Human Span-Of-Control in Cyber Operations: An Experimental Evaluation of Fan-Out

Vincent F. Mancuso

Gregory J. Funke

Monica B. Eckold

Adam J. Strang

HUMAN SPAN-OF-CONTROL IN CYBER OPERATIONS: AN EXPERIMENTAL EVALUATION OF FAN-OUT

Vincent F. Mancuso[1], Gregory J. Funke[2], Monica B. Eckold[2], Adam J. Strang[2]

[1]Oak Ridge Institute for Science and Education, Wright-Patterson AFB, OH
[2]Air Force Research Laboratory, Wright-Patterson AFB, OH

Modern cyber operations require operators to maintain supervisory control of remote computer agents. A current operational concern is the number of agents an operator can control at once. This issue resonates with similar "span-of-control" research conducted in UAV operations (e.g., Cummings & Mitchel, 2008). One way to identify operator span is via "fan-out," a numeric calculation that provides a span-of-control estimate based on system and environmental variables. However, fan-out is a mechanical representation that only accounts for task-characteristics and environmental variables, thus providing an upper bound of human performance that does account for cognition, workload, or work interruptions. The present study compares fan-out estimates against actual human performance in a supervisory control cyber task. Results are discussed and future research trajectories proposed.

Over the last decade, cyber security has become a primary concern for homeland security. This concern is only likely to grow as we continue to develop technologies that employ computer networks to manage our major private (e.g., banking) and government (e.g., nuclear power plant) assets.

Traditionally, cyber security research focuses mostly on computer science and engineering problems, such as the development of algorithms and systems to detect, identify, and mitigate specific threats that exist on computer networks. While this research is critical to our national defense, it does not acknowledge the critical role that human operators play in cyber security. Recently, the Human Factors community has recognized this critical research gap and in response has started an initiative to explore human-centered aspects of cyber operations. Current research has focused on identifying important dimensions of cognition within cyber operations, such as situation awareness (Giacobe, 2012), team knowledge structures (Mancuso & McNeese, 2012), and team collaboration (Rajivan et al, 2013). However, little research has focused explicitly on specific issues related to task load and individual operations management.

Current cyber operations exist within a complex system of human-machine interaction, where operators are tasked with monitoring the activities, efficacy, and progress of intelligent and autonomous computer systems (Tyworth et al., 2013). In these environments, cyber operators supervise intelligent systems as they execute tasks across the network. This task places significant cognitive demand on operators due to the need to maintain situation awareness while dividing attention across a set of dynamic tasks and managing a deluge of information. While novel within the context of human-centered cyber research, these environments share many commonalities with human supervisory control tasks.

In Human Supervisory Control (HSC) tasks, a human-in-the-loop provides an autonomous asset with high-level plans, instructions, and goal directives (Miller, 2004). In operations such as Unmanned Aerial Vehicles (UAVs), a task that is currently performed using manual control but will likely shift to HSC (at least partially) in next-gen operations, researchers have examined human performance in HSC simulations to improve system design, mitigate operator workload, account for individual differences, and identify/optimize human span-of-control. However, similar research has not been conducted on cyber operations despite that fact that current-gen work environments often involve HSC tasks.

Based on this, the purpose of this research was to conduct a preliminary exploration of the translation of previous HSC research to cyber operations. Specifically, in this paper we perform an experimental evaluation of the predictive span-of-control metric known as "fan-out." To assess fan-out's

potential application for cyber-operations, we will compare fan-out estimates against observed human performance in a cyber supervisory control task that requires human operators to control multiple automated agents across a simulated area network.

Fan-Out

Nehme et al. (2008) and Cummings & Mitchel (2008) developed a work-flow model of HSC tasks that parse performance into three components: service time, productive time, and wait time. Service time, also known as interaction time (IT) is the amount of time that it takes a human to service an autonomous asset. This value includes the amount of time it takes an operator to allocate the asset, determine the necessary inputs, and expresses those inputs to the asset via the interface (Olsen & Goodrich, 2003). Next, productive time, also known as neglect time (NT), represents the average amount of time an asset can operate without an operator's intervention (Olsen & Goodrich, 2003). Finally, wait time (WT) is the amount of time that an asset spends in the queue waiting to be serviced by the human. Maximum efficiency can be achieved by minimizing service and wait times, while maximizing productive time.

To help improve HSC systems for maximal performance, some of the variables mentioned above are combined to obtain computational estimates of fan-out (FO) according to the equation, $FO = (NT / IT) + 1$ (Dixon, Wickens & Chang, 2005; Miller, 2004; Olsen & Goodrich, 2003; Sheridan, 1992). From a linear throughput perspective, FO indicates the number of homogenous assets a single operator can control at once without interference or drag (Olsen & Wood, 2004). Thus, FO can (and has been) used as a theoretical upper bound for estimates of operator span-of-control. For example, Cummings & Mitchell (2008) found that operators performed at approximately 36% below fan-out estimates in a UAV simulation and speculated three reasons for sub-optimal performance: a) WT in the human decision-making queue (WTQ), b) interaction wait time (WTI), and c) wait time due to loss of situation awareness (WTSA). WTQ occurs when an asset goes unattended while an operator completes their decision making task. WTI, which is very similar to IT, includes time that the operator spends determining the appropriate action and communicating it to the asset. Finally, WTSA includes time that the operator spends away from the asset as they perceive elements in the environment, comprehends their meanings and makes future predictions of their status. While conceptually quite simple, capturing these metrics for the purpose of predicting human span of control in a computational model can be quite complex, especially WTQ and WTSA.

Based on such evidence, we expected that, assuming accurate inputs, that estimates of FO in a cyber HSC simulation (BotNET Operator Ratio Determination; BOARD v1.5) would correlate with human span-of-control. However, due to the inherent complexity of cyber-operations, we also expected that actual span-of-control will be significantly lower than the estimates generated by the fan-out equation. Together, these finds would allow us to pinpoint the theoretical and realistic span-of-control for HSC cyber operations.

Methods

Simulated Task Environment

BOARD is a human-in-the-loop scaled world simulation that is set in the context of cyber supervisory control operations. In this simulation, participants are tasked with remotely controlling computer agents using a command line interface. During the tasks, participants interact with three main components: a) the agent control window, b) the agent beacon monitor, and c) the mission commander messaging system (Figure 1).

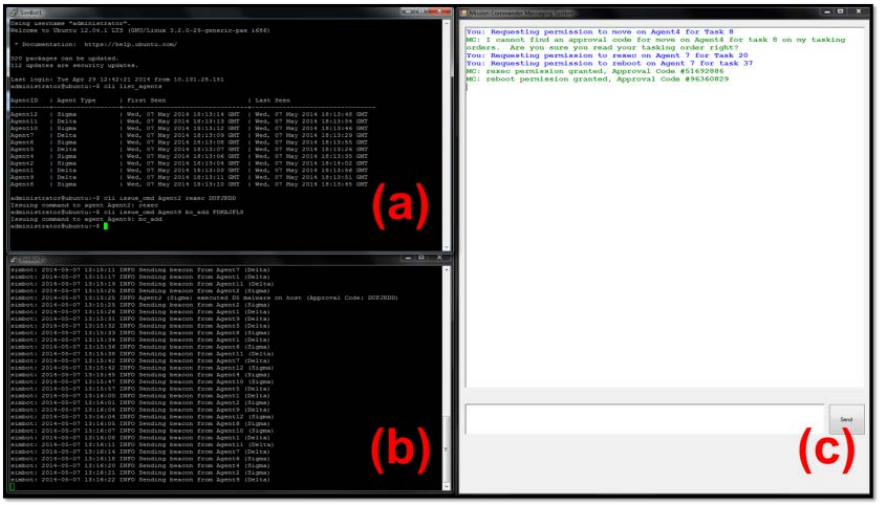

Figure 1. The BOARD simulation user interface depicting the agent control window (a), the agent beacon monitor (b), and the mission commander messaging system(c)

These interface elements allow participants to remotely control computer agents using a set list of commands, monitor the progress and state of varying missions, and communicate with a mission commander to obtain permission to execute restricted commands.

At the start of each experimental trial of BOARD performed in this experiment, participants received an Operating Tasking Order (OTO) print out and were given access to a set of computer agents. Each OTO comprised a mission set, where missions within the set were characterized by an assigned "agent type" and three steps that must be accomplished (in serial order) to successfully complete the mission (see Table 1 for an example). In each mission step, participants were provided with a command they must execute and an authorization code to run the command. In order to complete the mission, the participant needed complete all three steps in order, using an agent of the correct type. If they enter the incorrect authorization code, the command may not execute and they cannot complete the mission.

Table 1.
Example BOARD Operating Tasking Orders.

	Type	Step 1	Step 2	Step 3
1	Delta	getlog "yxdembfl"	grab "xtizjluw"	phish "sblukllp"
2	Sigma	getlog "hlhsggwg"	rename "lgxustfx"	bl_url "rahtdkcl"
3	Delta	bc_add "hihhohwv"	bc_add "upfyvedo"	rexec

In some cases, authorization codes were not provided to participants (e.g. Mission 3, Step 3 in Table 1). In these instances, the participant was forced to utilize the Mission Commander Messaging system to request an authorization code. This manipulation was included to represent cyber rules of engagement, which require command authorization before executing sensitive missions. After requesting the code, the mission commander responded with an authorization code in a variable amount of time (between 10 and 60 seconds). While participants waited for authorization, they were permitted to move on and complete other missions (but not subsequent steps in the same mission).

Measures

In this study, fan-out measures were customized for each participant from typing ability and average response time for agents during a pre-experiment simulation. Prior to the experimental task, each participant was asked to complete a typing test lasting two minutes. Words-per-minute results were adjusted based on errors, and then multiplied by 5 to calculate estimated characters per minute (CPM;

Arif & Stuezlinger, 2009). Using the equation proposed by Olsen & Goodrich (2003), IT was calculated by comparing the CPM to the average character length of agent execution commands (37 characters) and NT was calculated from the average time it took an agent to execute a command (45 seconds).

Observed operator span-of-control was calculated as the maximum number of agents operating in a one-minute window averaged across each 20 minute trial.

Participants

Twenty-one participants (17 Male, 4 Female), drawn from local universities and United States Air-Force agencies, were financially compensated for their participation (on-duty Airmen received no compensation outside of their regular duty pay). All participants were between the ages of 19 and 45 (M = 23.95, SD = 5.15) and had some level of experience with Command Line Interfaces (CLI) and/or computer programming languages. Four participants had previous cyber security experience (classes, professional experience, etc.).

Procedures

Experimental sessions were conducted in the AFRL Cyber Integrated Performance and Human Effectiveness Research (CIPHER) Laboratory. Prior to the task, participants completed three phases of training. First participants completed a short self-paced computer based training (CBT) which took approximately 15 minutes. Following CBT, the researcher guided the participant through a short practice scenario using the BOARD environment to familiarize themselves with the computer interface, task requirements, and rule. Following the guided training scenario, participants were instructed to complete a second training scenario independently, while the researcher observed their progress and corrected any mistakes that were made. After the participant had reached satisfactory performance on the independent training scenario, participants completed four experimental trials, each lasting 20 minute. During each trial, participants were provided a different number of autonomous agents (4, 8, 12 or 16) with which to complete their assigned missions. These manipulations were part of a larger investigation and are peripheral to the central question of the current study (i.e., evaluation of the accuracy of fan-out measures in cyber). As such, results pertaining to their effects on performance are being prepared for presentation elsewhere. Estimates of operator span-of-control presented herein are means calculated across the four scenarios. The total duration of the experiment was approximately 3 hours.

Results

Fan out estimations were calculated for each individual based on their typing speed (cpm; M = 264.29, SD = 70.33) and the average NT (45 seconds). Results showed no significant correlations between fan-out estimates and actual span-of-control, r = 0.19, p > .05 (Figure 2).

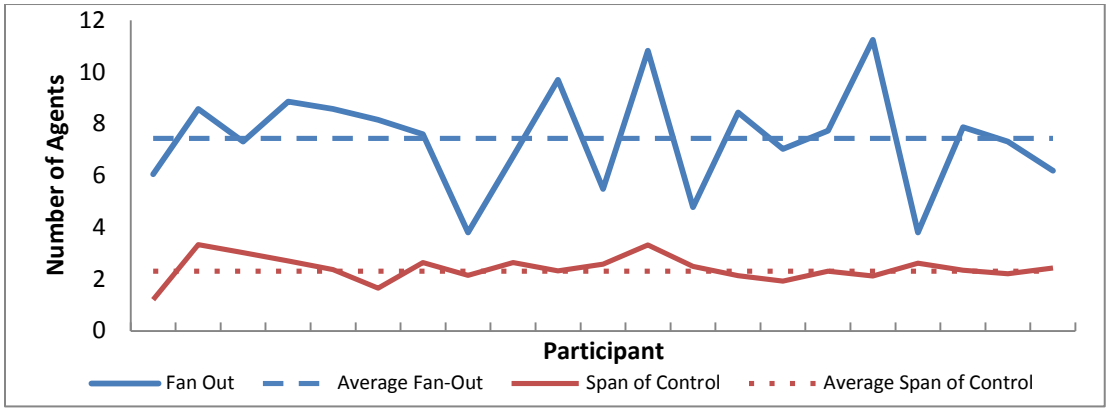

Figure 2. Participants Span-of-Control compared to Fan-Out Estimates

However, consistent with prior research, fan out estimates were found to be significantly higher than the actual span of control. A paired sample *t*-test revealed a significant difference between the predicted fan-out values ($M = 7.43$, $SD = 1.99$) and average span-of-control ($M = 2.27$, $SD = 0.49$), $t = 12.70$, $p < 0.05$.

Discussion

Previous research has presented fan-out as a model upper bound of human span-of-control in several HSC tasks (Crandall & Cummings, 2003). While it is to be expected that actual span-of-control will fall below this value, generally these metrics have been found to be somewhat correlated (Cummings & Mitchell, 2008). Surprisingly, we found that fan-out estimates were not significantly correlated with observed span-of-control in the BOARD cyber HSC simulation. However, we did find a consistent difference between the predicted values (via fan-out) and observed span-of-control. Given that observed span-of-control fell well below the estimates (close to a 100% difference, as compared to the 30-50% difference presented in other HSC tasks, e.g., Cummings & Mitchell, 2008), the current results suggest that the inherent complexity of cyber operations entails a higher cognitive load, reducing operator span well below fan-out estimates.

One caveat of these findings, however, is that fan-out was calculated from a relatively simplistic formulation proposed by Cummings & Guerlain (2004); other researchers have proposed more complex approximations. In their research, Mitchell, Cummings and Sheridan (2003) proposed the addition of wait times to the denominator of the equation. In their expanded formula, wait time is a combination of the time the asset spends in the queue before receiving instructions from the operator and time attributed to operator reorientation and activities supporting situation awareness. Another interpretation by Crandall and colleagues (2005) utilizes other metrics, such as neglect and interaction impact, to calculate an overall performance metric for each asset based on all possible values for interaction and neglect time.

Conclusion

Building from previous HSC research, the purpose of this study was to investigate how the metric of fan-out translated to cyber operations. Using a simulated autonomous agent control task, we evaluated participants' actual span-of-control, and compared them to the estimates calculated by a popular interpretation of the fan-out equation. Our findings indicated that fan-out did not provide an accurate representation of human performance in our task. Future research should focus on identifying the cognitive requirements of cyber work, so that more mature equations for predicting span of control can be developed.

References

Arif, A. S., & Stuerzlinger, W. (2009, September). Analysis of text entry performance metrics. In 2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH). Toronto, CA 26-27 September (pp. 100-105). IEEE.

Crandall, J. W., & Cummings, M. L. (2007). Identifying Predictive Metrics for Supervisory Control of Multiple Robots. IEEE Transactions on Robotics, 23(5), 942 - 951. doi:10.1109/TRO.2007.907480

Crandall, J. W., Goodrich, M. A., & Nielsen, C. W. (2005). Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics, 35(4), 438-439. doi:10.1109/TSMCA.2005.850587

Cummings, M. L., & Guerlain, S. (2004, September). An interactive decision support tool for real-time in-flight replanning of autonomous vehicles. AIAA 3rd Unmanned Unlimited Technical Conference, Workshop and Exhibit, Chicago, Il. doi: DOI: 10.2514/6.2004-6526

Cummings, M. L., & Mitchell, P. J. (2008). Predicting Controller Capacity in Supervisory Control of Multiple UAVs. IEEE Transactions on Systems, Man, and Cybernetics, 38(2), 451-460. doi:10.1109/TSMCA.2007.914757

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis. Human Factors, 47(3), 479-487. doi:10.1518/001872005774860005

Giacobe, N. A. (2013, September). A Picture is Worth a Thousand Alerts. In Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society, 30 September - 4 October (pp. 172-176). San Diego, CA, HFES.

Mancuso, V. F., & McNeese, M. D. (2012, September). Effects of Integrated and Differentiated Team Knowledge Structures on Distributed Team Cognition. In Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society, 22-26 October (pp. 388-392). Boston, MA. HFES

Miller, C. (2004). Modeling human workload limitations on multiple UAV control, Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA: HFES, 526

Mitchell, P. M., Cummings, M. L., & Sheridan, T. B. (2005, May). Management of multiple dynamic human supervisory control tasks. In 10th International Command and Control Research and Technology Symposium. MacLean, VA (pp. 1 - 11).

Nehme, C. E., Kilgore, R. M., & Cummings, M. L. (2008, September). Predicting the impact of heterogeneity on unmanned-vehicle team performance. In Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society, New York, NY 22-26 September (pp. 917-921). HFES.

Olsen Jr, D. R., & Wood, S. B. (2004, April). Fan-out: measuring human control of multiple robots. In Proceedings of the SIGCHI conference on Human factors in computing systems. Vienna, Au 24-29 April (pp. 231-238). ACM.

Olsen, D. R., & Goodrich, M. A. (2003, September). Metrics for evaluating human-robot interactions. In Proceedings of PERMIS, Gaithersburg, MD (p. 4-11).

Rajivan, P., Janssen, M. A., & Cooke, N. J. (2013, September). Agent-Based Model of a Cyber Security Defense Analyst Team. In Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society, 30 September - 4 October (pp. 314-318). San Diego, CA, HFES.

Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. Cambridge, MA: MIT Press.

Tyworth, M., Giacobe, N., Mancuso, V., McNeese, M., & Hall, D (2013). A Human-In-The-Loop Approach to Understanding Situation Awareness in Cyber Defense Analysis. EAI Endorsed Transactions on Security and Safety, 13(2), 1-10.