

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the
Affiliated Conferences, AHFE 2015

Trust as a human factor in holistic cyber security risk assessment

D. Henshel^a, M. G. Cains^a, B. Hoffman^b, T. Kelley^c

^a*School of Public and Environmental Affairs, Indiana University, Bloomington, IN*

^b*U.S. Army Research Laboratories, Aberdeen Proving Ground, MD*

^c*Psychological and Brain Sciences, Indiana University, Bloomington, IN*

Abstract

Holistic assessment of cyber security risks is a complex multi-component and multi-level problem involving hardware, software, environmental, and human factors. As part of an on-going effort to develop a holistic, predictive cyber security risk assessment model, the characterization of human factors, which includes human behavior, is needed to understand how the actions of users, defenders, and attackers affect cyber security risk. The work group developing this new cyber security risk assessment model and framework has chosen to distinguish between trust and confidence by using "trust" only for human factors, and "confidence" for all non-human factors (e.g. hardware and software) in order to reduce confusion between the two concepts within our model. We have developed an initial framework for how to incorporate trust as a factor/parameter within a larger characterization of the human influences (users, defenders and attackers) on cyber security risk. Trust in the human factors is composed of two main categories: inherent characteristics, that which is a part of the individual, and situational characteristics, that which is outside of the individual. The use of trust as a human factor in holistic cyber security risk assessment will also rely on understanding how differing mental models and risk postures impact the level trust given to an individual and the biases affecting the ability to give said trust.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of AHFE Conference

Keywords: Confidence; Cybersecurity; Expertise; Parameters; Model

1. Introduction

When considering digital networks as a space of conflict, traditional characteristics of conflict are overturned. Where in typical combat situations the defender has the advantage as they know the lay of the land and are able to set up adequate defenses ahead of time, the cyber attacker has the lead in digital realms. Attacks are easier to design, create, and launch from origins of the attackers' choosing, while cyber defense efforts instead are challenged with

predicting and detecting attacks while also attempting to develop new and effective defensive techniques. As a result, cyber defense is often very reactive in nature, where attackers set the pace in what becomes a chasing game [1,2]. Furthermore, computational modeling informing organizational planning and execution relies on a fairly complete understanding of a domain, which is an unrealistic goal for cyber security wherein the problem space is so large [3] exploits can be automatically generated [4] as to be nearly impossible to predict or initially comprehend [5].

Cyber security risk assessment has been narrow in focus and based on a business risk assessment approach [6,7,8]. However, given a defensive environment, cyber security risk needs to be more holistic, taking into account the user, information technology analyst, defender, and attacker. Cyber security risk assessment needs to consider the impacts well beyond the computers and network itself. To that end we have taken the 1996 Presidential Congressional Commission Framework for Risk Management [9] which incorporates standards from the environmental and human health risk assessment and framework and have applied those principles to the framing of cyber security risk assessment process. The adapted framework is as follows: context and problem formulation, system and human state assessment, threat assessment, characterization of risks posed to assets, agility options generated and decision-making process, action and agility, reassess state of system and humans, and finally determine where system is secure or additional security is needed. The successful implementation of this holistic cyber security risk assessment requires the development of measurement metrics for the information security attributes of confidentiality, integrity, and availability. A further challenge is determining what responses should be automated, and thus potentially subject to manipulation by attackers, and which decision-making remains fundamentally human. A behavioral component of the cyber security risk assessment accounts for the bounded rationality of human agents and for noisiness of the environment and decision-making process.

In developing a holistic cyber security risk assessment, the Army Research Laboratory Cyber Security Collaborative Research Alliance (CSEC CRA) aims to create a risk assessment framework that enables predictive and proactive defenses. The holistic assessment of cyber security risks is a complex multi-component and multi-level problem involving hardware, software, environmental, and human factors. As part of the on-going efforts to create this assessment model, the characterization of human factors, which includes human behavior, is needed to understand how the actions of users, defenders, and attackers affect cyber security risk. Trust and confidence are two similarly defined terms that are used to characterize the adherence to expected performance for hardware, software, and humans. This paper will focus primarily on trust given to defenders, but will also identify differences in characteristics between trust in defenders, trust in users, and trust in attackers.

1.1. Trust versus confidence

To increase clarity in the nascent stages, the work group developing this new cyber security risk assessment model and framework within the CSEC CRA has chosen to distinguish between trust and confidence by using "trust" only for human factors and "confidence" for all non-human factors (e.g. hardware and software). In this use of the terminology, then, there is confidence that a system or resource is functioning as expected, and there is trust placed in a person that they are performing their expected tasks and duties in a timely manner. Both confidence in non-human parameters and trust in humans can be considered as gradients.

This gradient of trust is affected by perceptions of all involved parties. In a business network system, for example, the IT manager and analyst trusts the users to use and interact with the network safely, while the user trusts the defender to keep the system hardened and free of malware. Within a cyber defense team, different members of the team have to trust the other defenders to effectively, efficiently, and accurately conduct their work as a part of the defense team. Within the context of business and marketing, two of the main elements of trust focus on reputation and credibility [10], two key components of human trust.

Characterizing the trust component of a holistic cyber security risk assessment allows for the incorporation of humans as positive and negative risk factors. Positive risk factors are factors that increase risk, while negative risk factors are factors that decrease risk. The degree to which a defender is a risk factor can be represented by the amount of trust given to the defender by superiors, the true intentions of the defender, and other inherent knowledge-based and behavioral characteristics.

With respect to understanding humans as risk factors, it is difficult to paint a clear picture of the attacker beforehand, aside from the obvious fact that these humans present positive risk factors. A malicious user or foreign party is not going to make themselves known prior to an attack. Poorly designed policies that work counter to trusted workers' goals [11], sparsely monitored systems that allow trusted users to circumvent established policies [12], or workers that are unaware of risks [13,14] have the potential to create insider threats. Insider threats are particularly insidious because most security concerns are focused on external threats [15]. Thus, insiders are more likely to go unnoticed even after an attack has been detected; remaining unseen until the analysis of the successful intrusion is complete.

In addition to malicious users--insiders or external--end users lack the ability to correctly evaluate potential risks [16]. While this inability to correctly identify risk is modulated to some extent by the end users' experience and perceived risk [17], the tools provided to users' to inform their decisions are inadequate [18]. Moreover, users lack awareness of the information available to them [19], do not have the practical security experience, have been conditioned to accept information presented to them in a digital manner [20], and have an inflated trust in digital entities which drives them to override warnings [21].

As a result, it is best to assume and plan for the worst with regards to both attackers and end users. Defenders must assume that their enemies are at least as clever as they are and that end users in their system are, in general, vulnerabilities in order to prepare as best as possible. Therefore, the onus falls on the defenders to generate and maintain trust amongst themselves and with their users to push toward maximizing positive factors with minimal negatives. How teams within cyber defense roles accomplish this relies on a combination of the individual defenders' skills, the communication of the team(s), and use of tools approaching an optimal result as best as possible.

The nature of cyber defense work requires human agents to sift through vast quantities of data, implying a set (or sets) of qualifications necessary to warrant trust in a person as a capable defender who can understand and interpret the various information presented by systems and communicated among defense teams. In observing and conducting a cognitive task analysis of cyber network defense (CND) professionals across seven organizations, D'Amico and Whitley [2] noted that while nuances and details differed, the overall missions and conduct overlapped and shared much in common. These analysts all had to sift through large amounts of data, drilling down from general information and alerts into the specific details of incidents and traffic in order to make judgment calls on what should be reported, what category or categories were relevant, and what actions were appropriate. Even assuming a level confidence (i.e. faith in the systems used) for all of the involved organizations, their observations highlight the relevance and importance of trusting these analysts to use their tools effectively, leverage communication appropriately, and produce accurate and timely reports in order to support defensive efforts.

1.2. Overview of the trust framework

We have developed an initial framework for how to incorporate trust as a set of factors or parameters within a larger characterization of the human components (users, defenders and attackers) within cyber security risk (Figure 1). The goal of the trust framework is to enable a quantitative or semi-quantitative characterization of trust in the humans who interact with networks within a cyber security risk assessment model.

Trust in the human factors is contributed by two main categories of factors: inherent characteristics, which are a part of the individual or "given" to the individual by the trust-giver, and situational characteristics, which are external to the individual. Inherent characteristics are further separated into two categories: behavioral – which captures rationality, benevolence/malevolence and integrity, and knowledge—which captures expertise and attention-related factors. Situational characteristics capture the degree of insider access which is determined by access level determined by user policy, software, and hardware. Trust itself is captured by reputation, based on public reputation and personal interactions, which can be broken down into credibility, perceived honesty, and predictability.

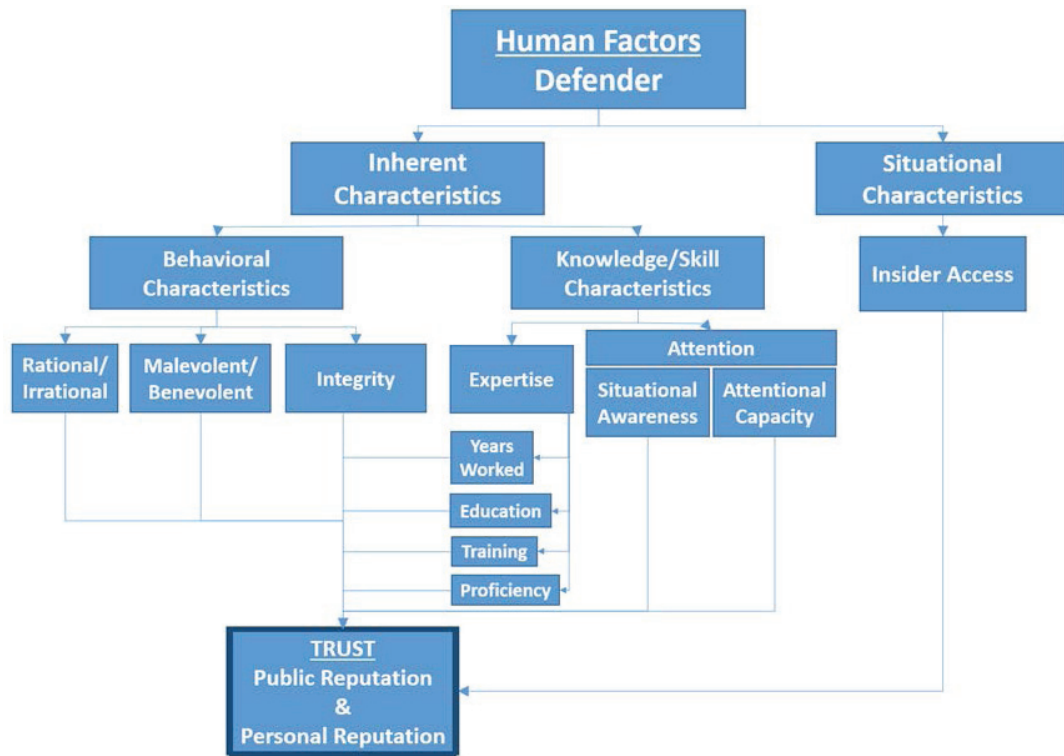


Fig. 1. Defender/analyst trust framework.

1.3. Inherent characteristics

Inherent characteristics affecting trust include behavioral and knowledge characteristics. The behavioral characteristics include intention - expressed as a scale of benevolence to malevolence, rationality, and integrity, both of which affect predictability. The knowledge characteristics include both the expertise factors (reflecting experience, education and other training, and measures of proficiency) and attention, which is affected by and contributes back to expertise, particularly experience and training. The attention parameter incorporates both situational awareness and attentional capacity. All human characteristics can change in a given individual over time. They are affected by experience, and both internal and external factors, although people start from a different baseline for each behavioral characteristic.

1.3.1. Behavioral characteristics: rationality factors

More rational people tend to behavior and respond more predictably which leads to increased trust. Rationality is affected by innate reasoning, ability, but is also affected by emotions and experiences, that is what is happening in one's life. For example most people behave more rationally when unstressed but can start to make more irrational decisions when highly stressed [22]. Physical factors can also affected rationality. For example, hypoglycemia can cause people to behave irrationally.

1.3.2. Behavioral characteristics: malevolence factors

Malevolence implies evil or malicious intent. Malevolent behavior can be rational or irrational. An intent to harm can be driven by irrational emotions and can be induced transiently or for whatever period of time the inducing stressor exist. For example, seemingly rational, calm people who apparently have no intent to hurt other people can be taken over by an irrationally, malevolent intent that is typically called road rage. In intergroup conflict between self-identifying groups such as those defined by ethnic, political, or religious characteristics, members can choose to

harm “other” rationally if they see it as protecting or promoting their own group when they perceive their group is threatened by “other” [23].

1.3.3. Knowledge characteristics: expertise

Knowledge characteristics are those that change over time and contribute a great deal through experience to building credibility. Expertise characteristics include education, training, experience as partial quantified by years worked in performed related tasks, and are measured by metrics assess proficiency. A detailed analysis of these measures are presented in Figure 2 and discussed below (section 2.).

1.3.4. Knowledge characteristics: attention

Attention has two contributing components: i) situational awareness, and ii) attentional capacity. Situational awareness is the extent to which a person perceives and understands the details and events occurring around them, and the ability to reason through the possible sequelae stemming from the known and possible dynamic changes [24]. Attentional capacity is reflected by the length of time a person can maintain highly attentive state, and the amount of detail a person can track consistently for an extended period of time. While expertise is critical to the development of accurate and well-reasoned responses to a given situation, such as when an analyst is faced with indicators of malware threats, attention is more critical to being able to apply that information under the typically stressful, extended attack scenario. Whereas experience and training increase both situational awareness and attentional capacity there is some inherent reasoning capacity (baseline) that make some people inherently highly skilled and effective at situational awareness processing and or having extensive attentional capacity

1.3.5. Situational characteristics: insider access

Insider access is not authorized without initial trust and the trust must be maintained in order to gain additional insider access. The additional trust is gained by continued evidence of various characteristics of expertise, perceived benevolence, and perceived rationality. One problem with insider access, is that it is not always given (by policy, hardware, or software) but can be taken or forced by physical, hardware, software, or other invasive techniques. The increasing acceptance of “Bring Your Own Device” of course increases the potential for less control of access.

2. Use of the trust framework within cyber security risk assessment

Increased trust is given to a defender who can effectively communicate with superiors and other defenders, is able to accurately log incident reports (minimize the number of false positive and false negative reports), is able to provide and relay information in a timely manner, and is able to use cyber defense tools as intended with competency.

2.1. Communication

Whenever two or more parties are involved in communication, if they establish a common ground and build upon shared mental models they will be more effective and efficient [25,26]. As a result, the interactions and collaborations shared among them should benefit and proceed as best as possible. Focusing on defender trust, we separate communication into the subcategories of accuracy, thoroughness or completeness, timeliness, and honesty. Effective communication relies on a defender to accurately convey information with thoroughness and in a timely manner, affected by the amount of information, how tied the information is to a specific context, and how well the parties involved in the communication understand one another. Whether or not communication is sufficiently thorough relies on how well the information is processed by the receiver with minimal need to repeat or revisit previous information [25]. For cyber defenders, timeliness is essential as any wasted time might give an attacker more opportunity to do damage or an attack the chance to escape detection. Lastly, honesty is important in building trust regardless of space, and dishonest communication will not only harm team effectiveness in the cyber domain but might harm how well and how accurately defensive efforts address intrusions.

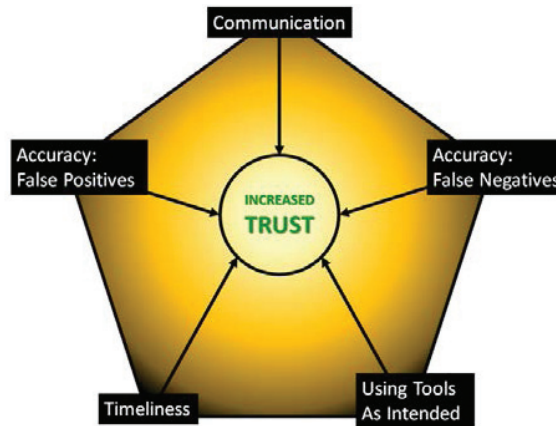


Fig. 2. Factors affecting trust in defender/analyst.

2.2. Accuracy

The detection accuracy of a defender can be measured by the percent of false positive and false negative incident reports filed by an individual defender. It would be expected that a skilled defender with expertise would have lower false positive and false negative rates than those of a less skilled defender. The ability to trust the work the defender produces will increase as the percentage of false positive and false negative incident reports decrease.

2.3. Timeliness

Just as in communication timeliness, the timeliness of a defender's actions is critical to achieving successful agility actions and mitigating the effects of attackers' actions. Timeliness considers: i) how quickly a defender is to detect intrusions, ii) how quickly a defender is to relay critical and time-sensitive information to their superiors, and iii) how quickly the defender chooses appropriate, effective agility or mitigation actions.

2.4. Using tools as intended

For cyber security, intended tool use focuses on the competent matching of tools and tool feature use to task completion, be it in the detection of intrusions, reactions to successful attacks, or hardening of the network against potential future attacks. There are two aspects of tool use that can be evaluated: the effectiveness of training for defenders and the performance of defenders in their day to day work.

Training for a cyber defender must include a combination of interacting with software and IT tools to establish familiarity with their function and relation to defensive tasks and learning and adhering to the policies that govern defense efforts. Training with software for cyber defenders takes two different forms: i) tool-based, focused on reviewing features and functions before running through exercises that use the discussed elements, and ii) narrative-based, where tool and software functions are discussed within the context of adversary tactics and techniques [27]. The evaluation of training can be seen in performance, as superior performance reflects superior training.

Performance can be directly measured by combining the timeliness and accuracy elements discussed above. All things considered equal in a cyber environment, if two defenders report on the same intrusion with different quickness and accuracy the root cause is likely performance differences with tools. These differences may reflect training issues and knowledge gaps. For example, Stevens-Adams et al. [27] separated defenders into three teams, one getting narrative training, one tool-based training, and the third with a mixture of members who either received one or the other; after five days of training and three days of exercise the narrative-based team scored the highest of

all teams, and the individuals with narrative-based training scored higher than the other individuals. Here, a controlled environment simulating real-world usage enabled team and individual performance to assess training and knowledge. Where these metrics reveal performance differences, the details of a CTA, constructing an understanding of the ways with which analysts complete work [2], also enable an evaluation of these methods and identify where tool use could be improved or where training should be re-tooled and updated.

3. Discussion: Building trust

In this context, trust in a rational and benevolent user is thought to be dependent on their historical compliance with cyber security policies, their level of security education, the number of successfully completed security awareness programs, experience in their occupation, and risk posture. It is important to note that using policy compliance as measure of trust would group benevolent users who used non-malicious means to circumvent onerous policies in order to complete tasks and duties. Trust in defenders is thought to be a function of historic effective intrusion prevention, intrusion detection accuracy and efficiency, and impact mitigation. Trust in an efficient defender is dependent upon on their education, skill, experience, and personal biases such as risk posture. Characterizing trust in attackers is a measure of consistency and predictability of their actions. Trust in the attacker enables one to better predict the risk itself in the network. The component of trust in attackers will be the most difficult of the three human factors to characterize due to general lack of attacker information availability.

The nature of cyber defense work requires human agents to sift through vast quantities of data, implying a set (or sets) of qualifications necessary to warrant trust in a person as a capable defender who can understand and interpret the various information presented by systems and communicated among defense teams. In observing and conducting a cognitive task analysis of CND professionals across seven organizations, D'Amico and Whitley noted that while nuances and details differed the overall missions and conduct overlapped and shared much in common [2]. These analysts all had to sift through large amounts of data, drilling down from general information and alerts into the specific details of incidents and traffic in order to make judgment calls on what should be reported, what category or categories were relevant, and what actions were appropriate. Even assuming a level confidence (i.e. faith in the systems used) for all of the involved organizations, their observations highlight the relevance and importance of trusting these analysts to use their tools effectively, leverage communication appropriately, and produce accurate and timely reports in order to support defensive efforts.

The proposed framework is built from a combination of literature discussions about trust as well as from discussions with working cyber defense team members. McKnight and Chervany [28] map the definitions and characteristics of trust to: competence, benevolence, integrity, and predictability, while grouping conceptual types of trust into: disposition, structural, affect or attitude, belief or expectancy, intention, and behavior. Hancock et al [29], in an analysis of factors affecting trust within human-robot interaction, identify the ability-based factors that affect trust as: attention capacity and engagement, situational awareness, expertise, competency, workload, and experience. Blausch [30] divides human user trust into behavioral or psychological components and privilege (insider access). The behavioral components include predictability in response or decision making, reliability or consistency, competence, and responsibility. In Blausch's categorization for information fusion, which is based heavily on Muir and Moray's categorization for process control [31], competence integrates the knowledge components, experience and reputation; yet there is a separate component of faith or belief in the person which incorporates historical experience, reputation, and historical loyalty.

Denning [32] identifies that trust needs to be based on a standard, and will usually be given within a limited domain. For example, a CND team member may be very experienced with one or a few tools, and so trusted to work with those tools on a team, but may (at a given point in time), not be trusted to take on a different tasks on the team. A good scanner of logs may not be a good forensics person without additional training and experience. Trust is dynamic, and as such needs to be continuously assessed. Experience and expertise can be readily quantified (see Figure 2 for expertise), and history of performance (accuracy, efficiency) can be assessed over time. Yet the reputation component is harder to quantify, and harder to assess continuously, as needed to be fit into a dynamic model. Further, trust once lost is harder to regain, as has been observed across multiple disciplines [33, 34, 35]

In the end this framework needs to be developed to the point where we can establish testable parameters that we can fit into the holistic cyber security risk assessment. The framework was developed at this point primarily from the literature as well as discussions with a broad range of cyber security researchers and practitioners. The CSec CRA research group is current developing models and experiments to validate the holistic cyber security risk model. Some studies are already underway to test and evaluate the human trust component of the broader framework.

Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

- [1] M. Branlat, A. Morison, D.D. Woods. Human Systems Integration Symposium, Vienna VA, 2011, pp. 10-25.
- [2] A. D'Amico, K. Whitley. VizSEC 2007: Proceeding of the Workshop on Visualization for Computer Science. Springer:Berlin, 2008, pp.19-37.
- [3] R. Anderson, in: Comput. Secur. Appl. Conf. 2001. ACSAC 2001. Proc. 17th Annu., IEEE, 2001, pp. 358–365.
- [4] S. Noreen, S. Murtaza, M.Z. Shafiq, M. Farooq. In Proceedings of the 11th Annual ACM conference on Genetic and evolutionary computation (2009) 1569-1576.
- [5] L. Bilge, T. Dumitras. In Proceedings of the 2012 ACM conference on Computer and communications security (2012) 833-844.
- [6] National Institute of Standards and Technology (NIST). Framework for Improving Critical Infrastructure Cybersecurity, 2014.
- [7] P. Mell, K. Scarfone, S. Ramonosky. CVSS:A complete guide to the Common Vulnerability Scoring System version 2, 2007.
- [8] C.J. Alberts, A.J. Dorofee. OCTAVE Method Implementation Guide: Version 2.0, 2001, Carnegie Mellon University.
- [9] Risk Commission. Presidential/Congressional Commission on Risk Assessment and Risk Management. A framework for environmental health risk management: Vol 1, 1997, US Government Printing Office, Washington, DC.
- [10] P. Herbig, J. Milewicz. J. Consum.Market. 12 (1995) 5-10.
- [11] C. Herley, Security (2009) 133.
- [12] H.-S. Rhee, C. Kim, Y.U. Ryu, Comput. Secur. 28 (2009) 816.
- [13] D. Liu, X. Wang, L.J. Camp, in: R. Dingleline, P. Golle (Eds.), Financ. Cryptogr. Data Secur., Springer Berlin Heidelberg, 2009, pp. 1–16.
- [14] D. Besnard, B. Arief, Comput. Secur. 23 (2004) 253.
- [15] D.B. Baker, in: Proc. 1992-1993 Work. New Secur. Paradig., ACM, 1993, pp. 126–130.
- [16] R. West, Commun. ACM 51 (2008) 34.
- [17] M. Arianezhad, L.J. Camp, T. Kelley, D. Stebila, in: Proc. Third ACM Conf. Data Appl. Secur. Priv. - CODASPY '13, ACM Press, New York, New York, USA, 2013, p. 105.
- [18] S.E. Schechter, R. Dhamija, A. Ozment, I. Fischer, in: 2007 IEEE Symp. Secur. Priv., IEEE, Los Alamitos, CA, USA, 2007, pp. 51–65.
- [19] J. Sunshine, S. Egelman, H. Almuhammedi, N. Atri, L.F. Cranor, in: Proc. 18th USENIX Secur. Symp., 2009.
- [20] B. Rainer, K. Stefan, (2010) 2403.
- [21] A. Hazim, A.P. Felt, R.W. Reeder, S. Consolvo, in: Symp. Usable Priv. Secur., 2014.
- [22] M.A. Staal, Stress, cognition, and human performance: A literature review and conceptual framework. NASA/TM—2004-212824 (2004).
- [23] E.R. Smith, D.M. Mackie, H.M. Claypool, in: Social Psychology (4th Ed), 2014, Psychology Press, pp 482 - 526.
- [24] A.A. Nofi. Defining and measuring shared situational awareness. No. CRM-D0002895. A1. Center for Naval Analyses, Virginia 2000.
- [25] A.Monk, A. in: J. Carroll (Ed), HCI Models, Theories, and Frameworks: Towards a Multi-disciplinary Science. Morgan Kaufmann, San Francisco, 2003, pp. 265-289.
- [26] J.E. Mathieu, T.S. Heffner, G.F. Goodwin, E. Salas, J.A. Cannon-Bowers Journal of Applied Psychology, 85 (2000) 273.
- [27] S.M. Stevens-Adams, A. Carbajal, A. Silva, K. Nauer, B. Anderson, T. Reed, J.C. Forsythe in: Foundations of Augmented Cognition (pp. 90-99), Springer Berlin Heidelberg.
- [28] D.H. McKnight, N.L. Chervany in: R. Falcone, M. Singh, Y.-H. Tan (Eds). Trust in Cyber-societies: Integrating the Human and Artificial Perspectives (Lecture Notes in Computer Science vol. 2246) Springer New York, 2001, pp 27-54
- [29] P.A. Hancock, D.R. Billings, K.E. Schaefer. Human Factors 53(2011) 517-527.
- [30] E. Blasch, in: M. Blowers, J. Williams, Machine Intelligence and Bio-Inspired Computation: Theory and Applications VIII, Proc. SPIE 9119, 2014, pp. 9119OL-1-11.
- [31] B. Muir, N. Moray, Ergonomics, 39 (1996) 429-460.
- [32] D.E. Denning. In Proceedings on the 1992-1993 workshop on New security paradigms (1993). 36-41.
- [33] D.L. McLain, K. Hackman Public Admin. Quart. 23 (1999) 152-176.
- [34] M.E. Schweitzer, J.C. Hershey, E.T. Bradlow, Org. Behavior. Human. Decision Proc. 101 (2006) 1-19.
- [35] P. Slovic, Risk Analysis 13 (1993) 675-682.