Stress and Workload Profiles of Network Analysis: Not All Tasks Are Created Equal

Eric T. Greenlee, Gregory J. Funke, Joel S. Warm, Ben D. Sawyer, Victor S. Finomore, Vince F. Mancuso, Matthew E. Funke and Gerald Matthews

Abstract Effective cyber defense depends upon intrusion detection, i.e., the process of monitoring, detecting, and reacting appropriately to cyber activity threatening network security. Intrusion detection requires the execution of multiple unique, interdependent network analysis tasks. The current study aimed to expand understanding of cyber defense by separately assessing task induced workload and stress for two key network analyst tasks, *triage analysis* and *escalation analysis*, which are the first and second lines of cyber defense, respectively. In separate studies, participants assumed the role of either a triage analyst or an escalation analyst, performed associated intrusion detection duties in simulated cyber task environments, and reported task induced workload and stress. Findings suggest that, even though triage and escalation analysts are both engaged in cyber defense, their tasks result in differentiable workload and stress profiles. This highlights the

E.T. Greenlee (✉)
National Research Council, Washington, DC, USA
e-mail: eric.greenlee.ctr@us.af.mil

G.J. Funke · J.S. Warm
Air Force Research Laboratory, Fairborn, OH, USA
e-mail: gregory.funke.1@us.af.mil

J.S. Warm
University of Dayton Research Institute, Dayton, OH, USA

B.D. Sawyer
AgeLab, Massachussets Institute of Technology, Cambridge, MA, USA

V.S. Finomore
United States Air Force Academy, Colorado Springs, CO, USA

V.F. Mancuso
Lincoln Lab, Massachussets Institute of Technology, Lexington, MA, USA

M.E. Funke
Naval Medical Research Unit, Dayton, OH, USA

G. Matthews
Institute for Simulation and Training, Orlando, FL, USA

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 501,
DOI 10.1007/978-3-319-41932-9_13

need for further human factors research examining operator performance and state across network analyst roles.

Keywords Cyber defense · Network analyst · Workload · Stress

1 Introduction

Military, industrial and commercial organizations have become increasingly reliant on the function of networked computer systems. Such cyber technologies have proven advantageous in each of these domains—enhancing individual and organizational capabilities. However, these benefits are not without risk. As former Chief Scientist of the U. S. Air Force Mark Maybury [1] cautioned, cyberspace is frequently contested, meaning that cyber assets are vulnerable to exploitation, intrusion, and attack. Consequently, it is crucial that adequate cyber defenses are deployed against such actions. The urgency of this need is highlighted in Maybury's report in which he suggests that the frequency, variety, and potential harm of cyber threats are likely to increase dramatically in the near future. In that report, Maybury recommended bolstering cyber defenses through scientific research and technological development. While many of his recommendations focus on optimization of technological capabilities (e.g., cloud computing, improvements to automated defense algorithms), the human side of cyber defense is also noted as an area for critical need for continued research. Specifically, research is required to assess the nature of human-computer interaction in the cyber domain and the effects that these interactions have on cyber operators' mental state and performance capabilities.

Initial research regarding human cyber defenders has focused on task analyses, which provide insight into the nature of tasks that those defenders are required to perform [e.g., 2]. Chief among them is computer network defense analysis, or *network analysis*, a multifaceted method of monitoring computerized networks to ensure their security. Toward this end, human network analysts are tasked with intrusion detection, i.e., "the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices" [3]. Simply stated, intrusion detection is the first line of defense against immediate cyber threats.

Network analysts are aided in this endeavor by automated *Intrusion Detection Systems* (IDS), which algorithmically inspect all network events and compare them to a database of known malicious activity. As the potential harm that undiscovered cyber attacks may cause is quite high, IDS systems typically exhibit a liberal bias. Network events that are even broadly *similar* to profiles of known malicious activity are marked as potentially suspicious and passed to human network analysts for further inspection. At this point, the process of intrusion detection requires human operators to execute multiple unique, interdependent network analysis tasks [2]. Two of these tasks, *triage analysis* and *escalation analysis*, are integral to successful

intrusion detection. Depending on the organization of a given cyber defense team, a single network analyst may perform one or both of these unique tasks [2]. They are discussed here as if performed by separate network analysts; such division of duty is typical in large organizations [2].

Triage analysis requires a network analyst to monitor the alerts that an IDS generates regarding potentially suspicious network activity. Triage analysts must evaluate the veracity of IDS alerts by inspecting relevant network sensor data. They determine whether the IDS alert represents truly suspicious network activity or if it is a false alarm. However, the high volume of data moving within and through modern networks, coupled with the bias mentioned above in IDS alerting, results in problems for analysts. Specifically, the number of alerts and false alarms generated by an IDS may be extreme, deluging analysts and forcing them to strategically sample alerts, rather than exhaustively investigating each one [4]. As a result, triage analysts typically have only a few minutes (often less) to evaluate each alert and determine whether each represents suspicious activity. If triage analysts find adequate evidence to support an IDS assertion of a suspicious network event, they will forward that alert to an escalation analyst. Otherwise, the alert will be discarded as a false alarm [2].

Escalation analysts are responsible for following up on the leads provided by triage analysts. This subsequent investigation typically involves a more thorough inspection of the network data. Escalation analysts have access to more sources of network data and may refer to external sources (e.g., the World Wide Web) to determine if malicious activity truly occurred. While escalation analysts are pressured to complete their alerts as quickly as possible (and in some cases within organizationally determined temporal limits), their work is more self-paced, unlike the work of triage analysts, where the pace of work is determined by the pace of incoming network traffic. If escalation analysts believe that there is enough evidence to confirm that an intrusion has occurred, they will recommend that an incident investigation be opened.

The procedural information provided by task analyses is supplemented by surveys of active cyber operators and experimental studies meant to assess the stress and workload associated with a network analyst's duties. Chappelle and colleagues [5] surveyed more than 500 active cyber operators, including computer network defense analysts. These operators reported a high degree of chronic, *occupational* stress and cited many organizational factors (e.g., leadership) and scheduling factors (e.g., shift-work) as sources of their distress. Other studies have shown that in addition to these occupational stressors, acute, *task-related* stress and workload are also potential concerns for cyber operators. Experimental simulations of cyber defense tasks have demonstrated that the performance of such tasks leads to elevated stress and reports of high workload [e.g., 6]. These extreme levels of stress and workload likely contribute to the high rate of burnout that has been reported by the cyber community [5]. Further, stress and overload are concerns because they may cause network defense analysts to lose situational awareness and reduce their ability to maintain network security [4].

To date, no studies have examined the possibility that stress and workload differ depending on the type of cyber defense task being performed. Instead, workload

and stress have often been assessed in a coarse manner by grouping multiple types of network analysts—or even cyber operators in general—into a single group [e.g., 5]. Some experimental studies have used simulated network analyst tasks to assess workload and stress more specifically, but none have drawn comparisons based on network analyst roles. Given the procedural differences between triage analysis and escalation analysis, it is possible that the degree and profile of task-related stress and workload differ between triage and escalation analyses. The present study was conducted to examine that possibility by comparing subjective stress and workload ratings from a simulated triage analysis task and a simulated escalation analysis task. The data used for these comparisons were generated in two different studies, one examining factors affecting operator state and performance in triage analysis and another examining factors affecting operator state and performance in escalation analysis. Given that these data were collected at different points of time, for different purposes, and using two different simulated network analysis tasks, we are treating the current analyses as exploratory. As such, our investigation focuses on description more so than direct inferential comparison of the two network analyst tasks. However, if task-related differences were apparent, they would encourage further research to explore task-related factors in network analysis.

2 Method

2.1 Participants and Design

This study utilized two data sets from two separate experiments: a study of triage analysis [7] and a study of escalation analysis [8]. Both samples consisted of college-age students and young adults recruited from Wright-Patterson Air Force Base and the surrounding area. Although the studies were completed at different times, both involved individual testing which was conducted in the same quiet, windowless laboratory room, using similar computer hardware and glare-controlled displays. Twenty-seven participants completed the triage analysis task and forty-six participants completed the escalation analysis task. Subject matter experts guided development of each simulated task in order to ensure validity. For the sake of brevity, methodological and task-related details are limited within the current report but are reported in greater detail elsewhere [7, 8]. Both experiments were approved by the WPAFB Institutional Review Board.

2.2 Measures of Stress and Workload

Stress was assessed using the Dundee Stress State Questionnaire (DSSQ) [9]. The DSSQ is a 96-item self-report measure of task engagement, distress, and worry.

Two different instruments were used to assess workload: the NASA-Task Load Index (NASA-TLX) [10] and the Multiple Resources Questionnaire (MRQ) [11, 12]. The NASA-TLX is a subjective measure that provides a global level of workload in addition to a workload profile based on six factors: mental demand, physical demand, temporal demand, performance, effort and frustration. All NASA-TLX items are rated from 0 (very low) to 100 (very high). The MRQ uses a scale ranging from 0 (no usage) to 100 (extreme usage) and participants are asked to rate demand on 17 MRQ items, each of which represents demand upon one independent, perceptual or cognitive resource.

The DSSQ, NASA-TLX, and MRQ were administered to participants who completed either the triage analysis task or the escalation analysis task (each detailed below). In both cases, a pre-task version of the DSSQ was completed prior to the task-training phase. After completion of their assigned network analyst task, participants in both experiments completed the NASA-TLX, the DSSQ, and the MRQ, in that order.

2.3 *Triage Analysis Task*

The triage analysis task required that participants monitor a simulated IDS display for the signature of an intrusion event. The display is illustrated in Fig. 1.

Each of the six rows in the IDS display represented one transmission, and each column represented a different piece of information about that transmission. The source IP address, source port, destination IP address, and destination port were each represented by a different column. The display cascaded, so that periodically, a new transmission appeared at the top of the list, while all other transmissions moved down one row and the oldest transmission disappeared from the bottom of the list. The rate of this periodic cascading (event rate) was either 8 or 16 updates per minute. In both event rates, there were two possible intrusion signatures, one contained in source information and one contained in destination information.

Source Addr.	Source Port	Dest. Addr.	Dest. Port
108.189.138.186	42	108.174.132.212	37
159.221.208.186	42	108.174.132.212	37
135.205.245.249	53	229.160.238.186	37
229.155.107.186	25	108.110.246.212	25
159.205.139.249	42	159.121.148.196	42
135.193.243.186	42	229.102.254.242	80

Fig. 1 Depiction of the simulated IDS display used for the triage analysis task. In this example, an intrusion signature has just appeared, since the *top row* and one other row (the *second row*, in this case) of the destination information section of the IDS display show the same destination IP address and destination port. See the definition of an intrusion signature presented below

If the *source information* (source IP address, source port) of a newly added transmission (top row) exactly matched the source information (source IP address, source port) contained in one of the older, displayed transmissions (lower five rows), participants were to report the intrusion with a keypress. An alternate intrusion signature was represented by cases in which newly added (top row) *destination information* (IP, port) exactly matched the destination information associated with one of the older, displayed transmissions. If a participant failed to detect an intrusion signature within three seconds of it appearing in the IDS, a miss was recorded. It should be noted that the probability of these signatures was manipulated, leading to a low intrusion rate (5 % of traffic) or a high intrusion rate (20 % of traffic). The two intrusion rates were combined factorially with the two event rates in a between subjects design.

All participants completed a 15-min training version of this triage analysis task. During training, auditory, verbal feedback was given to inform participants of all correct detections, misses, and false alarms. After training, participants engaged in the full experimental triage analysis task, which lasted 40 min. Feedback was removed during the full triage analysis task.

2.4 Escalation Analysis Task

The escalation analysis task was presented using a recently developed synthetic task environment, the Cyber Intruder Alert Testbed (CIAT) [8]. Within this platform, participants took on the role of escalation analysts who were charged with evaluating alerts that had been marked as potentially suspicious by a fictitious triage analyst. These alerts were depicted within an IDS display. Participants were free to select and investigate them in any order they chose and to take as much time as needed, i.e., the task was self-paced. To complete their task, participants were required to mark each of the alerts as either a "threat" or "not a threat" by pressing the corresponding button on their task display (see Fig. 2). Participants were instructed to make this determination by investigating each of the 45 presented alerts to determine whether it matched the signature of a known threat. To make an accurate decision, participants needed to collect information from multiple sources including the IDS display, packet capture software, the network list, and a signatures database. An alert was only to be confirmed as a threat if it matched *all* of the elements of the threat signature. With the exception of the response buttons and the IDS display, a participant was required to use display tabs to toggle between the multiple information sources within the CIAT display (see Fig. 2).

In 40 out of the 45 alerts, the threat signature was reflected by 4 to 5 elements. These 40 alerts were meant to provide insight into realistic escalation analysis. Ten of those 40 alerts matched all signature elements, meaning that they were to be confirmed as threats. The remaining five alerts were only intended as 'catch alerts,' i.e., they were designed to detect participants who were not engaged in the task. All of these five catch alerts were threats, but the signatures of these threats were

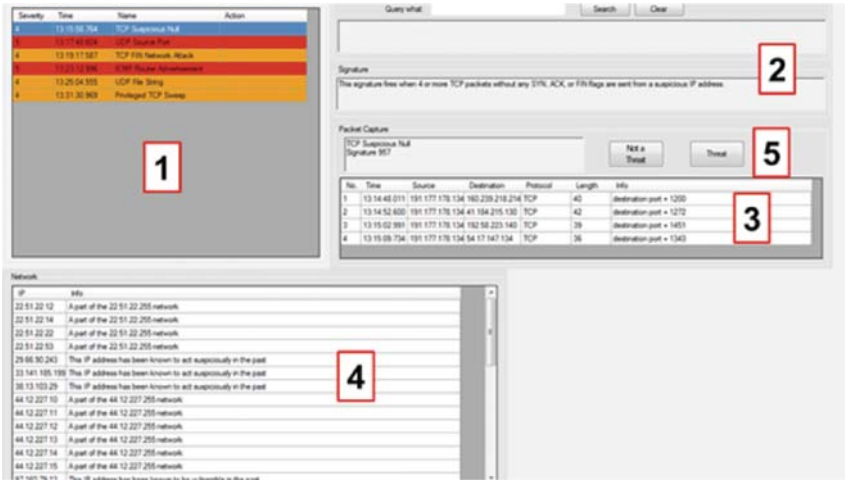

Fig. 2 Example of the CIAT interface. Represented in the figure are (*1*) the intrusion detection system (IDS), (*2*) the query and signatures database, (*3*) the packet capture software, (*4*) the network list, and (*5*) the participant response buttons (i.e., "Not a Threat" and "Threat"). Though these disparate components appear together in the figure to conserve space, during the experiment, each of the enumerated elements existed on separate "tabs" in the display, with the exception of the IDS and the response buttons that appeared below it

extremely simple relative to the other 40 trials. The catch alert signatures consisted of a single element, meaning that they could be accurately confirmed as a threat with minimal time and effort by any participant who was appropriately invested in the task. Participants were required to accurately identify at least four of these five catch alerts for their data to be analyzed.

Before engaging in the escalation analysis task, participants completed a training phase, which consisted of computerized instruction and practice. The purpose of this training phase was to familiarize them with the task and the CIAT interface. Like the full experimental escalation analysis task, the training version was self-paced. As practice, participants were presented with a list of three alerts. Each was designed to be unique, but similar to those used in the full escalation task (i.e., signatures consisted of 4–5 elements). For the first alert, the researcher demonstrated and verbally described the process of selecting an alert, investigating an alert, and making a correct threat determination. For the second alert, participants completing it were required to verbally describe the completion process as they did so, and were free to ask the researcher any questions they had. Lastly, the participants completed the third alert without any help from the experimenter and were still required to describe the process aloud while doing so. After completing all three practice alerts, participants began the full escalation task. On average, the full task required approximately one hour to complete.

In both the training version and the full, experimental version of the escalation analysis task, participants received one of two versions of the CIAT display.

Depending on their assignment to a between-subject condition, participants received either coordinated displays or uncoordinated displays. In the coordinated displays condition, selecting an alert in the IDS display automatically selected information relevant to that alert in each of the CIAT tabs. In the uncoordinated displays condition, selecting the alert had no effect on the information contained in the CIAT tabs, meaning that a participant had to perform a manual search of each information source to find relevant data.

3 Results

Data from each questionnaire (NASA-TLX, MRQ, DSSQ) and each task (triage analysis, escalation analysis) were analyzed separately. The primary factor of interest for each questionnaire was subscale. All analyses utilized an alpha of 0.05; the Box correction was applied to correct violations of the sphericity assumption; and the Bonferroni correction was used to adjust Type I error rate for post hoc, multiple comparisons.

The focus of these analyses was evaluating the workload and stress associated with triage analysis and escalation analysis as a whole. In that regard, we aggregated across all task-related experimental factors: event rate and intrusion probability for the triage analysis task, coordinated vs uncoordinated displays for the escalation analysis task.

Only two participants' data were excluded from analyses. Two participants in the escalation analysis task failed to identify four of the five catch alerts, indicating lack of engagement in the task. These participants' data were excluded, resulting in a final sample of 44 participants in the escalation analysis sample.

3.1 Workload: NASA-TLX

For each participant, NASA-TLX scores for each of the six subscales were derived using raw, unweighted ratings, as opposed to ratings weighted by subjective rankings of subscale importance. In the ANOVA analyses of the NASA-TLX, subscale was included as a factor with six levels (corresponding to the six items of the TLX). NASA-TLX scores for each of the subscales and each network analysis task are presented in Table 1.

Analysis of NASA-TLX scores from the triage analysis task revealed a significant main effect of subscale, $F(3.48,\ 90.44) = 36.61$, $p < 0.001$, $\eta_p^2 = 0.59$. Follow-up Bonferroni corrected *t*-tests revealed that ratings of mental demand ($M = 75.74$, $SE = 4.25$) temporal demand ($M = 69.82$, $SE = 4.99$), and effort ($M = 73.52$, $SE = 4.17$) were each greater than ratings for all other subscales ($p < 0.05$ in each case), but did not differ significantly from each other ($p > 0.05$ in

Table 1 Mean subscale scores and global score for NASA-TLX ratings in each network analysis task

Task	Subscale						
	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration	Global
Triage	75.74	16.67	69.81	32.96	73.52	42.96	51.94
Analysis	(4.25)	(3.54)	(4.99)	(4.97)	(4.17)	(5.48)	(2.73)
Escalation	79.55	12.43	30.30	20.52	67.73	29.73	40.04
Analysis	(2.46)	(2.24)	(3.69)	(2.46)	(2.80)	(3.27)	(1.71)

Standard errors are presented in parentheses

each case). Scores for mental demand, temporal demand, and effort were all significantly greater than 50 (i.e., the midpoint on the scale), as revealed by one-sample *t*-tests ($p < 0.05$). Scores above the midpoint are generally indicative of substantial workload [6]. As can be seen from Table 1, scores from all other subscales are below 50, descriptively. The only other notable subscale is frustration ($M = 42.96$, $SE = 5.48$), which was significantly greater than physical demand ($M = 16.67$, $SE = 3.54$).

The analysis of NASA-TLX scores from the escalation analysis task also revealed a significant main effect of subscale, $F(3.76, 161.79) = 116.41$, $p < 0.001$, $\eta_p^2 = 0.73$. Pairwise comparisons using Bonferroni corrected *t*-tests revealed that mental demand ($M = 79.55$, $SE = 2.46$) was rated as significantly higher than all other subscales, $p < 0.001$ in each case. Effort ($M = 67.73$, $SE = 2.80$) was also rated as a major element in escalation task engagement, second only to mental demand ($p < 0.001$) and greater than all other workload subscales ($p < 0.001$ in each case). Both mental demand and effort were rated as significantly greater than 50, $p < 0.001$ in each case.

3.2 Workload: MRQ

Prior to inferential analyses, scores for each MRQ subscale were evaluated separately to determine whether each resource was a considerable contributor to task workload. A resource (subscale) was only considered a notable contributor to task demand if at least 50 % of participants reported using that resource, i.e., workload was rated as greater than zero by at least half of the participants in a given network analysis task. Any subscale that failed to meet this criterion was excluded from subsequent data analyses. This method is one of two item reduction methods that are recommended by David Boles, the creator of the MRQ [12]. Mean subscale scores are presented in Fig. 3 for all subscales that represented a resource that contributed to task demand for either the triage analysis task or the escalation analysis task.

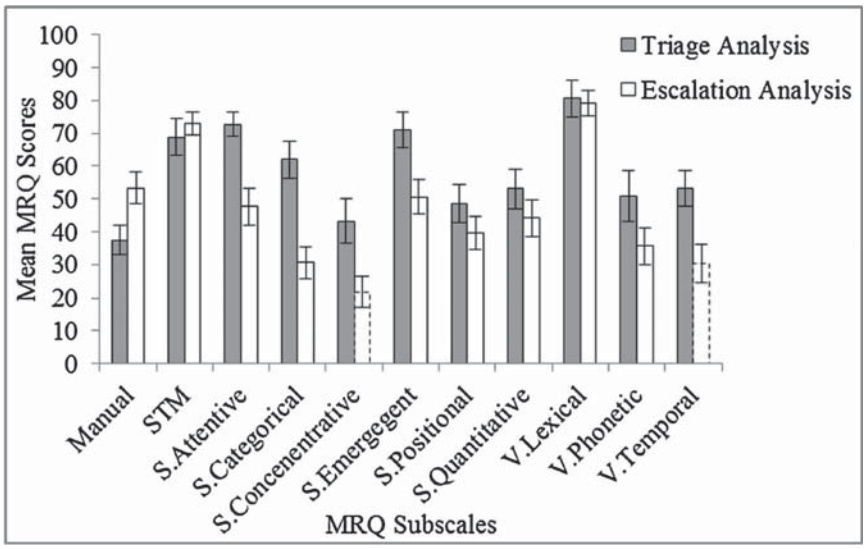

Fig. 3 Mean MRQ scores for each subscale and each network analysis task. Note that in two cases, a subscale score was only a significant contributor to workload in triage analysis. For those two subscales, escalation analysis scores are presented with a *dashed border*. Error bars are standard errors. Abbreviations: 'STM' = Short Term Memory; 'S.' = Spatial; 'V.' = Visual

For the triage analysis task, 11 of 17 subscales met the inclusion criterion set by the item reduction procedure. The global, mean workload across these subscales was 58.29 (*SE* = 3.59). These 11 subscales were subjected to further analysis using a repeated measures ANOVA with a single factor, subscale. This analysis showed that there were significant differences among mean subscale scores $F(5.97, 155.32) = 8.57$, $p < 0.001$, $\eta_p^2 = 0.248$. Pairwise comparisons indicated that triage analysis demanded significantly more of *short term memory*, *spatial attentive processing*, *spatial emergent processing*, and *visual lexical processing*, compared with other subscales, $p < 0.05$ in each case. Further, of the 11 subscales that were considered as notable contributors to the workload of the triage analysis task, only those four subscales produced demand scores that were significantly above the midpoint of the scale (i.e., 50; $p < 0.05$ in each case).

For the escalation analysis task, 9 of 17 subscales met the inclusion criterion. The included subscales were the same as those included in examination of the triage analysis task except the spatial concentrative subscale and the visual temporal subscale did not meet the inclusion criteria for the escalation analysis task. Global, mean workload for the 9 subscales was 50.37 (*SE* = 2.98). An ANOVA of these nine subscales revealed that there were mean differences in demand scores among subscales, $F(5.06, 217.45) = 14.99$, $p < 0.001$, $\eta_p^2 = 0.258$. Pairwise comparisons indicated that demands for *manual processes*, *short term memory*, *spatial emergent processing*, and *visual lexical processing* were greater than demands for other

subscales, $p < 0.05$ in each case. Moreover, one-sample *t*-tests revealed that only the demands for short term memory and visual lexical processing were significantly greater than the midpoint of the MRQ scale, $p < 0.05$ in each case.

3.3 Stress: DSSQ

Participant *task engagement*, *distress*, and *worry* scores, i.e., the three factors of the DSSQ, were computed using the procedure recommended by the developers of the scale [9], such that raw factor scores were transformed based on extant data representing a large normative sample. The result was standardized subscale scores ($M = 0$, $SD = 1$) for pre-task and post-task ratings of each of the three subscales. These values were used to calculate change scores (post-task minus pre-task) for each of the subscales, for each task. Mean DSSQ change scores are presented in Fig. 4 for each subscale and each network analysis task.

After computing change scores, the first step in analyzing data from the triage analysis task was determining whether the task-induced changes in stress were significant. To make this determination, change scores for each subscale were evaluated using separate, one-sample *t*-tests (versus zero). These analyses indicated that distress increased, $t(26) = 3.70$, $p = 0.001$, and task engagement decreased during triage analysis $t(26) = 2.54$, $p = 0.017$. Worry did not change significantly, $p > 0.05$.

Identical analyses were used to evaluate task-induced stress effects in escalation analysis. These analyses showed that worry decreased, $t(43) = 4.19$, $p < 0.001$, and distress increased during task performance $t(43) = 3.22$, $p = 0.002$. However, escalation task performance led to no significant changes in task engagement ($p > 0.05$).

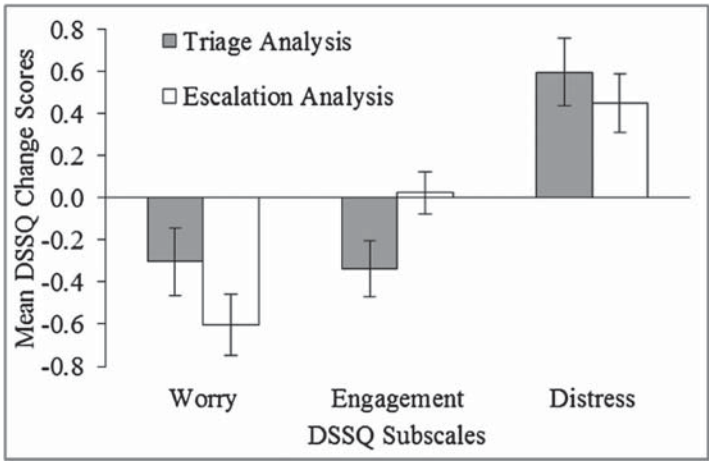

Fig. 4 Mean change scores for each DSSQ subscale and network analyst task. Error bars represent standard errors. *Note* "Engagement" in the figure refers to task engagement

4 Discussion

The present study was conducted to explore the possibility that different network analysis tasks may involve different task-related stressors and demands. Past studies in the cyber domain have utilized aggregated cyber operator samples, as if all network analysis tasks—or even cyber tasks, in general—are created equal [e.g. 5]. Results of the current study challenge this assumption, suggesting instead that procedurally distinct network analysis tasks elicit distinctive effects on operator state.

In terms of task workload, global ratings of task demands from the NASA-TLX and the MRQ indicated that triage analysis might be more demanding than escalation analysis. A more nuanced examination of the profiles of mental workload for each task allows for further comparison. Based on the NASA-TLX it appears that both network analysis tasks require a high degree of mental demand and effort. Additionally, temporal demand was a driving factor in task workload for triage analysis, but not for escalation analysis. These results suggest that both network analysis tasks are cognitively demanding, but triage analysts are further challenged by the constant flood of incoming network traffic and the pace at which it must be parsed.

The nature of these tasks was further elucidated by MRQ workload profiles, which indicated that both tasks recruit a similar set of perceptual, attentional, and memorial resources. Yet, there are some notable differences in workload profiles between triage analysis and escalation analysis. Participants in both tasks reported being taxed by demands upon short term memory and the need to recognize visual lexical information (e.g., words and digits within the IDS display). Further, both tasks appear to be similar in that there was a demand for participants to "pick out" critical visual information from "a highly cluttered or confusing background." Despite these similarities, it appears that triage analysis may place greater demands on multiple spatial processing resources. For example, upon perusal of Fig. 3, it will be evident that the perceived demand for spatial attention in triage analysis was approximately 50 % greater than that associated with escalation analysis.

Results of the DSSQ suggest that triage analysis and escalation analysis also invoke different stress responses. Triage analysis was associated with increased distress and decreased task engagement, while escalation analysis was associated with increased distress and decreased worry.

Note that the stress profile for triage analysis was identical to the typical stress response of vigilance tasks [13]. The NASA-TLX workload profile is also similar to the typical profiles of fast-paced vigilance tasks, suggesting the possibility that sustained attention, or vigilance, may play a major role in triage analysis [7]. In contrast, the self-paced nature of the escalation analysis may allow operators to manage the task to maintain an optimal level of workload and prevent task-induced stress [14].

These task-related differences are intriguing, but they should be considered with caution. The current study included only novice participants, who had no prior

cyber experience or expertise. In order to ensure the external validity of these findings and the real-world utility of potential solutions, future research should include expert analysts.

Although replication is needed, the current findings provide the initial demonstration that, though triage and escalation analysts are both engaged in cyber defense, their tasks result in differentiable workload and stress profiles. This suggests that it may be inappropriate to assume that all network analysts face the same task-related challenges or that there are universal solutions. Future research should assess specific needs and challenges associated with different network analyst tasks, so that human factors solutions can be appropriately tailored to augment cyber defense capabilities.

Acknowledgments This project was supported in part by grant no. F4FGA05076J003 from the Air Force Office of Scientific Research (Benjamin Knott, Program Officer).

References

1. Maybury, M.: Toward the assured cyberspace advantage: Air Force cyber vision 2025. IEEE Secur. Priv. **13**, 49–56 (2015)
2. D'Amico, A., Whitley, K.: The real work of computer network defense analysts: the analysis roles and processes that transform network data into security situation awareness. In: Goodall, J.R., Conti, G., Ma, K.-L. (eds.) VizSEC 2007: Proceedings of the Workshop on Visualization for Computer Security, pp. 19–37. Springer, Heidelberg, Germany (2007)
3. Scarfone, K., Mell, P.: Guide to Intrusion Detection and Prevention Systems (IDPS): Recommendations of the National Institute of Standards and Technology (Special Publication 800-94). National Institute of Standards and Technology, Technology Administration, U.S. Department of Commerce (2007)
4. Champion, M., Rajivan, P., Cooke, N., Janwala, S.: Team-based cyber defense analysis. In: IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, pp. 218–221 (2012)
5. Chappelle, W., McDonald, K., Christensen, J., Prince, L., Goodman, T., Thompson, W., Hayes, W.: Sources of Occupational Stress and Prevalence of Burnout and Clinical Distress Among US Air Force Cyber Warfare Operators (No. AFRL-SA-WP-TR-2013-0006). School of Aerospace Medicine, Wright-Patterson AFB, OH (2013)
6. Mancuso, V.F., Greenlee, E.T., Funke, G., Dukes, A., Menke, L., Brown, R., Miller, B.: Augmenting cyber defender performance and workload through sonified displays. Procedia Manufact. **3**, 5214–5221 (2015)
7. Sawyer, B.D., Finomore, V.S., Funke, G.J., Matthews, G., Mancuso, V.F., Funke, M.E., Warm, J.S., Hancock, P.A.: Cyber-vigilance: the human factor. Am. Intell. J. (2015)
8. Vieane, A., Funke, G., Mancuso, V., Greenlee, E., Dye, G., Borghetti, B., Miller, B., Menke, L., Brown, R.: Coordinated displays to assist cyber defenders. In: Proceedings of the Human Factors and Ergonomics Society 60th Annual Meeting. Sage Publications, Thousand Oaks, CA (in press)
9. Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J. Gilliland, K.,… Warm, J.S.: Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion **2**, 315–340 (2002)

10. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.S., Meshkati, N. (eds.) Human Mental Workload, pp. 239–250. North Holland Press, Amsterdam (1988)
11. Boles, D.B., Adair, L.P.: The multiple resources questionnaire. In: Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, pp. 1790–1794 (2001)
12. Boles, D.B., Dillard, M.B.: The measurement of perceptual resources and workload. In: Hoffman, R.R., Hancock, P.A., Scerbo, M.W., Parasuraman, R., Szalma, J.L. (eds.) The Cambridge Handbook of Applied Perception Research, vol. 1, pp. 39–59. Cambridge University Press, New York (2015)
13. Warm, J.S., Finomore, V.S., Vidulich, M.A., Funke, M.E.: Vigilance: A perceptual challenge. In: Hoffman, R.R., Hancock, P.A., Scerbo, M.W., Parasuraman, R., Szalma, J.L. (eds.) The Cambridge Handbook of Applied Perception Research, vol. 1, pp. 241–283. Cambridge University Press, New York (2015)
14. Xie, B., Salvendy, G.: Prediction of mental workload in single and multiple tasks environments. Int. J. Cogn. Ergon. **4**(3), 213–242 (2000)