Appendix to "STRAM: Measuring the Trustworthiness of Computer-based Systems"

JIN-HEE CHO, Virginia Tech, USA
SHOUHUAI XU, The University of Texas at San Antonio, USA
PATRICK M. HURLEY, US Air Force Research Laboratory
MATTHEW MACKAY, UK Defence Science and Technology Laboratory
TREVOR BENJAMIN, UK Defence Science and Technology Laboratory
MARK BEAUMONT, Defence Science and Technology Group, Australia

APPENDIX: STRAM - Measuring the Trustworthiness of Computer-based Systems

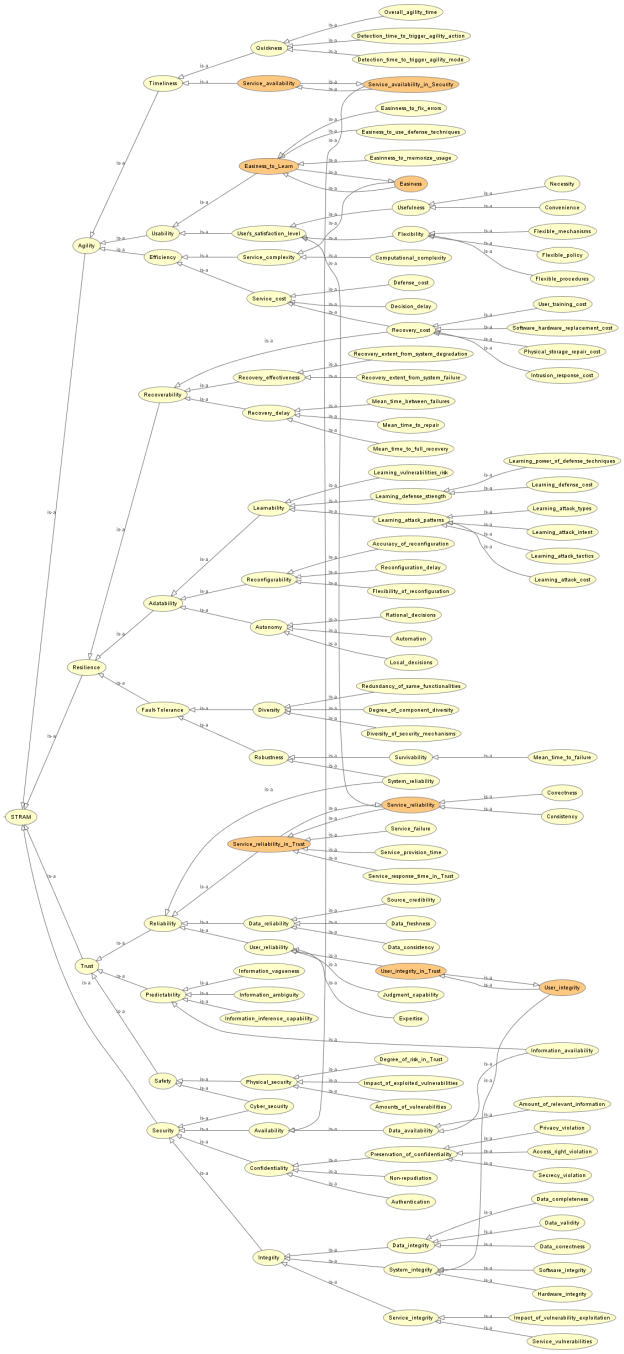

Fig. 1. STRAM Ontology: Key attributes of the four sub-metrics under STRAM.

Table I. Common metrics of vulnerability measurements.

VA Metric	Measurement of a metric	Note
True-Positive Rate (TPR)	$TPR = \frac{TP}{V} = \frac{TP}{TP+FN}$	TP: True Positives; FN: False Negatives; V: The number of vulnerabilities
False-Negative Rate (FNR)	$FNR = \frac{FN}{V} = \frac{FN}{TP+FN}$	
True-Negative Rate (TNR)	$TNR = \frac{TN}{\neg V} = \frac{TN}{FP+TN}$	$\neg V$: The number of non-vulnerabilities
False-Positive Rate (FPR)	$FPR = \frac{FP}{\neg V} = \frac{FP}{FP+TN}$	where $TPR+FNR = FNR+FPR = 1$
Accuracy ($\mathcal{A}$)	$\mathcal{A} = \frac{TP+TN}{TP+FN+FP+TN}$	
Precision ($\mathcal{P}$)	$\mathcal{P} = \frac{TP}{TP+FP}$	a.k.a. Bayesian detection rate
Recall ($\mathcal{R}$)	$\mathcal{R} = \frac{TP}{TP+FN}$	a.k.a. sensitivity and TPR
F-Measure	$F\text{-}Measure = \frac{2\times\mathcal{P}\times\mathcal{R}}{\mathcal{P}+\mathcal{R}} = \frac{2\times TP}{2\times TP+FP+FN}$	
Receiver Operating Characteristic (ROC)	ROC is shown with TPR vs. FPR for x-axis and y-axis, respectively	
Vulnerability Detection Operating Characteristic (VDOC)	VDOC is shown with TPR for x-axis and $\mathcal{P}$ for y-axis	
Relative Vulnerability Detection Power (RVDP)	$RVDP(d', \mathcal{D}', \mathcal{D}) = \frac{\lvert X_{d'} - \cup_{d\in\mathcal{D}} X_d \rvert}{V}$	A VA tool, d, and D is a set of VA tools; X_d be the set of vulnerabilities detected by $d \in \mathcal{D}'$; $d' \in \mathcal{D}'$ for $\mathcal{D} \subset \mathcal{D}'$
Collective Vulnerability Detection Power (CVDP)	$CVDP(\mathcal{D}') = \frac{\lvert \cup_{d\in\mathcal{D}} X_d \rvert}{V}$	where $\mathcal{D} \subseteq \mathcal{D}'$
Coverage ($\mathcal{C}$)	System components assessed by VA tools	

REFERENCES

W. H. Baker, L. P. Rees, and P. S. Tippett, "Communications of the acm," *Communications of the ACM*, vol. 50, no. 10, pp. 101–106, Oct. 2007.

H. Cam, "Risk assessmemt by dunamic representation of vulnerability, exploitation, and impact," in *Proceedings of SPIE*, vol. 9458, 2015.

Y.-Z. Chen, Z.-G. Huang, S. Xu, and Y.-C. Lai, "Spatiotemporal patterns and predictability of cyberattacks," *PLoS One*, vol. 10, no. 5, p. e0124472, 05 2015.

G. Da, M. Xu, and S. Xu, "A new approach to modeling and analyzing security of networked systems," in *Proc. 2014 Symposium and Bootcamp on the Science of Security (HotSoS'14)*, p. 6.

Y. Han, W. Lu, and S. Xu, "Characterizing the power of moving target defense via cyber epidemic dynamics," in *Proc. 2014 Symposium and Bootcamp on the Science of Security (HotSoS'14)*, 2014, pp. 10:1–10:12.

G. Heal and H. Kunreuther, "Modeling interdependent risks," *Risk Analysis*, vol. 27, no. 3, 2007.

J. Holt and S. A. Perry, *A Pragmatic Guide to Competency: Tools, Frameworks and Assessment*. BCS, The Chartered Institute, 2011.

X. Li, P. Parker, and S. Xu, "A stochastic model for quantitative security analysis of networked systems," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 1, pp. 28–43, 2011.

W. Lu, S. Xu, and X. Yi, "Optimizing active cyber defense dynamics," in *Proceedings of the 4th International Conference on Decision and Game Theory for Security (GameSec'13)*, 2013, pp. 206–225.

Table II. Common metrics of risk assessment measurements.

RA Metric	Measuring factor	Note and example results
Dynamics	Is a RA model static or dynamic?	Dynamic risk management, considering vulnerabilities, how they are exploited, and their impact, is critical to defenders' make decision in real-time [Cam 2015].
Dependence	Can a RA tool accommodate the dependence between the relevant random variables, if not stochastic processes, that need to be accommodated into the model?	Dependencies between cyber attacks affect the impact of risk associated with the independency [Da et al.; Xu et al. 2015a].
Inter-dependence	Can a RA tool accommodate the interdependence between system components (e.g., compromise of one component can cause the compromise of another component)? How does the interdependence dynamically evolve?	Interdependence of risk in one agent on risk in another agent can affect overall risk differently. Some studies model interdependent risks based on game theory and estimate expected outcome of risks that are interdependent to system components [Heal and Kunreuther 2007].
Scalability	Can a RA tool deal with a large-scale network?	For large-scale networks (e.g., millions of nodes), stochastic processes models are often intractable, leading to forcing mathematical approximations to the pertinent stochastic process models. Cybersecurity Dynamics models offer a good trade-off between mathematical tractability and faithfulness to the dynamics [Da et al.; Han et al. 2014; Li et al. 2011; Lu et al. 2013; Xu and Xu 2012; Xu et al. 2015a; Xu 2014a,b; Xu et al. 2012a,b, 2014, 2015b; Zheng et al. 2015].
Predictability	Can a RA tool offer predictions to provide cost-effective cyber defense (e.g., optimal or sub-optimal)?	Cybersecurity Dynamics and "grey-box" cybersecurity data analytics can offer the capabilities that capture dynamics of risk [Chen et al. 2015; Peng et al. 2016; Xu et al. 2017; Zhan et al. 2013, 2014, 2015].
Prescription	Can a RA tool offer prescriptive (e.g., control-theoretic or game-theoretic) instructions to guide the defender in adjusting its defense?	Cybersecurity Dynamics models aim to provide prescriptive instructions based on dynamics of risk estimated [Lu et al. 2013; Xu et al. 2012a].
Security measurements	What security metrics are can be described or derived from a RA tool?	Security metrics based on RA can be derived based on cost-effective analysis [Baker et al. 2007]. For example, how much risk is reduced after a certain security / defense mechanism is applied in a system?

C. Peng, M. Xu, S. Xu, and T. Hu, "Modeling and predicting extreme cyber attack rates via marked point processes," *Journal of Applied Statistics*, vol. 0, no. 0, pp. 1–30, 2016.

M. Xu and S. Xu, "An extended stochastic model for quantitative security analysis of networked systems," *Internet Mathematics*, vol. 8, no. 3, pp. 288–320, 2012.

M. Xu, G. Da, and S. Xu, "Cyber epidemic models with dependences," *Internet Mathematics*, vol. 11, no. 1, pp. 62–92, 2015.

M. Xu, L. Hua, and S. Xu, "A vine copula model for predicting the effectiveness of cyber defense early-warning," *Technometrics*, vol. 0, no. ja, pp. 0–0, 2017. [Online]. Available: http://dx.doi.org/10.1080/00401706.2016.1256841

S. Xu, "Cybersecurity dynamics," in *Proc. Symposium and Bootcamp on the Science of Security (HotSoS'14)*, 2014, pp. 14:1–14:2.

——, "Emergent behavior in cybersecurity," in *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security (HotSoS'14)*, 2014, pp. 13:1–13:2.

Table III. Team competence measurements in Red Teaming.

Team competence metric	Attributes or features	Measurements / tools
Team member expertise	Skill sets & competency	Experience, qualifications, portfolio of projects, functionalities, or peer standing (e.g., INCOSE / BCS SFIA competency framework [Holt and Perry 2011])
	Knowledge level	Knowledge for known system vulnerabilities, exploitability for known / unknown vulnerabilities
	Resource level	Computation / communication capability, a number of red team members, and/or time
Available techniques	Reverse Engineering	Reverse engineering tools and methods
	Side Channel	Side channel analysis tools and methods
	RF Emission	RF emission analysis tools and methods
	Documentation analysis	Document analysis tools and methods
	Tracing	Tracing tools and methods
	API analysis	API analysis tools and methods
	Protocol analysis	Protocol analysis tools and methods
	Static analysis	Static analysis tools and methods
	Dynamic analysis	Dynamic analysis tools and methods
Available tools	Name	The name of the tool (Full name and acronyms)
	Version	Version, date of issue, patch etc.
	Purpose	Provide a brief description of the tool including its purpose. Reason / application of the tool by the Red Team. How / Why used. Configuration of the tool for specific application / tests
	Vendor	Name the source of the tool, Name the developer of the tool
	Type	In-house / Open source / COTS / GOTS
	Maturity	Is the tool a prototype, under development / test, production quality or something else?
	Inputs	What data sources is the tool dependent on? Data format the tool requires e.g., .txt
	Deliverables	Describe the specific functionality provided by the tool / Output of the tool. Provide a rating identifying the assessed capability of the tool by function
	Implementation	Operating system / language implementation. Dependency on any other tools or conditions to run?
	Test Sets	List the tests (test sets) covered by the tool
	Gaps	Identify any known gaps in the capability of the tool. Identify any weaknesses in the tool and the areas it covers. List known bugs / unwanted features
	Effectiveness	Assess the effectiveness of the tool in achieving its purpose in comparison with other similar products
	Standards	Identify any standards to which the tool claims compliance
	Cost	Provide an estimation of the cost of tool / per license etc.
	Usability	Knowledge needed to use tool; difficulty of using the tool? Have the team give an estimate on this for comparison?

S. Xu, W. Lu, and L. Xu, "Push- and pull-based epidemic spreading in arbitrary networks: Thresholds and deeper insights," *ACM Transactions on Autonomous and Adaptive Systems (ACM TAAS)*, vol. 7, no. 3, pp. 32:1–32:26, 2012.

S. Xu, W. Lu, and Z. Zhan, "A stochastic model of multivirus dynamics," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 1, pp. 30–45, 2012.

S. Xu, W. Lu, L. Xu, and Z. Zhan, "Adaptive epidemic dynamics in networks: Thresholds and control," *ACM Transactions on Autonomous and Adaptive Systems (ACM TAAS)*, vol. 8, no. 4, p. 19, 2014.

S. Xu, W. Lu, and H. Li, "A stochastic model of active cyber defense dynamics," *Internet Mathematics*, vol. 11, no. 1, pp. 23–61, 2015.

Z. Zhan, M. Xu, and S. Xu, "Characterizing honeypot-captured cyber attacks: Statistical framework and case study," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1775–1789, 2013.

Table IV. Test results measurements in red teaming.

RT Test result category	Measuring outcome	Measurements
Successful attacks	Impact of vulnerability exploitation by successful attacks on valuable assets	***a.*** *None*: No impact of exploitation because a test result reports vulnerabilities only; PT is a special case of RT result with no impact; ***b.*** *Low*: Low impact of exploitation; the implementation of exploitation is hard due to the required local or physical system access; ***c.*** *Medium*: Medium impact of exploitation due to denial-of-service or very limited availability to the system, or only affects subsystems. In order to execute the exploit, the attacker may require local area connection or social engineering; ***d.*** *High*: High impact of exploitation; easy to exploit vulnerabilities with a little knowledge about system vulnerabilities; and ***e.*** *Critical*: Critical impact of exploitation, causing the root level access of the system; tools and information required for executing the exploit are widely available to attackers.
Defense capabilities	Effectiveness and efficiency of defense mechanisms in terms of prevention, detection, and recovery	***a.*** *Prevention capability* measures the skill level of the attacks that are prevented by the defense in terms of (1) little skilled (i.e., an attack is based on running attack tools available); (2) medium skilled (i.e., an attack requires a medium level of expertise or knowledge about local area connections); and (3) highly skilled (i.e., an attack requires high skills, requiring physical system access or social engineering). ***b.*** *Detection capability* measures the detection capability of defense mechanisms including (1) detection time (i.e., from attack launching time to detection time); (2) response time to detected attacks (i.e., from attack detection time to response time to the detected attack); and (3) effectiveness of detection mechanisms. ***c.*** *Recovery capability* measures the recovery capability of defense mechanisms including (1) automated recovery (i.e., automated reconfiguration without any manual procedures); and (2) autonomous recovery (i.e., autonomous reconfiguration based on collected evidence to maximize system reliability without any manual procedures).
Red teaming outcome	Significance of identified vulnerabilities by RT exercise	***a.*** *Number of vulnerabilities* identified by the RT; ***b.*** *Mission criticality* estimated by the impact of damage caused by successful attack by the RT; ***c.*** *Defense / recovery directions* suggested by the RT to patch vulnerabilities or fix system faults; and ***d.*** *Customer satisfaction level* evaluated based on an ordinal scale (e.g., 5 scales from 1 for extremely dissatisfied to 5 for surpassed expectations).

——, “A characterization of cybersecurity posture from network telescope data,” in *Proc. of the 6th International Conference on Trustworthy Systems (InTrust’14)*, 2014, pp. 105–126.

——, “Predicting cyber attack rates with extreme values,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1666–1677, 2015.

R. Zheng, W. Lu, and S. Xu, “Active cyber defense dynamics exhibiting rich phenomena,” in *Proc. 2015 Symposium and Bootcamp on the Science of Security (HotSoS’15)*, 2015, pp. 2:1–2:12.