workshop proceedings – a document that the authors of this paper participated in creating.

[5] A variety of problem domain taxonomies or descriptions may be useful. For example, the FITSAF provides a definition of the IS programmatic domain. The 16 May 2001 draft NIST publication, *Underlying Technical Models for Information Technology Security* (http://csrc.nist.gov/publications/drafts.html), provides a framework.

- Relevance. IA metrics should be useful to decision-makers. Considerable discussion related to IA metric stakeholders: decision-makers and the types of decisions IA metrics support, and individuals and organizations supplying inputs to IA metric evaluations.
- Effectiveness. It should be possible to evaluate the IA metric quickly enough, and with low enough costs, for it to be useful to the decision-makers who will use it.

Direct measurement of IS properties is desirable, but not always possible. The assessment process should include activities for validating the indicator, e.g., by correlating it against other indicators. For example, an indicator of an organization's IS program might be the quality of its documented plans. If an organization's commitment to information security is reflected in the size of its budget, an assessment of organizational assurance plans could be correlated with financial metrics.

IA metrics must evolve. A metric that is meaningful and useful today may be less relevant tomorrow, due to changes in technology, practice, or regulations. Organizational processes that apply IA metrics should include periodic re-evaluation of those metrics, and re-definition or orientation as needed. If metric evolution is not done deliberately, it will occur accidentally: the information that can be gathered will change in response to dynamic technology changes, and assessment techniques that involve expert judgment will evolve as expertise increases. Care must, therefore, be exercised in comparing metric values over extended periods of time.

IV. A Proposed Taxonomy

In order to develop an IA metrics program, it is useful to define a measurement classification framework. A taxonomy is a classification scheme that can serve as a crucial means for conducting any systematic study – to include a metrics program. There is no consensus taxonomy of IA metrics in the literature to our knowledge. We know from Villasenor [6] that there have been recent efforts by the DoD to develop such a taxonomy. In particular, the Air Force Research Laboratory (AFRL) Intelligent Information Systems

Branch (IFTD) is involved in such an effort [7]. In this paper, we suggest a taxonomy of IA metrics that can serve as a "cognitive infrastructure of IA assessment" [8] to assist in better understanding of the characteristics associated with different IA metrics. It may also provide a common frame of reference for classifying current and future IA metrics which will be useful in insuring organization coverage and for discussions surrounding the need and utility of the metrics.

A. Types of IA Metrics

IA metrics are essential for measuring the "goodness" of IA countermeasures, however there is no single system metric nor there is any "one-prefect" set of IA metrics for all. Which set of metrics will be most useful to an organization depends on one's IA goals, technical, organizational and operational needs, and the resources available. To investigate options for the IA metric selection process, we begin with a categorization of different forms of IA metrics. An IA metric can be objective/subjective, quantitative/ qualitative, static/dynamic, absolute/relative or direct/indirect. These categories are briefly described below:

- ***Objective/Subjective***: Objective IA metrics (e.g., mean annual down time for a system) are more desirable than subjective IA metrics (e.g., amount of training a user needs to securely use the system). Since subjectivity is inherent in information assurance, subjective IA metrics are more readily available.
- ***Quantitative/Qualitative***: Quantitative IA metrics (e.g., number of failed login attempts) are more preferable than qualitative IA metrics (e.g., FITSAF self-assessment levels) because they are discrete, objective values.
- ***Static/Dynamic***: Dynamic IA metrics evolve with time, static IA metrics do not. An example of a static IA metric can be the percentage of staff that received an annual security training refresher [9]. This metric can degrade in value if the content of the course does not change over time. A dynamic IA metric can be the percentage of staff who received training on the use of a current version of the

software package. Most metrics used in penetration testing are dynamic. Dynamic IA metrics are more useful than static because best practices change over time with technology. [10]

- ***Absolute/Relative***: Absolute metrics do not depend on other measures and either exist or not [9]. An example might be the number of SANS certified security engineers in an organization. Relative metrics are only meaningful in context - e.g., the number of vulnerabilities in a system cannot provide a complete assessment of the system security posture. The type and strength of countermeasures is also important in this context for making any decision about the system's IA posture.
- ***Direct/ Indirect***: Direct IA metrics are generated from observing the property that they measure - e.g., the number of invalid packets rejected for a firewall. Indirect IA metrics are derived by evaluation and assessment (e.g., ISO Standard 15408). It is normally preferred to measure behavior directly, but when that is not feasible, indirect measures are used to postulate the assurance posture.

IA is a triad of cooperation between the technology that provides assurance, the processes that leverage that technology, and the people who make the technology work [11] in operational use in the *real world*. IA metrics should encompass all - the product, the process and the people.

B. The Taxonomy of IA Metrics

In defining a classification scheme for IA metrics, we chose to investigate two issues - what can be measured with current technology and how we measure it. We also considered recent related research on IA metrics as reported in:

- Workshop on "Approaches to measuring security conducted" by the Computer System Security and Privacy Advisory Board (CSSPAB) from June 13-14, 2000.
- Workshop on "Information Security System Rating and Ranking" by Applied Computer Security Associates (ACSA) from May 21-23, 2001.

We selected these workshops because these were the only public workshops that were exclusively dedicated to the topic of information assurance/security measurements and represented the collective perspectives of more than forty researchers from different government, military, commercial, and private sectors. From these observations, we determined that the objective of assurance measurement could be grouped into two distinct categories - assessing an organization's IA posture or measuring the IA capabilities of systems or products. When grouping IA metrics for information systems and products within the same category, we refer to that which we wish to measure as a Technical Target of Assessment or TTOA (using the Common Criteria term). We often measure individual IA capabilities in products for systems and we generally take the additional step of assessing how the composition of different product's security strengths affect the overall IA capability. This category includes "technical objects" such as cryptographic algorithms.

The complete taxonomy is shown in Figure 2. At the root level we first classify IA metrics into the aforementioned categories: metrics for organizational security and metrics for the TTOA.

From an organizational perspective, we can refer to IA metrics as IA performance trends observed over time, based on repeatable measurements taken at regular intervals [10]. From the TTOA's perspective, we refer to the slightly modified definition by Connolly [12] which is that IA metrics are measures that gauge a TTOA's ability to protect, detect, and respond to IA attacks.

B.1. Metrics for Organizational Security. These metrics measure organizational programs and processes. Metrics for organizational security are used to provide feedback to improve the IA posture of the organization. Since different organization's infrastructure, objectives, and environmental settings can very diversified, IA metrics for organization are difficult to generalize.

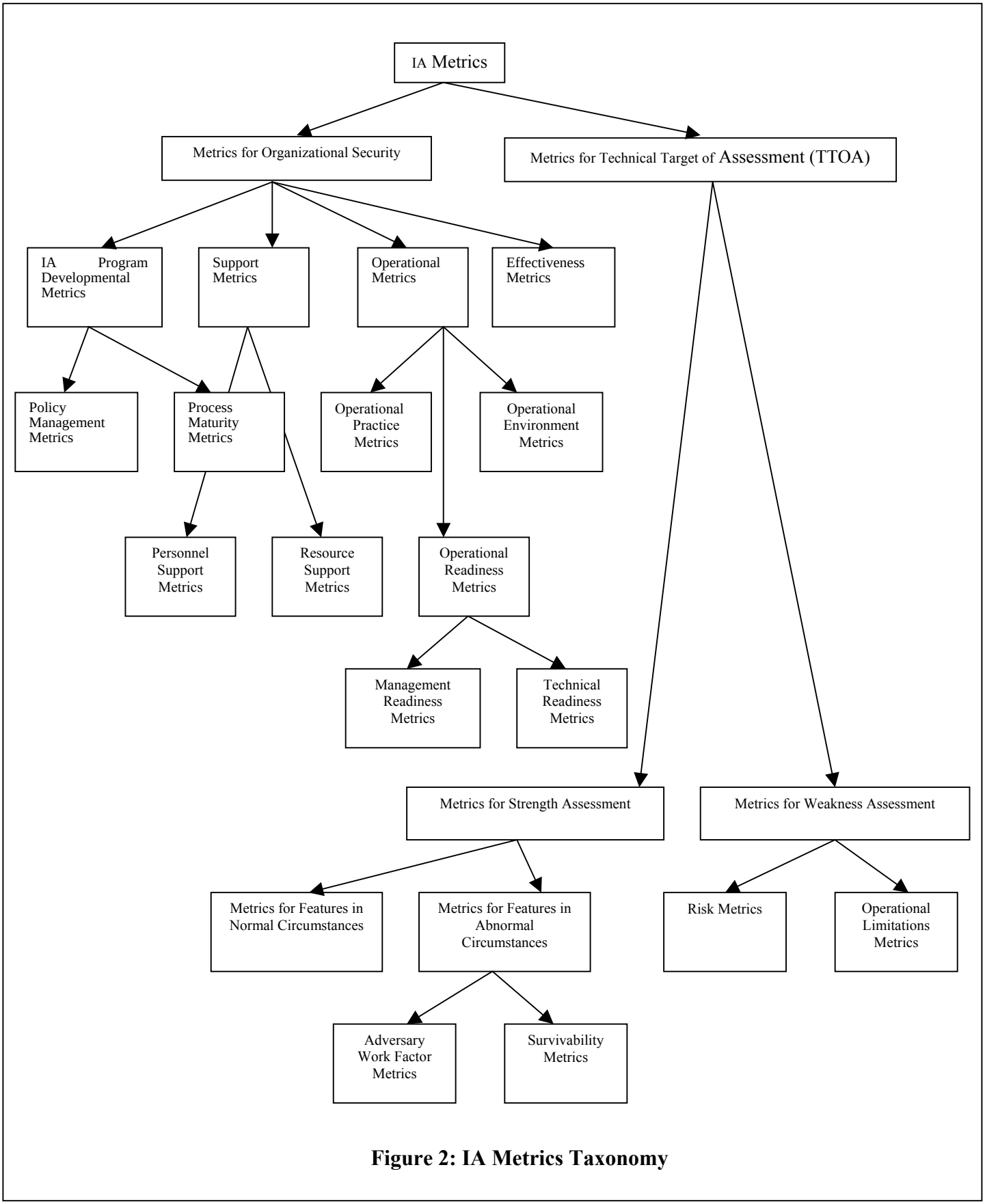

Figure 2: IA Metrics Taxonomy

We further classify metrics for organizational security into four categories based on what they measure - IA Program Developmental Metrics, Support Metrics, Operational Metrics, and Effectiveness Metrics. Each of these is briefly described below:

- **IA Program Developmental Metrics.** Organizational IA programs are a comprehensive set of program areas that together guide an organization's ability to provide information assurance [13]. IA program developmental metrics measure the extent that IA is effective in an organization by measuring if the organization has chosen the policies and process. These metrics can be further classified as policy management or process maturity.
    - **Policy Management Metrics.** Measures that management uses as security objectives for an organizational IA program. These metrics are specific to development of security strategy, policy, implementation of policy, and compliance with policy. An example of a Policy Management metric is the appraisal used by the Federal Information Technology Security Assessment Framework (FITSAF) that provides a self-assessment guide for organizations to use to measure the assurance of their security program. [14].
    - **Process Maturity Metrics.** These metrics assess the maturity of security practices in developing a system. They are used to measure the organizational security process framework required to develop a good information assurance program. Process maturity metrics concentrate on security engineering activities that span the life cycle of secured systems deployed by organizations. Examples here include the Common Criteria that measures process factors of systems by ranking them in one of the seven evaluation assurance levels (EAL's) - primarily by examining the artifacts of the development process [15]. Similarly, the systems Software Security Engineering Capability Maturity Model (SSE-CMM) measures developers' process and procedure based on artifacts [9].
- **Support metrics**. Measure an organization's support for security programs and processes in terms of personnel (e.g., awareness, training, experience) and resource (e.g., funding, technical resources).

- **Personnel Support Metrics.** People are a part of any process. Professionals and practitioners developing, operating, defending, attacking, or evaluating a system are critical components for ensuring higher standards for IA. ISC's Certified Information System Security Professional (CISSP) & Systems Security Certified Practitioners (SSCP) are good indicators of an individuals knowledge of best practices, their credibility as practitioners, and their exhibition of a sound working knowledge of security [16]. The number of CISSP professionals in an organization, for example, can indicate that an organization has experienced knowledgeable personnel support.
- **Resource Support Metrics.** Serve as indicators of organization's financial support and available resources for IA programs and processes. Such metrics help one to determine if budget allocation is adequate or proper resources are in place. An example of this type of metric can be the budget percentage allocated for security program as a percentage of annual organizational budgets.

- **Operational Metrics.** These are end-to-end measures of operational support in an organization. Operational metrics for organization's security program observe the working environment of the organization in terms of its security program and evaluate the organization's operational readiness and effectiveness in providing information assurance. The operational readiness metrics are sub-divided into three categories, operational readiness metrics, operational practice metrics, and operational environment metrics.
  - **Operational Readiness metrics**. This concept was drawn from the traditional military readiness measures of combat readiness [12]. The IA posture of an organization can be measured by how well its units (systems, departments) and individuals are prepared to perform their assigned tasks of operating the system in a proper manner. Readiness measures are internally self-assessed or externally assessed by third party. An example of the IA readiness metric exists in a current Joint Chief of Staff Instruction (CJCSI) as a self-assessment checklist

of IA related capabilities (e.g. “if adequate architecture for securing systems and networks is in place”) [12]. Operational readiness metrics can be further classified as management readiness related and/or technical readiness related.

- **Management Readiness Metrics**. Measures management’s support of information security processes in the organization – for example, commitment, personnel and resource management, and risk assessment of intellectual property. These metrics are mostly static, i.e., these are questionnaire-based assessments and are generated by reviews of organizational policy and procedures with respect to the operations by interviewing management [12]. An example is the frequency of regular audit trail reviews or operational procedure drills.
- **Technical Readiness Metrics**. Measure the readiness state of technical support that affects the organization’s ability to provide information assurance while performing operational missions. They can be static or dynamic. Risk assessment and vulnerability analysis are examples of static technical readiness measurements. Information Assurance Vulnerability Alerts (IAVA) by Defense Information System Agency (DISA) require organizations to use IA metrics to remediate known vulnerabilities of the technical resources, keep track of remediated systems and report compliance status [12]. Dynamic technical readiness assessments are more of a "live-play" exercise that simulates adversarial scenarios [12]. Red team threat based efforts apply a simulated task force to expose IA vulnerabilities, as a method to assess the readiness of DOD components. A specific example of this type would be the Information Design Assurance Red Team (IDART) methodology used by Sandia National Laboratories [17] which results in metrics such as attack percent completed, attack probability of success, and time/cost/skill in attacks.

- **Operational Practice Metrics.** Measure the security practices of people who directly or indirectly affect an organization's IA posture. These metrics assess culture and climate, awareness of existing policy, and socio-ethical awareness for example. An example might be the number of users with passwords in compliance with the local password management security policy.
- **Operational Environment Metrics.** Used for describing and measuring the security relevant aspects of the operational environment (i.e., external threats, conditions, objects) that affect the organization's security operations directly or indirectly. An example might be number of systems susceptible to a specific penetration technique.

- **Effectiveness Metrics.** Measure how effective the organization's IA program is in actually providing defense in-depth assurance. Examples include the number of malicious code incidents (measures protection), number of intrusions reported (measures detection), percentage of data recovered after security incident (response). The Air Force Information Warfare Center (AFIWC) and its one-line surveys (OLS) use quantitative effectiveness metrics such as the number of systems root or user privileges that were obtained as a percentage of the total number of systems [12]. The Air Force Communication Agency (AFCA) developed information protection metrics that measures compliance with and the effectiveness of information protection policy in organizations, e.g., number of intrusion attempts reported and number of reported successful intrusions with limited access or total control [18]. Another example might be the number of security incidents this month/number of security incidents previous month.

B.2. Metrics for Technical Target of Assessment (TTOA). This type of metric is intended to measure how much a technical object, system or product (collectively refered to as TTOA) is capable of providing assurance in terms of protection, detection and response. This type of metrics is often used in comparing or differentiating between alternative and competing TTOA, e.g. the EAL ratings of the Common Criteria, DITSCAP certification levels developed by DoD Information Technology Security Certification and

Accreditation Process. We further categorize metrics for TTOAs in two classes - metrics for measuring TTOA's strengths and its weaknesses.

- **Metrics for Strength Assessment**. The focus here is on how strong is the TTOA. The strength factor is further classified into two categories used for assessing the strengths of the TTOA in two categories based on the typical environment when there is no adversarial activity going on to compromise the TTOA and its capabilities and when there is some adversarial force working against the TTOA. We refer to these as normal and abnormal circumstances.
    - **Metrics for Features in Normal Circumstances**. These metrics measure the capabilities that the TTOA should have in order to provide information assurance under normal circumstances. They can be used for assessing the claimed features of a TTOA. For a firewall, metrics in this category might be the number of invalid packets a server can reject per second; for a cryptographic algorithm, this metric might be the number of clock cycles per byte encrypted, number of rounds, or something similar. The resilience assurance index is another example of this metric as it provides a way to evaluate systems in terms of the level of system expectations or assurances one expect from a system to provide defense to attacks [11].
    - **Metrics for Features in Abnormal Circumstances**. These metrics are used for measuring the TTOA's capabilities in the face of adversarial activities working to compromise the TTOA. They measure the TTOA's strength in resistance to and in response to attacks. Two further refinements of this classification are adversary work factor and survivability metrics.
        - **Adversary Work Factor Metrics**. Penetration testing is used to assess the strengths of systems [19] and the concept of Adversary Work Factor metrics was generated from penetration testing. The idea is, the stronger a system is the more likely it is to withstand attacks. Relative differences in adversary work factor can provide insight to relative assurance of information systems [20]. Adversary work factor is the amount of effort an adversary spends in order to compromise protective measure(s) of a system. It not only

incorporates technical factors, but also personnel and operational factors. SRI International developed an Adversary Work Factor metrics known as Red Team Work Factor metrics, which is an estimate of the effort required by a model adversary to achieve adversarial goals [20]. The metric is a function of preparation time, attack time, cost of resource and access, man-hours to break a security policy, and time to penetrate the system.

- **Survivability Metrics.** These metrics measure the TTOA's ability to deliver essential services in the presence of attacks and failures and to recover in a timely manner [21]. The Survivable Network Analysis (SNA) methodology was developed by the SEI CERT Coordination Center. This methodology utilizes statistical techniques for assessing of the survivable properties of systems. The analysis is carried out from the architectural level to the operational level. An example metric in SNA is actual survivability, which is quantitatively determined by the system's performance at the new state after attack against its normal performance level. SNA also looks at other metrics such as, expected survivability, average damage per unit time, and others [21].

- **Metrics for Weakness Assessment**. These metrics assess the weaknesses of the TTOA in terms of threats, vulnerabilities, risks, anticipation of losses in face of attack and any operational limitations of the TTOA. This classification of metric is sub-categorized into risk and operational limitation metrics.
    - **Risk metrics.** Risk metrics are those that measure threats, vulnerabilities and associated risks to the TTOA. Threat is an external or internal circumstance/event that may cause potential harm to the system. Vulnerability is a weakness of an information system or its components that could be exploited to violate assurances in systems. Risk is the probability that a particular threat will exploit a particular vulnerability of the system. The intelligent communities' INFOSEC Risk Management Methodology provides a consistent repeatable measurement method for determining IA risk of a system by observing and analyzing the threats, vulnerabilities and

significance levels. The result is a qualitative subjective measurement of the risk factor of the system [22].

- **Operational limitation metrics.** These metrics measure the impact of operational limitations that are generated by certain functionality or limitations that might restrict or affect the functionality of evident features of the TTOA. This metric is useful for evaluating competing products [23].

V. Conclusions

The workshop was successful in focusing attention on the area of metrics or measures for systems that have security or assurance as a requirement, but was not successful in coming to agreement on a set of measures to be used or even finding consensus in any particular approach. Nonetheless, there were observations that emerged from this workshop that the reader may be useful.

- There will be no successful single measure or metric that one can use to quantify the assurance present in a system. The problem is far too complicated and the stakeholder community far too diverse. Multiple measures will most certainly be needed and they will need to be refreshed frequently.
- Software and systems engineering are very much related to this problem. Quality of the software delivered, the architectures and designs chosen, the tools used to build systems, the specified requirements and more are all related to the assurance we are trying to quantify.
- Penetration testing is, today, a valid measurement method. It is imperfect and to some extent non-repeatable, but nonetheless, it is used in both Government and Commercial sectors. Several measures are suggested related to such testing – they include level of effort, numbers of vulnerabilities found, number of penetrations, number of vulnerabilities not found, and more.
- There are differences between the Government and the Commercial sectors. One is policy driven – the other is profit driven. One has the force of law behind it, the other has the force of stockholders driving it. This may result in different values placed on metrics or measures between the two sectors.

- Defense in depth and breadth is important. Knowing how to measure this defense is also important and a valid research area. There was no agreement on how to accomplish this measurement.
- Attempts to quantify and obtain a partial ordering of the security attributes of systems in the past have not been successful to a large degree (e.g., the TCSEC and the Common Criteria [2,3]). It remains to be seen if this will continue.
- Processes, procedures, tools, and people all interact to produce assurance in systems. Measures that incorporate all these are important. We believe Bodeau's work in Figure 1 very well characterizes this.

There is no definitive work in the area of a classification scheme or taxonomy. We believe the work we present here is a step forward in that direction. A desired property of any taxonomy is that its categories should be mutually exclusive and collectively exhaustive [24]. The taxonomy we offer does not guarantee these characteristics at the present time and it is hoped that the community can critique and build on that which is offered. We did, however, attempt to structure the categories in a way that met the desired properties. Overall, we suggest that the strengths of this proposed taxonomy are:

- The categories are accompanied by definitions such that an IA metric can find membership.
- The taxonomy is comprehensible making it suitable for general audience.
- The terminology of the taxonomy is consistent with established information systems terminology.
- The classification scheme presented provides an information security professional with a tool to held consider all areas needing measurement and suggestions for types of measures to employ.

This taxonomy is the beginning of the construction of a common framework for IA metrics. It is our hope that it will contribute to the community's larger effort. We welcome generous feedback in order to improve our initial work.

VI. Acknowledgements

Support by Harris Corporation is gratefully acknowledged. Work conducted at Mississippi State University was supported by the Army Research Laboratory contract DAAD 17-01-C-0011 and the National Science Foundation grant CCR-0085749.

VII References

[1] R. Vaughn, D. Dampier, and A. Siraj, "Information Security System Ranking and Rating", CrossTalk the Journal of Defense Software Engineering, May 2002, pp. 30-32.

[2] Department of Defense Standard, *Department of Defense Trusted Computer System Evaluation Criteria,* DOD 5200.28-STD, GPO 1986-623-963, 1985.

[3] ISO Standard 15408, The Common Criteria

[4] Vaughn, R. and Henning, R. "A Report from the Workshop on Information Security System Rating and Ranking" Proceedings of the DOD Software Technology Conference 2002, Salt Lake City, Utah, April 29 – May 2, 2002. http://www.stc-online.org

[5] G. Schuedel, and B. Wood,, "Adversary Work Factor as a Metric for Information Assurance", Proceedings of the New Security Paradigm Workshop, Sept 18-22, 2000, Ballycotton, Ireland.

[6] Villasenor, P. V. (SAIC, DIAP). 2001. DoD operational metrics. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VR from May 21-23, 2001.

[7] AFRL/IF. 2001. Information assurance metrics: Project overview. http://www.rl.af.mil/tech/programs/ia/ia12.html

[8] Shapiro, S. (The MITRE Corporation). 2001. The bull in the china shop: The "Merrill Lynch" IA Assessment manifesto. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[9] ISSRR. 2001. Proceedings Workshop on Information-Security-System Rating and Ranking (ISSRR) held in Williamsburg, VA, May 21-23, 2001. http://www.acsac.org/measurement/

[10] Bartol, N. (Booz•Allen & Hamilton). 2001. IA metrics development and implementation. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[11] McCallam, D. (Logicon). 2001. The case against numerical measures of information assurance. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[12] Connolly, J. (The MITRE Corporation). 2001. Information Assurance operational readiness metrics. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001

[13] King. G. (Computer Science Corporation). 2000. Best Security Practices: An overview. http://csrc.nist.gov/nissc/2000/proceedings/papers/022.pdf

[14] Swanson, M. (NIST). 2001. Self-assessment guide for information technology systems. http://csrc.nist.gov/publications/nistpubs/index.html

[15] Schneider, E. A. (Institute of Defense Analysis). 2001. Measurements of system security. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[16] ISC. 2001. International Information Systems Security Certification Consortium, Inc. http://www.isc2.org

[17] Sandia National Laboratories. 2001. About IDART. http://www.sandia.gov/idart/about.htm

[18] Air Force Communication Agency (AFCA). 1997. Information Protection metrics and measurements program. http://afpubs.hq.af.mil/pubfiles/af/33/afi33-205/afi33-205.pdf

[19] Downs, D. D. and R. Haddad. (The Aerospace Corporation). 2001. Penetration testing – The gold standard for security rating and ranking. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[20] Wood, B. J. and J. F. Bouchard. (Cyber Defense Research Center, SRI). 2001. Red team work factor as a security measurement. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[21] Moitra S.D. and S. L. Konda. (SEI). 2000. The survivability of network systems: An empirical analysis. CMU/SEI-2000-TR-021. http://www.sei.cmu.edu/publications/documents/00.reports/00tr021.html

[22] Kahn, J. J. (The MITRE Corporation). 2001. Certification of Intelligence Community Systems and measurement of residual risks. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VR from May 21-23, 2001.

[23] Bayuk, J. (ITG Security, Bear Stearns & Co. Inc.). 2001. Measuring security. In proceedings of the Workshop on Information-Security-System Rating and Ranking held in Williamsburg, VA, May 21-23, 2001.

[24] Lindqvist U. and E. Jonsson. 1997. How to systematically classify computer security intrusions. In Proceedings of the 1997 IEEE Symposium on Security and Privacy held in Oakland, CA from May 4-7, 1997.

[25] Vaughn, R. and Henning, R. "A Report from the Workshop on Information Security System Rating and Ranking" Proceedings of the 14th Annual Canadian Information Technology Security Symposium, Ottawa Canada, May 13-17, 2002.