SARR: A Cybersecurity Metrics and Quantification Framework (Keynote)

Shouhuai Xu

Laboratory for Cybersecurity Dynamics
Department of Computer Science, University of Colorado Colorado Springs
Email: sxu@uccs.edu Web: xu-lab.org

Abstract. Cybersecurity Metrics and Quantification is a fundamental but notoriously hard problem and is undoubtedly one of the pillars underlying the emerging Science of Cybersecurity. In this paper, we present an novel approach to addressing this problem by unifying Security, Agility, Resilience and Risk (SARR) metrics into a single framework. The SARR approach and the resulting framework are unique because: (i) it is driven by the *assumptions* that are made when modeling, designing, implementing, operating, and defending systems, which are broadly defined to include infrastructures and enterprise networks; and (ii) it embraces the *uncertainty* inherent to the cybersecurity domain. We will review the status quo by looking into existing metrics and quantification research through the SARR lens and discuss a range of open problems.

Keywords: Cybersecurity metrics, cybersecurity quantification, security, agility, resilience, risk, cybersecurity management

1 Introduction

Effective cybersecurity design, operations, and management ought to rely on quantitative metrics. This is because effective cybersecurity decision-making and management demands cybersecurity quantification, which in turn requires us to tackle the problem of metrics. For example, when a Chief Executive Officer (CEO) decides whether to increase the enterprise's cybersecurity investment, the CEO would ask a simple question: What is the estimated return, ideally measured in dollar amount, if we increase the cybersecurity budget (say) by $5M this year? Unfortunately, the status quo is that we cannot answer this question yet because cybersecurity metrics and quantification remains one of the most difficult yet fundamental open problems [10, 38, 32], despite significant efforts [35, 8, 21, 40, 6, 37, 33, 59, 7, 39, 3, 4, 30].

Our Contributions. In this paper, we propose a systematic approach to tackling the problem, by unifying Security, Agility, Resilience, and Risks (SARR) metrics into a single framework. The approach is *assumption*-driven and embraces the *uncertainty* inherent to the cybersecurity domain. Moreover, we evaluate existing cybersecurity metrics through the SARR lens and propose a range

of open problems for future research. Our findings include: (i) it is essential to explicitly and precisely articulate the assumptions made at the design and operation phases of systems; (ii) it is important to understand and characterize the relationships between cybersecurity assumptions, because they may not be independent of each other; (iii) uncertainty is inherent to cybersecurity because defenders cannot directly observe whether or not assumptions made at the design phase are violated in the operation phase; (iv) the current understanding of cybersecurity agility and resilience metrics is superficial, even if defenders can be certain about which assumptions are violated; (v) cybersecurity risk metrics emerge from the uncertainty inherent to assumptions.

Related Work. From a conceptual point of view, the present study corresponds to one pillar of the Cybersecurity Dynamics framework [53, 47, 48, 56], which aims to quantify and analyze cybersecurity from a holistic perspective (in contrast to the building-blocks perspective). This approach stresses the importance of considering the *time* dimension in cybersecurity, leading to *time-dependent* metrics and analysis methods (e.g., [26, 51, 57, 58, 11, 18, 49, 50, 54, 60]). The SARR framework is partly inspired by the STRAM framework [8], which systematizes security metrics, trust metrics, resilience metrics, and agility metrics. The SARR framework goes far beyond the STRAM framework [8] because STRAM does not present the underlying connections between the families of metrics. In contrast, SARR uses *assumptions* and *uncertainty* to unify families of metrics, and these two aspects play no roles in STRAM.

From a technical point of view, the present study focuses on characterizing *what* need to be measured, rather than *how* to measure because we treat the measurement of well-defined metrics as an orthogonal research problem. The latter can be challenging as well. For example, when we infer the ground-truth labels of files in the setting of malware detection, we often encounter the situation that malware detectors give conflicting information (e.g., one detector says a file is benign but another says the file is malicious) [31, 23, 13, 1, 2].

Paper Outline. Section 2 presents the SARR framework. Section 3 discusses the status quo in cybersecurity metrics and quantification research. Section 4 explores future research directions. Section 5 concludes the present paper.

2 The SARR Framework

2.1 Terminology

Abstractions and Views. Cyberspace is a complex system which mandates the use of multiple (levels of) abstractions to understand them. We use the term *network* broadly to include the entire cyberspace, an infrastructure, an enterprise network, or a cyber-physical-human network of interest. Networks can be decomposed horizontally or vertically, leading to two views:

– In the horizontal view, a network can be decomposed into many networked *devices*, which are combinations of hardware and software with computing

and networking capabilities. Devices include computers (e.g., servers, sensors and IoT devices), network devices (e.g., routers and switches), and cybersecurity devices which run (e.g.) intrusion detection systems and firewalls. The horizontal view is often used by cyber defense operators.

– In the vertical view, a network can be decomposed into layers of *components*, which are hardware or software sub-systems, possibly provided by different vendors. Examples of components include operating systems (e.g., Microsoft Windows vs. Linux), applications, and security functions (e.g., intrusion detection systems, malware detectors, and firewalls). We may treat *data* as components as well. Each component may be further divided into layers. For example, the TCP/IP stack can be seen as the communication component, which can be divided into layers of communication protocols. Each component may incorporate or integrate multiple *building-blocks*, such as the machine learning techniques employed by malware detectors. This distinction is important because building-block techniques are often carefully analyzed, components are often proprietary and analyzed only superficially, but networks are analyzed even less thoroughly, perhaps because they are very complex.

Design vs. Operation. In principle, the lifecycle of a network, device, component, or building-block can be divided into a *design* phase and an *operation* phase. The design phase deals with its modeling, design, analysis, implementation, and testing; for ease of reference, we refer to the entities that conduct these activities as *designers*. The operation phase deals with its installation, configuration, operation, maintenance, and defense in the real world; similarly, we refer to the entities that conduct these activities as *operators*. The design vs. operation distinction is important because there can be huge gaps between these two phases, which will be elaborated later.

Cybersecurity vs. Security Properties and Metrics. We use the term *security properties* to describe the standard notions of confidentiality, integrity, availability, non-repudiation, authentication, etc. We use the term *cybersecurity properties* to describe security, agility, resilience, risk and possibly other properties. This means that cybersecurity properties are much broader than security properties. Cybersecurity quantification indicates precise characterization of these cybersecurity properties. For this purpose, we need *cybersecurity metrics*. A metric is a function that maps from a set of objects (e.g., networks, devices, components or building-blocks) to a set of values with a scale (e.g., $\{0, 1\}$ or $[0, 1]$), reflecting security or cybersecurity properties of the objects [35].

2.2 SARR Overview

Figure 1 highlights the framework, which is driven by the *assumptions* that are made at the design and operation phases of a network, device, component or building-block. For a given set of assumptions, there are three kinds of scenarios according to a spectrum of *(un)certainty* in regards to the assumptions.

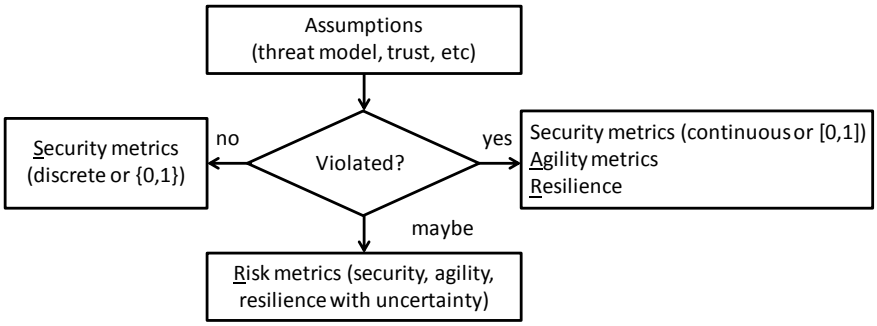

Fig. 1. The SARR framework is driven by *assumptions* and embraces *uncertainty*.

1. It is *certain* that the assumptions are not violated. This often corresponds to the analyses that are conducted at the design phase, where designers consider a range of security properties (e.g., confidentiality, integrity, availability, authentication, and non-repudiation) with respect to a certain system model and a certain threat model. Essentially, these security properties are often defined over a binary scale, denoted by $\{0, 1\}$, indicating whether a property holds or not under the system model and the threat model.
2. It is *certain* that some or all assumptions are violated. This often corresponds to the operation phase, where security properties may be partially or entirely compromised. Therefore, security properties may be defined over a continuous scale, such as $[0, 1]$ (e.g., the fraction of compromised computers in a network). In this case, detection of violations would trigger the defender to take countermeasures to "bounce back" from the violations, leading to the notion of *agility* and *resilience* metrics, which will be elaborated later.
3. It is *uncertain* whether assumptions are violated or not (i.e., assumptions may be violated). This naturally leads to *risk* metrics by associating uncertainties to security, agility and resilience metrics.

In the rest of the section we will elaborate these matters.

2.3 Assumptions

In order to tame cybersecurity, assumptions may be made, explicitly or implicitly, during the design and operation phases of a network, device, component or building-block. They are fundamental to cybersecurity properties.

Assumptions Associated with the Design Phase. At this phase, assumptions can be made with respect to system models, vulnerabilities, attacks (i.e., threat models) and defenses. For example, designers often use *system models* to describe the interactions between the participating entities, the environment and the interaction with it (if appropriate), the communication channels between the participating entities (e.g., authenticated private channel), and the trust that is embedded into the model (e.g., a participating entity is semi-honest

or honest). Designers use *threat models* with simplifying assumptions when specifying security properties, proposing systems architectures, selecting protocols and mechanisms, analyzing whether a property is attained or not under those assumptions. Programmers and testers detect / eliminate bugs and vulnerabilities in the course of developing software, while making various (possibly implicit) assumptions (e.g., competency of a bug/vulnerability detection tool).

Assumptions Associated with the Operation Phase. During this phase, various kinds of (possibly implicit) assumptions are often made (e.g., competency of configurations or defense tools). One example of assumptions that are often made at the design phase and then inherited at the operation phase is the attacker's capability. For example, Byzantine Fault-Tolerance (BFT) protocols, which can be seen as a building-block, work correctly when no more than one-third of the replicas are compromised [29]. However, there is no guarantee in the real world that the attacker cannot go beyond the one-third threshold, effectively compromising the assurance offered by these powerful building-blocks. This can be further attributed to the limited capabilities of cyber defense tools, such as intrusion detection systems and malware detectors.

2.4 Metrics When Assumptions Are Certainly Not Violated

Under the premise that assumptions are complete and are not violated, cybersecurity metrics may degenerate to security metrics in the sense that agility, resilience and risk may become irrelevant. Moreover, it may be sufficient to use binary metrics, namely $\{0, 1\}$, to quantify security properties. This serves as a starting point towards tackling cybersecurity metrics because it would be rare to ascertain in the real world that assumptions are certainly not violated and that the articulated assumptions are sufficient.

Metrics Associated with the Design Phase. At the design phase, we need to define metrics to precisely describe the desired security properties. Textbook knowledge would teach us that the desired properties include confidentiality, integrity, availability, authentication, non-repudiation, etc. However, they may not be sufficient. We advocate accurate and rigorous definitions (or specifications) of metrics, ideally as accurate and rigorous as the definitions given in modern cryptography [16]. This is important because when accurate and rigorous definitions are not given, it is not possible to conduct rigorous analysis to establish desired properties. This means that each security property must be precisely defined with respect to a system model and a threat model. For example, when we specify an availability property, we should specify it as a property of a *service* (e.g., the service offered at port #80) vs. *data* (e.g., a file in a computer) in the presence of some attack.

Metrics Associated with the Operation Phase. We need to define metrics to precisely describe the required security properties of a network, device, component, or building-block at the operation phase. For example, availability metrics at the operation phase may include service response time and service

throughput. Metrics associated with the operation phase are less understood than their counterparts associated with the design phase.

2.5 Metrics When Assumptions Are Certainly Violated

When assumptions are violated, some or all of the security properties are compromised. In order to describe how defenders respond to such violations of assumptions or compromises of security properties, agility and resilience properties emerge. Intuitively, agility quantitatively characterizes how *fast* a defender responds to cybersecurity situation changes [30, 8], and resilience quantitatively characterizes whether and how the defender can make the network, device, component or building-block "bounce back" from the violation of assumptions (i.e., correcting the violations) and the compromise of security properties (i.e., making them hold again). The state-of-the-art is that the notions of *agility-by-design* and *resilience-by-design* are less investigated and understood than *security-by-design*. Agility and resilience are inherently associated with the operation phase because (i) assumptions are the starting point of a design process and (ii) assumptions are violated in real-world operations but not at the design phase. When assumptions are violated, we propose quantifying *security*, *agility*, and *resilience* properties.

For quantifying security properties, examples of metrics are described as follows. (i) To what extent may an assumption have been violated? This may require quantifying the extent to which a network, device, component, or building-block is compromised. This is important for example when using BFT protocols to tolerate attacks, where the fraction of devices that are compromised (e.g., 35% vs. 50%) would make a difference in the defender's response to the attacks. (ii) To what extent is a security property compromised? This is important because a security property may not be *all-or-nothing*, meaning that a violation of assumptions may only cause a degradation of a security property. For example, when a network (or device) is compromised, the attacker may only be able to steal some, but not all, of the data sorted on the network (or device), causing a partial loss of the confidentiality property.

For quantifying agility, example metrics are described as follows. (i) How agile is the defender in detecting the violation of an assumption? One assumption can be that an employed intrusion prevention system can effectively detect a certain class of attacks. Another assumption can be that the attacker does not identify any 0-day vulnerability or use any new attack vector that cannot be recognized by defense tools. (ii) How fast do the desired security properties degrade because of the violation of assumptions? (iii) How quickly does the defender react to the violation of assumptions or successful attacks? (iv) How quickly does the defender bring the network to the required level of security properties?

For quantifying resilience, example metrics are described as follows. (i) What is the maximum degree of violation in terms of the assumptions or security properties that would make it possible for the defender to recover the network (or device or component) and its services without shutting down and re-booting it from scratch? In order to quantify these, we would need to quantify the maximum

degree of violation with respect to the assumptions that can be tolerated. (ii) Does a security property degrades gradually or abruptly when assumptions are violated? (iii) How does the degradation pattern, such as gradual vs. abrupt, depend on the degree of violations of the assumptions?

2.6 Metrics When Assumptions May Be Violated

The preceding two scenarios correspond to the two ends of the spectrum of (un)certainty about the assumptions being violated or not. In the real world, it is rare that the defender would be certain about whether an assumption is violated or not. As a consequence, it is rare for the defender to be certain about whether a security property is compromised or not. Since *uncertainty* is inherent to the cybersecurity domain, we have to embrace the uncertainty, meaning that cybersecurity metrics must be defined while bearing in mind the uncertainty factor. We use the term *risk* to accommodate the security, agility and resilience metrics that can cope with uncertainty. Some examples of risk metrics are described as follows. (i) What is the degree of certainty that a security property is compromised? In order to quantify this, the defender would need to quantify the degree of certainty that an assumption is violated. (ii) What is the degree of certainty when a defense tool flags an event as an attack (e.g., an incoming network connection is an attack or a file is malicious) or anomaly? This may be measured as the conditional probability (or trustworthiness), for example, Pr(the event is indeed an attack|a detector says an event is an attack). (iii) What is the degree of certainty that some software contains a zero-day vulnerability that is known to the attacker but not the defender? (iv) What is the degree of certainty about a threat model (e.g., attacker indeed cannot wage attacks that are not permitted by the threat model)?

Observation 1 *Uncertainty is inherent to cybersecurity, meaning that we must define cybersecurity metrics to help defenders quantify cybersecurity risks and make decisions in their cyber defense operations.*

3 Status Quo

In this section, we use the SARR framework as a lens to look into the cybersecurity metrics that have been proposed in the literature. For this purpose, we leverage survey papers [35, 8, 37] as a source of metrics, while considering more recent literature published after those survey papers (e.g., [30, 13]).

3.1 Assumptions

Assumptions are often articulated more clearly in building-block studies (e.g., cryptography) than the other settings of cybersecurity (e.g., what a chosen-ciphertext attacker can do exactly). However, there are still gaps that are yet to be bridged. First, assumptions may be stated *implicitly*. For example, cryptography assumes that cryptographic keys are kept secret, either entirely or at least for

most information of cryptographic keys (i.e., a partial exposure of cryptographic key may be tolerated). However, cryptographic keys in the real world can be compromised in their entirety (see, e.g., [20, 14]). As a consequence, the security property of digital signatures, known as unforgeability, under the assumption that the private signing keys are kept secret is compromised. This highlights the importance of coping with the presence of compromised cryptographic keys which have not been revoked yet [52, 12, 41]. Still, the trustworthiness of digital signatures has yet to be quantified given the uncertainty that the private signing keys or services may have been compromised without being detected.

Second, assumptions may be inadequate or incomplete. One example of inadequacy is the evolution from considering chosen-plaintext attacks to considering chosen-ciphertext attacks. One example of incompleteness is that earlier threat models simply did not consider the presence of side-channel attacks, which are however realistic. This is not surprising because cyber attacks evolve with time, meaning that threat models also evolve with time [53, 47, 48].

The preceding examples highlight the gaps between the validity of assumptions made at the design phase and the validity of these assumption in the real world. These gaps highlight the importance of explicitly and precisely articulating assumptions because violation of assumptions cause new properties and metrics to emerge (e.g., emergence of agility and resilience metrics). Moreover, the inevitable uncertainty causes the emergence of risk metrics.

Observation 2 *In order to tame cybersecurity, it is essential to explicitly and precisely articulate the assumptions that are made at the design phase and the operation phase. This is far from being achieved and is a big challenge.*

3.2 Security Metrics

In [35], four classes of security metrics are defined: those for quantifying vulnerabilities (including user/human, interface-induced, and software vulnerabilities), those for quantifying attack capabilities (including zero-day, targeted, botnet attacks, malware, and evasion attacks), those for quantifying the effectiveness of defenses (including preventive, reactive, proactive defense capabilities), and those for quantifying situations (e.g., the percentage of compromised computers at a point in time). It is concluded in [35], and re-affirmed in [55], that the problem "what should be measured" is largely open.

Observation 3 *Our understanding of what should be measured in cybersecurity is superficial.*

3.3 Agility Metrics

In a broader context, the existing metrics that can be adapted to measure agility are classified into the following categories [8]: those for quantifying timeliness (including detection time, overall agility quickness) and those for quantifying usability (including ease of use, usefulness, defense cost).

In the narrower context of attack-defense interactions, a novel family of agility metrics are proposed in [30] to quantify the co-evolution (or escalation) of cyber attacks and defenses. Unlike the classification used in [8], the agility metrics defined in [30] accommodate two dimensions of the attack-defense co-evolution, namely *timeliness* and *effectiveness*. Timeliness metrics describe how quickly an attacker is in terms of evolving its attacks in response to the defender's use of new strategy and/or techniques (and comparable metrics from the defender's perspective). These metrics include: *generation-time*, which is the time it takes an attacker (or defender) to evolve its strategies or techniques from one generation to another generation as observed by the defender (or attacker), where a generation may be a new version of a tool (e.g., a new version of malware detector); and *triggering-time*, which is the time it takes an attacker (or defender) to evolve into the next generation of strategy or techniques. Effectiveness metrics quantify how effective a new generation of attacks (or defenses) are, including: *evolutionary-effectiveness*, which describes the effectiveness of the attacker's (defender's) strategy or techniques with respect to defender's (or attacker's); *relative-generational-impact*, which is the effectiveness gain of the current generation of attack (or defense) over the past generation of attack (of defense).

Observation 4 *Our understanding of agility metrics are even more superficial than our understanding of security metrics.*

3.4 Resilience Metrics

By adapting the existing metrics that are defined in other contexts, resilience metrics may be classified into the following families [8]: those for quantifying fault-tolerance metrics (including mean-time-to-failure, percolation threshold, diversity), those for quantifying adaptability (including degree of local decision, degree of intelligent decision, degree of automation), and those for quantifying recoverability (including mean-time-to-full-recovery, mean-time-between-failures, mean-time-to-repair, and intrusion response cost). There are no systematic studies on resilience metrics.

Observation 5 *Our understanding of resilience metrics are even more superficial than our understanding of security metrics.*

3.5 Risk Metrics

Risk is often investigated in the setting of hazards and is often defined as a product of threat (which is a probability estimated by domain expert or other means), vulnerability (which is another probability estimated by domain expert or another means), and consequence (which is the damage caused by the threat when it happens) [22]. This means that risk is quantified as the expected or mean loss. However, this approach is not competent for managing the risk incurred by terrorist attacks [9] because it cannot deal with, among other things, the dependence between many events (e.g., cascading failures). This immediately implies

that this approach is not competent for cybersecurity risk management because there are many kinds of dependencies and interdependencies which make cybersecurity risks exhibit emergent properties [34, 17, 36, 46]. In order to deal with these problems, Cybersecurity Dynamics offers a promising approach, especially its predictive power in forecasting the evolution of dynamical situational awareness attained by first-principle analyses (e.g., [19, 27, 61, 11, 18, 26, 28, 45, 42, 49, 51, 50, 54, 60]) and data-driven analyses (e.g., [24, 15, 25, 44, 5, 43, 57, 58]).

4 Future Research Directions

In order to ultimately tackle the Cybersecurity Metrics and Quantification problem, we highlight some open problems that must be adequately addressed.

Taming Cybersecurity Assumptions. It would be ideal that (i) assumptions are always explicitly and precisely stated, (ii) assumptions are independent of each other, and (iii) assumptions made at the design phase are always valid at the operation phase. However, these are hard to achieve. Alternatively, we should characterize the relationships between related assumptions. For example, an authenticated private communication channel assumes the following: (i) authenticity of the communication parties, (ii) confidentiality of the communication contents, and (iii) integrity of the communication contents. These assumptions further rely on other, often implicitly made, assumptions. Specifically, the preceding assumptions (i)-(iii) would have to be based on the assumption that the communication parties are not compromised when cryptographic mechanisms are used to realize these assumptions; otherwise, assumptions (i)-(iii) are violated. Therefore, when the threat model assumes that the attacker cannot compromise any of the communication parties, the security guarantee rigorously proven in the abstract model may become irrelevant in the real world.

Bridging Design vs. Operation Gaps. There are several gaps between designers' views and defenders' views, especially in terms of their levels of abstractions. In particular, designers often deal with build-blocks and components, but defenders often deal with networks and devices. There are big gaps between these views. First, designers often make assumptions with the mindset that these assumptions will not be violated in the real world. As a consequence, the resulting cybersecurity properties are not only bound to the completeness and accuracy of the assumptions, but also bound to the premise that the assumptions are not violated in practice. Therefore, there is a big gap between the *certainty* of assumptions considered by designers and the *uncertainty* of assumptions being violated or not as perceived by defenders. Second, the network-level and device-level implications of the assumptions that are made when designing building-blocks and components are often unaddressed. This further amplifies the uncertainty encountered by defenders in the real world.

The preceding discussion would explain why security properties are often analyzed in academic research literature but not agility or resilience properties. Moreover, the preceding discussion would also explain why designers often focus on achieving preventive defense with no successful attacks. However, defenders

often deal with successful attacks, which break security properties by violating the assumptions made by designers. This explains why real-world cyber defenders need to leverage preventive defenses, reactive defenses, adaptive defenses, proactive defenses, and active defenses collectively in order to achieve effective defenses [47, 48]. This also explains why the motivating question mentioned in the Introduction cannot be answered yet, namely that the current cybersecurity metrics and quantification knowledge is not sufficient to answer the defender's question in regards to the return on cybersecurity investment.

Identifying and Defining Cybersecurity Metrics That Must Be Measured. As mentioned above, the current understanding of what should be quantified is superficial [35, 55]. It is important to define a comprehensive, ideally complete, suite of metrics under each of the security, agility, resilience, and risk pillars. Since the literature study is often geared towards designers' views, existing metrics are often defined for some purposes but rarely for the purposes of cyber defense operations. Since academic research is often geared towards that assumptions are not to be violated, there is a very limited body of knowledge that can help defenders achieve quantitative cyber defense decision-making and cybersecurity risk management. In order to bridge these gaps, one candidate approach is to leverage cybersecurity datasets to define cybersecurity metrics at multiple levels of abstraction: data vs. knowledge vs. application [48]. Using Medical Science as an analogy, data-level metrics may be defined to quantify building-block or "cell" level properties; "cell" level metrics may be leveraged to define sub-system or "tissue" level properties; "tissue" level metrics may be further leveraged to define "organ" level metrics; "organ" level metrics may be further leveraged to define "human body" level metrics. It should be mentioned that a higher level metric would not be any simple aggregation of some lever level metrics, because cybersecurity is largely about emergent properties [46, 55], meaning that the phenomenon observed at a higher level of abstraction is the outcome of interactions between its composing parts.

Seeking Foundations to Distinguish Good from Poor Metrics [35]. It would not be hard to define cybersecurity metrics, but it is certainly hard to define "good" cybersecurity metrics. This is because it is hard to define criteria or seek foundations to evaluate the competency or usefulness of cybersecurity metrics. In order to tackle this problem, we may need to conduct many case studies and define metrics at multiple levels of abstractions [55] before we can draw general insights along this direction. It would be ideal to conduct such case studies on some killer applications; two candidate killer applications are cyber defense command-and-control and quantitative cyber risk management [48].

Fostering a Cybersecurity Metrics Research Community. In order to tackle such a fundamental problem like cybersecurity metrics and quantification, it must take a community effort. This can be justified by how the basic medical science research has supported clinical healthcare practices. For example, the basic medical science research creates knowledge to help understand how the various kinds of metrics (e.g., blood pressure) would reflect a human being's health condition (e.g., presence or absence of certain diseases), and this

kind of knowledge is applied to guide the practice of medical diagnosis and treatment. Analogously, cybersecurity metrics research would need to identify, invent, and define metrics (e.g., "cybersecurity blood pressure") that reflect the cybersecurity situations and can be applied to diagnose the "health conditions" of networks or devices.

In order to accelerate the fostering of a research community, we can start with some "grass roots" actions. For example, when one publishes a paper, the author may strive to clearly articulate the assumptions that are needed by the new result. Moreover, the author may strive to define metrics that are important to quantify the progress made by the new result [35]. Furthermore, when we teach cybersecurity courses, we should strive to make students know that much research needs to be done in order to tackle the fundamental problems of cybersecurity metrics and quantification. For this purpose, we would need to develop new curriculum materials.

Developing a Science of Cybersecurity Measurement. Well defined cybersecurity metrics need to be measured in the real world, which would demand the support of principled (rather than heuristic) methods. This problem may seem trivial at a first glance, which may be true for some metrics in some settings. However, the accurate measurement of cybersecurity metrics could be very challenging, which may be analogous to the measurement of light speed or gravitational constant in Physics. To see this, let us consider a simple and well-defined metric: What is the fraction (or percentage) of the devices in a network that are compromised at a given point in time t? The measurement of this metric is challenging in practice when the network is large. The reason is that automated or semi-automated tools (e.g., intrusion detection systems and/or anti-malware tools) that can be leveraged for measurement purposes are not necessarily trustworthy because of their false-positives and false-negatives.

5 Conclusion

We have presented a framework to unify security metrics, agility metrics, resilience metrics, and risk metrics. The framework is driven by the assumptions that are made at the design and operations phases, while embracing the uncertainty about whether these assumptions are violated or not in the real world. We identified a number of gaps that have not been discussed in the literature but must be bridged in order to tackle the problem of Cybersecurity Metrics and Quantification and ultimately tame cybersecurity. In particular, we must bridge the assumption gap and the uncertainty gap, which are inherent to the discrepancies between designers' views at lower levels of abstractions (i.e., building-blocks and components) and operators' views at high levels of abstractions (i.e., networks and devices). We presented a number of future research directions. In addition, it is interesting to investigate how to extend the SARR framework to accommodate other kinds of metrics, such as dependability.

Acknowledgement. We thank Moti Yung for illuminating discussions and Eric Ficke for proofreading the paper. This work was supported in part by ARO Grant

#W911NF-17-1-0566, NSF Grants #2115134 and #2122631 (#1814825), and by Colorado State Bill 18-086.

Bibliography

[1] J. Charlton, P. Du, J. Cho, and S. Xu. Measuring relative accuracy of malware detectors in the absence of ground truth. In *Proc. IEEE MILCOM*, pages 450–455, 2018.

[2] J. Charlton, P. Du, and S. Xu. A new method for inferring ground-truth labels. In *Proc. SciSec'2021*.

[3] H. Chen, J. Cho, and S. Xu. Quantifying the security effectiveness of firewalls and dmzs. In *Proc. HoTSoS'2018*, pages 9:1–9:11, 2018.

[4] H. Chen, J. Cho, and S. Xu. Quantifying the security effectiveness of network diversity. In *Proc. HoTSoS'2018*, page 24:1, 2018.

[5] Y. Chen, Z. Huang, S. Xu, and Y. Lai. Spatiotemporal patterns and predictability of cyberattacks. *PLoS One*, 10(5):e0124472, 05 2015.

[6] Y. Cheng, J. Deng, J. Li, S. DeLoach, A. Singhal, and X. Ou. Metrics of security. In *Cyber Defense and Situational Awareness*, pages 263–295. 2014.

[7] J. Cho, P. Hurley, and S. Xu. Metrics and measurement of trustworthy systems. In *Proc. IEEE MILCOM*, 2016.

[8] J. Cho, S. Xu, P. Hurley, M. Mackay, T. Benjamin, and M. Beaumont. Stram: Measuring the trustworthiness of computer-based systems. *ACM Comput. Surv.*, 51(6):128:1–128:47, 2019.

[9] National Research Concil. *Review of the Department of Homeland Security's Approach to Risk Analysis*. The National Academies Press, 2010.

[10] INFOSEC Research Council. Hard problem list. http://www.infosec-research.org/docs_public/20051130-IRC-HPL-FINAL.pdf, 2007.

[11] G. Da, M. Xu, and S. Xu. A new approach to modeling and analyzing security of networked systems. In *Proc. HotSoS'14*, pages 6:1–6:12, 2014.

[12] W. Dai, P. Parker, H. Jin, and S. Xu. Enhancing data trustworthiness via assured digital signing. *IEEE TDSC*, 9(6):838–851, 2012.

[13] P. Du, Z. Sun, H. Chen, J. H. Cho, and S. Xu. Statistical estimation of malware detection metrics in the absence of ground truth. *IEEE T-IFS*, 13(12):2965–2980, 2018.

[14] Z. Durumeric, J. Kasten, D. Adrian, J. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, J. Beekman, M. Payer, and V. Paxson. The Matter of Heartbleed. In *Proc. IMC'2014*.

[15] Z. Fang, M. Xu, S. Xu, and T. Hu. A framework for predicting data breach risk: Leveraging dependence to cope with sparsity. *IEEE T-IFS*, 16:2186–2201, 2021.

[16] O. Goldreich. *The Foundations of Cryptography*, volume 1. Cambridge University Press, 2001.

[17] Yacov Y. Haimes. On the definition of resilience in systems. *Risk Analysis*, 29(4):498–501, 2009.

[18] Y. Han, W. Lu, and S. Xu. Characterizing the power of moving target defense via cyber epidemic dynamics. In *HotSoS*, pages 1–12, 2014.

[19] Y. Han, W. Lu, and S. Xu. Preventive and reactive cyber defense dynamics with ergodic time-dependent parameters is globally attractive. *IEEE TNSE*, accepted for publication, 2021.

[20] K. Harrison and S. Xu. Protecting cryptographic keys from memory disclosures. In *IEEE/IFIP DSN'07*, pages 137–143, 2007.

[21] J. Homer, S. Zhang, X. Ou, D. Schmidt, Y. Du, S. Rajagopalan, and A. Singhal. Aggregating vulnerability metrics in enterprise networks using attack graphs. *J. Comput. Secur.*, 21(4):561–597, 2013.
[22] Uwe Jensen. Probabilistic risk analysis: Foundations and methods. *Journal of the American Statistical Association*, 97(459):925–925, 2002.
[23] A. Kantchelian, M. Tschantz, S. Afroz, B. Miller, V. Shankar, R. Bachwani, A. Joseph, and J. Tygar. Better malware ground truth: Techniques for weighting anti-virus vendor labels. In *Proc. AISec*, pages 45–56, 2015.
[24] D. Li, Q. Li, Y. Ye, and S. Xu. Sok: Arms race in adversarial malware detection. *CoRR*, abs/2005.11671, 2020.
[25] D. Li, Q. Li, Y. Ye, and S. Xu. A framework for enhancing deep neural networks against adversarial malware. *IEEE TNSE*, 8(1):736–750, 2021.
[26] X. Li, P. Parker, and S. Xu. A stochastic model for quantitative security analyses of networked systems. *IEEE TDSC*, 8(1):28–43, 2011.
[27] Z. Lin, W. Lu, and S. Xu. Unified preventive and reactive cyber defense dynamics is still globally convergent. *IEEE/ACM ToN*, 27(3):1098–1111, 2019.
[28] W. Lu, S. Xu, and X. Yi. Optimizing active cyber defense dynamics. In *Proc. GameSec'13*, pages 206–225, 2013.
[29] N. Lynch. *Distributed Algorithms*. Morgan Kaufmann, 1996.
[30] J. Mireles, E. Ficke, J. Cho, P. Hurley, and S. Xu. Metrics towards measuring cyber agility. *IEEE T-IFS*, 14(12):3217–3232, 2019.
[31] J. Morales, S. Xu, and R. Sandhu. Analyzing malware detection efficiency with multiple anti-malware programs. In *Proc. CyberSecurity*, 2012.
[32] David Nicol, Bill Sanders, Jonathan Katz, Bill Scherlis, Tudor Dumitra, Laurie Williams, and Munindar P. Singh. The science of security 5 hard problems (august 2015). http://cps-vo.org/node/21590.
[33] S. Noel, , and S. Jajodia. *A Suite of Metrics for Network Attack Graph Analytics*, pages 141–176. Springer International Publishing, 2017.
[34] J. Park, T. P. Seager, P. S. C. Rao, M. Convertino, and I. Linkov. Integrating risk and resilience approaches to catastrophe management in engineering systems. *Risk Analysis*, 33(3):356–367, 2013.
[35] M. Pendleton, R. Garcia-Lebron, J. Cho, and S. Xu. A survey on systems security metrics. *ACM Comput. Surv.*, 49(4):62:1–62:35, 2016.
[36] S.L. Pfleeger and R.K. Cunningham. Why measuring security is hard. *Security Privacy, IEEE*, 8(4):46–54, July 2010.
[37] A. Ramos, M. Lazar, R. H. Filho, and J. J. P. C. Rodrigues. Model-based quantitative network security metrics: A survey. *IEEE Communications Surveys Tutorials*, 19(4):2704–2734, 2017.
[38] National Science and Technology Council. Trustworthy cyberspace: Strategic plan for the federal cybersecurity research and development program. https://www.nitrd.gov/SUBCOMMITTEE/csia/Fed_Cybersecurity_RD_-Strategic_Plan_2011.pdf, 2011.
[39] L. Wang, S. Jajodia, and A. Singhal. *Network Security Metrics*. Springer, 2017.
[40] L. Wang, S. Jajodia, A. Singhal, P. Cheng, and S. Noel. k-zero day safety: A network security metric for measuring the risk of unknown vulnerabilities. *IEEE TDSC*, 11(1):30–44, 2014.
[41] L. Xu, L. Chen, Z. Gao, X. Fan, K. Doan, S. Xu, and W. Shi. KCRS: A blockchain-based key compromise resilient signature system. In *Proc. BlockSys*, pages 226–239, 2019.
[42] M. Xu, G. Da, and S. Xu. Cyber epidemic models with dependences. *Internet Mathematics*, 11(1):62–92, 2015.

[43] M. Xu, L. Hua, and S. Xu. A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, 59(4):508–520, 2017.
[44] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu. Modeling and predicting cyber hacking breaches. *IEEE T-IFS*, 13(11):2856–2871, 2018.
[45] M. Xu and S. Xu. An extended stochastic model for quantitative security analysis of networked systems. *Internet Mathematics*, 8(3):288–320, 2012.
[46] S. Xu. Emergent behavior in cybersecurity. In *Proc. HotSoS*, pages 13:1–13:2, 2014.
[47] S. Xu. Cybersecurity dynamics: A foundation for the science of cybersecurity. In *Proactive and Dynamic Network Defense*, pages 1–31. 2019.
[48] S. Xu. The cybersecurity dynamics way of thinking and landscape (invited paper). In *ACM Workshop on Moving Target Defense*, 2020.
[49] S. Xu, W. Lu, and L. Xu. Push- and pull-based epidemic spreading in networks: Thresholds and deeper insights. *ACM TAAS*, 7(3), 2012.
[50] S. Xu, W. Lu, L. Xu, and Z. Zhan. Adaptive epidemic dynamics in networks: Thresholds and control. *ACM TAAS*, 8(4), 2014.
[51] S. Xu, W. Lu, and Z. Zhan. A stochastic model of multivirus dynamics. *IEEE Transactions on Dependable and Secure Computing*, 9(1):30–45, 2012.
[52] S. Xu and M. Yung. Expecting the unexpected: Towards robust credential infrastructure. In *Financial Crypto*, pages 201–221, 2009.
[53] Shouhuai Xu. Cybersecurity dynamics. In *Proc. HotSoS'14*, pages 14:1–14:2, 2014.
[54] Shouhuai Xu, Wenlian Lu, and Hualun Li. A stochastic model of active cyber defense dynamics. *Internet Mathematics*, 11(1):23–61, 2015.
[55] Shouhuai Xu and Kishor Trivedi. Report of the 2019 satc pi meeting break-out session on "cybersecurity metrics: Why is it so hard?", 2019.
[56] Shouhuai Xu, Moti Yung, and Jingguo Wang. Seeking foundations for the science of cyber security. *Information Systems Frontiers*, 2021/04/28.
[57] Z. Zhan, M. Xu, and S. Xu. Characterizing honeypot-captured cyber attacks: Statistical framework and case study. *IEEE T-IFS*, 8(11), 2013.
[58] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. Predicting cyber attack rates with extreme values. *IEEE T-IFS*, 10(8):1666–1677, 2015.
[59] M. Zhang, L. Wang, S. Jajodia, A. Singhal, and M. Albanese. Network diversity: A security metric for evaluating the resilience of networks against zero-day attacks. *IEEE Trans. Inf. Forensics Secur.*, 11(5):1071–1086, 2016.
[60] R. Zheng, W. Lu, and S. Xu. Active cyber defense dynamics exhibiting rich phenomena. In *Proc. HotSoS*, 2015.
[61] R. Zheng, W. Lu, and S. Xu. Preventive and reactive cyber defense dynamics is globally stable. *IEEE TNSE*, 5(2):156–170, 2018.