Towards a Theoretical Framework for Trustworthy Cyber Sensing

Shouhuai Xu
Department of Computer Science
University of Texas at San Antonio
shxu@cs.utsa.edu

ABSTRACT

Cyberspace is an indispensable part of the economy and society, but has been "polluted" with many compromised computers that can be abused to launch further attacks against the others. Since it is likely that there always are compromised computers, it is important to be aware of the (dynamic) cyber security-related situation, which is however challenging because cyberspace is an extremely large-scale complex system. Our project aims to investigate a theoretical framework for trustworthy cyber sensing. With the perspective of treating cyberspace as a large-scale complex system, the core question we aim to address is: What would be a competent theoretical (mathematical and algorithmic) framework for designing, analyzing, deploying, managing, and adapting cyber sensor systems so as to provide trustworthy information or input to the higher layer of cyber situation-awareness management, even in the presence of sophisticated malicious attacks against the cyber sensor systems?

Keywords: Cyber security, trustworthy cyber sensing, trustworthy sensing information, attack-resilient cyber sensing, complex networks, complex systems, attack-defense dynamics

1. INTRODUCTION

Cyberspace, a human-made extremely large-scale complex system, evolved from the Internet that was originally designed for use by trusted individuals who are not supposed to intentionally or unintentionally exploit the software vulnerabilities in the software systems. As many enterprise and personal information systems are now networked together, cyber attacks — driven by the motivation or incentive of making (for example) profits — have emerged and evolved rapidly. The rich connectivity in cyberspace (e.g., any computer could attack any other) prompts us to consider cyber security from a holistic or whole-system perspective, rather than from a component-oriented perspective. Despite decades of studies in (among other things) network security and cryptography, this perspective has yet to be explored and calls for efficient and effective cyberspace management so as to significantly mitigate, if not completely prevent, the damage of cyber attacks. For this purpose, it is our belief that competent cyber sensing is inevitable.

In this position paper, we explore the problem of trustworthy cyber sensing including its objectives, design principles, architecture, modeling and analysis methodologies. Inspired by the practice of monitoring the natural ecosystem so as to understand its evolution dynamics and to make better decisions, we propose to monitor cyberspace so as to understand the dynamics of the cyber ecosystem and to make better decisions in the presence of cyber threats.

Paper outline. Section 2 discusses the objectives and challenges of trustworthy cyber sensing. Section 3 proposes a set of principles for designing trustworthy cyber sensing systems. Section 4 presents an architecture of trustworthy cyber sensing systems. Section 5 discusses the modeling and analysis of trustworthy cyber sensing systems. Section 6 presents a concrete example about trustworthy cyber sensing. Section 7 briefly discusses related prior work. Section 8 summarizes the paper with a discussion on our on-going project in trustworthy cyber sensing.

2. OBJECTIVES AND CHALLENGES OF TRUSTWORTHY CYBER SENSING

As cyberspace has become the battlefield between cyber attackers and cyber defenders, we must equip the defenders with competent decision-making systems or processes, for which cyber sensing is motivated to make defenders aware of the dynamics in cyberspace (i.e., cyber situational awareness) and to provide early warning of cyber attacks. Inspired by USAF Colonel John Boyd's conceptual framework known as the OODA loop (Observation, Orientation, Decision, and Action; see `http://en.wikipedia.org/wiki/OODA_loop`), we propose the concept of "cyber defense OODA Loop," which is highlighted in Figure 1. Basically, after the initial design and deployment of cyber sensing systems, raw information about cyberspace is collected (observation) and then refined to produce higher-level or higher-quality cyber situational information (orientation). Based on the refined situational information, decision may be made to defend against potential threats, to disrupt on-going attacks, or to adapt the cyber sensing system itself.

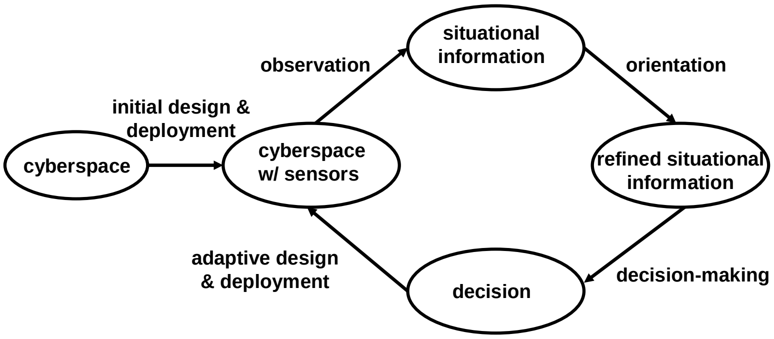

Figure 1. Cyber sensing in the larger framework of cyber defense OODA Loop

Because the output of cyber sensing systems will be used in higher-layer decision-making systems, we would want cyber sensing systems to be trustworthy and, at the same time, privacy-friendly (i.e., without breaching the privacy of honest people). It would be ideal that cyber sensing systems have zero false-positive and zero false-negative, which is however over costly to achieve, if possible at all. As an analogy, we mention that there are still crimes in the physical world after thousands of years of evolution and development in laws and deterrence. Therefore, it is more reasonable to require cyber sensing systems to detect major threats well before they cause actual damage, where by "major threats" we mean attacks whose consequences are potentially above some thresholds (e.g., in terms of the amount of money that will be lost, or in terms of the size of the population that will be affected). In addition, we may have to lower our expectation for trustworthiness because of some possible false positives and false negatives.

There are many technical challenges that we must address before achieving trustworthy cyber sensing. The challenges are caused by that cyberspace is an extremely large-scale complex system that consists of many heterogeneous sub-systems, which are managed by many players (or shareholders) that are not necessarily cooperative. For example, players may not want to expose the whole structure of their autonomous systems to the others (e.g., an ISP might not want to let any other player know certain malware traffic originates from which computer that is connected to its access network). Moreover, cyberspace is extremely dynamic; for example, the number of compromised computers in cyberspace today might be very different from the number of compromised computers in cyberspace tomorrow. Moreover, it is desired to achieve trustworthy cyber sensing without breaching any privacy of the honest people, which is however difficult as demonstrated by what we have witnessed in the subfield of privacy-preserving data mining in the past decade.

3. DESIGN PRINCIPLES FOR TRUSTWORTHY CYBER SENSING

What are the principles that can guide us in designing trustworthy cyber sensing systems? The following principles, which address the needs of being *holistic* and *trustworthy*, are believed to be useful.

- Holistic: In order to most effectively defend cyberspace, we need holistic information about cyberspace. In order for the defender to "connect the dots" so as to detect attacks that might not be detected from non-holistic perspectives, cyber sensing system should provide the necessary information from the following perspectives.

  - Cross-layer sensing: At each single layer, we can observe relevant activities that can be part of an attack process. Cross-layer sensing allows the defender to more effectively and accurately detect attacks because it would detect malicious activities that may be deemed as benign at each individual layer. This is a kind of combining vertical views for more holistic information.
  - Cross-network sensing: By combining the views at different network nodes, we can detect attacks that may not be recognized by looking at each individual node alone. This is a kind of combining horizontal views, possibly cross multiple autonomous systems, for more holistic information.
  - Multi-resolution sensing: Because cyberspace is an extreme large-scale complex systems, we need multi-resolution sensing so as to observe the cyberspace at different granularity. For example, at core routers level, DDoS (Distributed Denial of Service) attacks may not be easy to detect because the traffic may not reach the bandwidth limit; at a higher resolution (e.g., per-destination or per-port), however, it is possible to detect DDoS at both edge routers and core routers. This allows us to consider and flexibly zoom in and out macroscopic, mesoscopic, and microscopic cyber sensing.

- Trustworthy: A trustworthy cyber sensing system should have the following characteristics.

  - Robust sensing: A cyber sensing system should be robust, meaning that even if some sensors have been compromised, the sensing system is still able to catch malicious events. This means that there should be no single point of failure in a cyber sensing system. This also means that we need to make it as difficult as possible for the attacker to figure out where the sensors are, what they are monitoring for, and how they are monitoring. This privacy- or anonymity-based defense offers another layer of protection.
  - Adaptive sensing: The above discussion already suggested that the cyber sensing system itself should be adaptive, ideally in an automated fashion. This is especially relevant when some cyber sensors may have been compromised, or the cyber sensing system may have overlooked some important aspects in the currently deployed design. Moreover, it would be ideal if cyber sensing systems can self-adapt to monitor suspicious activities they have not seen in the past (e.g., communications using certain port numbers are abruptly increasing).
  - Proactive sensing: If a cyber sensing system is detected by the attacker, the attacker might be able to evade it. Proactive sensing disrupts the evasion by dynamically changing the deployment of the sensor systems. This is especially important when the cyber sensing systems are partially compromised as the time goes by.
  - Networked sensing: A cyber sensing system itself should be networked together so as to achieve better resilience against attacks. For example, if some cyber sensors can detect that a certain sensor has been malfunctioning or compromised, the compromised sensor can be replaced with a new or different sensor. This kind of distributed processing capability enabled by networked cyber sensing is crucial in many real settings.

The above discussion reflects that cyber sensing exhibits kinds of emergent properties in terms of both *function* and *trustworthiness*.

4. ARCHITECTURE OF TRUSTWORTHY CYBER SENSING

The above design principles suggest the following architecture for trustworthy cyber sensing (see also Figure 2): application-oriented, infrastructure-oriented, and cross-layer which should be seamlessly integrated together to facility holistic cyber sensing.

Application-oriented cyber sensing might be able to detect attacks that are otherwise more difficult to detect via infrastructure-oriented cyber sensing. For example, when attackers “legitimately” abuse applications that require little or none authorization, the resulting malicious activities are perfectly camouflaged into the traffic

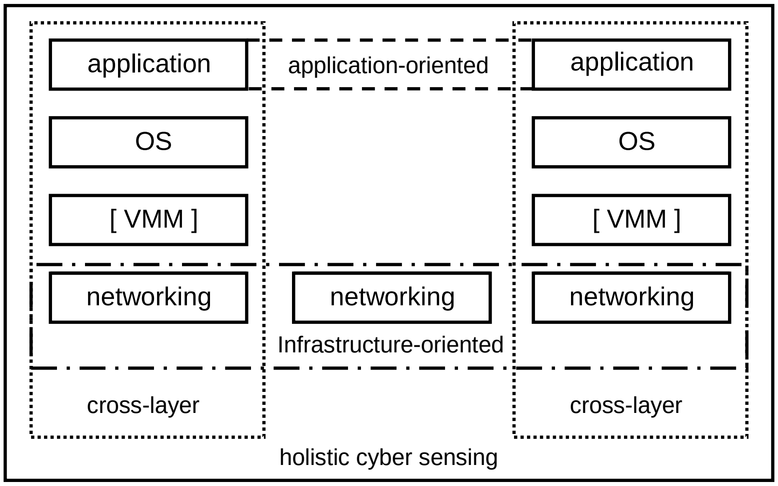

Figure 2. A architecture for trustworthy cyber sensing

of honest users. A concrete example is that in the recent incident of botnets using social network web sites (e.g. `twitter.com`) as their command-and-control mechanism,[1] the command and control messages are perfectly legitimate application flows because the attacker(s) own the accounts that did not require much authorization (beyond being able to register an account).

Infrastructure-oriented cyber sensing goes beyond traditional network-level intrusion detection and attack correlation because it allows to recover a more holistic picture about the potential or present threats in cyberspace. For example, it helps address the following questions: Have activities corresponding to which port numbers changed abruptly? Have flows targeting at which destination increased abruptly, and where are the flows originated from?

Host-oriented cross-layer sensing allows the defender to be aware the suspicious activities in both user-land and kernel-land, including the activities in the host OS (Operating System). Since computer malware (including virus, bot, Trojan Horses, rootkits) has become a major attack vector in cyberspace, host-oriented cross-layer sensing is an important tool in detecting such threats. It goes beyond traditional signature-based anti-virus software because signature-based defense can be easily evaded, and should rely very much on behavior-based techniques. For example, we have been conducting a systematic study on analyzing and exploiting malware behaviors so as to detect unknown or zero-day attacks and facilitate early warning. Initial results show that we can already detect malware attacks that can evade popular anti-virus software defenses.[2]

The aforementioned application-oriented, infrastructure-oriented, and host-oriented cross-layer cyber sensing must be seamlessly integrated into a single holistic cyber sensing system so as to address questions such as the following: How do we know the fraction of compromised computers at any point in time? What is going on in the underground and/or black market where stolen information may be traded or new attack techniques are solicited? Which popular servers are disseminating malware via drive-by download, and have recently visited by which clients computers? Which p2P (peer to peer) networks are actively used for sharing malware? Which DNS names have abruptly become hot (e.g., possibly have been compromised and then abused as a botnet command-and-control center)? How should we integrate the situation information from the application- and host-oriented cyber sensing?

To understand and manage to what extent questions like those mentioned above can be addressed, we need to model, in particular, the trustworthiness of cyber sensing systems. In the next section we report our initial exploration towards the ultimate goal.

5. MODELING AND ANALYZING TRUSTWORTHY CYBER SENSING

In this section we discuss modeling and analyzing cyber sensing systems. Because how holistic being enough depends on the need, we here only discuss the modeling and analysis of trustworthiness, which requires the specification of threat model, the identification of relevant attributes, and the invention of analysis methodologies.

5.1 Threat Modeling

At a high level, we consider two types of attacks: attacks against cyberspace and attacks against the cyber sensing system, where the former may be independent of the cyber sensing systems but the later will have to be specific to the cyber sensing systems in question. In either case, the attacker aims to compromise many and/or important nodes without being detected (i.e., the attacker might always wants to remain stealth). While DDoS attacks are relevant in many cases, many future attacks may go beyond undermining connectivity and availability, and would aim to breach the secrecy and integrity of information that is stored or flows in complex networks by compromising, for example, the software systems and/or the underlying cryptographic authentication infrastructure.

5.2 Attributes of Trustworthy Cyber Sensing

In order to manage, quantify, minimize or mitigate the damage of cyber attacks, we need to identify relevant trustworthiness attributes. This is challenging not only because cyberspace is an extremely large-scale complex system, but also because each node itself may be a complex system (e.g., each computer executes millions of lines of software program that might be vulnerable to attacks). We will identify and formalize trustworthiness attributes that can be classified into the following three types of capabilities: monitoring, early-warning and situation-awareness, and adaptation. These capabilities correspond to the "observation, orientation & decision, and action" phases of the cyber defense OODA Loop mentioned above.

- Monitoring capabilities:

  Attributes of this type capture the capabilities of cyber sensing systems in monitoring the presence of cyber threats or attacks, and are related to the following questions. What are the fundamental characteristics of cyber sensing systems that are necessary and/or sufficient for surviving large-scale complex networks from various degrees of cyber attacks? What kinds of assumptions we need to make? What are the fundamental characteristics of cyber sensing systems that can themselves survive various degrees of cyber attacks?

- Early-warning and situation-awareness capabilities:

  Attributes of this type are related to the following questions. To what extent a cyber sensing system can diagnose that cyberspace is under an attack of certain degree? What is the availability of connectivity or services under certain attacks? What is the probability that a node is compromised at a certain point in time? How does the probability depend on other system parameters (e.g., node degrees)? How can we predict the cascading behavior of an on-going cyber attack so that we can act in advance?

- Adaptation capabilities:

  Attributes of this type are related to the following questions. To what extent a cyber sensing system can (automatically) adapt its configurations so as to accomplish missions in a cost-effective or even optimal fashion? To what extent a cyber sensing system can automatically recover from attacks? How can we adapt the network to minimize the damage of an on-going cascading attack?

Attributes capturing the above capabilities may be in general, or may be specific to missions that is planned or undertaken.

5.3 Analyzing the Trustworthiness of Cyber Sensing

We need novel methodologies for measuring and analyzing attributes of cyber sensing systems, and for characterizing the roles of other system parameters. In particular, we need dynamical system and/or algorithmic models to characterize and analyze for each attribute a family of functions:

$$f_{trustworthiness\_attribute} : cyberspace \times cyber\_sensing\_system \times threat\_model \mapsto trustworthiness\_measure.$$

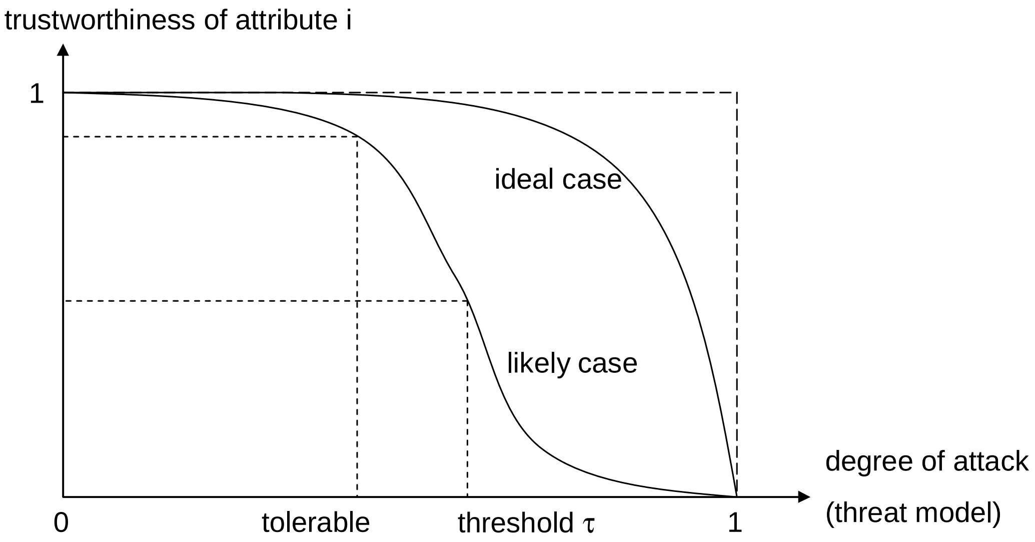

Figure 3. Ideal case vs. likely case

In general, we may want to achieve, for each attribute, the "ideal case" shown in Figure 3; in practice, we might often achieve the "likely case" in Figure 3. In the likely case, two relevant concepts are "attack threshold τ", which captures the "phase transition" as in the context of Physics, and the "degree of tolerable attack" — above it often requiring to adapt the system. Note that the above discussion applies to both the trustworthiness of cyber sensing system attributes and the trustworthiness of cyber security attributes.

6. AN EXAMPLE

As hinted above, we are devising a holistic cyber sensing system prototype, which aims to accommodate various system components, including their potential vulnerabilities and the interdependencies between them. This allows us to conduct theoretical analyses (e.g., algorithmic and dynamics-based "what if" analyses), which can offer useful insights for guiding the design of real-life cyber sensing systems. For example, let us consider a networked system, which can be as large as all the connected systems in cyberspace. We know that computers will be compromised soon or later. This suggests us to ask an important question: At time t, what is the portion of compromised computers? Recently, we considered a general model for describing and understanding the realistic push- and pull-based attack-defense dynamics in cyberspace.[3] Within this model, we show that the index mentioned above can be estimated, either through computation or through simulation, as long as we are given the parameters (which capture the attack and defense capabilities, respectively). This, while conceptually interesting, is not practical in many realistic circumstances. Very recently, we have moved a significant step by showing that we can still estimate the global index mentioned above by local monitoring *without* knowledge of the parameters, namely monitoring only a small number of nodes.[4] We believe this is a significant step towards ultimate trustworthy cyber sensing. We will investigate further by considering the attacks against the sensing systems because if the attack is able to evade from the sensing systems (i.e., by not compromising certain nodes in a complex network), the sensing system becomes useless.

7. RELATED WORK

Monitoring Internet traffic was successfully used for network performance management,[5] but such monitoring is not sufficient for security-oriented cyber sensing. In this paper we take a novel and fresh *top-down* approach to exploring trustworthy cyber sensing. Nevertheless, some content of this paper was inspired by previous studies. In particular, the design principles for trustworthy cyber sensing systems were influenced by Saltzer and Schroeder's design principles for protection mechanisms.[6] The concepts of host-oriented cross-layer and infrastructure-oriented cyber sensing were influenced by numerous previous studies (e.g.,[7–11]). The concept of application-oriented cyber sensing was inspired by our recent study.[1] The concept of robust monitoring and its algorithmic aspect was inspired by our on-going study.[12] The importance of keeping sensor location private was investigated.[13]

8. SUMMARY

In this position paper we took a fresh *top-down* approach to exploring the problem of trustworthy cyber sensing, including its objectives, design principles, architecture, modeling and analysis methodology. The content presented in this paper is rather preliminary and reflects our current understanding. As our investigation advances, the framework will be improved and enhanced. Moreover, there are many fundamental questions we hope to answer: What can and cannot be detected via cyber sensing, under what circumstances? What are the fundamental impact of wide deployment of cloud computing infrastructure and services? Such theoretic studies will have to be coupled with empirical analysis based on real-life datasets, which are not easy to get in practice. Nevertheless, we are making progresses in both perspectives. For example, we have implemented an initial prototype system for cross-layer, integrated host-oriented and application-oriented cyber sensing system. We are in the process of enhancing it based on the design principles mentioned above. We are also investigating infrastructure-oriented cyber sensing.

ACKNOWLEDGMENTS

This work was supported in part by AFOSR and ONR.

REFERENCES

[1] Kartaltepe, E., Morales, J., Xu, S., and Sandhu, R., "Social network-based botnet command-and-control: Emerging threats and countermeasures," (2010 (manuscript in submission)).

[2] Morales, J., Al-Bataineh, A., Xu, S., and Sandhu, R., "Analyzing and exploiting network activity behaviors of malware," (2010 (manuscript in submission)).

[3] Li, X., Parker, P., and Xu, S., "Towards quantifying the (in)security of networked systems.," in [*21st IEEE International Conference on Advanced Information Networking and Applications (AINA'07)*], 420–427 (2007).

[4] Xu, S., Lu, W., and Zhan, Z., "Push- and pull-based epidemic spreading in networks: Thresholds and quantitative insights," (2010 (manuscript in submission)).

[5] Feldmann, A., Greenberg, A., Lund, C., Reingold, N., Rexford, J., and True, F., "Deriving traffic demands for operational ip networks: methodology and experience," *IEEE/ACM Trans. Netw.* **9**(3), 265–280 (2001).

[6] Saltzer, J. and Schroeder, M., "The protection of information in computer systems," *Proceedings of the IEEE* (1975).

[7] Lakhina, A., Crovella, M., and Diot, C., "Diagnosing network-wide traffic anomalies," in [*SIGCOMM'04*], 219–230 (2004).

[8] Xie, Y., Sekar, V., Maltz, D., Reiter, M., and Zhang, H., "Worm origin identification using random moonwalks," in [*IEEE Symposium on Security and Privacy*], 242–256 (2005).

[9] Collins, M. and Reiter, M., "Finding peer-to-peer file-sharing using coarse network behaviors," in [*11th European Symposium on Research in Computer Security (ESORICS'06)*], 1–17 (2006).

[10] Gu, G., Porras, P., Yegneswaran, V., Fong, M., and Lee, W., "Bothunter: Detecting malware infection through ids-driven dialog correlation," in [*Proceedings of the 16th USENIX Security Symposium (Security'07)*], (2007).

[11] Gu, G., Zhang, J., and Lee, W., "Botsniffer: Detecting botnet command and control channels in network traffic," in [*Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08)*], (2008).

[12] Lu, S., Xu, S., and Zhan, Z., "Algorithms for optimal and robust monitoring," (2010 (manuscript in preparation)).

[13] Bethencourt, J., Franklin, J., and Vernon, M., "Mapping internet sensors with probe response attacks," in [*Proceedings of the 14th conference on USENIX Security Symposium*], 13–13 (2005).