SAND2018-8543C

Human Factors in Security

Ann E. Speed*, Bryana L. Woo†, Camron G. Kouhestani ‡, Jaclynn J. Stubbs§, and Gabriel C. Birch¶
Sandia National Laboratories
1515 Eubank SE, Albuquerque, NM, USA
*aespeed@sandia.gov †blbaird@sandia.gov ‡cgkouhe@sandia.gov §jstubbs@sandia.gov ¶gcbirch@sandia.gov

***Abstract*—Physical security systems (PSS) and humans are inescapably tied in the current physical security paradigm. Yet, physical security system evaluations often end at the console that displays information to the human. That is, these evaluations do not account for human-in-the-loop factors that can greatly impact performance of the security system, even though methods for doing so are well-established. This paper highlights two examples of methods for evaluating the human component of the current physical security system. One of these methods is qualitative, focusing on the information the human needs to adequately monitor alarms on a physical site. The other of these methods objectively measures the impact of false alarm rates on threat detection. These types of human-centric evaluations are often treated as unnecessary or not cost effective under the belief that human cognition is straightforward and errors can be either trained away or mitigated with technology. These assumptions are not always correct, are often surprising, and can often only be identified with objective assessments of human-system performance. Thus, taking the time to perform human element evaluations can identify unintuitive human-system weaknesses and can provide significant cost savings in the form of mitigating vulnerabilities and reducing costly system patches or retrofits to correct an issue after the system has been deployed.**

***Index Terms*—physical security system, alarm station operator, human factors**

I. Introduction

Current physical security systems (PSSs) function under a human-in-the-loop paradigm. PSS architecture follows the idea of automating all aspects of system detection, then using human analysts for assessment and response. The PSS has sensors that perform automatic intruder detection, a network that automatically communicates the sensor output to the operator station, and an operator console that displays the sensor alarm information and associated camera views to allow the operator to make the final decision regarding what actions to take. This current PSS paradigm can be improved, especially with respect to the human-system interaction component. Determining what elements involving human interactions within the PSS can be improved is often challenging, as this require information regarding the baseline performance of a system that involves a human component. To this end, we consider two methods for determining baseline activities and performance of human-system interactions with the PSS; a qualitative method that outlines operator tasks, and a quantitative test that enables characterizing PSS operator performance as a function of nuisance/false alarms.

II. Operator Situation Awareness

SA is the human operators ability to accurately understand current states, and adequately assess near-future states of a system to take appropriate action [1]. Specifically, Endsley (1995) has identified four levels of situation awareness an operator can have [2]:

1) No SA
2) SA level 1: basic perception of the environment (e.g., noticing alarms)
3) SA level 2: developing a big-picture understanding of the current state of the facility
4) SA level 3: the ability to anticipate likely events in the near future

Accuracy at the higher levels is partially dependent on accuracy at the lower levels of SA, however, many human-in-the loop systems function in a way that limits the human ability to get beyond Level 1 SA because of large numbers of individual sensors that, in aggregate, produce large numbers of nuisance and/or false alarms. Our overarching hypothesis is that if we can design an interface and associated algorithms with the end-user goals in mind, we can meet requirements for acknowledgement and assessment of every alarm while at the same time reducing bias due to the lopsided NAR/FAR to real alarm rates (called the prevalence effect [3]–[5]), and enabling the PSS operator to make accurate, rapid decisions in the face of an actual attack without imposing additional cognitive load during these high-stress times. However, adding algorithmic elements or chaining user interfaces requires understanding the major tasks which PSS operators are engaged in with respect to the current system, as well as characterization of the baseline performance of PSS operators.

As an example, in one current project, the analysis of the operators role within the PSS began with a goal-directed task analysis (GDTA) [1]. This type of analysis focuses on first identifying the goals the operator must achieve in performing the job, the decisions necessary for achieving those goals, and the information the operator needs to make those decisions effectively. The task analyst performing the GDTA first reviews relevant documents regarding the job, which may include training documents and SOPs. After reviewing relevant documents, the task analyst interviews several individuals who have acted in the operator role, and may interview other subject matter experts in the domain. Using these interviews, a systematic description of how the operator makes goal-directed decisions was created, and information needed to

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

make those decisions was identified. The resulting GDTA enables the identification of the source of prior decision errors, and provides guidance for mitigating those errors in future system instantiations. An example of a GDTA performed for a PSS is shown in Figure 1.

Several types of errors are critical to understand in physical security contexts, and can be understood in terms of different levels of situation awareness. Errors at SA level 1 may include a lack of awareness of an alarm or group of alarms. Errors at SA level 2 can include a misunderstanding of the information being provided at the operator console, leading to a mode error, or misunderstanding the state of a system. An example SA 2 error in physical security systems may be when an operator interprets all alarms to be nuisance/false alarms rather than actual intrusions. Errors at SA level 3 can include misjudging the rate at which an attack is taking place, and, as a result, dispatching response forces to the incorrect location. Causes for SA errors are numerous and range from critical information being hidden from the operator to too much irrelevant information being presented to the operator, causing cognitive overload. Understanding the critical information needs under different circumstances, and designing the operator interface, and underlying algorithms, to support those information needs, can reduce errors at all levels of situation awareness.

In terms of the current work, we have developed a detailed goal and decision hierarchy, have identified the information needed to support those decisions and goals, and have begun identifying the information that operators feel is most critical to supporting SA at all levels. As displayed in Figure 1, this effort resulted in a list of 10 sub-goals subsumed by one overall goal: Act as eyes, ears, triage, and communications center for facility and responding forces. Underneath these 10 sub-goals were 12 critical decisions PSS operators need to make. Associated with those 12 critical decisions were 63 separate pieces of information our SMEs indicated they needed in order to make these 12 critical decisions.

An interesting outcome of this analysis is that one of the core decisions the PSS operator must make, whether an alarm is nuisance/false or real, is a very complex decision and one that must be made quite rapidly. In the unlikely event that it is an example of a real alarm indicative of an unauthorized intruder, response forces must be deployed appropriately and with the highest level of SA possible. In the event the alarm indicates a malicious intruder, intent on doing harm, this decision can be even more difficult as such an intruder may try to spoof the system, or occlude their activities from the PSS operator. Identifying attempts to occlude activity are a challenge for any human and are beyond the reach of current algorithmic solutions because of the functionally infinite ways an adversary might attempt to breach a physical facility

Furthermore, mapping these results to the principles **Endsley & Jones** identify for enhancing situation awareness has led to some interesting insights. First, in this domain, designing an interface to support PSS operator situation awareness is not as simple as identifying individual pieces of information to display on an interface. For example, one piece of information both experts indicated to be particularly important was the exact location of an alarm (e.g., which building and where on the building). Another critical piece of information they indicated they needed to know was what type of sensor was alarming (e.g., microwave or pan-tilt-zoon camera). What they are actually wanting to know, in terms of situation awareness is: Does the spatiotemporal pattern of current alarms define a pathway from the perimeter to the critical asset we are protecting? If we were to simply display location information on the interface along with sensor type information, the display could rapidly turn into an interface full of pinpoints of light. However, if we incorporate temporal information about when how long ago individual alarms occurred, in addition to information about where they occurred, we can provide the PSS operator with a better understanding of whether a rash of alarms is due to weather (wind, rain, which should display as a random spatiotemporal pattern), due to an animal that has crossed into the facility (which should display as a path, but one that meanders or is unlikely to head in a direct way toward a critical asset), or due to an adversary with a specific intent.

Another interesting insight happened when we watched the developer of the current interface interact with that interface to execute current standard operating procedure (SOP) while attempting to respond to multiple, rapidly occurring alarms. Specifically, the SOP dictated that the PSS operator had to acknowledge every alarm as it came through. The interface was set up to allow the PSS operator to indicate the cause of the alarm and other important information by navigating a series of pull-down menus and check boxes. When multiple alarms occurred in rapid succession, this intensive method for classifying every alarm before moving onto the next rapidly overwhelmed the developer. The lesson here was that there need to be some way to enable the PSS operator to follow the SOP without compromising his or her ability to make the key decision about whether or not the facility was under attack. Being buried in the many menus was preventing just that had we not paid attention to the requirements of the SOP in addition to the critical decisions the PSS operator was being asked to make, this interface weakness would have been overlooked.

Qualitative studies and observations like those done for the current project clearly yield information that no empirical study would nothing substitutes for asking the task performers, in a structured manner, how they do their jobs, what information they need, and what their goals are. This kind of assessment necessarily includes other qualitative information such as SOPs. However, this is only the first step in the process of designing interfaces and underlying algorithms to support an operators rapid hypothesis testing and Level 2 / Level 3 SA. Once initial strategies are developed for enhancing SA, the only way to determine whether any of the strategies actually impacts performance is to develop methods to objectively measure performance.

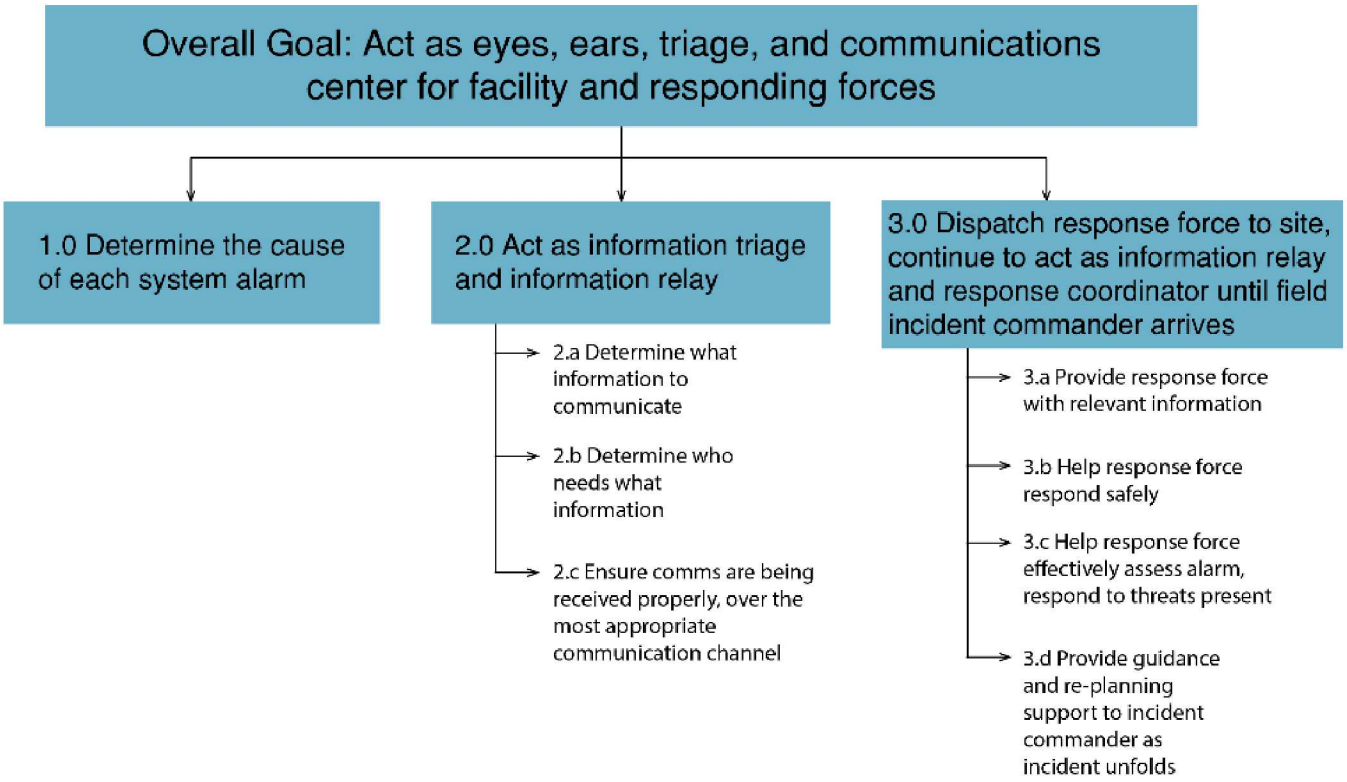

Fig. 1. Example top-level goals and decisions, derived from a GDTA for a physical security operator. This qualitative analysis enables identification of decision error sources.

III. Objectively Measuring Performance

Once a detailed understanding of the goals, decisions, and information needs of the job is developed, methods for determining how existing and new interfaces support operator performance can be developed. In terms of the current work, this will take the form of experiments performed using an approach we call near operational environments. This means that we will attempt to replicate the conditions under which PSS operators make time-critical decisions. Because our specific application concerns the PSS console operator, we anticipate being able to feed sensor data to the PSS console interface in a way that allows us to control for several variables. This allows us to know ground truth, and enables collection of data such as decision time, operator interactions with the software interface (e.g., which information is accessed and in what order), as well as overall decision accuracy. Post-event information can also be collected to enable us to assess the operators understanding of the scenario and to assess where additional information was needed. All these data, taken together, will enable us to determine the relative impact of the new interface compared to the original interface in terms of SA support at all levels, the amount of cognitive load imposed on the operator, and will enable us to identify additional gaps in the information provided to the operator.

One example of an empirical assessment is a current experiment we are performing to determine the effects of different false alarm rates relative to the frequency of targets something that impacts levels 2 and 3 SA. One perspective on the false alarm problem is that operators need to assess every single alarm, both to provide culpability and to keep humans alert, especially if actual targets are infrequent. However, experimental evidence suggests that tasks with low target prevalence in the face of high false alarms (e.g., 5% of alarms are actual targets), leads to an operator tendency to automatically label alarms as non-threats, thus causing them to miss targets more often than they should. Called the prevalence effect, this result has been observed in multiple settings, both using generic tasks and using domain-specific tasks (Wolfe, et al., used an X-Ray screening task similar to that found at the airport passenger checkpoint).

The study we are currently conducting attempts to characterize the performance curve of target prevalence, manipulating the ratio of targets to non- target trials. Specifically, subjects participate in one of four target prevalence conditions: 50/50 in which half of the trials contain targets, 1/10 in which 1 out of ever 10 trials contains a target (on average), 1/100 in which one out of every 100 trials contains a target (on average) and 1/1000 in which only the final trial of the experiment contains a target. In each of the conditions, there are a total of 1000 trials each subject sees, and in each condition the last trial is a target. The rationale for this manipulation is to see how target frequency impacts the ability for a subject to detect a target after a relatively long period of time (each subject participates for about 2.5 hours with only two short

Fig. 2. Probability of detection as a function of target prevalence condition. The 1/1000 condition was not included on this graph because only one subject out of 13 detected the lone target in that condition. There are 14 subjects in each condition except for the 1/1000 condition, which has 13 subjects. Error bars represent the standard error of the mean.

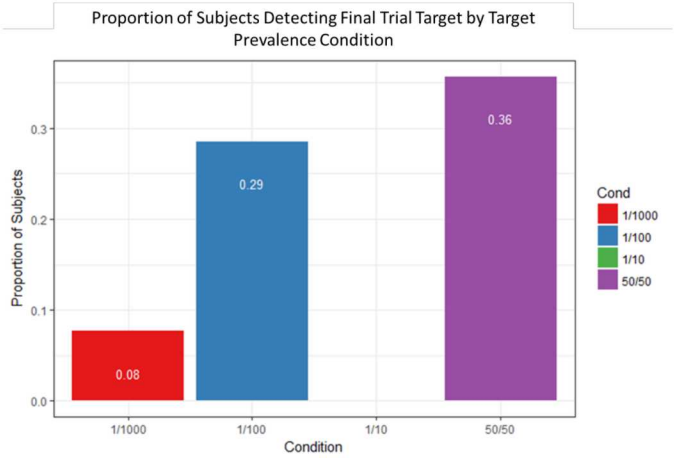

Fig. 3. Proportion of subjects who detected the T target in the final (i.e., 1000th) trial as a function of target prevalence condition. There are 14 subjects in each condition except for the 1/1000 condition, which has 13 subjects.

breaks). The critical condition, with 1 target out of 1000 trials, is intended to replicate the conditions experienced by a PSS operator who deals with numerous false or nuisance alarms before ever seeing a real alarm. We used a traditional visual search task, in which subjects (current student interns) search for perfect T's amidst offset T's and L's. Preliminary results are shown in Figure 2.

Each condition currently has 14 subjects (except for the 1/1000 condition which has 13) although we are attempting to acquire data from at least 25 subjects for each condition. It is notable that of the 13 subjects in the 1/1000 condition, only one subject found the target T in the final condition (see Figure 3 for the proportions of subjects in each condition who detected the final target). While this particular experiment doesnt fall into the category of near operational paradigms, it provides rationale for us to conduct a similar study using more realistic (and more costly) stimuli with PSS operators as subjects. It also illustrates the benefit of performing a controlled study based on this work and prior literature [3]–[5]. The implication is that arguing that PSS operators should be required to clear every single alarm may actually be harmful in terms of target detection and overall adjudication accuracy as the nuisance/false alarm rate increases relative to the number of real alarms.

Each condition currently has 14 subjects, although we are attempting to acquire data from at least 25 subjects for each condition. It is notable that of the 14 subjects in the 1/1000 condition, only one subject found the target T in the final condition (see Figure 3 for the proportions of subjects in each condition who detected the final target). While this particular experiment doesnt fall into the category of near operational paradigms, it provides rationale for us to conduct a similar study using more realistic (and more costly) stimuli with PSS operators (rather than student interns) as subjects. It also illustrates the benefit of performing a controlled study based on this work and prior literature [3]–[5], it is clear that arguing that PSS operators should be required to clear every single alarm is increasingly harmful in terms of target detection and overall adjudication accuracy as the nuisance/false alarm rate increases relative to the number of real alarms.

This study was a simple proof-of-concept designed to justify the cost and effort needed to conduct a more realistic study a near-operational study. Such an experiment would use people with PSS alarm center experience and would likely use a simulation environment to generate alarm scenarios. Specifically, we would manipulate the features of the user interface based on the GDTA we conducted. We could have several conditions, including a baseline of the interface as it currently stands. Experimental conditions would include different methods for displaying the information needs identified during the GDTA. We would also manipulate the false alarm to target ratio to determine whether the patterns we see in data using novices and a domain-independent task carry over to experts functioning in their domain. Results from such a study can have significant impact on the design of analytic algorithms and interface design, along with impact on standard operating procedures for PSS operators in the future.

IV. Conclusion

The results of well-designed human-in-the-loop evaluations, especially if they include both qualitative and quantitative components, provide insight into why humans interact with systems sub-optimally. Often, these human weaknesses are not obvious. We all have insights into what helps or hurts our decision-making, but those insights are not always correct. Objective methods for assessing information needs and threat detection often highlight surprising effects that can have significant implications for the overall security posture of a PSS.

Acknowledgment

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineer-

ing Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525.

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] M. R. Endsley and D. G. Jones, *Designing for situation awareness: An approach to user-centered design (2nd ed.)*. CRC press, 2012.

[2] M. R. Endsley, "Measurement of situation awareness in dynamic systems," *Human factors*, vol. 37, no. 1, pp. 65–84, 1995.

[3] A. T. Biggs, S. H. Adamo, and S. R. Mitroff, "Rare, but obviously there: Effects of target frequency and salience on visual search accuracy," *Acta psychologica*, vol. 152, pp. 158–165, 2014.

[4] S. R. Mitroff and A. T. Biggs, "The ultra-rare-item effect: Visual search for exceedingly rare items is highly susceptible to error," *Psychological Science*, vol. 25, no. 1, pp. 284–289, 2014.

[5] J. M. Wolfe, T. S. Horowitz, M. J. Van Wert, N. M. Kenner, S. S. Place, and N. Kibbi, "Low target prevalence is a stubborn source of errors in visual search tasks." *Journal of Experimental Psychology: General*, vol. 136, no. 4, p. 623, 2007.