Measuring and Mitigating the Costs of Attentional Switches in Active Network Monitoring for Cybersecurity

Sean W. Kortschot, **Dusan Sovilj**, **Greg A. Jamieson**, **Scott Sanner**, and **Chelsea Carrasco**, University of Toronto, Toronto, Ontario, Canada, and **Harold Soh**, National University of Singapore, Singapore

Objective: The authors seek to characterize the behavioral costs of attentional switches between points in a network map and assess the efficacy of interventions intended to reduce those costs.

Background: Cybersecurity network operators are tasked with determining an appropriate attentional allocation scheme given the state of the network, which requires repeated attentional switches. These attentional switches may result in temporal performance decrements, during which operators disengage from one attentional fixation point and engage with another.

Method: We ran two experiments where participants identified a chain of malicious emails within a network. All interactions with the system were logged and analyzed to determine if users experienced disengagement and engagement delays.

Results: Both experiments revealed significant costs from attentional switches before (i.e., disengagement) and after (i.e., engagement) participants navigated to a new area in the network. In our second experiment, we found that interventions aimed at contextualizing navigation actions lessened both disengagement and engagement delays.

Conclusion: Attentional switches are detrimental to operator performance. Their costs can be reduced by design features that contextualize navigations through an interface.

Application: This research can be applied to the identification and mitigation of attentional switching costs in a variety of visual search tasks. Furthermore, it demonstrates the efficacy of noninvasive behavioral monitoring for inferring cognitive events.

Keywords: attentional processes, cybersecurity, adaptive automation, visual search, interface evaluation

Address correspondence to Sean W. Kortschot, MIE, University of Toronto, 5 King's College Circle, Toronto, ON M5S 3G8, Canada; e-mail: sean.kortschot@mail.utoronto.ca.

HUMAN FACTORS
Vol. 60, No. 7, November 2018, pp. 962–977
DOI: 10.1177/0018720818784107

As dependence on networked systems has increased, the global vulnerability to cyber crime has grown (Goutam, 2015). This trend is mirrored by the growing number of yearly cyber attacks (Ben-Asher & Gonzalez, 2015), with the net economic cost of data breaches expected to exceed $2 trillion by 2019 (Moar, 2015). A significant investment in cybersecurity measures is therefore necessary for the protection of organizational activities and finances. Despite computational safeguards such as antivirus software and firewalls (Alrajeh, Khan, & Shams, 2013), any sufficiently large network remains vulnerable to attacks (Ahmad, Hadgkiss, & Ruighaver, 2012). In these instances, active network monitoring (ANM) is necessary.

ANM is the process of detecting, diagnosing, and mitigating the effects of network intrusions or attacks (D'Amico, Whitley, Tesone, O'Brien, & Roth, 2005). This task is extremely complex given the magnitude and distribution of the networks (Mitropoulos, Patsos, & Douligeris, 2006), the coordination required between operators (Tyworth, Giacobe, & Mancuso, 2012; Werlinger, Muldner, Hawkey, & Beznosov, 2010), significant time pressure and stakes (Khan, Gani, Abdul Wahab, Shiraz, & Khan, 2016), and the difficulty of determining an appropriate course of action for each unique attack (Werlinger et al., 2010). These challenges necessitate heavy reliance on decision support systems (DSSs). A typical cybersecurity DSS detects anomalous or malicious behavior by comparing both the current overall state and individual activities of the network against both the expected activity under nonattack conditions and known attacks (Ashfaq & Khayam, 2011). These relationships are then depicted in tables, which often provide scores for elements in the network, and graphs that support understanding of patterns of data and alerts.

The meaning of an element's score differs from system to system, but it is generally a function of the probability and the cost of malicious behavior within that element (Ashfaq & Khayam, 2011). For example, moderately anomalous behavior in an important server would elicit a higher score than highly anomalous behavior from a low-level email account. Operators combine scores with network graphs to understand how alerts are distributed to identify functional relationships (Franke & Brynielsson, 2014). Operators then use this higher order understanding of the system to determine how to allocate their attention (Hopf, Boehler, Schoenfeld, Heinze, & Tsotsos, 2010; Olszewski, 2014; Pfleeger & Caputo, 2012).

Operators are required to repeatedly determine where to focus their attention among a multitude of viable options. This near-constant demand for attention allocation, coupled with extreme information quantity, is likely to result in *cognitive bottlenecks*, which are performance-limiting constraints in the information flow between the system, the human, and the situation (Dorneich, Whitlow, Ververs, & Rogers, 2003). Bottlenecks have been studied in perception (Salvucci & Taatgen, 2008), response selection (Nobre & Kastner, 2014), and memory (Borst, Taatgen, & van Rijn, 2010), as well as across domains like driving (Donmez, Boyle, & Lee, 2006) and command and control (Dorneich, Mathan, Ververs, & Whitlow, 2007). While several cognitive bottlenecks are likely to be implicated in ANM, we focus on bottlenecks stemming from visual *attentional switching* here. Attentional switching is the process of moving one's focus from one point to another and consists of three phases: disengagement from current fixation point, shifting to a new location, and engagement of a new fixation point (Posner & Presti, 1987). We concentrate our analysis on the cognitive costs of the disengagement and engagement phases of the attentional switch.

Although engagement and disengagement are separated by the shifting phase, they are closely related. For example, engagement with a new target is hindered if participants need to first disengage from an initial target (Duncan, 1980). Ettwig and Bronkhorst (2015) corroborated this, showing that even when a previously perceived stimulus is masked, switching attention to a new information stream is hindered, a finding explained as a need to disengage from the masked stimuli even after it has disappeared. Dombrowe, Donk, and Olivers (2011) found that eye saccade accuracy and speed were significantly hindered when participants were asked to scan a series of targets of different colors, illustrating the potential performance decrements incurred from attentional switches. Longman, Lavric, and Monsell (2017) showed that allowing participants to prepare for an upcoming switch lessened the cost of switching but did not eliminate it. The common thread in the above research is that there is a clear cost incurred from disengaging from a previous attentional fixation and engaging with a new one. In real-world settings, where a large number of attentional switches are necessary, the cumulative cost of disengaging and engaging with targets could be detrimental to performance. We investigated this relationship in an ANM setting that more closely approximated real-world tasks than did the paradigms used to attain the evidence presented above.

This paper describes two studies, whose respective principal aims were to (1) examine how attentional switches impact operator behavior in ANM and (2) determine the efficacies of two interventions aimed at mitigating the negative effects of attentional switches.

OVERALL METHOD

Experimental Platform, Participants, and Scenario Design

Our experiments were conducted on a platform that allowed for the representation and visualization of a communication network, provided users with the tools to inspect elements within that network, and allowed users to tag elements as normal, suspicious, or malicious (see Figure 1; Kortschot et al., 2017). The network was populated with the 2015 version of the Enron Email Corpus (https://www.cs.cmu.edu/~./enron/), which is a public dataset of real emails. The visualization of the network was simple, with nodes representing users and edges representing emails. Each email in the network had a score, which was the probability that it contained malicious content. Users could

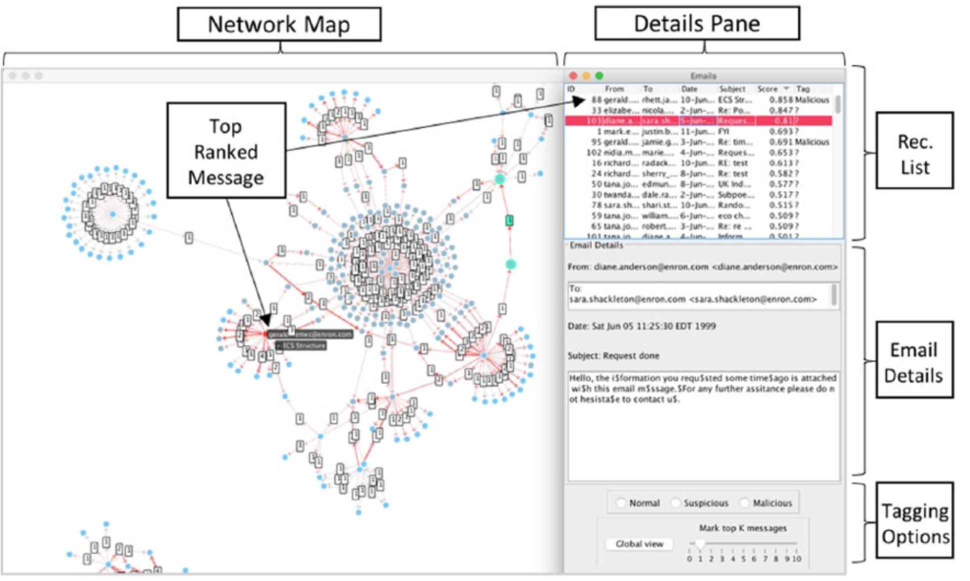

Figure 1. The experimental platform developed for the present study. The network map is in the left pane, while the details pane is on the right. Note that the Top-K slider (bottom-right) is currently set to 1, which reveals the top ranked message on the network map.

zoom in on areas of the network and inspect an email by either selecting it from a list or directly on the network map. They could also click on users (i.e., nodes) in the network map and view all of their outgoing emails. The Top-K slider displayed 0 to 10 of the highest scored emails. For example, if the Top-K slider was set to 1, it would highlight the single email in the network map with the highest score (see Figure 1).

We used a machine learning text classifier to determine the score for every communication in the network. This classifier was trained to detect anomalous content in the networks used in our experiments. Over the course of each trial, it then learned to increase the scores of emails similar to those tagged as malicious and decrease the scores of emails similar to those tagged as normal. Scores were represented in two ways. First, a recommendation list was presented beside the network map that showed the score. Second, the score was represented by the degree of redness of the edges representing the emails, with more redness indicating a higher score. Scores were updated after each tag applied by the user.

A logging system recorded all click actions as well as what object was clicked on. In addition to the action itself, the state of the *viewport* (i.e., the portion of the network map currently visible in the left panel of Figure 1) at the time of the action was also recorded. This included information such as the visible nodes, the percent of the network that was in view, and the center position of the viewport within the network. From these details, we were able to derive all zooming and panning behavior for later statistical analyses.

A population of qualified security or network operations center operators was not available to us. In their place, we recruited a sample of engineering students as participants. The skills and competencies of these participants imposed a significant limitation on the complexity of both the simulated network and the experimental tasks. We scaled down the complexity of both the network and the tasks by using a small, static network and by asking participants to perform a relatively simple search and inspection task. This was aimed at emulating the cognitive demands of expert operators in full-scale systems, which hinges on the assumption that the cognitive tolerance of novice operators is a fraction of that of experts. A formal analysis of the accuracy of this scaling procedure was not conducted, and therefore, this represents a limitation in our study.

In both experiments, participants attempted to uncover a chain of malicious emails sharing common characteristics that were indicative of their maliciousness. These chains were island-hopping in nature, meaning that the target email jumped between adjacent users. This encouraged interaction first with the recommendation list to identify the first email in a chain and then with the network map to explore adjacent areas of the network.

All research herein complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the University of Toronto. Informed consent was obtained from each participant.

Passive Data Monitoring (PDM)

We employed a PDM approach with logged interaction data to form inferences regarding attentional switching. PDM derives data from interactions that are inherent to the task so that subsequent analyses can determine if patterns in user interaction align with certain cognitive events or states. PDM is an alternative to identifying cognitive events by actively collecting biometric data through methods such as eye tracking or electroencephalogram (EEG; Palmius et al., 2016). For example, engagement with novel stimuli has been found neurologically to be represented in the dorsal stream (Janczyk & Kunde, 2010). Therefore, one could outfit participants with EEG and infer engagement periods by detecting when the subject was experiencing these neurological processes. However, the prospect of outfitting civilian network operators who work long, sedentary shifts with intrusive EEG equipment is unrealistic. The efficacy of PDM has been demonstrated in map navigation (Aoidh, Bertolotto, & Wilson, 2012) and depression onset detection (Palmius et al., 2016).

PDM pairs well with the ANM domain for three reasons. First, integrating a robust logging program to passively monitor user interactions is relatively easy since the vast majority of the user's interactions are done at the desktop (Goodall, Lutters, & Komlodi, 2009). Second, information about the system state and the content of displays is also readily available and time-stamped (Corchado & Herrero, 2011). Finally, ANM is a highly interactive domain (Werlinger et al., 2010), allowing for sufficient operator behavior to be recorded to make reliable inferences.

EXPERIMENT 1

Motivation

The principal objective of our first experiment was to examine the behavioral impact of attentional switches in ANM. Our secondary objective was to evaluate the efficacy of PDM in ANM.

Methods

Participants. We recruited 18 engineering students via an emailed advertisement (9 male, 9 female, $M_{age} = 21.5$, $SD = 2.89$). None of the participants had prior experience in cybersecurity or prior knowledge of the experimental platform, paradigm, or hypotheses. Participants were paid CAD 30.00 for 2 hours of participation.

Experimental platform. The version of the platform used in Experiment 1 is the same as that presented in Figure 1 with a slightly lower contrast color scheme. Participants had the option to sort the recommendation list by any of its columns. Participants were able to use the Top-K message slider to provide spatial context for the emails that had the highest scores associated with them, which was useful for determining if there were any clusters of recommendations in the network map.

Participants were able to explore the layout of the network by zooming and panning (Supplementary Material S1; the online supplementary material is available with the manuscript on the *HF* website). Zooming allowed users to expand the network such that a smaller portion of it occupied a larger portion of the viewport (i.e., zooming in). Panning allowed the users to remain zoomed in and to click and drag the network map to bring adjacent areas into their viewport.

Scenario design. Each scenario was constructed around one of four worm attacks. Worms are self-replicating codes that propagate through adjacent machines (Li, Salour, & Su,

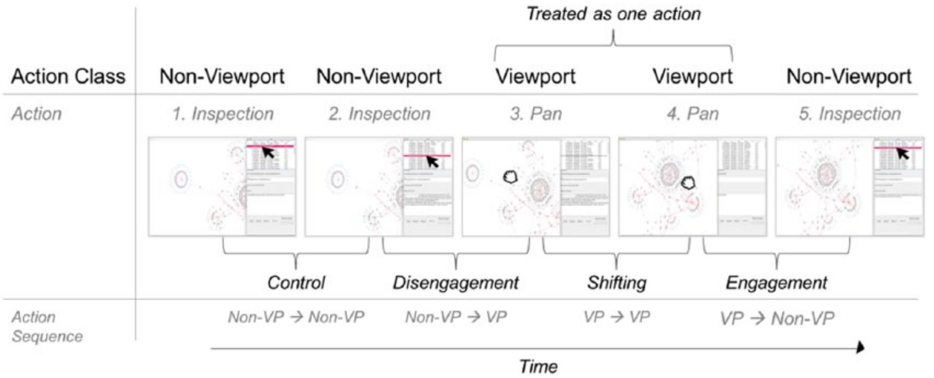

Figure 2. Timeline illustrating how the different sequences of action-class are treated in the data analysis. The figure illustrates how a viewport-action (i.e., Actions 3 and 4) changes the content of the viewport. The action latencies of interest are those of the second action in each of the action sequences (Actions 2, 3, and 5).

2008). They represent a fairly intuitive class of cyber attack that novices could grasp given sufficient training. Each worm was marked by a unique characteristic that reflected a real-world scenario. The first worm was marked by some of the text within the body of the email being replaced with punctuation marks. The second was marked by the text inviting users to click on a suspicious website URL. The third asked users to update their passwords for a financial server. The fourth worm had some text that was out of order. While the marker was consistent within each worm, the actual text and subject lines differed such that the exact same email was not simply being forwarded along. This forced the users to read the content of the emails rather than identify their structure at a glance. Users were tasked with flagging all of the emails in the chain as malicious.

Experimental design. Two machine learning algorithms for updating the scores of emails were evaluated. However, it was not hypothesized that they would have any effect on attentional switches. Our results confirmed this, and therefore, the alternative algorithms will not be described further here.

No experimental manipulations used in Experiment 1 were pertinent to the goal of understanding attentional switching in ANM. Instead, we divided all actions within trials into two labels: viewport-actions and non-viewport-actions. Viewport-actions included panning and zooming inputs. Non-viewport-actions included tagging, sorting, and inspecting actions. There are four possible sequences of these two action-classes, which align with the three phases of an attentional switch, plus a control condition (see Figure 2). Each of the action-classes represents an independent variable and were analyzed as follows:

- Control: the time it took participants to initiate non-viewport-actions that were preceded by non-viewport-actions.
- Disengagement: the time it took participants to initiate viewport-actions that were preceded by non-viewport-actions.
- Shifting: Consecutive viewport-actions were treated as one action, with the initiation time being the time of the first action in the sequence, and the completion time being the completion time of the last action in the sequence. This was done because multiple pans or zooms were typically required for the participants to achieve their desired viewport.
- Engagement: the time it took participants to initiate non-viewport-actions that were preceded by viewport-actions.

The underlying rationale for using action latency as a measure of the impact of attentional switches is that if users experienced a cost from an attentional switch caused by either disengaging from their previous viewport or engaging with their new viewport, that should be reflected by an increase in action latency, during which the disengagement or engagement processes would be occurring.

We used completion time as the sole performance metric due to ceiling effects with tagging accuracy (average of 90% accuracy across participants and trials). Completion time was measured as the time of the final malicious tag in each trial. If participants failed to tag all emails in a worm, the completion time for that trial was set to the maximum trial length of 15 minutes.

Given our use of PDM over gathering biometric data (e.g., eye-tracking), our operational definition of an attentional switch deviates from that of Posner and Presti (1987), as we solely examined extraviewport attentional switches. This operationalization sacrifices precision for practicality and accepts that we are missing attentional switches within viewports, asking the question of whether we can still characterize the effects of attentional switches in spite of this loss of precision.

Procedure. Participants were led through a PowerPoint presentation describing the components of the platform and their corresponding control actions (Supplementary Material S2). It also introduced them to the recommendation table and gave a high-level description of the machine learning algorithms driving the adaptations in the interface.

Participants were then introduced to the experimental task—searching for and tagging worms that consisted of between 5 and 8 emails hidden within the network. They were given examples of the four markers of worms that they would be looking for and told that each worm would be defined by one of these markers. Their principal objective was to find the origin of the worm, and their secondary objective was to tag the rest of the emails in the worm as malicious. A strategy for how best to combine information in the recommendation table with the network map was described and demonstrated to participants. Participants were not trained to criterion. Instead, the experimenter judged their competency with the platform prior to advancing from the training phase. We do not feel that this represents a significant limitation in our research as our analyses focused on the execution of individual actions rather than overall performance. Each participant completed four trials, and each trial had a new worm.

Results

All actions with latencies of 0 seconds were removed from analysis, as these represented double clicks. Following this, the top and bottom 1% of action latencies were removed from the dataset to remove both long pauses that were not representative of any engagement or disengagement processes and inadvertent clicks (Aguinis, Gottfredson, & Joo, 2013). This removal process retained actions with latencies between 0.015 seconds and 10.08 seconds.

Participants performed an action every 0.64 seconds ($SD = 1.27$ seconds) on average. The data were heavily skewed, so we used a Wilcoxon Rank-Sum test to identify differences in action latencies between the control actions and the actions corresponding to disengagements and engagements (see Figure 2). We found significant disengagement delays ($M = 1.45$ seconds; $z = 28.60$, $p < .0001$) and engagement delays ($M = 1.53$ seconds; $z = 35.42$, $p < .0001$) compared to control actions ($M = 0.57$ seconds). Figure 3 illustrates the distribution of action latencies across control actions, disengagements, and engagements.

We also compared average disengagement and engagement delays against completion time and did not find a significant relationship for either, $F(1, 16) = 1.23$, $p > .05$; $F(1, 16) = 2.75$, $p > .05$, indicating that the best and worst performers were equally susceptible to costs resulting from attentional switches.

Discussion

The results from Experiment 1 revealed both a significant disengagement delay preceding and a significant engagement delay following viewport-actions. This increase is large, nearly tripling action latency for both disengagement and engagement. This provides strong evidence that at a microlevel (i.e., individual actions), viewport movement hinders performance, and that limitations in human attentional switching capacity represent a cognitive bottleneck in ANM.

We did not find a significant relationship between either disengagement or engagement

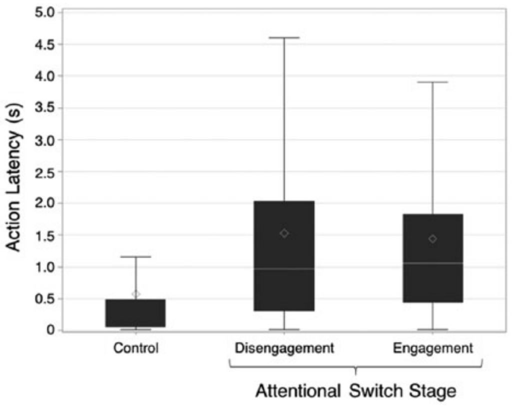

Figure 3. Boxplots showing the distribution of action latencies across control actions, disengagements, and engagements. All outliers have been removed from this graphic.

delay and performance. While this does not eliminate the limitation imposed by our novice participant pool, it does show that these delays are not artifacts of skill. Future studies should compare the effects of attentional switches between experts and novices.

Beyond identifying the microperformance impacts of viewport movements, Experiment 1 provided evidence for the effectiveness of PDM in ANM by successfully identifying the presence of a fundamentally cognitive phenomenon through behavioral indices. We believe that these methods can be extended to real-time monitoring of operator cognitive states, which can trigger adaptations to the user interface.

EXPERIMENT 2

Motivation

Our first experiment demonstrated that we could characterize micro performance impacts of attentional switches in ANM through PDM. Our second experiment builds on these results by assessing the cumulative impact of attentional switches in ANM and by seeking to facilitate attentional switches by improving the *visual momentum* in the interface. Visual momentum is the ease of extracting and integrating information when operators move to a new point in a display (Bennett & Flach, 2012; Woods, 1984). Moving to a new location requires operators to disengage from their previous location, shift to the new location, and then engage with that new location. This process aligns with the three phases of an attentional switch described by Posner and Presti (1987), which suggests that attentional switches may be influenced by visual momentum. Woods (1984) posited that a key to increasing visual momentum is to provide the user with context that will allow them to more easily discern where they are in the interface relative to their previous location. We therefore implemented interventions that were aimed at providing this context to users when they moved to a new point in a display under the hypothesis that this would lessen the disengagement and engagement delays resulting from an attentional switch.

Methods

Participants. Nineteen participants (11 male, 8 female; M_{age} = 23.16, SD = 3.02) were recruited via an emailed advertisement. The participants were engineering students with no prior experience in cybersecurity or prior knowledge of the experimental platform or paradigm. None of the participants were involved with the first experiment. Participants were paid CAD 40.00 for 2 hours of participation.

Experimental platform. The main difference between the experimental platform used in our second experiment and the one used in our first was that we added a minimap, which allowed users to see where their current viewport fell within the larger network (see Figure 4). The minimap could also show how a tag impacted scores throughout the entire network by showing widespread color changes beyond the user's current viewport. We built navigation control into the minimap so that the users could effectively *jump* to new points in the network by clicking on the corresponding point in the minimap, allowing for more efficient network navigation.

We overlaid recommendation boxes on the minimap as a new method for displaying recommendations. These boxes suggested areas for inspection rather than individual emails. The recommendation boxes were a function of the average score of the emails within each possible box but with a minimum size such that the system did not simply recommend individual

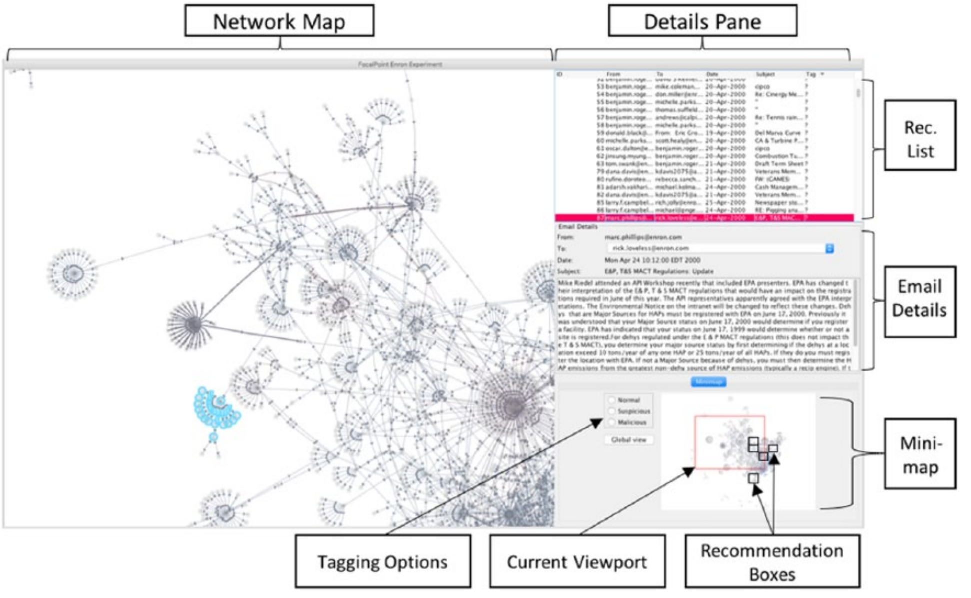

Figure 4. The experimental platform used for Experiment 2. The main changes that are visible are the size of the network and the presence of the minimap. Note that the recommendation boxes have been darkened for the purpose of this figure.

emails with high scores. Users could click on the recommendation box and their viewport would shift to that area. These were added to provide the users with additional capacity to use the minimap for navigation.

Our efforts to improve the visual momentum in the interface centered on the degree of context that was provided to the users when they navigated to a new area of the network via the minimap. This context was addressed through two interventions. First, we tried to increase the similarity between successive viewports when a user navigated via the minimap by assigning a proximity coefficient to the recommendation boxes such that they clustered around the user's current viewport. Second, we tried to provide users with a better understanding of the directionality and distance of their minimap navigations by implementing a sweeping transition wherein the screen would smoothly pan to a new location in a continuous motion. These two interventions were, respectively, contrasted by conditions where the recommendation boxes had no proximity coefficient and where the screen simply updated in discontinuous cut when users navigated to a new location in the interface via the minimap. In each user trial, the interface had one of the two recommendation methods (proximity or global) and one of the two transitional methods (sweep or snap). These are summarized below and demonstrated in Supplementary Materials S3.1 and S3.2:

- Recommendation Method:
  - Proximity conditions: Recommendation boxes are the product of both the malicious content in the box and the proximity of the box to the user's current viewport.
  - Global conditions: Recommendation boxes are only the product of the malicious content in the box.

- Transitional Method:
  - Sweep conditions: Continuous panning transitions when navigating to new areas of the network via the minimap.
  - Snap conditions: Discontinuous cut when navigating to new areas of the network via the minimap.

In addition to adding features associated with the minimap, we also increased the size of the network from 150 to 500 emails. This decreased the signal-to-noise ratio to better emulate real-world operating scenarios and encouraged the use of the minimap. We also removed the score column from the recommendation table and the Top-K slider. These decisions were based on pilot testing and sought to encourage reliance on the network map, which would be consistent with how network operators are observed to behave (Werlinger et al., 2010). A full description of the experimental platform used in Experiment 2 is provided in Kortschot et al. (2017).

Scenario design. We modified the scenarios from our first experiment in order to force participants to read the content of the email rather than simply scanning it for some of the characteristics that they had been warned about. In the modified scenarios, participants sought to identify a chain of users discussing a suspicious subject. The nodes representing the users who were involved in the target conversations were distributed over a greater portion of the network compared to the first experiment, requiring more navigation. Each scenario retained the underlying island-hopping structure of the first experiment but had *linearly connected* emails, meaning that users who received an email in the conversation never responded to the person who sent them that email, instead sending a new email to a new member of the chain.

Each of the four scenarios involved users talking about something that was either illegal or against company protocol. The two illegal scenarios involved users discussing an embezzlement scheme or discussing leaking confidential information to the press, respectively. The two protocol violation scenarios involved users requesting and sharing passwords through email or discussing confidential information (with no mention of leaking it to the press), respectively.

Experimental design. The two intervention dimensions (i.e., recommendation and transitional method) resulted in a 2×2 within-subjects experimental design wherein participants completed four trials, each with a different pairing of recommendation and transitional method, and with a new scenario. All trials and conditions were randomized and counterbalanced to account for any learning effects.

Measures. Both interventions were centered on use of the minimap. As such, our analyses focused on minimap navigations rather than viewport navigations. The breakdown of action-classes was identical to Experiment 1 in that we examined disengagement and engagement delays on either side of a shift (see Figure 2). However, instead of the shift action being panning and zooming, we focused on jumping via the minimap. Using this method, we observed how the abovementioned interventions influenced the different phases of an attentional switch. Control actions remained identical to Experiment 1.

In addition to examining the effects of individual attentional switches, we also sought to characterize the cumulative effect of these switches on task performance. To do this, we examined the average action latency, the total viewport movement, and the number of switches—all across an entire trial—and compared them to the completion time of that trial. Completion time was measured as the time of the last malicious tag given by the user and was used as the principal performance metric. Once again, we observed ceiling effects with accuracy, with only 25 false tags out of the 380 total tags over the experiment.

Procedure. The procedure of Experiment 2 was similar to Experiment 1 with some key differences. Training was delivered through a narrated video of a modified version of the PowerPoint from the first experiment (Supplementary Material S4). Participants were encouraged to pause the video and ask questions of the experimenter, who also demonstrated some of the concepts on a sample network during planned pauses. The sole experimenter judged when participants were ready for experimentation. We do not believe that this presents a significant limitation as we were again focused on microbehavioral measures related to switch cost rather than examining experimental performance alone.

Prior to each trial, the participants were alerted to what marker they would be searching for. This was done after pilot testing revealed that participants had substantial difficulty finding the target conversations, which looked very similar to benign emails at first glance. Following each trial, participants completed a NASA-TLX workload rating scale (Hart & Staveland,

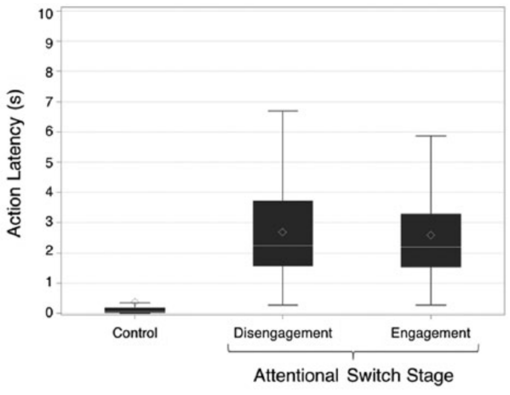

Figure 5. Boxplots showing the distribution of action latencies across control actions, disengagements, and engagements. All outliers have been removed from this graphic.

1988) and a system usability scale (SUS; Brooke, 1996). These were included to assess the relative differences in usability or workload resulting from the experimental manipulations and not to compare against industry standards.

Results. We used the same outlier removal procedures from Experiment 1 to eliminate long pauses that were not reflective of disengagement or engagement processes as well as double clicks. Again, a Wilcoxon rank-sum test was used for several of our analyses due to skewed data and an inability to fit mixed models to those data. Relative to the control condition (see Figure 2; $M = 0.38$ seconds, $SD = 0.93$ seconds), we found significant disengagement delays ($M = 2.68$ seconds, $SD = 1.51$ seconds; $z = 30.91$, $p < .0001$) and significant engagement delays ($M = 2.60$ seconds, $SD = 1.51$ seconds; $z = 33.24$, $p < .0001$), resulting from attentional switches via minimap navigation. Figure 5 illustrates these results.

We did not find a significant relationship between disengagement and engagement delays, $F(1, 447) = 0.30$, $p > .05$, indicating that pausing longer prior to making a jump did not alleviate the engagement delay following that jump.

Generalized linear mixed models on disengagement and engagement delays between experimental conditions found a significant effect of transitional method on disengagement delay, with sweeping transitions ($M = 2.22$ seconds, $SE = 0.03$ seconds) shortening disengagement delays relative to snapping transitions ($M = 2.60$ seconds, $SE = 0.03$ seconds), $F(1, 374) = 8.25$, $p < .001$. The sweeping transitions also had a significant reduction in engagement delays ($M = 2.21$ seconds, $SE = 0.03$ seconds) compared to the snap conditions ($M = 2.46$ seconds, $SE = 0.03$ seconds), $F(1, 442) = 4.69$, $p < .05$. Figure 6 illustrates these results. We did not find an effect of recommendation method on disengagement delays, $F(1, 374) = 0.01$, $p > .05$, or engagement delays, $F(1, 442) = 0.03$, $p > .05$.

Having again observed microperformance decrements of attentional switching, we then examined the cumulative, macroperformance impacts. To do this, we ran mixed linear models examining the relationship between completion time, average action latency, total viewport movement, and total number of actions. Completion time was measured as the time of the fifth and final malicious tag in each trial. If participants failed to tag all five emails, the completion time was set to the maximum trial length of 15 minutes. A trial's average action latency was the average amount of time between successive actions across a trial. The total movement within a trial was the summation of both zooming and panning behavior, rescaled between 0 and 1, since zooming and panning were measured on different scales.

There was a significant negative relationship between the total number of actions and both the average action latency within trials, $F(1, 48) = 26.54$, $p < .0001$, and the total movement within a trial, $F(1, 48) = 7.59$, $p < .01$. There was also a significant positive relationship between total number of actions and completion time, $F(1, 48) = 55.12$, $p < .0001$. The average action latency across a trial had a positive relationship with the total viewport movement within that trial, $F(1, 48) = 10.69$, $p < .01$, but not with the completion time of that trial, $F(1, 48) = .02$, $p > .05$. We did not find a significant relationship between the total viewport movement in a given trial and the completion time of that trial, $F(1, 48) = 1.02$, $p > .05$. Figure 7 shows the correlation matrix between the abovementioned variables.

There was no effect of either the recommendation method, $F(1, 46) = 0.43$, $p > .05$, or the transitional method, $F(1, 46) = 0.15$, $p > .05$, on

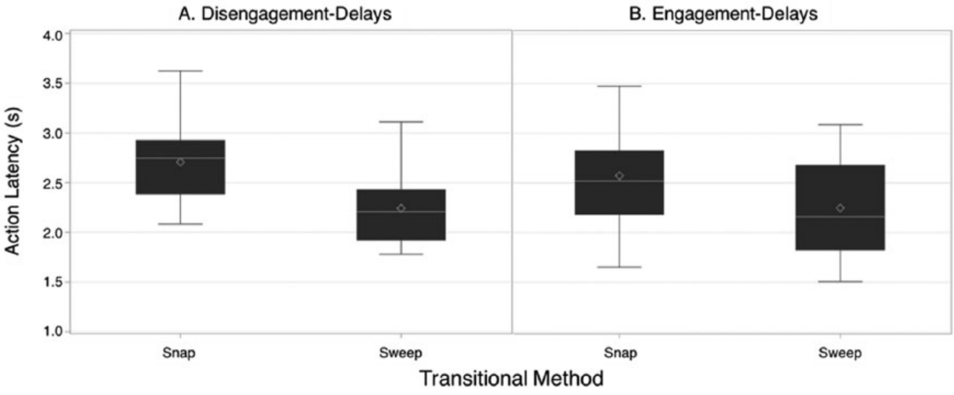

Figure 6. (A) Boxplots showing the differences in disengagement delays between the two transitional methods used in Experiment 2. (B) Boxplots showing the differences in engagement delays between the two transitional methods used in Experiment 2. All outliers have been removed from this graphic.

completion times. However, there was a significant interaction effect between recommendation method and transitional method, $F(1, 46) = 7.15$, $p = .01$, indicating that the proximity recommendation method yielded faster completion times when using snap transitions but slower times when using sweeping transitions.

We also found that the transitional method significantly impacted the average action latency across a trial, $F(1, 46) = 7.49$, $p < .01$, with snap conditions eliciting shorter latencies on average (M = 0.41 seconds, SD = 0.04 seconds) than sweep conditions (M = 0.47 seconds, SD = 0.04 seconds). We did not find that either the recommendation method or the transitional method had any impact on the total movement within a trial, $F(1, 46) = 0.07$, $p > .05$; $F(1, 46) = 1.99$, $p > .05$.

We computed a single overall workload measure for each trial (Byers, Bittner, & Hill, 1989) and found that neither transitional method, $F(1, 48) = .20$, $p > .05$, nor recommendation method, $F(1, 48) = 2.31$, $p > .05$, were significant predictors of workload. However, we did find a significant interaction effect, $F(1, 48) = 5.00$, $p < .05$, indicating that sweeping transitions yielded lower workload ratings in the global recommendation method but higher workload in the proximity recommendation method. There were no significant impacts on SUS scores across transitional method, $F(1, 48) = .06$, $p > .05$, recommendation method, $F(1, 48) = .04$, $p > .05$, or the interaction between the two, $F(1, 48) = 3.02$, $p > .05$.

Discussion

Our results replicated the findings from Experiment 1, showing that participants succumbed to both disengagement and engagement delays when moving to a new viewport. These delays were longer in Experiment 2, likely as a result of the added capacity to make larger jumps.

The panning transition was designed to increase visual momentum by allowing users to grasp where they have navigated to relative to where they previously were (Woods, 1984). In the snap conditions, it may have been difficult for the operators to grasp any directionality of the transition, and it therefore took additional time to engage with the new viewport following a jump. We also found that the sweeping conditions facilitated disengagement processes by showing a significant reduction in disengagement delays relative to the snap conditions. We believe this result stems from improved visual momentum reducing the degree to which participants needed to prepare for switches. We did not find an effect of recommendation method on engagement delays, contradicting the transitional method finding. We had anticipated that disengagement and engagement delays would be lessened in the proximity recommendation

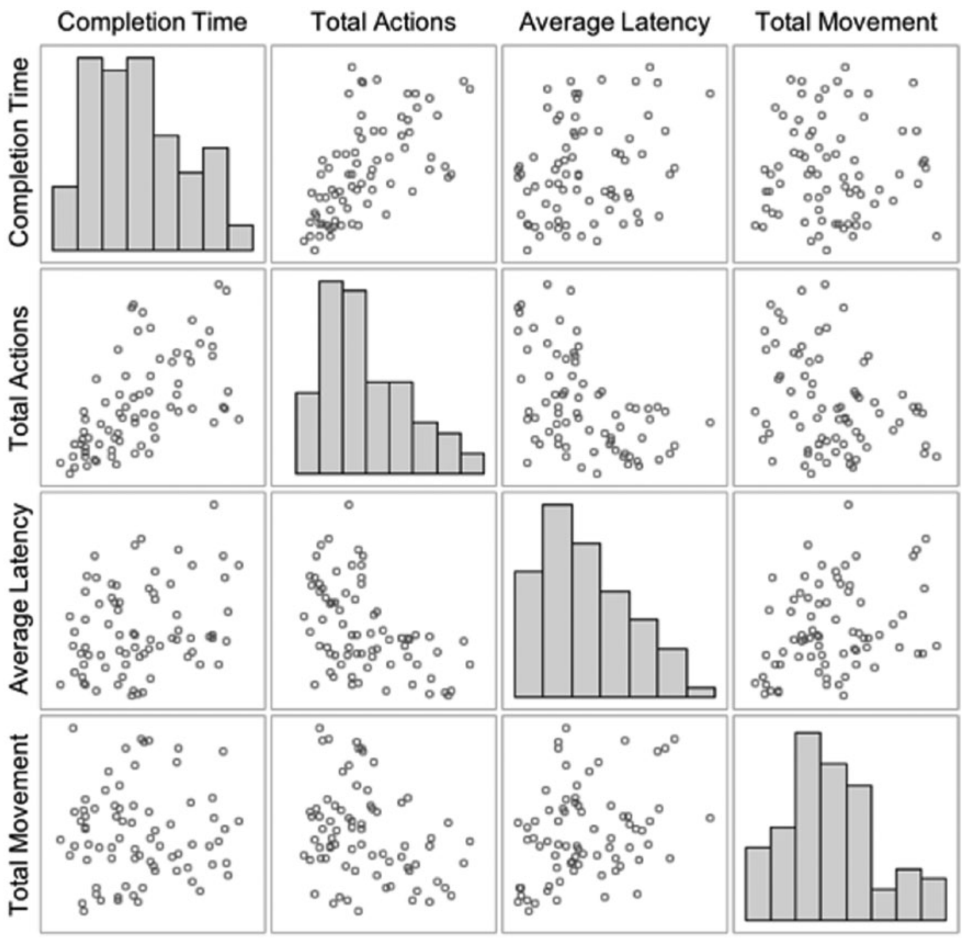

Figure 7. A correlation matrix showing the relationships and distributions between completion time, total number of actions, average action latency, and total movement. Each point represents a single trial.

conditions due to new viewports being closer to prior ones, but this was not borne out by evidence. This suggests that in spite of the shared context, a screen update requires significant disengagement and engagement irrespective of the distance between successive screens.

Our second experiment also characterized the cumulative effect of attentional switches across an entire trial. We found a significant negative relationship between action count and action latency, indicating that as people sped up their actions, they performed more of them. Interestingly, we did not find a significant relationship between average action latency and completion time, which shows that as people sped up their actions, they did not complete trials faster. These findings show that improving the speed at which a user interacts with the system on an action-by-action basis does not necessarily improve the speed that they complete a task. Therefore, the wastefulness of actions needs to be considered when attempting to improve operator temporal performance.

The significant relationship between the total viewport movement in a trial and the average action latency from that trial indicates that the more people move throughout a network, the slower their actions become. It is consistent with our microfindings to attribute this to the cumulative effect of disengagements and engagements across a trial. Surprisingly, we did not find a relationship between total movement and completion time or between average action latency and completion time. However, based on the strong correlation between action latency and movement, we suspect that this relationship does exist. This hypothesis needs to be studied further to make concrete conclusions and design recommendations.

We found a significant interaction effect between recommendation method and transitional method on completion time, showing that in the sweep conditions, the global recommendation method yielded the fastest completion times, whereas in the snap conditions, the proximity method yielded faster completion times. We believe the most likely explanation for this effect is that in the proximity conditions the minimap navigations were to closer areas, as the recommendation boxes were clustered around the user's current viewport. Participants may have therefore needed less context when navigating to those proximal areas, thereby reducing the need for a contextualizing transition. We also found that across a trial, the snap conditions elicited shorter average action latencies than the sweep conditions, indicating that the snap conditions accelerated the rate at which users interacted with the system. It is unclear why this effect occurred and should therefore be studied at greater depth in future experiments.

Workload was not significantly impacted by either transitional method or recommendation method. We did find a significant interaction, suggesting that in the global recommendation conditions, sweeping yielded lower workload scores, whereas in the proximity recommendation conditions, snapping yielded lower scores. Again, we believe that this is the result of the recommendation boxes being further away in the global recommendation conditions and therefore the sweeping transition lowering the cognitive load of navigating with the minimap. In the proximity recommendation conditions, where minimap navigations were closer, the sweeping transition was less necessary.

We did not find a significant effect of either transitional method or recommendation method on system usability. This suggests that participants were tolerant to the various interventions. This nonsignificant result is telling. As participants did not find the interventions (i.e., sweeping, proximity) to be less usable compared to the controls (i.e., snapping, global) in spite of their relative novelty, we believe that new interventions aimed at lessening the cognitive load of operators should be sought after, regardless of their potential novelty to users.

GENERAL DISCUSSION

Taken together, the results of our two studies demonstrate the adverse effects of attentional switching in ANM at both the micro- and macrolevels. We provide uncommon evidence that the findings from the laboratory studies summarized in much of the attentional switching literature generalize to a more complex task environment. We also showed that the disengagement and engagement phases of an attentional switch outlined by Posner and Presti (1987) translate to navigational tasks. Our results suggest that attentional switches may be prevalent in real-world tasks and that the cumulative effects of these switches could be detrimental to performance.

It should be noted that trials in our study were limited to 15 minutes. The net effects of attentional switches are likely greater when they accrue across a full work shift. This suggests that attentional switches represent an important aspect of operator performance and should therefore be considered in the design of future DSSs. Future work should examine the costs of attentional switches associated with real-world, nonnavigational tasks such as opening a new window or switching to a different application over the course of full operator shifts.

A novel finding from our experiments was the self-imposed preparation that participants demonstrated. Previous literature has demonstrated that increasing preparation time facilitates attentional switching (e.g., Sohn & Anderson, 2001), but rarely do these experiments allow for participants to assume these preparatory periods voluntarily. Recently, Longman et al. (2017) showed that voluntary preparations reduced the costs of a switch. We did not find a significant relationship between the duration of disengagement delays and their corresponding engagement delays. However, the fact that participants were consistently incurring such a significant disengagement delay suggests that this delay either serves a purpose or is the result of a limitation in human cognition. Future studies should examine whether this behavior is exhibited across other attentional switches and whether or not this relationship can be facilitated through design.

The studies reported here also demonstrate the practicality and effectiveness of PDM within ANM. The cost of attentional switching has been difficult to study in real-world scenarios largely due to the need for highly sensitive time measures. However, if the user's interactions are easily logged, PDM can overcome this issue. Beyond PDM's ability to generalize time-sensitive measures like attentional switch detection to real-world domains, we believe that it represents a promising method for inferring cognitive states. There has been a recent trend in the literature toward biometric readings for cognitive measurement, largely motivated by the improved access and accuracy of these measures (Verwey, Shea, & Wright, 2015). Although increasingly accessible, biometric readings are still far more invasive than PDM (Balakrishnan, Durand, & Guttag, 2013). PDM can be implemented on many systems and imposes virtually no additional load on the operator. It is therefore highly feasible for domains where operators work long, sedentary shifts and wear minimal equipment. Future studies should continue to develop methods of inferring cognitive phenomena and states through the use of logged operator interaction data.

While the current study demonstrated the efficacy of PDM for inferring cognitive events (i.e., disengagements and engagements), future studies should look at inferring cognitive states (e.g., stress, boredom, etc.). These are less likely to have direct behavioral manifestations and clear onsets and offsets and would therefore demand machine learning algorithms to be run on more continuous data streams such as cursor position. However, in spite of the increased difficulty, the potential utility of this research is extensive and can be directly applied to triggering adaptive interface behaviors based on operator measurement through PDM (Feigh, Dorneich, & Hayes, 2012).

CONCLUSION

As machine learning and reasoning become more prevalent in ANM, the psychological demands of the human operators need to be considered to optimize the joint human-machine task performance. We have shown here that these demands can be successfully incorporated into the interfaces and algorithms that drive the DSS through the use of PDM.

Our study focused solely on attentional switching within ANM. However, with the growing size and complexity of both computer networks and the attacks on those networks, the psychological demands imposed on the operator will continue to grow. Therefore, in order to realize the full potential of the machine learning driving DSSs, a full spectrum of cognitive states should be studied and incorporated into the design of future systems.

ACKNOWLEDGMENTS

This project was partially supported by Uncharted Software Inc., the Ontario Centres of Excellence (Grant 23884), and the Natural Science and Engineering Research Council (Grant CRDPJ 500832-16). We would like to thank Scott Langevin and Scott Ralph at Uncharted Software, as well as Yihao Du and Alexis Morris for their contributions to these grants.

KEY POINTS

- Attentional switches are both preceded and succeeded by increased action latency relative to actions that do not involve a change of context, during which operators are likely preparing for or recovering from an attentional switch, respectively.
- Additional movement over the course of a trial decreases the mean action latency over that trial. Decreases in action latency are correlated with decreases in completion time.
- Providing context during attentional switches lessens both the disengagement and engagement load resulting from those switches.
- Passive data monitoring is an effective and non-invasive alternative to biometric data gathering

for inferring cognitive phenomena in operators in highly interactive domains.

SUPPLEMENTAL MATERIAL

The online supplementary material is available with the manuscript on the *HF* Web site.

REFERENCES

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301. http://doi.org/10.1177/1094428112470848

Ahmad, A., Hadgkiss, J., & Ruighaver, A. B. (2012). Incident response teams—Challenges in supporting the organisational security function. *Computers and Security*, *31*(5), 643–652. http://doi.org/10.1016/j.cose.2012.04.001

Alrajeh, N. A., Khan, S., & Shams, B. (2013). Intrusion detection systems in wireless sensor networks: A review. *International Journal of Distributed Sensor Networks*, *2013*, 1–7. http://doi.org/10.1109/SURV.2013.050113.00191

Aoidh, E. M., Bertolotto, M., & Wilson, D. C. (2012). Towards dynamic behavior-based profiling for reducing spatial information overload in map browsing activity. *Geoinformatica*, *16*, 409–434. http://doi.org/10.1007/s10707-011-0137-4

Ashfaq, A. B., & Khayam, S. A. (2011). Evaluation of contemporary anomaly detection systems (ADSs). In J. A. Zubairi & A. Mahboob (Eds.), *Cyber security standards, practices and industrial applications: Systems and methodologies* (pp. 90–112). Hershey, PA: Information Science Reference. http://doi.org/10.4018/978-1-60960-851-4.ch006

Balakrishnan, G., Durand, F., & Guttag, J. (2013). Detecting pulse from head motions in video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3430–3437. Retrieved from https://dl.acm.org/citation.cfm?id=2516035

Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. *Computers in Human Behavior*, *48*, 51–61. http://doi.org/10.1016/j.chb.2015.01.039

Bennett, K. B., & Flach, J. M. (2012). Visual momentum redux. *International Journal of Human Computer Studies*, *70*(6), 399–414. http://doi.org/10.1016/j.ijhcs.2012.01.003

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(2), 363–382. http://doi.org/10.1037/a0018106

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and Raw Task Load Index (TLX) correlations: Are paired comparisons necessary? In A. Mital & B. Das (Eds.), *Advances in industrial ergonomics and safety* (vol. 1., pp. 481–488). London: Taylor & Francis.

Corchado, E., & Herrero, Á. (2011). Neural visualization of network traffic data for intrusion detection. *Applied Soft Computing Journal*, *11*(2), 2042–2056. http://doi.org/10.1016/j.asoc.2010.07.002

D'Amico, A., Whitley, K., Tesone, D., O'Brien, B., & Roth, E. M. (2005). Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*, 229–233. http://doi.org/10.1177/154193120504900304

Dombrowe, I., Donk, M., & Olivers, C. N. L. (2011). The costs of switching attentional sets. *Attention, Perception, and Psychophysics*, *73*(8), 2481–2488. http://doi.org/10.3758/s13414-011-0198-3

Donmez, B., Boyle, L. N., & Lee, J. D. (2006). The impact of distraction mitigation strategies on driving performance. *Human Factors*, *48*(4), 785–804.

Dorneich, M. C., Mathan, S., Ververs, P. M., & Whitlow, S. D. (2007). An evaluation of real-time cognitive state classification in a harsh operational environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*, 146–150. http://doi.org/10.1177/154193120705100401

Dorneich, M. C., Whitlow, S. D., Ververs, P. M., & Rogers, W. H. (2003). Mitigating cognitive bottlenecks via an augmented cognition adaptive system. *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, *1*, 937–944. http://doi.org/10.1109/ICSMC.2003.1243935

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272–300. http://doi.org/10.1037/0033-295X.87.3.272

Ettwig, J. F., & Bronkhorst, A. W. (2015). Attentional switches and dual-task interference. *PLoS ONE*, *10*(3), e0118216. http://doi.org/10.1371/journal.pone.0118216

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(6), 1008–1024. http://doi.org/10.1177/0018720812443983

Franke, U., & Brynielsson, J. (2014). Cyber situational awareness—A systematic review of the literature. *Computers & Security*, *46*, 41. http://doi.org/10.1016/j.cose.2014.06.008

Goodall, J. R., Lutters, W. G., & Komlodi, A. (2009). Developing expertise for network intrusion detection. *Information Technology & People*, *22*(2), 92–108. http://doi.org/10.1108/09593840910962186

Goutam, R. K. (2015). Importance of cyber security. *International Journal of Computer Applications*, *111*(7), 14–18.

Hart, S. G., & Staveland, L. E. (1988). Developent of a multidimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, The Netherlands: Elsevier.

Hopf, J.- M., Boehler, C. N., Schoenfeld, M. A., Heinze, H.- J., & Tsotsos, J. K. (2010). The spatial profile of the focus of attention in visual search: Insights from MEG recordings. *Vision Research*, *50*(14), 1312–1320. http://doi.org/10.1016/j.visres.2010.01.015

Janczyk, M., & Kunde, W. (2010). Does dorsal processing require central capacity? More evidence from the PRP paradigm. *Experimental Brain Research*, *203*(1), 89–100. http://doi.org/10.1007/s00221-010-2211-9

Khan, S., Gani, A., Abdul Wahab, A. W., Shiraz, M., & Khan, I. A. (2016). Network forensics: Review, taxonomy, and open challenges. *Journal of Network and Computer Applications*, *66*, 214–235. http://doi.org/10.1016/j.jnca.2016.03.005

Kortschot, S. W., Sovilj, D., Soh, H., Jamieson, G. A., Sanner, S., Carrasco, C., . . . Ralph, S. (2017). *An open source adaptive user interface for network monitoring*. Paper presented at the IEEE International Conference on Systems, Man and Cybernetics, Banff, Canada.

Li, P., Salour, M., & Su, X. (2008). A survey of internet worm detection and containment. *IEEE Communications Sur-*

veys and Tutorials, *10*(1), 20–35. http://doi.org/10.1109/COMST.2008.4483668

Longman, C. S., Lavric, A., & Monsell, S. (2017). Self-paced preparation for a task switch eliminates attentional inertia but not the performance switch cost. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(6), 862–873. http://doi.org/10.1037/xlm0000347

Mitropoulos, S., Patsos, D., & Douligeris, C. (2006). On incident handling and response: A state-of-the-art approach. *Computers and Security*, *25*(5), 351–370. http://doi.org/10.1016/j.cose.2005.09.006

Moar, J. (2015). *The future of cybercrime & security: Financial and corporate threats & mitigation*. Retrieved from https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion

Nobre, A. C., & Kastner, S. (Eds.). (2014). *The Oxford handbook of attention*. Oxford, UK: Oxford University Press.

Olszewski, D. (2014). Fraud detection using self-organizing map visualizing the user profiles. *Knowledge-Based Systems*, *70*, 324–334. http://doi.org/10.1016/j.knosys.2014.07.008

Palmius, N., Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Geddes, J. R., Goodwin, G. M., & De Vos, M. (2016). Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering*, *64*(8), 1–12. http://doi.org/10.1109/TBME.2016.2611862

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. *Computers & Security*, *31*, 597–611. http://doi.org/10.1016/j.cose.2011.12.010

Posner, M. I., & Presti, D. E. (1987). Selective attention and cognitive control. *Trends in Neurosciences*, *10*(1), 13–17. http://doi.org/10.1016/0166-2236(87)90116-0

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101–130. http://doi.org/10.1037/0033-295X.115.1.101

Sohn, M.- H., & Anderson, J. R. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology: General*, *130*(4), 764–778. http://doi.org/10.1037/0096-3445.130.4.764

Tyworth, M., Giacobe, N. A., & Mancuso, V. (2012). Cyber situation awareness as distributed socio-cognitive work. *Cyber Sensing 2012, 8408*(Level 2), 1–9. http://doi.org/10.1117/12.919338

Verwey, W. B., Shea, C. H., & Wright, D. L. (2015). A cognitive framework for explaining serial processing and sequence execution strategies. *Psychonomic Bulletin & Review*, *22*, 54–77. http://doi.org/10.3758/s13423-014-0773-4

Werlinger, R., Muldner, K., Hawkey, K., & Beznosov, K. (2010). Preparation, detection, and analysis: The diagnostic work of IT security incident response. *Information Management & Computer Security*, *18*(1), 26–42. http://doi.org/10.1108/09685221011035241

Woods, D. D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, *21*, 229–244. Retrieved from http://ac.els-cdn.com/S0020737384800437/1-s2.0-S0020737384800437-main.pdf?_tid=a6312d00-238a-11e7-9c20-00000aacb360&acdnat=1492446585_63d6b2c8437eedd810baaa400937658b

Sean W. Kortschot is a PhD candidate at the University of Toronto. He received a master's of applied science from the University of Toronto in 2016.

Dusan Sovilj is a postdoctoral fellow at the University of Toronto. He received a PhD from Aalto University School of Science in Computer Science.

Greg A. Jamieson is a professor in mechanical and industrial engineering at the University of Toronto. He received a PhD from the University of Toronto in 2002.

Scott Sanner is an assistant professor in industrial engineering at the University of Toronto. He received a PhD in computer science from The University of Toronto in 2008.

Chelsea Carrasco is a master of applied science candidate at the University of Toronto. She received a BASc in industrial engineering in 2014 from the University of Toronto.

Harold Soh is an assistant professor in computer science at the National University of Singapore. He received a PhD from Imperial College in London in Computer Science.

Date received: May 31, 2017
Date accepted: May 27, 2018