Report from Dagstuhl Seminar 14491

Socio-Technical Security Metrics

Edited by

Dieter Gollmann[1], Cormac Herley[2], Vincent Koenig[3], Wolter Pieters[4], and Martina Angela Sasse[5]

1   TU Hamburg-Harburg, DE, `diego@tu-harburg.de`
2   Microsoft Research, Redmond, US, `cormac@microsoft.com`
3   University of Luxembourg, LU, `vincent.koenig@uni.lu`
4   TU Delft & University of Twente, NL, `w.pieters@tudelft.nl`
5   University College London, GB, `a.sasse@cs.ucl.ac.uk`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 14491 "Socio-Technical Security Metrics". In the domain of safety, metrics inform many decisions, from the height of new dikes to the design of nuclear plants. We can state, for example, that the dikes should be high enough to guarantee that a particular area will flood at most once every 1000 years. Even when considering the limitations of such numbers, they are useful in guiding policy. Metrics for the security of information systems have not reached the same maturity level. This is partly due to the nature of security risk, in which an adaptive attacker rather than nature causes the threat events. Moreover, whereas the human factor may complicate safety and security procedures alike, in security this "weakest link" may be actively exploited by an attacker, such as in phishing or social engineering. In order to measure security at the level of socio-technical systems, one therefore needs to compare online hacking against such social manipulations, since the attacker may simply take the easiest path. In this seminar, we searched for suitable metrics that allow us to estimate information security risk in a socio-technical context, as well as the costs and effectiveness of countermeasures. Working groups addressed different topics, including security as a science, testing and evaluation, social dynamics, models and economics. The working groups focused on three main questions: what are we interested in, how to measure it, and what to do with the metrics.

1 Executive Summary

Dieter Gollmann
Cormac Herley
Vincent Koenig
Wolter Pieters
Martina Angela Sasse

Introduction

Socio-technical vulnerabilities

Information security, or cyber security, is not a digital problem only. Humans have been termed "the weakest link", but also physical access plays a role. Recent cyber attacks cleverly exploit multiple vulnerabilities of very different nature in the socio-technical systems that they target. For example, the StuxNet attack relied both on Industrial Control System (ICS) vulnerabilities and on the physical distribution of infected USB sticks, allowed by the business processes in the target facilities [8]. With new developments such as cloud computing, the attack surface of the systems only increases, and so do the options for potential attackers. At any company in the service supply chain, there may be malicious insiders or benevolent employees who fall victim to social engineering, and they influence the security of the system as a whole significantly. In order to compare and prioritize attacks and countermeasures, for example in terms of risk, the different types of vulnerabilities and threats need to be expressed in the same language. The seminar on "Socio-technical security metrics" aims at developing cross-domain metrics for this purpose.

Defining metrics

The idea of defining information security in terms of risk already appeared quite a while ago [2, 10]. Since then, many metrics have been proposed that aim to define attacks and attack opportunities in information systems in quantitative terms (see e.g. [7, 12]). Often, likelihood and impact of loss are mentioned as the key variables, from which risk can then be calculated. Furthermore, notions of vulnerability, difficulty, effort, cost, risk for the attacker, and many more, show up in the literature.

Even in a purely technical setting it is not always clear how all these different concepts are related. Still, including the human element forms a particular challenge, which deserves a separate event and a better integrated community. Too often it is thought that models of humans in the social sciences and models of technology are fundamentally incompatible. This inhibits progress on some very relevant questions: How does sending a phishing message compare to an SQL injection, in terms of the above mentioned variables? Or do we need additional notions in the technical models to express the human elements, or in the social science models to express the technical ones?

We thus need unified – or at least comparable – metrics that apply to all types of vulnerabilities. In order to represent socio-technical attacks, the key concepts need to apply to very different types of actions in an attack, including technical exploits and social engineering alike. This requires knowledge on technical infrastructures, social science, and actual incidents. To enable meaningful socio-technical security metrics, key features to be addressed in the seminar are outlined below.

Multi-step attacks

Cyber attacks, like StuxNet, tend to consist of multiple steps, combining technical and social or organizational vulnerabilities. Attack trees [17] are often used to represent possible multi-step attacks on systems, and they can be annotated with quantitative metrics. It has also been proposed to develop formal analysis techniques and simulations ("attack navigators") that generate such trees based on a model of the socio-technical system at hand [5, 16]. By defining methods to calculate metrics for attacks from metrics for steps, one can compare the attacks in terms of the metrics, e.g. difficulty. However, next to methods for prediction, one would also want to be able to estimate the relevant parameters for the model based on observed events. For example, if one observes a set of successful and unsuccessful attacks, what does that say about the difficulty of the steps involved, and how does that influence the prediction of possible future events? Statistical methods from social science may assist here [15].

Estimating metrics from data

Data is thus key to developing good metrics, but obtaining them requires care. Given the data that is typically available in organizations already, including enterprise architecture, network logs, and potentially even organizational culture, how to obtain the right metrics from that data? What could be the role of "Big Data" in improving security metrics? And how to acquire additional data in tailor-made experiments? From the modeling point of view, a distinction can be made here between bottom-up approaches, leveraging existing data, and top-down approaches, defining targeted data collection methods and experiments. A good example on the social side are the phishing studies by Jakobsson & Ratkiewicz [6]. On the technical side, intrusion detection systems may constitute an important source of data.

Attacker models

As security threats originate from attackers and not from nature, attacker models are key for security metrics [9]. Attackers will adapt their strategies to the security situation, and also to newly deployed countermeasures. We therefore need meaningful and measurable features of attackers that can be used as a basis for the metrics. For example, the motivation of an attacker may determine the goal of the attack, the resources available to an attacker may determine the number of attacks that he can attempt, and attacker skill may determine the likelihood of success. Costs of an attack as well as risk of detection influence attacker behavior [3]. Again, the theoretical and empirical basis of such models needs to be carefully studied, and (security) economics may provide important insights here.

Countermeasures

All these aspects come together in one final goal: supporting investments. In order to estimate the cost-effectiveness of security measures (also called ROSI, for return on security investment), one would need metrics for both the risk prevented by the countermeasures, and of their cost. The former could be calculated based on the properties discussed above. The latter, however, is far from trivial by itself, as costs not only involve investment, but also operational costs. Operational costs, in turn, may include maintenance and the like, but an important factor in the total cost of ownership is impact on productivity. Security features may increase the time required to execute certain tasks, and people have a limited capacity for complying with security policies. If security is too cumbersome or misdirected,

people will find workarounds, and this may reduce the effect of the measures on risk [1]. Thus, metrics for countermeasure cost form an important topic in itself, requiring input from the human factors and usable security domains.

Another application area for the metrics would be selection among alternative system designs. For example, if two vendors offer the same equipment or service, but one is much cheaper, how to take security risk into account when making this decision? Both vendors as well as customers may be interested in security metrics from this point of view. However, metrics would need to be designed carefully in order to avoid creating perverse incentives, tweaking systems to score high on the metrics without actually being "better".

Communities

In order to develop meaningful metrics for socio-technical security, participants from the following communities were invited:
- Security metrics and data-driven security, for obvious reasons;
- Security risk management, to provide input on suitable risk variables to be included;
- Security economics, to build upon economic theories of behavior of both attackers and defenders;
- Security architectures, to get relevant data on information system architecture and incidents;
- Formal methods, to analyze attack opportunities in complex systems;
- Social / crime science, to understand attacker behavior and the influence of controls;
- Human factors, to understand the impact of security controls on users.

Main findings

Paraphrasing some ancient philosophical questions (what is there, what can we know, what should we do), we can structure the main outcomes of this seminar as follows:
1. What properties are we interested in?
2. What can we measure?
3. What should we do with the measurements?

What properties

One of the main outcomes of the seminar is a much better view on which types of security metrics there are and for which purposes they can be used.

This leads to a distinction between metrics that exclude the real-life threat environment (type I) and metrics that include the real-life threat environment (type II). Metrics describing difficulty or resistance are typically of type I. They give a security metric that is independent of the actual activity of adversaries, or of the targets that they might be after. For example, which percentage of the people fall for a simulated phishing mail. This is similar to what Böhme calls "security level" [4]. The threat environment is often specified explicitly in such metrics, and the metrics may thus enumerate threat types. However, they do not estimate their occurrence rates, and in fact the occurrence rate is often controlled. In the phishing case, the researchers control the properties and occurrence of the phishing e-mails, and describe the e-mail (controlled threat) in their results.

Metrics describing loss (risk) or incidents are typically of type II. They describe undesired events that happen based on interaction of the system with a threat environment (activity of

adversaries), and their consequences. For example, the number of infected computers of a particular Internet Service Provider [18].

An illustration of this difference is the following. Consider two systems, system A and system B [13]. In system A, a locked door protects € 1,000. In system B, an identical locked door protects € 1,000,000. Which system is more secure? Or, alternatively, which door is more secure? One might say that system A is more secure, as it is less likely to be attacked (assuming the attacker knows the system). On the other hand, one might say that the doors are equally secure, as it is equally difficult to break the lock. The former argument is based on including an evaluation of the threat environment, the latter on excluding it.

Obviously, when trying to derive type II metrics from type I metrics, one needs metrics on the threat environment as well. For example, when one wants to calculate risk related to phishing attempts, and one knows how likely one's employees are to fall for phishing mails based on their sophistication, then one also needs information on the expected frequency of phishing mails of certain levels of sophistication in order to calculate the risk. Such models of the threat environment may be probabilistic or strategic (game-theoretic), representing non-adaptive and adaptive attackers, respectively. Probabilistic models, in turn, may be either frequentist (based on known average frequencies) or Bayesian (based on subjective probabilities). The various points of view have not been fully reconciled up to this point, although integration attempts have been made [14].

Another consideration is the integration of security metrics from different domains: digital, physical and social. Often, there are different styles of type I metrics, which one would like to integrate in a single type II metric representing the level of risk in a socio-technical system (e.g. an organization). Digital metrics may represent difficulty as required skill (e.g. CVSS), physical metrics may use required time (e.g. burglar resistance), and social metrics may use likelihood of success (e.g. likelihood of success of phishing attempts). Integration of these metrics is still an open challenge.

What measurements

The seminar discussed methods applied in different scientific communities for measurement purposes. Some of those methods rely on quantitative indicators, some rely on qualitative indicators, and some combine both. A further distinction can be made between subjective and empirical metrics, e.g. expert judgements versus monitoring data. Hereafter, and for the purpose of illustration, we have drawn a non-comprehensive list of such methods. They can be applied individually or in a complementary way, covering one measure or combined measures. A specific usage we consider underrepresented so far is the combination of methods in an effort to augment the measurement quality, or to provide information about the validity of a new measure. This approach has often been referred to, during the seminar, as triangulation of measures.

These are social methods discussed in the seminar:
- semi-structured interviews; in-depth interviews; surveys;
- observations of behavior;
- critical incident analysis;
- laboratory experiments; field experiments;
- expert / heuristic analysis / cognitive walkthrough;
- root cause analysis.

These are technical methods discussed in the seminar:
- security spending;

- implemented controls;
- maturity models;
- incident counts;
- national security level reports;
- service level agreements.

It is important to assess which type of metric (type I or type II) is produced by each of the techniques. For example, penetration testing experiments produce type I metrics, whereas incident counts produce type II. Maturity models and national security level reports may be based on a combination of type I and type II metrics. In such cases, it is important to understand what the influence of the threat environment on the metrics is, in order to decide how the metrics can be used.

What usage

Security metrics can contribute to answering questions about a concrete system or questions about a design (hypothetical system), and questions about knowledge versus questions about preferences. Here, we focus on a simpler distinction, namely between knowledge and design questions. In the case of knowledge questions, metrics are used to gather information about the world. In the case of design questions, metrics are used to investigate a design problem or to evaluate the performance of a design, such as a security control. In terms of knowledge questions, a typical usage discussed is a better understanding of the human factor in security. In terms of design, possible questions are how much security feedback a system should give to users or operators, or how to provide decision support for security investment.

Security metrics may have several limitations. In particular, many metrics suffer from various forms of uncertainty. It may be unclear whether the metrics measure the right thing (validity). Even if this is the case, random variations may induce uncertainty in the values produced (reliability). It is therefore important to understand the implications of such uncertainties for decisions that are made based on the metrics. Triangulation may contribute to the reduction of uncertainty. In some cases, quantitative metrics may not be possible at all, and qualitative methods are more appropriate.

Another limitation is that stakeholders may behave strategically based on what they know about the metrics (gaming the metrics). If stakeholders are rewarded when their security metrics become higher, they may put effort into increasing the metrics, but not "actual security". Even if the metrics are valid under normal circumstances, this needs not be the case under strategic behavior.

Conclusions

Security is difficult to measure, which should not be a surprise to those involved. However, to understand security in today's complex socio-technical systems, and to provide decision support to those who can influence security, rigorous conceptualisation, well-defined data sources and clear instructions for use of the metrics are key assets. This seminar laid the foundations for understanding and applying socio-technical security metrics.

In particular, we strove for clarity on (a) the different types of security metrics and their (in)compatibility, (b) the different sources and methods for data extraction, and (c) the different purposes of using the metrics, and the link with types, methods and sources. Several papers are planned as follow-up activities, as described in the reports of the working groups (Section 4). On many topics there are different views, which may not always be compatible, as was clear from the panel discussion (Section 5). Future follow-up seminars would be very valuable to address the open problems (Section 6).

References

 1 A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: Managing security behaviour in organisations. In *Proc. of the 2008 Workshop on New Security Paradigms*, NSPW'08, pp. 47–58, New York, NY, USA, 2008. ACM.

 2 B. Blakley, E. McDermott, and D. Geer. Information security is information risk management. In *Proc. of the 2001 New Security Paradigms Workshop*, pp. 97–104, New York, NY, USA, 2001. ACM.

 3 A. Buldas, P. Laud, J. Priisalu, M. Saarepera, and J. Willemson. Rational choice of security measures via multi-parameter attack trees. In *Critical Information Infrastructures Security*, volume 4347 of *LNCS*, pp. 235–248. Springer, 2006.

 4 R. Böhme. Security metrics and security investment models. In Isao Echizen, Noboru Kunihiro, and Ryoichi Sasaki, editors, *Advances in Information and Computer Security*, volume 6434 of *LNCS*, pp. 10–24. Springer, 2010.

 5 T. Dimkov, W. Pieters, and P. H. Hartel. Portunes: representing attack scenarios spanning through the physical, digital and social domain. In *Proc. of the Joint Workshop on Automated Reasoning for Security Protocol Analysis and Issues in the Theory of Security (ARSPA/WITS'10)*, volume 6186 of *LNCS*, pp. 112–129. Springer, 2010.

 6 P. Finn and M. Jakobsson. Designing ethical phishing experiments. *Technology and Society Magazine, IEEE*, 26(1):46–58, 2007.

 7 M. E. Johnson, E. Goetz, and S. L. Pfleeger. Security through information risk management. *IEEE Security & Privacy*, 7(3):45–52, May 2009.

 8 R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *Security & Privacy, IEEE*, 9(3):49–51, 2011.

 9 E. LeMay, M. D. Ford, K. Keefe, W. H. Sanders, and C. Muehrcke. Model-based security metrics using adversary view security evaluation (ADVISE). In *Proc. of the 8th Int'l Conf. on Quantitative Evaluation of Systems (QEST'11)*, pp. 191–200, 2011.

10 B. Littlewood, S. Brocklehurst, N. Fenton, P. Mellor, S. Page, D. Wright, J. Dobson, J. McDermid, and D. Gollmann. Towards operational measures of computer security. *Journal of Computer Security*, 2(2–3):211–229, 1993.

11 H. Molotch. *Against security: How we go wrong at airports, subways, and other sites of ambiguous danger.* Princeton University Press, 2014.

12 S. L. Pfleeger. Security measurement steps, missteps, and next steps. *IEEE Security & Privacy*, 10(4):5–9, 2012.

13 W. Pieters. Defining "the weakest link": Comparative security in complex systems of systems. In *Proc. of the 5th IEEE Int'l Conf. on Cloud Computing Technology and Science (CloudCom'13)*, volume 2, pp. 39–44, Dec 2013.

14 W. Pieters and M. Davarynejad. Calculating adversarial risk from attack trees: Control strength and probabilistic attackers. In *Proc. of the 3rd Int'l Workshop on Quantitative Aspects in Security Assurance (QASA)*, LNCS, Springer, 2014.

15 W. Pieters, S. H. G. Van der Ven, and C. W. Probst. A move in the security measurement stalemate: Elo-style ratings to quantify vulnerability. In *Proc. of the 2012 New Security Paradigms Workshop*, NSPW'12, pages 1–14. ACM, 2012.

16 C. W. Probst and R. R. Hansen. An extensible analysable system model. *Information security technical report*, 13(4):235–246, 2008.

17 B. Schneier. Attack trees: Modeling security threats. *Dr. Dobb's journal*, 24(12):21–29, 1999.

18 M. J. G. Van Eeten, J. Bauer, H. Asghari, and S. Tabatabaie. The role of internet service providers in botnet mitigation: An empirical analysis based on spam data. OECD STI Working Paper 2010/5, Paris: OECD, 2010.

2   Table of Contents

3   Overview of Talks

3.1   Metrics for Security Awareness?

Zinaida Benenson (Universität Erlangen – Nürnberg, DE)

The usefulness of measures for raising security awareness in organizations and for the general public is controversially discussed in the current IT security research and practice. The differences in opinions range from publishing detailed guidelines for panning and conducting security awareness campaigns to reasoning that most security awareness measures are pointless. Measuring the effectiveness of security awareness interventions is an important tool for resolving this debate. Unfortunately, approaches from the computer science and information systems literature are not sufficiently well developed to fulfill this task. Moreover, the state of the art does not clearly define security awareness, which makes measuring anything connected to this concept even more difficult, if not impossible.

An attempt to characterize the existing security awareness definitions according to three orthogonal dimensions "Knowledge about threats", "Knowledge about protection mechanisms" and "Behavior" is presented in this talk. Its purpose is to understand what security awareness actually means and what is missing in the current research on this topic. A preliminary version of this systematization can be found in the joint work with Norman Hänsch [1].

References
1   Hänsch, Norman and Benenson, Zinaida. *Specifying IT Security Awareness*. 1st Workshop on Security in Highly Connected IT Systems (SHCIS), collocated with 12th International Conference on Trust, Privacy, and Security in Digital Business (TrustBus), IEEE, 2014

3.2   The National Role of CS Metrics

Kas P. Clark (Ministry of Security and Justice – The Hague, NL)

The Dutch government needs the help of the research community to develop better, quantitative cyber security metrics for use at the national level. What data do we need and how can we combine it together to form coherent, relevant cybersecurity indicators?

3.3   Normative Security

Simon N. Foley (University College Cork, IE)

The increasing scale and complexity of modern computer systems means that the provision of effective security is challenging, as well as being prohibitively expensive. Consequently, security tends to regulate those activities perceived to be critical, with the assumption that other unregulated activities, whether known or unknown, are not of significance. An added complication with security regimes that are overly strict, is that such unregulated activities can become the means of getting things done in the system. However, the difficulty is that these side-activities often lead to the compromise of security in a system. While security controls may provide monitoring and enforcement of the critical activities related to the security policy, little may be known about the nature of the other activities.

Normative security seeks to view a system as a society in which security is achieved by a combination of legislative provisions and normative behaviors. Drawing solely on legislative provisions is insufficient to achieve a just and orderly society. Similarly, security regimes that focus solely on security policies and controls are insufficient. Our position is that systems have analogous normative behaviors – behavioral norms – whereby the security of a system is based not only on the regulation of what is perceived to be its security critical activities, but also on the orderliness of its unregulated activities.

Using this analogy we are exploring how current theories about social norms in society can provide insight into using normative behavior in systems to help achieve security. We are investigating how these behavioral norms, representing potentially unknown side-activities, can be revealed by mining detailed system logs. The assumption is that, absent other information, adherence to past normative behavior can be taken as some indication of continuing orderliness. However, we note that these behavioral norms can be used to gauge the order or disorder in a system and, therefore, adherence to past normative behavior may also indicate a continuation of disorderliness

3.4   Socio-Technical Security Metrics

Aleksandr Lenin (Technical University – Tallinn, EE)

The talk outlines the socio-technical metrics used by the so-called "Failure-Free" models for quantitative security analysis, describes the problems obtaining quantitative input data from expert estimations, as well as suggests approaches that may be used to deal with the complexities of socio-technical security metrics.

References

1   Buldas, A., Lenin, A.: *New efficient utility upper bounds for the fully adaptive model of attack trees.* In Decision and Game Theory for Security – 4th International Conference,

GameSec 2013, Fort Worth, TX, USA, November 11–12, 2013. Proceedings, pages 192–205, 2013.

2   Jürgenson, A., Willemson, J.: *On fast and approximate attack tree computations*. In IS-PEC, pages 56–66, 2010.

3   Lenin, A., Buldas, A.: *Limiting adversarial budget in quantitative security assessment*. In Decision and Game Theory for Security – 5th International Conference, GameSec 2014, Los Angeles, CA, USA, November 6–7, 2014. Proceedings, pages 153–172, 2014.

4   Lenin, A., Willemson, J., Sari, D. P.: *Attacker Profiling in Quantitative Security Assessment Based on Attack Trees*. In Simone Fischer-Hübner and Karin Bernsmed, editors, 19th Nordic Conference on Secure IT Systems, NordSec 2014, Tromsø, Norway, October 15–17, 2014. Proceedings, volume 8788 of Lecture Notes in Computer Science, pages 199–212. Springer, 2014.

5   Pieters, W., Hadziosmanovic, D., Lenin, A., Montoya, L., Willemson, J.: *Poster Abstract: TREsPASS: Plug-and-Play Attacker Profiles for Security Risk Analysis*. In Proceedings of the 35th IEEE Symposium on Security and Privacy, 2014. Poster and Extended Abstract.

3.5   Attack Trees and Socio-Technical Trees

Sjouke Mauw (University of Luxembourg, LU)

In this presentation I sketched two tree-based modelling formalisms: attack trees and socio-technical trees. I briefly highlighted their syntax, semantics and pragmatics.

3.6   Security-Related Behavior and Economics

Frank Pallas (KIT – Karlsruher Institut für Technologie, DE)

Main reference F. Pallas, "An Agency Perspective to Cloud Computing," in J. Altmann, K. Vanmechelen, O.F.
        Rana (eds.), "Economics of Grids, Clouds, Systems, and Services – Proc. of 11th Int'l Conf.
        GECON 2014," LNCS, Vol. 8914, pp. 36–51, Springer, 2014.
        URL http://dx.doi.org/10.1007/978-3-319-14609-6_3

The application of economic theories and concepts to the field of information security has led to important findings and insights throughout the past years. In particular, the role of the yearly Workshop on the Economics of Information Security [1] deserves explicit mention here.

From an economic perspective, achieving better information security does in most cases require cooperation between different players who pursue different goals. This cooperation, in turn, is hallmarked by information asymmetries, externalities and therefore often counter-productive incentives that lead to unfavourable security outcomes for all involved parties. In particular, this is the case when ultimate security-related decisions and activities are delegated from one party to another one which is assumed to have better capabilities and/or better situational knowledge allowing for more appropriate outcomes. This does, for example, apply to all security instruments focused on individual users as well as to most scenarios of cloud computing.

---

[1] See http://econinfosec.org

As laid out in the talk, economic agency theory provides valuable insights on the fundamental characteristics shaping such settings, the respective conflicts of interests and the reasonableness of different countermeasures. Socio-technical security metrics, in turn could be employed to diminish agency-related inefficiencies. In particular, they could in the future play an important role in the context of signalling (e.g. audit certificates), screening (e.g. inspections), and monitoring.

3.7 Comparison of Cloud Provider Security

Sebastian Pape (TU Dortmund, DE)

Suppose, you want to start a new service and have the task of selecting a cloud provider. How do you determine which one is most secure? How do you decide which data is helpful for the selection? There already exist some approaches, but each of them (more or less) has its own utility function. How do you decide which approach returns the best results for ranking / comparison? Obviously the solution depends on the requirements of the tenant. Is it possible to come up with a 'requirement independent' ranking?

3.8 Metrics for Security Behaviour in Organisations

Simon Parkin (University College London, GB)

In this short presentation I discuss monitoring of security behaviour in organisations, looking at user compliance and non-compliance, and the appropriateness of the security implementation for users and the business. I then discuss directions for measurement, including articulating incentives and costs for organisations to measure security behaviour, and the approachability and packaging of socio-technical expertise for practitioners and business in education and tools.

3.9 Metrics in social engineering experiments

Wolter Pieters (TU Delft, NL)

In social science, experimental setups can provide information on the risk associated with attack steps that involve social engineering, i.e. the manipulation of people. Typically, the threat environment is controlled by executing carefully scripted actions, which may involve phishing e-mails but also real-life interaction. One example of such an experiment is the "persuasion and security awareness experiment", in which we measured the success rate of obtaining physical credentials from university employees. It was found that a combined

awareness intervention was effective in reducing this success rate. Other possible metrics may include the time taken until success, or the stage in which the attack succeeds (if the script supports multiple stages). In this way, social science experiments with controlled threat environments can provide information on the difficulty of social engineering attack steps, and the effect of interventions. Because the threat environment is controlled, the metrics obtained are independent of actual attacker activity, which is not the case in studies that measure actual victimisation.

3.10   Metrics for Security of Cooperating Systems

Roland Rieke (Fraunhofer SIT – Darmstadt, DE)

Systems of systems that collaborate for a common purpose are called cooperating systems. They are characterised by freedom of decision and loose coupling of their components. Typical examples of cooperating systems are electronic health systems, vehicular ad hoc networks, distributed air traffic management systems, telephone systems, and electronic money transfer systems.

In this talk, three problems with respect to security metrics for cooperating systems have been addressed, namely, (1) abstract representation of security information, (2) security information quality, and (3) elicitation, linkage, and management of security information.

References
1   Roland Rieke. Abstraction-based analysis of known and unknown vulnerabilities of critical information infrastructures. *International Journal of System of Systems Engineering (IJSSE)*, 1(1/2):59–77, 2008.
2   Roland Rieke, Luigi Coppolino, Andrew Hutchison, Elsa Prieto, and Chrystel Gaber. Security and reliability requirements for advanced security event management. In Igor Kotenko and Victor Skormin, editors, *Computer Network Security*, volume 7531 of *Lecture Notes in Computer Science*, pp. 171–180. Springer Berlin Heidelberg, 2012.
3   Roland Rieke and Zaharina Stoynova. Predictive security analysis for event-driven processes. In *Computer Network Security*, volume 6258 of *LNCS* (pp. 321–328). Springer.
4   Roland Rieke, Jürgen Repp, Maria Zhdanova, and Jörn Eichler. Monitoring security compliance of critical processes. In *Parallel, Distributed and Network-Based Processing (PDP), 2014 22th Euromicro International Conference on*, pp. 525–560. IEEE Computer Society, Feb 2014.
5   Roland Rieke, Julian Schütte, and Andrew Hutchison. Architecting a security strategy measurement and management system. In *Proceedings of the Workshop on Model-Driven Security*, MDsec'12, pp. 2:1–2:6, New York, NY, USA, 2012. ACM.

3.11 Three challenges with respect to measurement from a risk perspective

Ketil Stolen (SINTEF – Oslo, NO)

One challenge is the gap between the data or information required by methods and tools for risk analysis put forward in main stream academic publications and the data available in practice.

A second challenge is the communication of risk relevant information among humans. What scales are best suited for what purpose? In particular, how should we measure likelihood and uncertainty?

A third challenge is the validation of risk models. How to determine that a risk model is sufficiently reliable?

3.12 Ideas for Socio-Technical Security Metrics

Axel Tanner (IBM Research GmbH – Zürich, CH)

To induce discussions, four ideas for potential socio-technical security metrics are presented:

- Graph based: as many models used or proposed for security analysis, e.g. in TREsPASS, are based on graph structures, with additional data on nodes and edges, could we use graph characterising parameters, like connectivity or centrality, possibly in combination with data like 'value' on nodes and 'resistance' on edges to build and define security relevant metrics?
- Coverage based: many processes and operations happen in every organisation – can we measure what part of these is covered by operations and automated security policies? Or what part is covered by information flowing into tamper-proof log files?
- Reality gap: out of the security policies covering processes and operations in an organisation – how many and to which degree are these actually fulfilled in reality?
- Time to detect: in case of a breach of a security policy – how long will it take to detect this non-compliance?

3.13 Susceptibility to Social Engineering

Sven Übelacker (TU Hamburg-Harburg, DE)

In my short presentation I discussed my research on factors influencing the susceptibility to social engineering attacks which I try to categorise via existing research. Beside factors like socio-demographics, knowledge, impulsiveness, or stressors, I focused on the question: How big is the impact of personality traits on this susceptibility? I talked about my ongoing work on a scenario-based social engineering questionnaire including many of the aforementioned factors.

3.14 How should we measure implementation complexity?

Jan Willemson (Cybernetica AS – Tartu, EE)

One of the weakest points in computer security are the implementations. The amount of potential mistakes correlates with complexity of the application. Should we acknowledge application complexity as one source of insecurity? Should we design a measure for this?

Some examples of implementation complexity:

- Highway speed limits do not guarantee the globally optimal outcome (e.g. that the total time needed for everyone to get home is minimal), but they have a virtue of being easy to follow and easy to verify.
- The definition of safe elliptic curves by Dan Bernstein and Tanja Lange includes several criteria that are designed to minimize the risk of getting the implementation wrong.

3.15 Playing poker for fun, profit and science

Jeff Yan (Newcastle University, GB)

I propose to use poker as a new instrument for studying the psychology of deception, which is fundamental to many security and cybercrime problems such as social engineering. Poker enables the studies of a wide range of deceptive behaviours, and in these settings, observable, measurable and computable metrics are often available. Moreover, poker offers better ecological validity than trust games that have been widely used in economics studies.

I also explore how to inform cyber security with poker research, and discuss experiments designed for this purpose.

4 Working Groups

4.1 Models, Economics and Threats – Working Group Report

Tristan Caulfield, Kas Clark, Trajce Dimkov, Carrie Gates, Cormac Herley, Mass Soldal Lund, Sjouke Mauw, Roland Rieke, and Jeff Yan

In the cyber security domain, policy makers in both the public and private sectors make decisions regarding which project to fund, which legislation to propose and how to increase the overall resilience of their respective society or company given their finite resources. These decisions are made based on the best information available that given moment. Generally speaking, these decisions are based on qualitative metrics, such as expert or public opinion.

Some policy makers, including the Dutch Parliament, have officially asked that the existing metrics are supplemented with quantitative metrics. The underlying assumption is that quantitative metrics are more reliable as they are impartial and less susceptible to

anecdotal evidence. This working group is interested in exploring the available metrics and creating a framework to organize and evaluate them. To this end, this working group will: (1) identify relevant socio-technical security metrics, (2) estimate the desired properties of these metrics and (3) define a taxonomy to organize and correlate these metrics.

4.2 Social Dynamics Metrics – Working Group Report

Zinaida Benenson, Sören Bleikertz, Simon N. Foley, Carlo Harpes, Stewart Kowalski, Gabriele Lenzini, Daniela Oliveira, Simon Parkin, Shari Lawrence Pfleeger, Paul Smith, and Sven Übelacker

Introduction

Individuals continually interact with security mechanisms when performing tasks in everyday life. These tasks may serve personal goals or work goals, be individual or shared. These interactions can be influenced by peers and superiors in the respective environments (workplace, home, public spaces), by personality traits of the users, as well as by contextual constraints such as available time, cognitive resources, and perceived available effort.

All these influencing factors, we believe, should be considered in the design, implementation and maintenance of good socio-technical security mechanisms. Therefore, we need to observe reliable socio-technical data, and then transform them into meaningful and helpful metrics for user interactions and influencing factors.

More precisely, there are three main questions that the group discussed:

1. What data do we need to observe and what of this data we actually can observe and measure?
2. How can we observe and measure?
3. What can we do with the results of the observations?

What do we need to (and can) observe?

General data and metrics for individuals and groups

The discussion converged towards the idea of observing elements of behavior, not knowledge or attitudes, as the latter are not considered reliable indicators of security-related behavior. These observations can focus on behavior at an individual or group level.

Additionally to observing behavioral elements, e.g. patterns, we need to understand which factors influence people's behavior and trigger actions, and how to measure them. Among possible factors are personality traits, including irrationality and heuristics in decision-making. For example, deception susceptibility (or resilience) is an important personality trait that has been studied in the psychology of persuasion. The group also discussed measuring moral dimensions and risk perception. Other possible metrics include habits and cognitive abilities.

Collecting socio-demographic data such as age and sex is also important in order to know how these characteristics relate to other observable data.

Data and metrics for organizations

In the context of organizations, the group discussed what data and metrics can be used to indicate, estimate and quantify the security culture of the organization, its behavioral norms, risk perception (from the organization point of view), national differences, and the level of commitment of the group and individuals to the organization's goals. Metrics relating to individuals' capacity to expend effort for organization security without a perception of personal benefit (their 'compliance budget' is also important, as the employees can experience a number of draws from the security mechanisms on their available effort).

Data and metrics for adversaries

In the adversary domain, the group discussed metrics on attackers' risk perception of the possibility of being caught. Other possible metrics are organizational defense strengths as perceived by the attackers and attack resource costs (time, money and cognitive effort).

Data and metrics for employees

Collecting data on employees' privilege levels and roles is important, as this information helps to identify potentially dangerous deviations in behavior. Regarding employee activity, especially in the context of insider attacks, important metrics discussed were artifact collection rate per employee (number of artifacts per hour, week, and day), number and size of files transferred (flash drive, other local machines, remote machines), and number of artifacts printed.

Unintentional mistakes (such as accidentally printing out a document that is not allowed to be printed) or intentional workarounds in cases where security measures are perceived as impediments to task execution (such as sharing of login credentials) can mislead inferences about the prevalence of malicious insider behavior, indicating misconfigured systems and unusable security mechanisms instead. Therefore, it is important to develop metrics that can reduce false positives and lead to adjustments of security mechanisms. These metrics are especially related to organizational culture in terms of learning from mistakes, how mistakes are treated and reported, and also to individual metrics such as level of commitment and risk perception.

Limitations and future work

The group did not have enough time to discuss metrics for the general public and society, and also for special groups such as software developers, system administrators or managers, leaving these metrics to future work.

How can we observe and measure?

Most of the mentioned data can be collected using qualitative and quantitative methods from social sciences and psychology, although some data have technical nature (such as artifact collection rates). Quantitative methods include field and laboratory experiments, large-scale observations of behavior and carefully designed surveys, whereas qualitative methods include semi-structured interviews and in-depth observations (e.g., ethnography). Quantitative methods can be used to collect descriptive statistics as well as to test hypotheses.

Researchers and practitioners should pay attention to the constraints and limitations of the respective methods, such as external and internal validity and generalizability. Observations in organizations are challenging because it will usually take time before a relevant number of

events is collected. However, this approach represents probably the most natural means of measuring security behaviors.

What can we do with the results of the measurements?

Good social dynamics metrics support decision-making in an organization, improve its security processes and promote visibility of security policies. They also help with the communication of security needs at the high level of the organization, thus influencing the company's security budget.

Provide appropriate security

Security provisioning over time is important – stable metrics can allow baseline measurement of security before changes are made. This allows managers and providers to objectively measure elements of behavior over time to determine if end-users are being adequately supported.

Communicate human factors evidence

Social dynamics metrics can provide a common language that has the potential to support engagement with technology-oriented security researchers and practitioners. This common language would communicate the value of considering human factors in the design and implementation of future security solutions. Further, this common language would help to better frame the expectations and requirements for security training programs and security policies within organizations. We need both, social and technical metrics, as only a combination of them can provide enough supporting evidence for the design of better security processes.

Understand the appropriation of security

Social dynamic metrics also help discovering optimal levels of feedback about the security state of a system and of the control that the users can and should have over the security means. One possibility is the personalization of user engagement in security depending on their personality traits and experience (at the individual or per-task level, depending on the qualities of a task or group of tasks). Some people may wish to defer choices about security to the technology and receive minimal feedback (we call this *black box security*), whereas some other people may wish to have a lot of control and detailed feedback (*white box security*).

Next steps

The group discussed the importance of studying metrics for specific domains and producing a generalized framework for social security metrics, as some metrics will be valid for several domains or perspectives. A subset of the working group agreed to continue research within this area and to consolidate findings towards producing publications that support researchers and practitioners. This group includes Zinaida Benenson, Carlo Harpes, Stewart Kowalski, Gabriele Lenzini, Daniela Oliveira, Simon Parkin, and Sven Übelacker.

4.3   Testing, Evaluation, Data, Learning (Technical Security Metrics) – Working Group Report

Rainer Böhme, Michel Van Eeten, Simon Foley, Dina Hadžiosmanović, Aleksandr Lenin, Sebastian Pape, and Wolter Pieters

Questions and objectives

The WG session started by brainstorming potential research questions around the topics of security evaluation and testing using security metric. Some of the questions were:

- What are different types of (technical) security metric?
- What kind of outcomes can we expect using different types of security metric?
- What kind of metric can be used for evaluation/testing purposes?
- What kind of data is required for using specific types of security metrics?

The WG then focused on two concrete objectives: *(i)* identify different dimensions to characterise a security metric and *(ii)* discuss the existing metrics with respect to the identified dimensions.

Properties of security metric

Böhme [2] presents a framework characterising security levels with different indicators mapped across *the level of abstraction* (as concrete or abstract) and *the amount of probabilistic nature in measuring* (as deterministic or probabilistic). The framework represented an excellent starting point for the WG discussion on different types of security metrics. For example, *security spending* represents an abstract but deterministic measure of security investment (i.e., as it represents the total spending). By contrast, specific *protection measures* represent concrete and deterministic measure (i.e., as they provide concrete technical checklists which can be directly related to security vulnerabilities). In this context, *incident counts* represent concrete yet probabilistic measure (i.e., as it reasons on the level of security based on the outcomes).

During the WG session, we introduced another aspect of the characterisation: *inclusion of threat environment.* More specifically, indicators like protection measures and penetration testing do not consider specific threat environment into the measure (as they mainly focus on the system itself). On the other hand, incident counting implicitly includes the specific threat environments (i.e., by referring to attackers and specific attack vectors).

Security metrics in practice

To understand how metrics used in practice map to the theoretical framework, the WG focused on discussing several approaches for measuring the security level:

Security service level agreement The agreements are typically used by security-service providers to indicate the scope and the character of the provided security service. Commonly, the agreements include guarantees on service availability (e.g., 99%), the maximum time for incidents response (e.g., 30 minutes) and repair time (e.g., 3 business days) [1].

Maturity models Organisations use maturity models to evaluate the overall level of security awareness and technological implementation of the organisation. Common approaches include maturity models like O-ISM3 [4], OpenSAMM [6] and BSIMM [3]. The models use

combinations of checklists to indicate the estimated level of security in the organisation or in software. For example, ''have a lightweight approach to risk classification and prioritization" corresponds to the first (lowest) maturity model regarding architecture design, while "build proactive security architecture" corresponds to the fourth (highest) maturity level in architecture design.

Government-driven security level assessment Different countries annually publish a general, nation-wide, report on the cyber security level. One such example is the Cyber Security Assessment in the Netherlands, published by National Cyber Security Centre (NSCS) [5]. As the input information, the report uses incidents across different industry and public domains to draw the threat landscape, current trends and predictions.

Observations

With respect to the security metric framework, the WG discussions on the existing security metrics resulted in the following observations:

- Security service level agreements and maturity models represent metrics which weakly include threat environment into consideration (by focusing on protection measures) while security assessment reports largely include the threat environment (by using incident statistics).
- Metrics which do not consider the threat environment focus on security controls (e.g., protection measures).
- Metrics which consider the threat environment focus on evaluating the existing security controls (e.g., incidents indicate the accuracy of protection measures).
- Risk distribution in metrics is directly related to the inclusion of the threat environment. For example, security metrics in service level agreements focus on specifying controls (e.g., response time), and avoiding risk of guaranteeing the level of attack impact.
- A desirable security metric should include indicators across the whole spectrum of measurements (i.e., w/ and w/o threat environment).

Follow up

As the follow up activity, the participants of the WG agreed to revision the initial framework by introducing more practical examples and case studies, and exploring new possibilities for measuring the security level (e.g., measuring botnet/malware mitigation, measuring outcomes of penetration testing).

References

1 Avira. Service Level Agreement for Avira Managed Email Security of Avira. http://www.avira.com/en/service-level-agreement, 2011.

2 R. Böhme. Security metrics and security investment models. In I. Echizen, N. Kunihiro, and R. Sasaki, editors, *Advances in Information and Computer Security*, volume 6434 of *Lecture Notes in Computer Science*, pp. 10–24. Springer Berlin Heidelberg, 2010.

3 B. Chess and B. Arkin. Software security in practice. *IEEE Security & Privacy*, 9(2):89–92, 2011.

4 The Open Group. *Open Information Security Management Maturity Model (O-ISM3)*. Van Haren Publishing, 2011.

5 NCSC. Cyber Security Assessment Netherlands. Technical report, National Cyber Security Centre, 2014.

6 OWASP. Software Assurance Maturity Model.
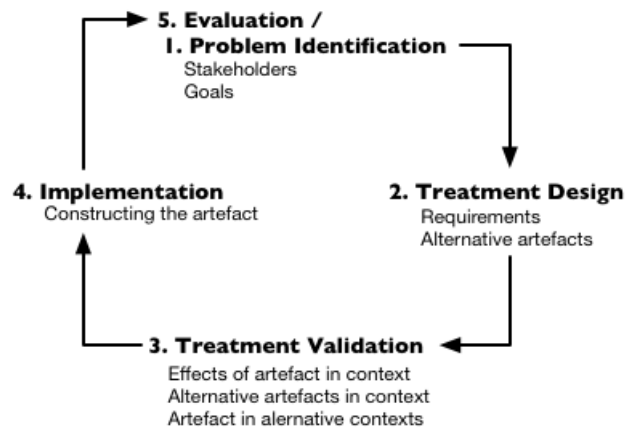
Figure 1 An illustration of the design cycle used in design science.

4.4   Security as a Science – Working Group Report

Roeland Kegel, Vincent Koenig, Frank Pallas, Kai Rannenberg, Ketil Stølen, and Axel Tanner

This working group initially focused on identifying the type of research that is done in computer science in general, in an attempt to answer the question *what kind of science does computer security as a research field practice?* The discussions that followed from this question led to the conclusion that security research involves answering both knowledge questions and design questions:

- **Knowledge questions** are the archetypical form of science designed to answer specific questions by way of experimentation. The added value of answering these questions comes in the form of knowledge about the world as it is.
- **Design questions** are challenges: A call to change the world by introducing a new artefact in a certain context. This is something designed to improve the context into which it is introduced, such as a faster search algorithm, a test methodology or a new software system.

To define a scope for the discussion that fits the aim of the seminar, we then focused on design science and the role it plays in answering the following questions: *what can we measure, how do we measure it, and what can we do with these measurements?* We found that we could map these questions to elements of the design cycle illustrated by Wieringa [1] (see Figure 1). The design cycle corresponds closely with the engineering cycle of investigation, design, implementation and evaluation.

Metrics and the Design Cycle

The design cycle uses metrics in different stages of the cycle. Below is a summary of these relationships:

- **Problem investigation**: In this stage, problems and stakeholders are identified. Completion of this step results in a clearly defined problem. As such, the question *what do we measure?* can be answered using the results of this step.

- **Treatment design**: In this stage, the solution (new artefact) is designed to solve or mitigate the problem. The result of this step is a blueprint of the solution. By considering this proposed solution, we can find guidelines on *how to measure*, since the implementation guides what measurements are possible/practical.
- **Implementation**: In this stage, the proposed solution is inserted into the context. This corresponds to an *application* of the chosen security metrics.
- **Evaluation**: Finally, in this stage we consider the effects of applying this solution to the context. Using the measurements performed in the previous step, we can now consider the question *what to do with these measurements*.

Having identified the relationship of these questions and the design cycle, we can now reason about the issues with using the design cycle with security.

Problems and Pitfalls when using the Design Cycle

We identified three problems with the use of this cycle within the context of security research:

- **Invalidation by anecdote**: Often, a proposed treatment for a problem is invalidated by anecdotes, the availability of which being random. As a result, the random absence of anecdotes (i.e., the absence of proof for invalidation) might be confounded with the proof that no such anecdote exists (i.e., the proof of absence of arguments that could invalidate a treatment). Systematic evidence, supported by metrics, should however be sought for invalidation: a single counter-example to a security measure will lead to redesign of the treatment and only the proof of absence of such counter-examples will validate a design.
- **Skipping problem identification**: After the proposed treatment is deemed unsatisfactory, the problem is often not reconsidered. The treatment is immediately adapted to incorporate a defense to the anecdote. However, such counterexamples might be indicative of an incorrect problem investigation. Care has to be taken not to skip this step without due thought being given to the matter.
- **Problem considered static in following cycles**: When an iteration of the design cycle is complete and problems are identified with the current solution, often the problem definition is not reconsidered. Subsequent iterations of the process should consider whether the gathered evidence suggests that the problem identification needs updating (because of changing requirements, or identified shortcomings).

We feel that these problems are typical in security research. Good socio-technical security metrics can offer valuable support to mitigate these problems, especially for systemising the invalidation by anecdote rather than relying on random availability of anecdotes.

Conclusions

We feel that metrics should play a larger role in supporting treatment validation, lowering the reliance on randomly available anecdotes to validate (often expensive to implement) treatments. Additionally, we feel that metrics can play a vital role in reassessing whether the solution has proven successful. Finally, we are interested in the question of whether the design cycle is an effective methodology to use in the *development* of these metrics, rather than just the evaluation. To this end, as future work, we intend to use a case study in order to further investigate the interactions between design science and security metrics.

References

1   R. J. Wieringa. The design cycle. In *Design Science Methodology for Information Systems and Software Engineering*, pages 27–34. Springer Berlin Heidelberg, 2014.

5 Panel Discussion

At the end of the seminar, a panel discussion was organized with the following participants:
- Sasse, Martina Angela (moderator)
- Gates, Carrie
- Herley, Cormac
- Pfleeger, Shari Lawrence
- Stølen, Ketil

The panel helped identify fundamental differences in metrics, as well as open problems. It discussed a variety of key points as represented hereafter.

Definitions

The first discussion focused on defining what we need when intending to measure human behaviour in a security context. The panel suggests defining
- what behaviours we can expect to see;
- what triggers behaviours;
- what the range of behaviours is;
- what behaviours we want to encourage or discourage;
- what the differences between individual and group behaviours are;
- what triggers for sharing are;
- what attitudes lead to what behaviours.

The panelists identify an additional challenge which is understanding when people want to be in control, and when they want 'to be taken care of' in terms of security.

Data and research methods

The second point of discussion regarded the difficulty to rely on the 'right' data and the right methods for producing such data. There is a gap between the data that is required, and what is available – one reason being that data capture techniques originate from safety and process industry, where capturing data is much simpler than in cyber security.

The panel focused on the difficulty of getting reliable data; they formulated the following problems and recommendations:
- Use metrics that are as explicit as possible;
- People collecting data need hands-on experience of risk analysis – this is currently often confused with requirements analysis;
- Predict risk level after changes have been implemented;
- Combine risk analysis with other techniques to check risk model;
- Use two risk models – before and after;
- Combine with other measures, e.g. vulnerability scans, to check predictions – program and functional testing.

The panel agreed that there are many ways of measuring risk, e.g. attack trees; the ISO 27000 2-factor measure of risk consequence and risk likelihood; or by quantifying the ability of threat – e.g. OWASP risk rating methodology.

Transition to practice

It is felt that research methods can contribute to gathering reliable data. The transfer from research to practice however is difficult and might be very slow. To illustrate how distant research and practice might sometimes be, the panel provides a set of statements meant to describe the "industry view", as opposed or distant to the research view:

- "Social metrics are hard and expensive, which is why we don't do it";
- "Security awareness – we assume that it works if people pass the test" (we want to believe it works);
- "Testing is hard and expensive to do" – technical responses are easy and cheap, and work 'well enough' – so people buy them – measuring staff 'grumpiness' is not easy and cheap;
- "We prefer capital expenditure to consultancy" – results need to be easy and cheap to measure;
- "It's very hard to resist a good test" – people test and measure what's easy to test and measure;
- "Standards drive adoption".

In addition, the following observations were made:
- 'Best practices' are not quickly updated;
- Gartner and other influencers have a lot of power – everybody wants to be 'best of breed & forward looking' quadrant;
- As far as socio-technical security metrics are concerned, the phrase "garbage in, garbage out" applies;
- The industry approach to measurement is insufficiently mature – it's a vacuum that research could fill;
- Honest and authoritative tests and criteria are needed.

The usage of metrics is another important point of discussion and the panelists feel that metrics and numbers are used to justify decisions already made. It is thus unsure *why* we want the measure. The answer ought to be: we should spend on things that have value for the system overall, not just to prevent something bad from happening, which is also the argument of Harvey Molotch [1]. We must combat exceptionalism – 'security is special' – as this seems to be an excuse for not working along scientific principles.

Also, we should not measure proxies or shortcuts. Examples:
- Studies on how many users give their passwords for a chocolate bar – these are numbers that don't tell us very much;
- Mechanical Turk (mTurk) studies: the composition of participant groups is often limited, and motivation of those who participate for very little money may be to complete the study as quickly as possible.

The panelists feel there are too many of this type of debatable study – and bad data drives out good, myths about user behaviour are perpetuated. These hide the need to consider if technology and policies are actually working.

Finally, the panel agrees on a series of concrete recommendations and take-home messages:
- Be honest about what you don't know;
- Throw out data that is not grounded, and start with what is left over;
- Triangulate your measurements or results with other metrics;
- Look at multiple metrics, especially context metrics, to understand what is causing changes in metrics;

- Relate your measurements to business metrics to understand cost and benefit of doing the measurements;
- Question how good the metric is. Are any of the insights actionable?

Conclusion

We need to work together to develop better studies, experimental paradigms, data collection and analysis techniques, and standards of proof and evidence.

References

1   H. Molotch *Against security: How we go wrong at airports, subways, and other sites of ambiguous danger.* Princeton University Press, 2014.

6   Open Problems

Despite interdisciplinary efforts, progress in socio-technical security is still slow. Research and practice are relying on security approaches that are felt to be unsatisfactory, but we are currently lacking demonstrably better alternatives. This seminar has made important contributions to advancing understanding, developing ontologies and identifying key issues, but much more research is needed in this domain. The following open problems have been identified:

- Reconciling metrics based on cost, time, and probability;
- Analysing security of complex systems based on attack step metrics;
- Relation with economic metrics;
- Relation with privacy metrics;
- Application to specific domains, such as critical infrastructures;
- Simulation of socio-technical systems;
- Defining "good" policies that are not only oriented towards liability but well grounded in what really happens in companies; that also rely on an understanding of human behavior rather than a prescription of behavior only;
- Triangulation of metrics;
- A clear definition of what socio-technical means, as opposed to the sum of two systems with different rules and concepts.

In particular, we recommend a follow-up seminar on analyzing the security of complex socio-technical systems based on metrics.

7   Relation with previous seminars

This seminar builds on the Insider Threat series (08302, 10341, 12501) and the seminar Secure Architectures in the Cloud (11492). However, this new seminar is focused on risk and security metrics, which is a specialized topic that can be of value to a broader community, and as such does not belong to the series.

Other related seminars include Verifying Reliability (12341), and From Security to Dependability (06371). Furthermore, a GI Dagstuhl Research Seminar on Dependability Metrics was held Oct. 30 – Nov. 1, 2005 (http://link.springer.com/book/10.1007/978-3-540-68947-8/

	Sunday 30-11	Monday 1-12	Tuesday 2-12	Wednesday 3-12	Thursday 4-12	Friday 5-12
7:30 – 8:45		Breakfast	Breakfast	Breakfast	Breakfast	Breakfast**
9:00 – 10:30		Introduction	Pitches	Mid-seminar review	Pitches	Summary
10:30 – 10:45		Short break	Short break	Short break	Short break	Short break
10:45 – 12:15		Pitches	Working groups	Working groups	Working groups	Follow-up activities
12:15 – 13:30		Lunch	Lunch	Lunch	Lunch	Lunch
14:00 – 15:30		Working groups	Results from WGs	Social event	Results from WGs	Departure
15:30 – 16:00	Arrival*	Short break	Short break		Short break	
16:00 – 17:45		Tutorials	Demo session		Panel discussion	
18:00 – 19:00	Dinner	Dinner	Dinner		Dinner	
19:30 – 21:00	Arrival*	Social event	Meet / socialise		Meet / socialise	
21:00 -	Wine and cheese	Wine and cheese	Wine and cheese	Wine and cheese	Wine and cheese	

Colour coding:

Non-seminar	Working groups
Food and drinks	Special sessions
Talks	High-level

* Dagstuhl staff present 15:00 - 19:00

** Please check out before 9:00

Figure 2 The programme of the seminar.

page/1). These seminars covered broader classes of metrics, not specifically focused on security or socio-technical integration. The present seminar brings together the socio-technical security angle from the Insider Threat series and the focus on metrics of the above mentioned seminars.

In the Lorentz Center in the Netherlands, a related seminar took place on Formal Methods for the Informal World (http://www.lorentzcenter.nl/lc/web/2013/531//info.php3?wsid= 531). In this seminar, formal models of socio-technical systems were discussed, although not primarily focused on cyber security.

8   Programme overview / Organisation

In an effort to foster exchange among the participants and take full advantage of the Dagstuhl seminar concept, the organizers purposefully defined a program without long ex cathedra talks (Figure 2). The aim was twofold: (1) put emphasis on short presentations, involving a broad variety of people, each followed by sufficient discussion time; (2) avoid the style of presentations that are given in other contexts and that focus more on reporting rather than on sharing new ideas, visions, questions. As a result, the program included the following activities:

- 24 pitches (short talks, focusing at new ideas, visions, major questions);
- 3 tutorials (same objective than pitches, with increased talking and discussion time; suited for topics that are felt shared across the participants);
- 3 demo sessions (focus on concrete use-cases, video material, etc.);
- 1 panel discussion;
- 4 working groups (parallel break-out sessions; see Section Working Groups).

Furthermore, the parallel activities have been complemented by plenary sessions in order to present results to the entire group of participants and facilitate discussing those results.

Participants

Zinaida Benenson
Univ. Erlangen-Nürnberg, DE

Sören Bleikertz
IBM Research GmbH –
Zürich, CH

Rainer Böhme
Universität Münster, DE

Tristan Caulfield
University College London, GB

Kas P. Clark
Ministry of Security and Justice –
The Hague, NL

Trajce Dimkov
Deloitte – Einthoven, NL

Simon N. Foley
University College Cork, IE

Carrie Gates
Dell Research, CA

Dieter Gollmann
TU Hamburg-Harburg, DE

Dina Hadziosmanovic
TU Delft, NL

Carlo Harpes
itrust – Berbourg, LU

Cormac Herley
Microsoft Corp. – Redmond, US

Roeland Kegel
University of Twente, NL

Vincent Koenig
University of Luxembourg, LU

Stewart Kowalski
Gjøvik University College, NO

Aleksandr Lenin
Technical University –
Tallinn, EE

Gabriele Lenzini
University of Luxembourg, LU

Mass Soldal Lund
Norwegian Defence Cyber
Academy – Lillehammer, NO

Sjouke Mauw
University of Luxembourg, LU

Daniela Oliveira
University of Florida –
Gainesville, US

Frank Pallas
KIT – Karlsruher Institut für
Technologie, DE

Sebastian Pape
TU Dortmund, DE

Simon Parkin
University College London, GB

Shari Lawrence Pfleeger
Dartmouth College Hanover, US

Wolter Pieters
TU Delft and University of
Twente, NL

Kai Rannenberg
Goethe-Universität Frankfurt am
Main, DE

Roland Rieke
Fraunhofer SIT – Darmstadt, DE

Martina Angela Sasse
University College London, GB

Paul Smith
AIT – Wien, AT

Ketil Stolen
SINTEF – Oslo, NO

Axel Tanner
IBM Research GmbH –
Zürich, CH

Sven Übelacker
TU Hamburg-Harburg, DE

Michel van Eeten
TU Delft, NL

Jan Willemson
Cybernetica AS – Tartu, EE

Jeff Yan
Newcastle University, GB