Exploring Human-Technology Interaction in Layered Security Military Applications

Amanda Wachtel(✉), Matthew Hoffman, Craig Lawton, Ann Speed, John Gauthier, and Robert Kittinger

Sandia National Laboratories, Albuquerque, USA
{awachte, mjhoffm, crlawto, aespeed, jhgauth, rskitti}@sandia.gov

Abstract. System-of-systems modeling has traditionally focused on physical systems rather than humans, but recent events have proved the necessity of considering the human in the loop. As technology becomes more complex and layered security continues to increase in importance, capturing humans and their interactions with technologies within the system-of-systems will be increasingly necessary. After an extensive job-task analysis, a novel type of system-of-systems simulation model has been created to capture the human-technology interactions on an extra-small forward operating base to better understand performance, key security drivers, and the robustness of the base. In addition to the model, an innovative framework for using detection theory to calculate d' for individual elements of the layered security system, and for the entire security system as a whole, is under development.

Keywords: System-of-systems · Layered security · Human-technology interaction · Human performance factors · Modeling and simulation · Detection theory

1 Introduction

Historically, systems of systems (SoS) modeling efforts have focused on depicting physical systems and the connections (whether physical connections or communications related) between them. Man-made systems are modeled in detail, while humans mainly play a supportive role when their inclusion is absolutely necessary. For example, the human may be included in the model to provide maintenance to a system. Yet the maintenance time is based on a specified distribution or even a static number, giving little consideration to variation in individuals or circumstances. Due to this idyllic treatment of humans in SoS modeling, many failures seen in real life are missed in the modeling realm.

The danger of failing to include the human in the loop becomes apparent when looking at historical cases. Take for example the security breach at NNSA's Y-12 National Security Complex, which arose in part from distractions, improper technology use, false alarm fatigue, and poor assumptions [1]. Unknown outsiders with unclear intentions can also pose a threat as seen at Patrol Base Bushmaster when an outsider charged the base with a vehicle outfitted with a 2,000 lb vehicle-borne improvised

F.F.-H. Nah and C.-H. Tan (Eds.): HCIBGO 2016, Part II, LNCS 9752, pp. 241–250, 2016.
DOI: 10.1007/978-3-319-39399-5_23

explosive device (VBIED) that killed and injured many of the soldiers [2]. Other examples of human failures in security systems abound. In fact, Jarret Lafleur et al. performed an analysis of 23 heists and found that "A common thread of all defeat methods is that they attack segments of the security system in which humans are in the loop…" [3]. Whether looking at the recent failure of the layered security system at Y-12, successful heists, or the devastating bomb attack on Patrol Base Bushmaster, the conclusion is that the opportunity for a major breakdown in SoS safety or security is due in large part to the human in the loop.

2 Motivation

Since humans account for the majority of uncertainty in an SoS and in layered security, the role of human-technology interaction must be understood. As technology becomes increasingly sophisticated, uncertainties around human performance will grow dramatically. In addition, human inefficiency can lead to high costs, increased logistics, and increased vulnerability. We are focusing our modeling efforts on a first attempt at understanding the effects of human-technology interactions on SoS level performance. From there, this understanding will be incorporated into SoS engineering processes using empirical, data-driven methods. Only when the impact of humans and their interactions with the physical systems present in an SoS framework is fully accounted for will we be able to predict, assess, and improve performance and human efficiency in SoS models and be able to accurately evaluate the effects of potential organizational, doctrinal, or system changes.

3 Use Case

Our first step in beginning to capture human-technology interaction in an SoS was to define a representative use case to focus on. For the initial modeling effort, the use case focused on a patrol base (now classified as an extra-small forward operating base) in the Middle East and the threat posed by vehicle-borne improvised explosive devices (VBIEDs) as motivated by the attack on Patrol Base Bushmaster. The research team is comprised of individuals with expertise in the fields of SoS modeling and industrial-organizational psychology. To develop a conceptual model, all human entities, tasks, and the corresponding systems most relevant to the use case were identified. From there work was done to qualitatively characterize interdependencies between both systems and also humans and systems.

It was also key to identify variables relevant to the performance of the tasks. In simulated deployment environments such as the Navy SEAL's Hell Week, factors such as fatigue and environmental, physical, and psychological stressors were shown to degrade human performance to a greater degree than that caused by intoxication, sedatives, or hypoglycemia [4]. Variables were included in the SoS model to capture factors that impact the ability of human-technology couplings to perform tasks. Such variables include, but are not limited to, the number of tasks a human is required to perform, fatigue, stress, distractions, and environmental conditions.

4 Data Collection

To validate the SoS modeling approach we are engaging with subject matter experts (SMEs) who have experience in layered security on military bases. Focusing on key aspects of an extra-small forward operating base (FOB) such as vehicle checkpoints and the tactical operations center (TOC), a detailed task analysis was conducted. The SMEs filled in details such as the duration of each task, how many humans are needed for each task, what types of technology are used for the task, and task interdependencies. They have also given critical input as to which human and external variables may affect specific instances of human-technology interaction.

4.1 Job-Task Analysis

To begin to analyze the tasks humans perform on FOBs, duties were broken down into guard duties and duties performed within the Tactical Operations Center (TOC) on the base. The guard duties were further broken down to include tasks performed by guards outside of base (i.e. at a checkpoint), at the base gate, and inside the base near critical infrastructure. To model these tasks in an SoS model, tasks are considered to be associated with a location on, or outside of, the FOB. The external guards perform their tasks at a roadblock, the gate guards at the base gate, the internal guards near a target of value, and the surveillance, command, and control tasks are performed inside the TOC. Examples of tasks that would occur at the base gate are checking the driver's identification and checking the vehicle for contraband items and VBIEDs. These examples will be revisited below to help explain the more complex metrics included in the job-task analysis. Each location's tasks were vetted with the SMEs.

Each task listed includes multiple dimensions of data used in the model. Basic metrics include the frequency of the task, how long the task takes to complete, the number of people required to complete the task, and any technologies/equipment used to carry out the task. The nature of the task, whether physical, mental, or both, is also included to begin to capture the human element. Many variables will ultimately be included in the model that act upon the human's ability to successfully complete the task. These include (among many others) factors such as heat, fatigue, time of day, physical injury, dehydration, length of shift, and hunger. Some factors have a greater impact to human performance when the task is more physical in nature, such as not being able to use a complicated, hand-held user interface in extreme cold, while other factors have a greater impact on primarily mental tasks, such as not being able to accurately identify an ID as fake when working long shifts for months on end with little sleep. Which factors impact which tasks was determined primarily through SME input.

In addition to factors that impact human performance, SMEs also helped determine factors that would impact technology functionality. Depending on the technology or equipment in question, these factors could include heat, humidity, rain, lightning, cold, and high wind/sandstorms. It is conceivable that these factors could at times cause the technology to malfunction or become unusable to the humans relying on that technology to complete a task. For example, a sandstorm could prevent a guard at a roadblock from using binoculars to see a vehicle approaching at a distance, while

lightning could temporarily disable the communications network and prevent a guard from relaying a threat to the TOC.

While some conditions affect only the human or only the technology, it is clear that many important effects are interactive; i.e., they only occur when a person is using a particular kind of equipment under certain physical and mental conditions. These can range from the cognitively and technologically banal (such as gloves reducing dexterity in cold weather) to the complex (such as a soldier having difficulty using a weapon technology's complex interface during a cognitively demanding combat situation). When including humans in the SoS model, it was of primary importance to capture every time a human is interacting with technology and every way that interaction can be changed by affecting the human, the technology, or both.

4.2 Business Rule Data Elicitation

In addition to the data-driven approach to the job-task analysis outlined above, work is being done to capture the business rules needed for the model by interviewing SMEs about life on base. The data elicitation helps clarify the chain of events from the time a threat is detected to the time when a response team is sent. It also allows the team to understand which tasks and/or technologies pose the greatest difficulty to personnel and how the chain-of-command plays out both daily and in heightened-security situations. A few examples of questions used for business rule elicitation include:

- What are the hardest things for new people to come up to speed with? What are the most common mistakes?
- What equipment do you find most frustrating to use?
- Under what situations do you feel particularly stressed or confused? Bored?
- How often and in what environments do you train? How relevant do you feel the training is to the actual job?
- How many people are on a shift? Are teams usually the same people?
- How is information passed during changes in personnel?
- How much sleep do you get each night? How often do you go between sleep? How often is your sleep disrupted?
- How long does it take for a response team to get there after being dispatched?

Note that the same question may be asked multiple times in various ways to try to be able to work around topics that may elicit strong reactions when worded a certain way. For example, someone who has been on a FOB may not want to admit to having had difficulty using a certain technology, but when asked which technologies new recruits have the most problems with, they are free to respond without commenting on their own personal capabilities.

5 System-of-Systems Model

An SoS model to capture these human-technology interactions is now in the intermediate stages of development. The model is built in FlexSim [5], an off-the-shelf discrete event simulation software traditionally used to build manufacturing models, but which offers flexibility through custom scripting and can thus handle a wide array of models and logic. The simulation is portioned off into a series of tasks, each requiring a specified number of people using given technologies to complete the tasks as discussed in the previous section. Various threats such as malicious outsiders (or insiders), contraband objects, suspicious activity outside the fence, and VBIEDs are randomly generated as the simulation progresses. The model tracks how many times the humans and layered security system are able to correctly detect, assess, and respond to these threats (Fig. 1).

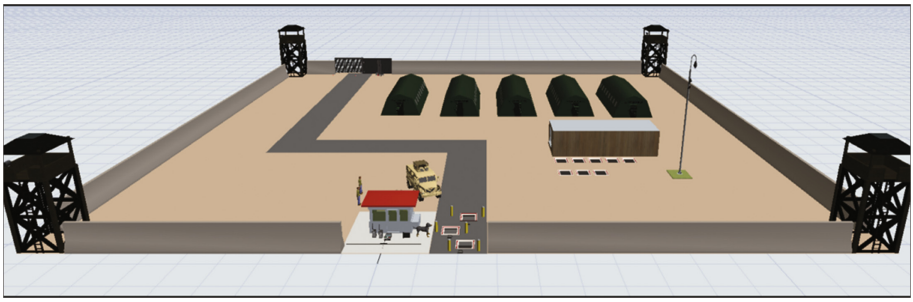

Fig. 1. FlexSim model of extra-small forward operating base

5.1 Human-Technology Performance

Each soldier has a probability distribution associated with their ability to successfully complete the task(s) to which they are assigned. These performance factors are largely based on existing studies and SME input. The factors not only capture a soldier performing a task, they capture them performing the task *using a certain technology*. This means that the likelihood of completing a task can be impacted by changes to the soldier, changes to the technology, or both.

The factors impacting performance may be external, such as a dust storm, or inherent to the system, such as a soldier suffering from extreme fatigue. In either case, the human may or may not have control over the changing situation, yet their performance may be impacted. Factors such as fatigue are obvious culprits but there are seemingly innocuous factors that have an equally large influence. Take for example a soldier using a mirror on a stick and a flashlight to search underneath a vehicle for IEDs. Once the sun sets, the soldier is actually able to better use the technology since the light from the flashlight is now concentrated on the mirror and bounces up onto the undercarriage of the vehicle. Whereas daylight created additional shadows and glare, night has created ideal conditions for this specific task.

To capture the impacts of external and inherent factors on performance, the baseline performance metrics are scaled up or down based on whether the factor is expected to degrade or improve performance. Using Wincek and Haight's human error rate formulas [6] as a mathematical basis for our framework, we have modified the human error rate equation to create the following calculation for human performance.

$$AHP = BHP * HPM_1 * HPM_2 * \cdots * HPM_n \quad (1)$$

where AHP is the adjusted human performance, BHP is the baseline human performance, and HPM is the human performance modifier. HPM is greater than one when the factor increases human performance and less than one when the factor decreases human performance. Each factor that impacts performance on the task in question would be represented by a unique HPM factor. An HPM factor may capture an effect to the human performing the task, to their technology, or to the combination.

To illustrate this method, this equation can be applied to the vehicle search at night example. (Note that these values are notional and used only for illustrative purposes.) The baseline human performance value for finding a contraband or explosive item on the vehicle during the day is assumed to be 0.8 when the vehicle is driven by an outsider. One human performance modifier is applied to account for the time of day, giving the equation

$$AHP_{VehicleSearch} = 0.8 * HPM_{TimeofDay} \quad (2)$$

where $AHP_{VehicleSearch}$ is the adjusted human performance for the vehicle search task, 0.8 is the baseline human performance, and $HPM_{TimeofDay}$ is the human performance modifier based on time of day. If the search takes place during daylight hours, the human performance modifier is 1 to maintain the baseline human performance. However if the search occurs at night, a scale factor of 1.1 is used to increase the performance on the task by ten percent. This methodology is applied to all human-technology performance factors for the external and inherent impacts of interest. Each technology also has inherent reliability data such as how many hours it can be used before needing new batteries, and how long its average lifespan is, resulting in multiple failure modes for each technology in the model. More sophisticated performance adjustment methods could be considered and used within the framework if warranted (and supported by data); however, given the data available, using the product of HPM factors is sufficient to allow exploration of interesting conditional and interactive model behaviors.

5.2 Communications

Additionally, each task is linked into the communications network which can also experience various types of failures. The communications network is primarily used to relay detected threats, assess the situation, and take action to nullify the threat. Each guard task defined in the job-task analysis is capable of communicating with the TOC to report a threat. Communications in the model are handled according to the following communications hierarchy (Fig. 2).

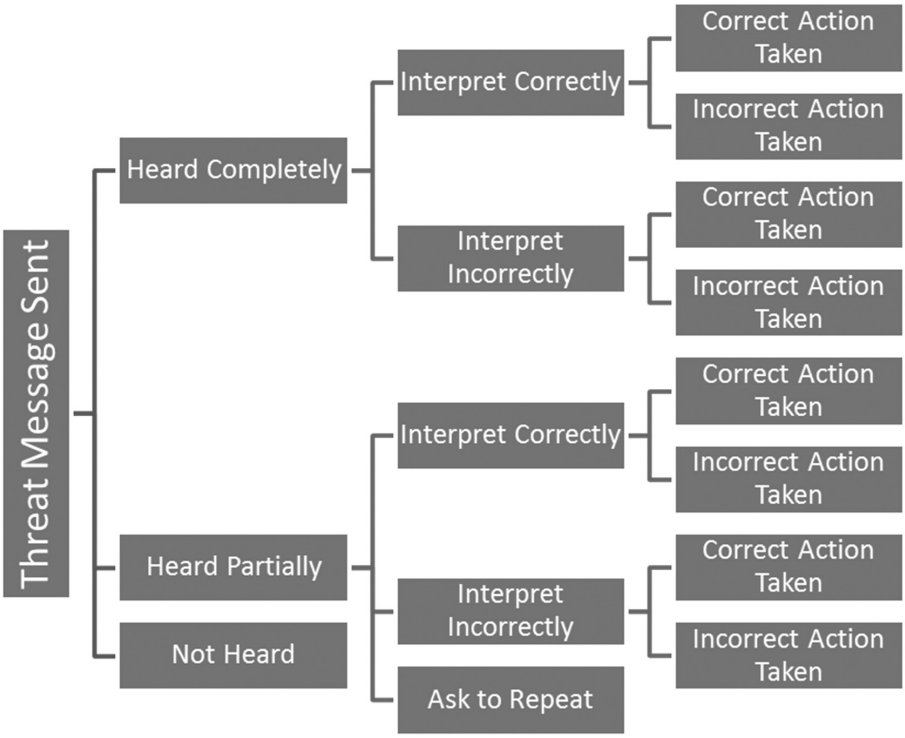

Fig. 2. Communications logic flow

For every threat message that is radioed in to the TOC, the message can be heard completely, heard partially, or not heard at all. Messages that are heard can either be interpreted correctly or incorrectly with the additional option of asking the sender to repeat the message if only part of the message is heard. If the message was never received by the TOC, the sender tries again after a defined time interval. Whether the message is interpreted correctly or incorrectly, the TOC personnel can still take either the correct or the incorrect action. Imagine the scenario where the soldier in the TOC has been trained for a different task or has simply received inadequate training for the TOC. A threat is radioed in and they understand the message, but due to lack of training they inadvertently activate the wrong response. Similarly, a message could be heard incorrectly and the correct action could still be taken by dumb luck. Of course these scenarios are less likely to occur, which are reflected in their probabilities relative to the other scenarios. Each item in the communications logic flow is assigned a probability in the model and the model automatically chooses the communication path when a threat message is sent based on those allocations.

5.3 Model Outputs

The intention is to use the completed extra-small FOB model to understand the range of SoS performance, key drivers in security failures and successes, the base's robustness to unusual combinations of conditions, and conditions that enable particularly low success rates. Through running different scenarios we hope to identify whether there are recurring weaknesses in the layered security system or if the weaknesses are situation-dependent. Moreover, the model provides a framework to assess SoS-wide impacts of improving human-technology interaction through implementation of more intuitive technology interfaces and operations, or through an increased emphasis on training.

6 Detection Theory for Layered Security

Alongside development of the SoS model, a mathematical framework for calculating the "goodness" of a layered security system using detection theory is also being developed. Output metrics from the simulation will feed the probability equations in the framework. Building upon the classic definition of d' in Macmillan and Creelman's *Detection Theory* [7]

$$d' = z(\text{hit probability}) - z(\text{false alarm probability}), \tag{3}$$

where z is the number of standard deviations from the mean, a methodology was created for calculating d' for the four main components of layered security: detection, delay, communications, and response. Each of the four security-area-specific d' values will then be rolled-up into a d' representative of the layered security system as a whole.

The area-specific d's vary based on the domain-specific terms that define what constitutes the hit probability and the false alarm probability. Tables 1, 2, 3 and 4 are given below to help the reader understand what is meant by a hit and a false alarm in each instance.

Table 1. Terminology for detection d'

	Response		
Attack	Detected	Undetected	# Trials
Yes	Hit	Miss (false neg)	n
No	False alarm (false pos)	Correct rejection	m

Table 2. Terminology for delay d'

	Response		
Incident	Delayed	Insufficient delay	# Trials
Adversary	Hit	Miss (false neg)	n
Friendly	False alarm (false pos)	Correct rejection	m

Table 3. Terminology for communications d‘

	Response		
Attack	Communicated	Not communicated	# Trials
Yes	Hit	Miss (false neg)	n
No	False alarm (false pos)	Correct rejection	m

Table 4. Terminology for response d‘

	Response		
Attack	Adequate	Insufficient	# Trials
Yes	Hit	Miss (false neg)	n
No	False alarm (false pos)	Correct rejection	m

Once the terminology for hits and false alarms has been specified for each area, probabilities for events happening in serial and parallel can also be defined. For detection, delay, communications, and response the events are considered to be sensors, barriers, messages, and responses, respectively. For example the probability of a hit for sensors in parallel can be defined as

$$P_{hit} = \min\left(P_{hit,s1}, P_{hit,s2}\right) \text{OR} \frac{P_{hit,s1} + P_{hit,s2}}{2}, \tag{4}$$

where $P_{hit,s1}$ is the hit probability of the first sensor and $P_{hit,s2}$ is the hit probability of the second sensor. The minimum would be used to calculate the “worst case” hit probability while the average would be used to calculate the “average case” hit probability. The probability of a false alarm for sensors in parallel can be defined as

$$P_{FA} = 1 - (1 - P_{FA,S1})(1 - P_{FA,S2}), \tag{5}$$

where $P_{FA,S1}$ is the probability of a false alarm for sensor one and $P_{FA,S2}$ is the probability of a false alarm for sensor two.

Similar equations have been derived for events in series for multiple sensors, and in series and parallel for the other three areas. As mentioned, the d’ values for the individual areas are then combined into a base level d’ to measure the effectiveness of the layered security system. The individual d’ are aggregated statistically according to the event path through the security system. The result is a single measure of the sensitivity—the ability to respond appropriately to a stimulus—of the physical security system.

7 Future Work

While the methodology employed to capture humans within a systems-of-systems model has shown great potential, many data gaps still exist that must be addressed. The human performance factors used in the model have been obtained from studies with

situations as close as possible to those of the tasks being performed in the base camp, but the existing studies may or may not accurately capture the intricacies of using the specific military technologies in question. Ideally, more studies involving military personnel in deployment scenarios would be needed to further refine and validate the model.

Acknowledgements. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000 (SAND2016-1471 C).

References

1. U.S. Department of Energy Office of Inspector General: Inquiry into the Security Breach at the National Nuclear Security Administration's Y-12 National Security Complex. Special report DOE/IG-0868, DOE (2012)
2. Stars and Stripes. http://www.stripes.com/news/a-moment-that-changed-everything-1.88432
3. Lafleur, J.M., Purvis, L.K., Roesler, A.W.: The Perfect Heist: Recipes from Around the World. SAND2014-1790, Sandia National Laboratories (2014)
4. Lieberman, H.R., Bathalon, G.P., Falco, C.M., Morgan III, C.A., Niro, P.J., Tharion, W.J.: The fog of war: decrements in cognitive performance and mood associated with combat-like stress. Aviat. Space Environ. Med. **76**(7), Section II (2005)
5. FlexSim. https://www.flexsim.com
6. Wincek, J.C., Haight, J.M.: Realistic human error rates for process hazard analyses. Process Saf. Prog. **26**(2), 95–100 (2007). American Institute of Chemical Engineers
7. Macmillan, N.A., Creelman, C.D.: Detection Theory: A User's Guide. Psychology Press, Abingdon (2004)