Journal of Information Warfare

Volume 16, Issue 2
Spring 2017

Contents

Understanding Operator Fatigue, Frustration, and Cognitive Workload in Tactical Cybersecurity Operations

CL Paul and J Dykstra

Research Directorate
National Security Agency, U.S.A.
E-mail: clpaul@tycho.ncsc.mil; jdykstra@ltsnet.net

Abstract: *While the human factors of mission critical systems such as air traffic control and weapons systems have been extensively studied, there has been little work on cyber operations. As with any system, the perfect storm of complex tasks in a high-risk environment takes an incredible toll on human operators, leading to errors, decreased performance, and burnout. An extensive study of tactical cyber operations at the National Security Agency found that operator fatigue, frustration, and cognitive workload significantly increase over the course of an operation. A discussion of these findings helps us understand the impact that the high-stress, high-risk environment of tactical cyber operations has on its operators.*

Keywords: *Cyber Operations, Cognitive Workload, Fatigue, Frustration, Burnout, Human Factors, Cybersecurity*

Introduction

Cybersecurity operations are a mission-critical service for the safety and business continuity of companies and organizations in the digital world. From red team network penetration testing to real-time defensive monitoring, evolving technology and threats to the network make cybersecurity operations high-value, complex, and difficult. This environment is considerably high-risk, and success or failure can greatly affect the mission or reputation of an organization. Research and development for cybersecurity operations has heavily focused on technological means of achieving a more secure enterprise. However, it is the human experts who play the most critical role in the deployment, configuration, monitoring, and operation of networks.

The National Security Agency (NSA) coordinates, directs, and performs highly specialized activities to protect U.S. government information systems and to produce foreign signals intelligence. One of NSA’s missions is to defend the Department of Defense Information Network (DODIN), National Security Systems (NSS), and other critical U.S. government systems. Intelligence analysts and network operators work together around the clock to detect, assess, and prevent foreign threats to networks. In addition to its headquarters in Maryland, NSA has cryptologic centers in Colorado, Georgia, Hawaii, and Texas that also conduct foreign signals intelligence, cyberspace operations, and information assurance operations.

NSA recruits and hires computer network operators to both defend U.S. military networks and to exploit the networks of foreign adversaries. For these jobs, NSA seeks people with competencies that include operating systems and network analysis, network penetration testing, intrusion detection and incident response, digital forensics, critical thinking and problem solving. Since new operators must complete extensive training and satisfy an ongoing skills certification process, the investment in operator training and certification makes early career burnout and turnover very costly.

Additionally, the NSA provides direct intelligence and technical support to the U.S. military through U.S. Cyber Command. Cyber Command's mission is to conduct integrated tactical cyber operations in defense of the nation (U.S. Department of Defense 2015). By 2018, Cyber Command plans to staff the Cyber Mission Force with teams that will defend the United States and its interests (National Mission Teams), defend Department of Defense networks and systems against threats (Cyber Protection Teams), and provide operational planning and support to Combatant Commands (Combat Mission Teams). These cyber operations specialists will be responsible for executing a variety of offensive and defensive cyberspace operations in support of military objectives.

There is a long tradition of human factors research focused on studying risk, error, and burnout in a variety of domains. In a meta-analysis of cognitive workload, Grier (2015) found that air traffic control, command and control, and medical tasks were among the most demanding cognitive domains. These tasks rely heavily on cognitively demanding tasks that utilize attention, memory, and visual perception. An analysis of friendly fire incidents among U.S. troops in the 1991 Gulf War found that stress and anxiety combined with poor situation awareness were major factors that contributed to these incidents (U.S. Army 1992). Work by Richter *et al.* (2005) found that vigilance tasks are greatly affected by task length, and that longer vigilance leads to increased fatigue and errors. In a study of short-haul airline pilots, Powell *et al.* (2007) found a strong relationship between fatigue and flight length; they further determined that the time a flight began had a significant impact on the amount of flight-induced fatigue.

However, the relationship between these stressors in a high-risk environment cannot always be tied to a specific outcome. Kowalski-Trakofler and Vaught (2003) found in their study of emergency responders that judgment is not always compromised under stress and that the effects of stress on judgment and cognition are highly contextual. In 2005, van der Linden *et al.* found that burnout in a variety of knowledge work domains is closely related to mental exhaustion due to stress. Later related work by Hopstaken *et al.* (2014) found a strong connection between mental exhaustion and task performance, regardless of the rewards. The insights gained from these related domains are informative, but their results can only be generalized to cybersecurity so far.

Although there is little work specifically studying the factors that contribute to error and burnout in our type of tactical cybersecurity operators, related work has investigated contributing factors in more defensive, reactive cybersecurity environments. An Air Force Research Laboratory study of cyber warfare operators (Chappelle *et al.* 2013) attributed occupational stress to shift work, shift changes, and hours worked, while cyber defense stressors such as tasks related to defending networks from real-time attacks were not listed as primary stressors. Although they did not study stress specifically, Sawyer *et al.* (2014) found that cognitive workload of cyber defense was greatly affected by the length of vigilance required to complete a task. Along those lines, a study of incident responders reported increased stress from a combination of urgency to respond and the hyper vigilance associated with not wanting to give away position or strategy to a network intruder who might still be present on the network (Haney & Paul 2016). In a study of more specific network defense tasks, Greenlee *et al.* (2016) identified differences in cognitive workload and stress between tasks related to initial network intrusion triage and subsequent escalation analysis.

It is the operators in a tactical role who are closest to the network and who take on the most risk that are the most concerning. Network complexity, adversary skill, and adversary knowledge contribute to cybersecurity risk (Cowley, Greitzer & Woods 2015) and are all qualities of the environment in which operators work. While high risk is undertaken in search of high reward, the cost of risk goes beyond the outcome of the mission and extends to the impact on human capital. Operator error and mission success are easily measured; however, operator fatigue, frustration, cognitive workload are equally important. These factors can contribute to technical errors or personnel burnout (Sundaramurthy *et al.* 2015) and ultimately lead to an increased mission risk from insufficient and inexperienced staffing due to constant turnover.

Dynamic environments are those where the nature of a task may change independently of the participant’s actions. Domains such as cybersecurity are incredibly complex given the high number of independent environmental variables. Abnormal situations in complex environments create acute conditions of stress that often subside when the situation returns to normal. While a pilot’s stress is closely tied to the protection of life and property, a cybersecurity operator’s stress is largely related to the protection against discovery and safety of sensitive tools and techniques.

The goal of this research has been to understand cybersecurity operator stress and to measure fatigue, frustration, and cognitive workload. Unlike related work that has focused mostly on reactive cyber defense situations, this research is focused on operators who work missions that are more proactive in nature. Previous reports of this work described factors specific to the NSA environment (Dykstra & Paul 2015). This paper focuses on providing a generalized picture of operator fatigue, frustration, and cognitive workload so that others with similar cybersecurity missions can learn from this study.

Study Methodology

The goal of this study was to develop a deeper understanding of the human factors in cyber operations. Study participants were cybersecurity specialists in an NSA mission office focused on tactical computer network operations from both the civilian (NSA) and military (U.S. Cyber Command) services. These operators were highly trained and skilled at using a variety of tools and techniques to achieve a specific goal or effect on a network. Most participants (85%) had at least two years of experience as an operator.

Because studying the cyber mission environment is particularly challenging (Paul 2014), a methodology that required low time cost and was nonintrusive was chosen in order to minimize disruption to the missions studied. Pre- and post-operation assessments were designed to capture the effects an operation has on the fatigue, frustration, and cognitive workload of an operator. The assessment was a two-page survey that featured three types of questions:

- Samn-Perelli Fatigue Scale (SPFS) (Samn & Perelli 1982) to measure pre- and post-operation fatigue ($\text{Fatigue}_{\text{Post-op}} - \text{Fatigue}_{\text{Pre-op}} = \Delta\text{Fatigue}$) on a 20-point scale.
- NASA Task Load Index (TLX) (Hart 2006) to measure cognitive workload across six factors: Mental Demand, Physical Demand, Time Pressure, Self-assessment of Performance, Effort, and Frustration. This was a post-operation assessment made on a 20-point scale reported as individual factors and as a raw adjusted score (RTLX, Formula 1). Pre-operation frustration was also measured to baseline post-operation frustration ($\text{Frustration}_{\text{Post-op}} - \text{Frustration}_{\text{Pre-op}} = \Delta\text{Frustration}$).
- Additional operation-specific information about the mission environment was also collected (not reported in these results), and an open text area to collect general comments from participants was also provided.

In this paper, RTLX is the sum of cognitive workload factors *f* divided by the product of survey scale *s* and total number of factors *n* that is then multiplied by 100.

$$RTLX = \frac{f_1 + f_2 + \cdots + f_n}{s * n} * 100 \qquad (1)$$

This method normalizes the RTLX mean for different scales (for example, 7-point versus 20-point) and different cognitive work models (for example, all versus a subset of factors), which makes the task of comparing RTLX values across different studies easier.

The assessment was delivered in both paper-based and web-based forms to adjust to the needs of the different data collection environments. Participation in the study was voluntary. First time participants completed a short demographics survey before the operation began. At the start of the operation, operators completed the pre-operation questions, then conducted their operation, and finally completed the post-operation section. Operators completed one survey per operation and could submit multiple surveys for additional operations they conducted over the course of the data collection period.

A total of 126 operators across NSA provided 361 operation assessments. Assessments from NSA headquarters and from three of the four NSA cryptologic centers were collected at different periods within a 14-month window. The data collection period at each site lasted three to five weeks. Partially completed assessments were accepted; however, this data was excluded from pairwise comparisons that required complete cases.

Results

Operator Fatigue and Frustration increased significantly over the course of an operation (**Figure 1**). Fatigue increased by 16% (*SD* = 22%), *student's t*(359) = -13.92, $p < .001$, while Frustration increased by 12% (*SD* = 27%), *Student's t*(336) = -8.51, $p < .001$. There was a weak positive relationship between ΔFatigue and ΔFrustration, *r*(334) = .248, $p < .001$; that is, as Fatigue increased, Frustration also slightly increased.

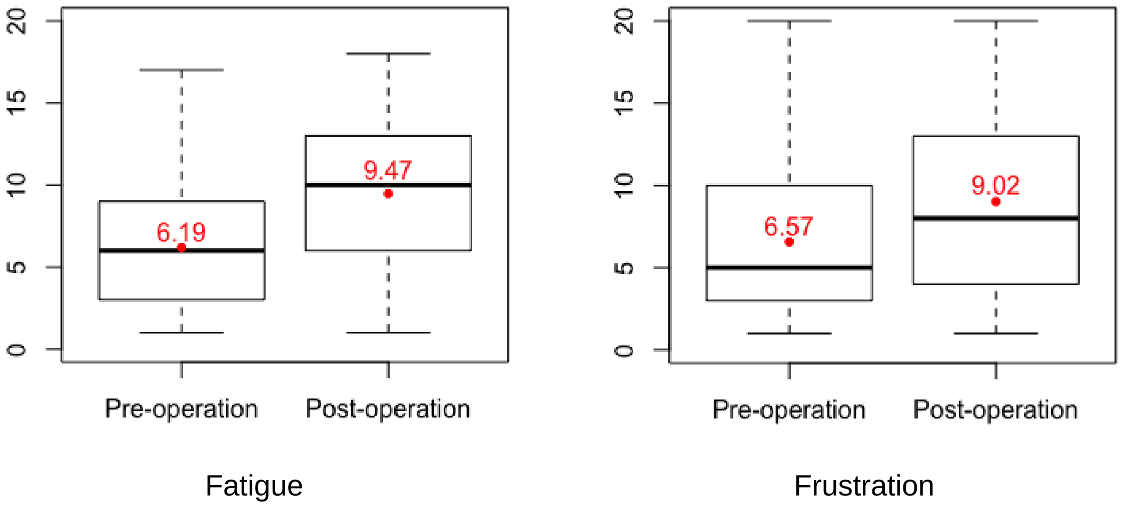

Figure 1: Boxplot with mean points of pre- and post- operation Fatigue and Frustration pairs

In addition to Fatigue and Frustration, operator cognitive workload was assessed across six factors (**Figure 2**). The calculated RTLX score was 44.55 (*SD* = 28.1).

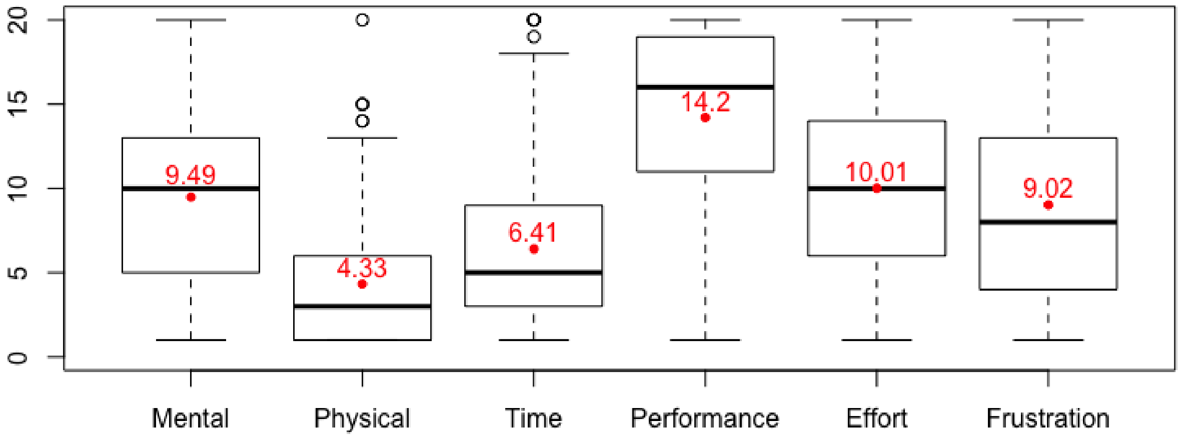

Figure 2: Boxplot with mean points of TLX cognitive workload factors

There were moderate to strong positive relationships between all cognitive workload factors except for Performance (**Table 1**). The strongest relationship was between Mental Demand and Effort; that is, as Mental Demand increased, Effort also increased. The only correlation with Performance was a weak negative relationship with Frustration; that is, as Frustration increased, self-assessments of Performance slightly decreased.

	Mental	Physical	Time	Perf.	Effort	Frust.
Mental	—	.479*	.547*	-.034	.686*	.468*
Physical	.479*	—	.541*	-.012	.486*	.334*
Time	.547*	.541*	—	-.022	.509*	.429*
Perf.	-.034	-.012	-.022	—	-.009	-.315*
Effort	.686*	.486*	.509*	-.009	—	.469*
Frust.	.468*	.334*	.429*	-.315*	.469*	—

Table 1: Pearson coefficient correlations (*r*) between cognitive workload factors (complete cases, $N = 350$, * $p < .001$)

Both ΔFatigue and ΔFrustration had significant relationships with cognitive workload factors (**Table 2**); that is, larger differences in Fatigue or Frustration were indicative of higher cognitive workload.

	Mental	Physical	Time	Perf.	Effort	Frust.
ΔFatigue	.263**	.225**	.162*	-.078	.227**	.173**
ΔFrustration	.238**	.194**	.201**	-.184**	.277**	—

Table 2: Pearson coefficient correlations (*r*) between cognitive workload factors and ΔFatigue (complete cases, $N = 349$, * $p < .01$, ** $p < .001$) and ΔFrustration (complete cases, $N = 328$, ** $p < .001$)

In the case of ΔFatigue, the relationships were quite weak and mirror what was found in the cognitive workload model, especially with the lack of a significant relationship with Performance. However, the relationship between ΔFrustration and Performance, although quite weak, is the first time Performance was seen to be influenced by other factors.

In addition to understanding the relationship between Fatigue, Frustration, and cognitive workload during an operation, researchers also investigated the effects of Operation Length on the operator. The average amount of time that an operator contributed to an operation was 5.12 hours (*SD* = 2.0 hours).

There was a weak positive relationship between the length of an operation and how much operator ΔFatigue and ΔFrustration increased (**Table 3**). Shorter operations had smaller increases in ΔFatigue and ΔFrustration while longer operations had greater effects.

	ΔFatigue	**ΔFrustration**
Operation Length	.361*	.210*

Table 3: Pearson coefficient correlation (*r*) between Operation Length and ΔFatigue and ΔFrustration (complete cases, *N* = 312, **p* < .001)

Operations longer than five hours had 11% higher ΔFatigue, *Welch's t*(270.85) = -4.45, *p* < .001, and 9% higher ΔFrustration, *Welch's t*(265.88) = -2.89, *p* = .004, than operations shorter than five hours.

The length of an operation had a weak to moderate positive relationship with all operator cognitive workload factors except for Performance (**Table 4**). Longer operations resulted in higher Mental Demand, Physical Demand, Time Pressure, Effort, and Frustration. There was no relationship between Operation Length and Performance; that is, longer operations did not result in higher or lower operator performance.

	Mental	**Physical**	**Time**	**Perf.**	**Effort**	**Frust.**
Operation Length	.376*	.253*	.271*	.032	.296*	.176*

Table 4: Pearson coefficient correlation (*r*) between Operation Length and cognitive workload factors (complete cases, *N* = 324, * *p* < .001)

Just as was found with ΔFatigue and ΔFrustration, operations longer than five hours had significant effects on all cognitive workload factors, except for Performance, *Welch's t*(278.39) = 0.32, *p* = .747. Operations longer than five hours had 11% higher Mental demand, *Welch's t*(281.76) = -4.00, *p* < .001, 8% higher Physical demand, *Welch's t*(266.84) = -3.59, *p* < .001, 10% higher Time pressure, *Welch's t*(278.01), *p* < .001, 11% higher Effort, *Welch's t*(283.45) = -3.89, *p* < .001, and 9% higher Frustration, *Welch's t*(282.15) = -2.97, *p* = .003.

Discussion

Because this is the first study of its kind, there was no baseline to qualify observed fatigue, frustration, and cognitive workload. However, these results can be compared to results from other similar work.

While no work that investigates fatigue in cybersecurity was found, other fatigue research provides some context for these observations. Researchers studying workers at a mining processing plant found increases between 14% (day shift) and 17% (night shift) after a 12-hour shift (Baulk *et al.* 2009). Other work investigating the effects of four-hour vigilance tasks saw a 10% increase in fatigue (Richter *et al.* 2005). This study's pre- and post-operation fatigue observations (ΔFatigue *M* = 16%) aligns most with a study in short-haul pilots who reported an increase in fatigue of 14% at the 5-hour flight mark (Powell *et al.* 2007), the mean length of time an operator was active on an operation in this study.

In the case of cognitive workload, Grier's (2015) meta-study of NASA TLX results surveyed more than 200 publications that covered a wide range of cognitive tasks. The current study's RTLX score ($RTLX = 44.55$, $SD = 28.1$) aligns with the mean reported RTLX scores in this meta-study ($RTLX = 45.29$, $SD = 14.9$), as well as other related work in network defense. This includes Champion *et al.'s* 2012 study of incident response with CyberCog ($RTLX = 56.94$, $SD = 21.4$) and Greenlee *et al.*'s 2016 report on network triage ($RTLX = 51.94$, $SD = 14.2$) and incident escalation ($RTLX = 40.04$, $SD = 11.6$) tasks.

The measurement of performance was a self-assessment of how well the operator was able to complete his or her task rather than an unbiased measure of mission success. The subjective nature of the judgment does not always overlap with the objective reality, but that is not the purpose or value of this factor. Subjective performance measures are indicative of a person's 'self-efficacy', which is the judgment of ability or capability to perform a task (Marakas, Yi Va & Johnson 1998). Self-efficacy greatly influences a person's level of persistence during a task, the amount of effort he or she expends during a task, his or her commitment to a goal, and his or her ability to work independently. Negative influences on self-efficacy have direct consequences on task performance and anxiety. Self-efficacy is also strongly linked to locus of control and the feelings of control over one's destiny (Judge *et al.* 2005). Tactical cyber operations reside within a very complex system of people, tradecraft, and infrastructure. While an operator may do everything right, other factors in the environment that are out of his or her control may lead to a mission failure. This lack of control has short-term effects on frustration and job performance and long-term effects on job satisfaction.

Some effects of this were observed within this study. The first was the negative relationship between Performance and Frustration ($r = -.315$, $p < .001$), with increases in Frustration as Performance declined. However, this relationship became quite weak when looking at the relationship between change in Frustration (ΔFrustration) and Performance ($r = -.184$, $p < .001$). While Operation Length was a significant influence on other cognitive workload factors, Performance was unaffected ($r = .032$, not significant). The difference in size and type of effects suggests that factors outside the operation were also having an effect on the operator. It is important to remember that operators can carry frustrations from home or other office duties into the operational environment. While this type of stress is a normal part of life and most people have coping strategies, the high-stakes nature of the tactical cyber operations environment may amplify the effects of this stress.

Subjectively, the increases in operation Fatigue and Frustration seem high (16% and 12% respectively). The factors that may influence Fatigue and Frustration are well-known, such as Operation Length. However, the questions are, for the difficulty and complexity of tactical cyber operations, is this increase significant, and should it be concerning? Frustration is a contributing factor to burnout. At the same time, these tactical operations are long and complex, and some amount of fatigue and frustration is to be expected. What are some things that can be done to minimize fatigue and frustration? Should researchers even focus on minimizing fatigue and frustration?

The moderate to strong relationships between Mental Demand, Physical Demand, Time Pressure, Effort and Frustration, and the lack of a relationship between these factors and Performance provides a practical model of cognitive work in tactical cybersecurity operations. Additionally, knowing that these factor relationships exist gives researchers methodology options in the future; that is, if only a subset of these factors can be measured, assumptions can be made about other factors based on the few that can be measured.

There are several independent factors regarding a cyber operation: the target, the objective, the tools, the infrastructure, the operator, and so on. This differs greatly across the different cyber missions within NSA, let alone across other government and non-governmental organizations. However, a consistent factor about an operation that can always be measured is its length. The length of an operation matters and should be considered not only for assessing risk to the mission, but also for its effects on operators in terms of long-term fatigue, frustration, cognitive wear and tear, and ultimately, burnout and turnover. Operations are long and complex, so it is no surprise that there are significant increases in fatigue.

Conclusions

The operations studied had a variety of mission goals, utilized a variety of tools and techniques, and met with a range of successes. While many of these details are classified (and what ultimately makes this work so valuable to the NSA mission), what researchers are able to share in this paper can be applied to those conducting other types of tactical cyber operations, such as developing red-team tradecraft. As with most early research, these results pose more questions than answers about the nature of tactical cybersecurity operations and the effects on their operators. While this research has only scratched the surface, it offers critical insights to the complex, high-risk environment in which these operators work on a daily basis. In the NSA environment, knowledge of the factors affecting operator fatigue, frustration, and cognitive workload have resulted in more informed policies and tradecraft. Interventions that aim to mitigate negative effects from critical human factors now have a baseline against which to can compare and measure effectiveness. It is hoped that the application of these results will help reduce operator burnout and workforce turnover, ultimately increasing operator efficacy and lowering mission risk.

References

Baulk, SD, Fletcher, A, Kandelaars, KJ, Dawson, D & Roach, GD 2009, ‘A field study of sleep and fatigue in a regular rotating 12-h shift system’, *Applied Ergonomics*, vol. 40, pp. 694-98.

Champion, M, Rajivan, P, Cooke, NJ & Jariwala, S 2012, ‘Team-based cyber defense analysis’, *Cognitive Methods in Situational Awareness and Decision Support: Proceedings of the IEEE International Multi-Disciplinary Conference*, pp. 218-21.

Chappelle, W, McDonald, K, Christensen, J, Prince, L, Goodman, T, Thompson, W & Hayes, W 2013, *Sources of occupational stress and prevalence of burnout and clinical distress among U.S. Air Force cyber warfare operators, Technical Report no. AFRL-SA-WP-TR-2013-0006*, Air Force Research Lab, Wright-Patterson AFB, Dayton, OH, U.S.A.

Cowley, J, Greitzer, FL & Woods, B 2015, 'Factors influencing network risk judgments: a conceptual inquiry and exploratory analysis', *Security Informatics*, vol. 4, no. 1, np, available at < http://link.springer.com/article/10.1186/s13388-015-0016-x>.

Dykstra, J & Paul, CL 2015, 'Stress and the cyber warrior: cognitive workload in a computer operations center', *Journal of Sensitive Cyber Research and Development*, vol. 3, no. 1, pp. 1-23.

Greenlee, ET, Funke, GJ, Warm, JS, Sawyer, BD, Finomore, VS, Mancuso, VF & Funke, ME 2016, 'Stress and workload profiles of network analysis: not all tasks are created equal', *Advances of Human Factors in Cybersecurity: Proceedings of the AHFE International Conference on Human Factors in Cybersecurity*, pp. 153-66.

Grier, RA 2015, 'How high is high? A meta-analysis of NASA-TLX global workload scores', *Human Factors and Ergonomics Society Annual Meeting: Proceedings*, vol. 59, no. 1, pp. 1727-31.

Haney, J & Paul, CL, 2016 'Toward integrated tactical operations for red/blue network defense teams', *Military Communications Conference, Proceedings of the IEEE Conference*, restricted section technical program.

Hart, SG 2006, 'NASA-task load index (NASA-TLX): 20 years later', *Human Factors and Ergonomics Society Annual Meeting: Proceedings*, vol. 50, no. 9, pp. 904-08.

Hopstaken, JF, van der Linden, D, Bakker, AB & Kompier, MAJ 2014, 'A multifaceted investigation of the link between mental fatigue and task disengagement', *Psychophysiology*, vol. 52, no. 3, pp. 305-315.

Judge, TA, Erez, A, Bono, JE & Thoresen, CJ 2005, 'Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?', *Journal of Personality and Social Psychology*, vol. 83, no. 3, 693-710.

Kowalski-Trakofler, KM & Vaught, C 2003, 'Judgment and decision making under stress: an overview for emergency managers', *International Journal of Emergency Management*, vol. 1, no. 3, pp. 278-89.

Marakas, GM, Yi, MY & Johnson RD 1998, 'The multilevel and multifaceted character of computer self-efficacy: toward clarification of the construct and an integrative framework for research', *Information Systems Research*, vol. 9, no. 2, June 1998, pp. 126-63.

Paul, CL 2014, 'Human-centered study of a network operations center: experience report and lessons learned', *Security Information Workers: Proceedings of the ACM CCS Workshop*, pp. 39-42.

Powell, DMC, Spencer, MB, Holland, D, Broadbent, E & Petrie, KJ 2007, 'Pilot fatigue in short-haul operations: effects of number of sectors, duty length, and time of day', *Aviation, Space, and Environmental Medicine*, vol. 78, no. 7, pp. 698-701.

Richter, S, Marsalek, K, Glatz, C & Gundel, A 2005, 'Task-dependent differences in subjective fatigue scores', *Journal of Sleep Research*, vol. 14, pp. 393-400.

Samn, SW & Perelli, LP 1982, *Estimating aircrew fatigue: a technique with application to airlift operations, Technical Report no. SAM-TR-82-21*, School of Aerospace Medicine, Brooks AFB, TX, U.S.A.

Sawyer, B, Finomore, V, Funke, G, Mancuso, V, Funke, M, Matthews, G & Warm, J 2014, 'Cyber vigilance: effects of signal probability and event rate', *Human Factors and Ergonomics Society Annual Meeting, Proceedings*, vol. 58, no. 1, pp. 1771-75.

Sundaramurthy, SC, Bardas, AG, Case, J, Ou, X, Wesch, M, McHugh, J & Rajagopalan, SR 2015, 'A human capital model for mitigating security analyst burnout', *Usable Privacy and Security: Proceedings of the USENIX Symposium*, pp. 347-59.

U.S. Army 1992, 'Fratricide: reducing self-inflicted loss', *CALL Newsletter*, no. 92-4, Fort Leavenworth, KS, U.S.A.

U.S. Department of Defense 2015, 'The Department of Defense cyber strategy,' April 2015, viewed 1 Feb 2017, <https://www.defense.gov/cyberstrategy>.

van der Linden, D, Keijsers, GPJ, Eling, P & Van Schaijk, R 2005, 'Work stress and attentional difficulties: an initial study on burnout and cognitive failures', *Work & Stress*, vol. 19, no. 1, pp. 23-36.