Cyber Trust and Suspicion: A Human-Centric Approach

Hongbin Wang

School of Biomedical Informatics, University of Texas Health Science Center at Houston
Hongbin.Wang@uth.tmc.edu

Abstract. Conventional wisdom has regarded cyberspace security as a pure technology issue – sophisticated information techniques, tools, and policies are a must in order to detect and defeat threats. At a more foundational level, however, it is now clear that cyberspace security is also, if not more, a human-social phenomenon - how human operators, be they everyday internet users or national intelligence analysts, perceive and make sense of cyber events "closes the loop" and is therefore essential for the ultimate success (or failure) of cyberspace security. In this position paper we argue for the need of studying cyber trust and suspicion from a human-centric approach. Based on a principled abduction-based framework, the results will answer a full range of fundamental questions regarding cyber trust and suspicion.

Keywords: Cybersecurity, Trust and Suspicion, Human Belief Revision, Psychology, Computational Modeling.

1 Introduction

"… it's now clear this cyber threat is one of the most serious economic and national security challenges we face as a nation. It's also clear that we're not as prepared as we should be, as a government or as a country…."
--- President Obama, 2009, on Securing Our Nation's Cyber Infrastructure

With the rapid advances of information technology, cyberspace, a space of 0's and 1's, has become as real as our physical space. At the same time, cyberspace security is increasingly becoming a serious challenge [1]. Cyber attacks, such as identity theft, cooperate espionage, password sniffing, DDoS (Distributed Denial of Service), stuxnet, and email spamming, to name a few, have presented grave threats to human everyday life as well as national security. Conventional wisdom has regarded cyberspace security as a pure technology issue – sophisticated information techniques, tools, and policies are a must in order to detect and defeat threats (for defense) and develop and deliver attacks (for offense). At a more foundational level, however, it is now clear that cyberspace security is also, if not more, a human-social phenomenon - how human operators, be they everyday internet users or national intelligence analysts, perceive and make sense of cyber events "closes the loop" and is therefore essential for the ultimate success (or failure) of cyberspace security. Unfortunately, the significant role of the human operations in cyber security cycles has largely been ignored or less

C. Stephanidis (Ed.): Posters, Part II, HCII 2013, CCIS 374, pp. 759–763, 2013.

understood thus far. This is particularly true with regard to cyber trust and suspicion, two fundamental concepts in cyberspace security. Is the email just received trustworthy? Is the network activity pattern normal? The bottom line is that a cyber attack (e.g., worm or sabotage) is more damaging and harmful if it is stealthy and with disguise, and disasters occur when a non-trustworthy source is trusted. One central question in cyberspace security is therefore to understand how cyber trust and suspicion are represented, measured, monitored and managed. Any security-oriented algorithms and systems have to have some form of trust and suspicion management built-in, though often implicitly [2,3]. It is critical to realize that trust and suspicion are fundamentally psychological constructs and human traits. Automated trust and suspicion management systems through sophisticated computer algorithms are certainly desirable and have been quite successful [4]. We have to accept, however, that it is humans (but not machines) that trust and suspect, and that the computer algorithms have to be based on sound theorization of human trust and suspicion intuition in order to be useful. Such a theorization has the potential to make automated solutions even more powerful, robust, and realistic by inserting key human factors such as motivation, intention, attention, perception, belief, and emotion into the picture. In addition, in cases when computer algorithms are inconclusive, it is human operators' trust and suspicion insights and intuitions that often *connect the dots and close the loop*.

1.1 Cyber Trust and Suspicion

Trust and suspicion are fundamental concepts in many fields including philosophy, literature, law, and psychology [5,6]. They are often used to describe a person's relationship to another person, to a thing, to a factor/belief, or to nature. In ancient Greece, skepticism philosophers argued to assert nothing and suspend judgment. With the development of modernity and technology the concepts of trust and suspicion become even more relevant (rather than obsolete) [7]. In reality, for a piece of information, people can choose to trust, to distrust, or, very often, to be anywhere between. Being uncertain is simply a basic fact of human conditions.

Trust and suspicion are naturally loaded concepts. Dictionary definitions of "trust" link the term to "confidence", "reliability", "credibility", "predictability", and "benevolence". Trust is distinguished from "distrust" in that distrust is not equal to lack of trust, which is more related to the concept of "suspicion". Suspicion is a cognition or disposition of doubt, which often results from co-existence of conflicting beliefs or lack of evidence, and may lead to more vigilance and information seeking [8].

In a classic review of the concepts of trust and suspicion [9,10], Deutsch defines "trust" as follows: "An individual may be said to have trust in the occurrence of an event if he expects its occurrence and his expectation leads to behavior which he perceives to have greater negative motivational consequences if the expectation is not confirmed than positive motivational consequences if it is confirmed" [10]. And he defines "suspicion" as follows: "An individual may be said to be suspicious of the occurrence of an event if the disconfirmation of the expectation of the event's occurrence is preferred to its confirmation and if the expectation of its occurrence leads to behavior which is intended to reduce its negative motivational consequences" [10].

It is clear from these definitions that trust and suspicion are closely linked to motivations and subsequent decision making. A trust-minded person, compared to a suspicious person, is more willing to take risks in an uncertain environment and therefore is more likely to be caught off-guard if something goes wrong. It is in this sense trust and suspicion are relevant and important factors in cyberspace security [e.g., 3,11,12]. Cyberspace fundamentally alters the dynamics of inter-personal and human-machine relationships. Internet and social media allow never-met-before people to know each other and become "friends." Communications become so fast and cheap that everybody is exploded with information. In these situations, what do "trust" and "suspicion" mean? When we say we trust an email message, do we trust the message itself or trust the person who sends the message, or trust something else? We can use sophisticated machine learning techniques to mine past data and develop algorithms to tell us the precise likelihood and consequence of such trust in the past, however we may still be uncertain about the intention/implication of the message and the sender, and the action we should take. Needless to say, all these factors are interwoven and together they form the landscape of today's cyberspace security. A foundational understanding of the underlying dynamics of cyber trust and suspicion is clearly needed for achieving better cyber security. It can only be acquired when we close the loop between information systems and human operators and study their interactions.

1.2 Psychometrics of Cyber Trust and Suspicion

Not much work has been done in understanding how trust and suspicion work in cyberspace domains with humans in the loop. Relevant work often focuses on answering questions such as: What are they? How to measure them? What affects them? Can they be exploited [2,3,11-15]? Findings in these efforts are essential in our effort to develop a comprehensive computational model of cyber trust and suspicion that can capture the dynamics between human operators and systems. One major thrust is to find a credible way to measure trust or develop a cybertrust indicator [e.g., 13]. Jian *et al*. [3] explored the possibility of establishing an empirically tested (rather than theoretically driven) scale for measuring trust (human-human trust, human-machine trust, and trust in general) in computerized systems. Their results show that trust and distrust are better treated as the opposite ends of a single continuum. Barelka and colleagues [14] examined the relationship of trust and suspicion in IT domains. Using sophisticated statistical techniques, they found that trust in automation was best characterized by two orthogonal dimensions (trust and distrust) and trust and distrust were independent from IT suspicion. Interestingly, these results seem to be at odds with a recent functional brain imaging study, which shows that trust and distrust have distinct neural correlates in the brain [15]. More specifically, the study, using fMRI technology, shows that trust is associated with the brain's reward, prediction, and uncertainty areas (e.g., caudate, anterior paracingulate cortex, and the orbitofrontal cortex), while distrust is associated with the brain's intense emotions and fear of loss areas (e.g., the insular cortex and amygdala). The results support a 2-dimention view of trust-distrust relation and suggest that the brain uses distinct regions to represent trust (credibility and benevolence) and distrust (discredibility and malevolence).

Overall, the study provides insightful supporting evidence for the claim that trust and distrust are qualitatively distinct phenomena and distrust is not just the absence of trust.

2 Abductive Approach to Cyber Trust and Suspicion

The benevolence-malevolence dimension underlying trust indicates that a person who trusts is willing to be vulnerable to another person who is being trusted based on the belief that whatever the trustee does will not harm the trustor. Such a motivational or intentional inference is possible because of a critical human mental function called theory of mind (ToM), which refers to a person' ability to perceive and reason about others' mental states such as beliefs, desires, intentions, and feelings [16].

Inference for motivation or intention is therefore critical for cyber trust and suspicion. In an earlier effort we explored how such an inference can help a cyber attacker to deliver completely covert attacks [17]. Consider the following scenario: It is 12am and that John, an analyst, is working on a sensitive document on his computer and you have delivered a virus to his computer in order to take a peek. Ideally, you would like your operation is completely invisible to John, but unfortunately, one inevitable side effect of your virus is that John's computer becomes slow, which *John eventually notices and starts to become* ***suspicious***. Then John receives an alerting pop-out message informing him that the antivirus software on his computer has started scanning as scheduled and that so far no virus has been found. John now understands why his computer becomes slow, is relieved, and continues to work on his document, without realizing your peeking eyes.

Though hypothetical, this example highlights an important aspect of cyber trust and suspicion, which has to do with an understanding of how a human operator reasons and explains unexpected observations and if and when the operator becomes suspicious. We call this an abductive approach since it is based on a powerful inference type called abduction [18]. The general form of abduction is shown below,

A fact C is observed,
H can explain C;
Hence, H may be true.

Charniak and McDermott [19] characterize abduction as modus ponens turned backward. Modern researchers often regard abduction as a complex process of finding a best explanation for a set of observations [20,21]. Since "explaining" is an inevitable aspect of human everyday activities, abductive reasoning is almost ubiquitous. In battlefields, commanders have to infer the enemy's motivations based on observations and intelligence and then take proper actions. In cyberspace security, operators may have to infer if an attack has occurred given observations. We therefore argue that an abduction-based framework provides a psychologically plausible and computationally tractable solution for understanding and modeling cyber trust and suspicion.

Acknowledgments. This work is supported by an AFOSR grant (FA9550-12-1-0457, 26-0302-50-65) and an ONR grant (N00014-08-1-0042).

References

1. Andress, J., Winterfeld, S.: Cyber warfare: tachniques, tactics and tools for security practitioners. Syngress, Waltham (2011)
2. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. Int. J. Hum.-Comput. Stud. 58, 697–718 (2003)
3. Jian, J., Bisantz, A.M., Drury, C.G.: Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics 4, 53–71 (2000)
4. Liu, L., Chen, S., Yan, G., Zhang, Z.: BotTracer: Execution-based Bot-like Malware Detection. In: Proceedings of the 11th Information Security Conference, Taipei, China (2008)
5. Barber, B.: The logic and limits of trust. Rutgers University Press (1983)
6. Misztal, B.A.: Trust in Modern Societies: The Search for the Bases of Social Order. Polity Press, Cambridge (1996)
7. Sztompka, P.: Trust: Cultural Concerns. International Encyclopedia of the Social & Behavioral Sciences, 15913–15917 (2002)
8. Sinaceur, M.: Suspending judgment to createvalue: Suspicion and trust in negotiation. Journal of Experimental Social Psychology 46, 543–550 (2010)
9. Deutsch, M.: The effect of motivational orientation upon trust and suspicion. Human Relations 13, 123–139 (1960)
10. Deutsch, M.: Trust and suspicion. J. of Conflict Resolution 2, 265–279 (1958)
11. Bagherir, E., Ghorbani, A.A.: Exploiting trust and suspicion for real-time attack recognition in recommender applications. In: Etalle, S., Marsh, S. (eds.) Trust Management. IFIP, vol. 238, pp. 239–254. Springer, Heidelberg (2007)
12. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum Factors 46, 50–80 (2004)
13. Leon, P.G., Cranor, L., McDonald, A.M., McGuire, R.: Token Attempt: The Misrepresentation of Website Privacy Policies through the Misuse of P3P Compact Policy Tokens. CMU-CyLab-10-014, CMU (2010)
14. Lyons, J.B., Stokes, C.K., Eschleman, K.J., Alarcon, G.M., Barelka, A.J.: Trustworthiness and IT suspicion: an evaluation of the nomological network. Hum Factors 53, 219–229 (2011)
15. Dimoka, A.: What does the brain tell us about trust and distrust? evidence from a functional neuroimaging study. MIS Quarterly 34, 373–396 (2010)
16. Baron-Cohen, S.: Mindblindness: An essay on autism and theory of mind. MIT Press, Cambridge (1995)
17. Wang, H., Sun, Y.: An Abductive Approach to Covert Interventions. In: Proceedings of the 34th Annual Conf. of the Cognitive Science Society, Cognitive Science Society (2012)
18. Fann, K.T.: Peirce's theory of abduction. Martinus Nijhoff, The Hague (1970)
19. Charniak, E., McDermott, D.: Introduction to artificial intelligence. Addison-Wesley Publishing Company, Reading (1985)
20. Thagard, P.: Conceptual revolutions. Princeton University Press, Princeton (1992)
21. Josephson, J.R., Josephson, S.G.: Abductive inference: Computation, Philosophy, Technology. Cambridge University Press, Cambridge (1994)