Cyber Situational Awareness Testing

Joel Brynielsson[1,2(✉)], Ulrik Franke[3], and Stefan Varga[2,4]

[1] FOI Swedish Defence Research Agency, 164 90 Stockholm, Sweden
joel.brynielsson@foi.se
[2] KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
[3] SICS Swedish Institute of Computer Science, Box 1263, 164 29 Kista, Sweden
ulrik.franke@sics.se
[4] Swedish Armed Forces Headquarters, 107 85 Stockholm, Sweden
stefan.varga@mil.se

Abstract. In the cyber security landscape, the human ability to comprehend and adapt to existing and emerging threats is crucial. Not only technical solutions, but also the operator's ability to grasp the complexities of the threats affect the level of success or failure that is achieved in cyber defence. In this paper we discuss the general concept of situation awareness and associated measurement techniques. Further, we describe the cyber domain and how it differs from other domains, and show how predictive knowledge can help improve cyber defence. We discuss how selected existing models and measurement techniques for situation awareness can be adapted and applied in the cyber domain to measure actual levels of cyber situation awareness. We identify generic relevant criteria and other factors to consider, and propose a methodology to set up cyber situation awareness measurement experiments within the context of simulated cyber defence exercises. Such experiments can be used to test the viability of different cyber solutions. A number of concrete possible experiments are also suggested.

Keywords: Situational awareness · Measurement technique · Experimental design · Cyber defence exercise

1 Introduction

In cyber security it is seldom straightforward to get a sense of the threat landscape as a whole in order to really know "what is going on"[1]. Still, to understand an immediate threat or a detected attack not only in itself but also in terms of the surrounding threats and its strategic implications will most likely be the key to effectively be able to deal with more elaborate forms of cyber threats. To understand the roots and causes underlying a threat and to be able to put this

[1] To know "what is going on" is a phrase used by Endsley [12] in order to provide an informal and intuitive definition of the situational awareness concept.

B. Akhgar and B. Brewster (eds.), *Combatting Cybercrime and Cyberterrorism*,
Advanced Sciences and Technologies for Security Applications,
DOI 10.1007/978-3-319-38930-1_12

information in an overall cyber arena context, is what cyber situational awareness[2] (CSA) is about. Such CSA will help the decision-maker/analyst to better understand the organisational implications, and how to assess and act given that a threat or an attack has been detected. As identified in previous work [20], CSA is considered to be the part of situational awareness which concerns the "cyber" environment, whilst at the same time acknowledging that acquiring and upholding CSA requires that external factors concerning, e.g., the physical environment, the political dimension, etc., need to be taken into account.

The cyber threat is omnipresent in today's connected world, and the necessity to uphold a high level of CSA naturally follows in many operational applications. Examples include the importance for IT departments to be able to distinguish between "background noise," e.g., attack attempts with slim chances of success, and more advanced attempts with potentially severe effects, and for intelligence personnel to understanding a cyber attack strategically in terms of its political implications. Related to the sought for operational CSA capacity, it follows that the ability to acquire and maintain a high level of CSA is also something that ought to govern educational endeavours. Moreover, the usefulness of solutions for tackling the cyber threat—be it technology, processes, or policies—is also closely related to CSA since the level of CSA that a solution provides, is a measure of its usefulness. As a consequence, it is important to develop reliable and valid measures of, and ways to measure, CSA so that, e.g., relevant training goals can be stated and cyber solutions can be evaluated.

The present paper presents an overview of existing situation awareness measurement techniques, and exemplifies how these techniques can be used for CSA measurement. The paper is structured as follows. Section 2 introduces the reader to the area of CSA and provides the necessary background regarding situational awareness. Then, Sect. 3 reviews the area of situational awareness measurement, and discusses measurement design from a cyber perspective. Next, Sect. 4 discusses experiment design considerations in general and how to perform measurement through using cyber defence exercises (CDXs) in particular, which is followed by a practical example of how to setting up a CDX for being able to train for a diversion attack. Finally, Sect. 5 concludes the paper.

2 Background

The purpose of this section is to frame the concept of situation awareness and its development. Situation awareness existed before [8] the publication of Mica R. Endsley's seminal article entitled "Toward a Theory of Situation Awareness in Dynamic Systems" [12], but a wider acceptance of the theories undoubtedly seem to have gained traction in the academic community thereafter as manifested by increasing numbers of research papers on the subject [40]. The reason for studying situation awareness, SA, in the first place is the assumption that good SA contributes to better system design, which in turn ultimately leads to

[2] In this paper we use the terms "situation awareness" and "situational awareness" interchangeably.

better decisions, actions and more successful mission outcomes. There are several proposed models for SA, but many of those appear to view the SA construct differently, and most models focus on the process of acquiring SA from the view of an individual operator as opposed to the multiple individual perspective where acquiring of shared or team SA is emphasised [45]. There are, however, theories that specifically aim to describe and measure phenomena such as team awareness, shared situation awareness and distributed shared awareness, DSA, and the like [1,44]. According to Artman [1], team members in a studied military command and control setting created SA at least by their interactions with the environment through active monitoring, negotiation with other team members, and by use of artefacts. Thus, when situation awareness theories involve groups or teams, a social dimension is also added.

According to Stanton et al. [52], three models and their associated theoretical perspectives dominate. Besides Endsley's three-level model, here: Endsley's model, there is the perceptual cycle model [50] and the activity theory model of Bedny and Meister [2]. In short, the perceptual cycle model emphasises that situation awareness is dependent on the task environment and that situation awareness is externally-directed, that goals and criteria for performance must be explicit in the environment and that the cyclic nature, as suggested by the name of the model, is due to the assumption that knowledge influences behaviour, which in turn sometimes affects and modifies the environment [50]. The activity model, which is a significantly larger construct than Endsley's model, gives that situation awareness can not be viewed in isolation, and that other behavioural concepts tied to human activity have to be understood as well [2]. To summarise, all three models of situation awareness build upon the assumption that the operator has to have a cyclic iterative interaction with the environment, but the perceptual cycle model emphasises the need for interaction with regard to perception, and the activity theory model emphasises the interplay via performed actions. We will not elaborate further on the perceptual cycle model or the activity theory model in this paper.

Endsley's model of situation awareness has found its use and gained widespread acceptance during the years as reflected in the contemporary literature, even if the scientific rigour of some of its theoretical underpinnings or different definition issues are questioned by some [4,5,19,48]. The formal definition of SA, due to Endsley [8], is that it denotes a person's "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." In addition, the person, or operator, also has to have an understanding of the relevant parameters of the system itself [11].

Endsley's model emerged from the aviation domain. She submits that the above mentioned definition merely specifies the scope of the situation awareness construct, and that the elements for different aircrafts or, indeed, systems, have to be determined [10] for each domain. She also proposed a methodology, situation awareness requirements analysis, for the task of determining those elements for the air-to-air combat fighters domain [10]. Other areas for which

relevant elements have been identified include, for example, en route air traffic control [15] and command of infantry platoons [35]. The proposed methodology includes the consecutive steps of conducting unstructured interviews with subject matter experts, SMEs, followed by a goal-directed task analysis in which goals, sub-goals and SA requirements to meet those goals are determined. In the next phase a structured questionnaire is submitted to another group of SMEs in order to add an objective assessment to the goals identified in previous phases. Each item is then rated depending on its criticality to reach the sub-goals. The resulting battery of questions about the identified parameters, is intended for the measurement of all three levels of situation awareness. To have a set of questions that reflects the relevant aspects of situation awareness is a critical prerequisite needed to perform further measurements of an operator's, or a team of operators', SA.

2.1 Evaluation of Cyber Threat Insight

As indicated above, situational awareness is often defined following Endsley [8] as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future." As suggested by Endsley in later work [12], this definition can be seen as delineating ascending levels of awareness ranging from (1) mere basic perception of important data, over (2) interpretation and combination of data into knowledge, to (3) the ability to predict future events and their implications.

In this paper we define cyber situational awareness to be the part of situational awareness which concerns the "cyber" environment. In other words, CSA is what enables system administrators and incident managers to swiftly and appropriately respond to cyber attacks and other incidents pertaining to their operations. However, to acquire and uphold appropriate CSA requires a full understanding of the threat in order to be able to plan strategically for appropriate actions concerning, e.g., training undertakings, possible insider threats, etc. Hence, CSA needs to be understood not only in itself but also with respect to external factors concerning, e.g., the physical environment, the political dimension, etc.

It is easy to see that lack of appropriate CSA makes victims more vulnerable to cybercrime (CC). This is all the more true today, when many crimes also have an IT aspect in them. For example, in June 2011, enterprise networks in the port of Antwerp, Belgium, were hacked by drug traffickers, so as to facilitate their smuggling operations alongside legitimate goods delivered in containers. By manipulating the dispatching of containers upon arrival, the smugglers were able to retrieve the containers holding drugs before the legitimate container owners did. The operation was exposed only when port workers started to notice containers disappearing for no apparent reason. Once the criminal operation was exposed, the police seized over two tons of cocaine and heroin, and more than a million euros [17].

Another example which is interesting to reflect upon from the perspective of CSA is the digital bank attack tactics exposed by Symantec in 2012: distributed

denial of service (DDoS) attacks are no longer just a blunt tool that causes a lot of annoyance, but less harm. Rather, attackers have started to use DDoS attacks as diversions, in order to draw the attention of system administrators away from a more sophisticated attack[3]. This kind of tactic really emphasises the need not only to perceive lots of data (e.g., by means of intrusion detection systems, etc.) but also to correctly interpret it in order to predict what the adversary will do next. In other words, countering these new and sophisticated attacks hinges on proper CSA.

3 Measurement of Awareness Level

The formal definition of situational awareness according to Sect. 2.1 has gained acceptance during the years and is widely used throughout the contemporary literature. *Testing* of situational awareness, however, has not matured into an equally well-defined tool set. Endsley's definition suggests that situation awareness can be reached in a gradual manner where the understanding on higher levels to some extent depends on the awareness on lower levels, but not in a linear way [14]. To test to what extent there is an understanding of the situation in terms of these levels typically requires that specific measurement solutions are developed in order to account for the specific domain. It follows that the validity of situational awareness measurement, and of CSA measurement as a means to evaluate cyber solutions, is closely related to (1) the measurement design, taken together with (2) the application of interest.

Concerning measurement design, many more or less elaborate and valid methods to measure SA exist. Hence, to determine whether it is possible to evaluate/test a cyber solution in terms of achieved CSA then amounts to identifying whether the cyber solution, in itself or a part of it, lends itself to CSA measurement, and, if so, to identifying a suitable activity where CSA can be measured using existing SA measurement techniques. Depending on the need, this activity can, e.g., be a small-scale exercise or a full-scale CDX using an exercise design where it is possible to perform relevant training whilst at the same time evaluating to what extent the cyber solution has resulted in individual understanding of the overall cyber situation. To measure the obtained CSA the exercise is typically frozen at randomly selected times and subjects are queried as to their perception of the situation at the time (queries on specific data or data criteria). The reasoning behind the randomly selected times of breaks is that it will not be possible for the subject to mentally prepare for the queries. Hence, it needs to be stressed that SA (and thereby CSA) is a distinct and unique phenomenon which applies to individuals' mental models in a universal sense. It refers to the availability of a comprehensive and coherent situation representation of what is currently known, and which is continuously being updated based on the individual's recurring assessment of the situation.

[3] http://www.zdnet.com/article/symantec-data-stealing-hackers-use-ddos-to-distract-from-attacks/.

As indicated, the three levels to be measured and distinguished between during CSA measurement consist of perception, comprehension, and projection. From a cyber security perspective, the perception level thus concentrates on the perception of cyber environment changes including, e.g., noticing an intrusion detection system alarm, whilst the comprehension level focuses on the understanding of what this actually means in terms of, e.g., a website defacement attack, a new kind of friendly user behaviour, etc. Finally, the projection level signifies a more in-depth understanding of the situation in that one is also able to make predictions concerning the forthcoming development of the situation to make informed decisions regarding how to act in order to manage the situation. For the purpose of constituting a means for assessment of cyber solutions, it is necessary that the cyber solution—be it a technical tool, a methodology, or something else—lends itself to testing with regard to understanding of some aspect of the cyber environment along the lines of perception, comprehension, and projection.

The objective for all kinds of measurement is to be able to compare an object or event with another. Stanley Smith Stevens, who made contributions to the field of measurement theory, states that it for measurement is essential that "numbers are assigned to aspects of objects or events according to one or another rule or convention" [53]. Accordingly it follows, when we have those numbers, that they have to be compared to something. For SA, the operator's SA has to be compared to, ideally, an objective truth in order to be able to rate the level of SA. Parasuraman et al. [39] claim, without further comment, that there is such a "ground truth" against which the SA can be compared, while Dekker et al. [5] vehemently argue against the feasibility of acquiring such a "ground truth" as unattainable since it requires an aperspectival, e.g., extracorporeal, objectivity. As we have established that the forms of situation awareness are highly context dependent, the question of *what* constitutes the situation, and what the relevant aspects are, therefore arises.

To address that problem, however, there are a number of techniques that are developed with specific SA target domains in mind. The techniques are asserted to inherently provide a sufficiently good "ground truth" and they also to some extent prescribe how and what to measure. Further, Salmon et al. [47] make the point that most measurement techniques are, consequently, developed in line with corresponding specific models.

According to an excellent inventory of situation awareness measurement methodologies for C^4I (command, control, communications, computers and intelligence) environments, made by Salmon et al. [46], such domains include military, aviation, air traffic control, nuclear power plants, and also a few techniques intended for generic use. Their inventory contains an analysis of 17 different measurement techniques suitable for measurement of military C^4I. One of the proposed techniques is the situation awareness requirements analysis [10], an integral part of SAGAT [9] which we will dwell further into below. Following the Salmon et al. categorisation [46], the remaining 16 techniques can be grouped into self-rating techniques, probe techniques, observer rating techniques, performance measures, process indices, and combinations thereof:

Self-rating techniques: CARS [37], MARS [34], SARS [58], SART [54], SA-SWORD [57].
Probe techniques: Sacri (freezing on-line probe) [25], SAGAT (freezing on-line probe) [9,11], SALSA (freezing on-line probe) [23], SPAM (real-time probe) [7].
Observer rating techniques: SABARS [34].
Performance measures: performance measures can be collected both by measuring explicit and implicit performance.
Process indices: eye tracker, verbal protocol analysis.
Combinations: QUASA [36], C-SAS [6], SASHA [29].

In addition, we also have CAST [22], which is designed to measure team SA. CAST can arguably be classified as a combined observer rating and performance measuring technique.

Endsley's definition suggests that ascending levels of perception, comprehension, and projection, also called level 1, 2, and 3 respectively, as derived from her definition, can be reached [14], but, as we have seen, to test to what extent those levels have been achieved often requires that specific measurement solutions are developed [47].

Endsley asserts that (good) SA can be seen as a factor that increases the probability for good performance, but does not guarantee it [11]. By measuring situation awareness, good design choices for systems can be made, which in turn ultimately increases the probability for the operator to make good decisions and avoid bad ones [13]. In order to develop useful measurement techniques she sought to ensure the validity and reliability of a technique by (1) establishing metrics that solely measure the construct that the technique claims to measure, (2) providing the required insight using sensitivity and diagnosticity measures, (3) utilising a well-balanced probing method in relation to its purpose, and (4) not substantially altering the construct during the process.

In her quest, Endsley reviewed and analysed several existing techniques. She concluded that physiological techniques such as electroencephalographic measurements as well as eye tracking are inadequate to measure situation awareness by themselves. With regards to performance measures she submits that a global performance measure may be useful for obtaining a "bottom line measure," but that performance measures otherwise are hard to conclusively tie to situation awareness as performance may be affected by many other factors than that of situation awareness [11]. Another technique, external task measures, which involves artificially changing or removing pieces of information as proposed by Sarter and Woods [48] was also deemed inadequate. She regards embedded task measurement, i.e., the measurement of specific subtasks, as a possible way to gain information that can be used to infer conclusions about overall situation assessment. An identified potential problem, though, is that the achieved SA for the measured subtask may not correspond to the level of overall SA. The observer rating technique was also discarded as being insufficient in itself to measure situation awareness because it, according to Endsley, probably does not provide an unbiased assessment of the operator's situation awareness. Further

techniques were also reviewed by Endsley who eventually arrived at the conclusion that a probe technique best met her requirements, according to above, for a measurement technique. In the following we elaborate further on three selected techniques, namely SAGAT, SART, and QUASA, due to their popularity and proven validity.

A standard technique suggested by Endsley [9], is the situation awareness global assessment technique (SAGAT). As depicted above, SAGAT may be classified as a probe technique, or more specifically as a freezing on-line probe technique. SAGAT includes queries about all situation awareness requirements as discussed above, including level 1, 2, and 3 components, system functioning and status, as well as relevant features of the external environment [11]. SAGAT suggests that operators are intermittently queried concerning carefully chosen state parameters at random points of time during a dynamic situation. The SAGAT protocol prescribes that a number of questions are asked for each of the three situational awareness levels in order to determine to which degree the subject is currently aware of the situation for each level. A commonly occurring setting in which SAGAT is typically used is in a simulator, such as a flight simulator, that simulates real-life situations. For querying the subject, the simulation is typically frozen so that the SAGAT questions can be asked whilst the simulation is at rest. The underlying idea is to remove all relevant information from the operator (e.g., the operator's displays) before the questions are asked. The answers are then compared to the states of the selected variables in the simulation, and the more accurate the answer, the better. Examples of states of variables that are asked for in the context of aviation [10] include own heading, own location, aircraft heading, G level, fuel level, weapon quantity, etc. Although SAGAT is intrusive, Endsley reports that the performance during the continuation of the simulation is not affected if the probing questions are answered within, at the most, five to six minutes [11].

Another wide-spread, versatile and easy to use measurement technique for SA is Taylor's [54] situation awareness rating technique, SART. SART uses self-rating. The protocol requires the subject to rate to what degree he or she perceives (1) a demand on operators resources, (2) supply on operator resources, and (3) understanding of the situation, on a set of bipolar Likert scales. The ratings are then combined in order to provide an overall SA measurement score [16].

The quantitative analysis of situational awareness technique (QUASA) [36] is a combined self-rating and probe technique. QUASA is performed via probe statements that state a proposition as of the current state of parameters in, e.g., a simulation to which the subjects have to agree or disagree, e.g., "true or false?," thus the probe. Then, the subject has to rate to what degree of confidence the prior assessment was made using a scale with five degrees, hence the self-rating part of the technique. As a third question, the subject is then asked "Which teams will mostly answer this probe correctly?" The idea behind QUASA is to take advantage of concepts from signal detection theory, i.e., the analogue of the detection and the consecutive step of determination of the quality (of the signal). Further, QUASA aims to measure the "actual situation awareness" as acquired

via cognition, and "perceived situation awareness" as sensed by metacognition. In experiments made within a military context (operational net assessments), it was shown that the technique provided insights into individual's situation awareness, but also regarding levels of sensitivity and biases in groups which may be useful information as well [36].

In a comparative study of the three situational awareness measurement techniques SAGAT, SART and CDM (Critical Decision Method, which is not further mentioned in this paper) within the context of a military planning task, it was shown that SAGAT level 2 (comprehension) showed a significant correlation relative to task performance as opposed to any other of the analysed techniques [47]. Another interesting conclusion was that no significant correlations between SAGAT and SART were found, indicating that the techniques may have measured different variables, as opposed to the stated intent not to do so, which is also the same conclusion that Endsley et al. made in a comparative analysis in 1998 [16]. Furthermore, Salmon et al. [47] make the important remark that success of SAGAT as a measurement technique is dependent on the ability to find relevant elements of situation awareness a priori, which is why they see SAGAT primarily as useful for measuring situation awareness in linear and deterministic settings.

3.1 The Cyber Domain

The U.S. Army Field Manual 3–38 entitled "Cyber Electromagnetic Activities" [56] defines cyberspace in terms of a man-made construct of systems of systems in that many small and diverse systems comprise the structure as a whole. These systems exist in the physical world. Cyberspace, which continually evolves, facilitates the use and exploitation of information, human interaction, and intercommunication through computers and telecommunication systems. Cyberspace and the electromagnetic spectrum, EMS, have converged into a global interdependent network, emphasising that the environment is not confined to a specific physical place. In order to successfully tackle cyber issues it is therefore asserted that a holistic approach involving physical infrastructure, data networks, and the EMS is suitable.

It seems, as given by the discussion hitherto, that there currently is no situation awareness measurement technique that is suitable for all domains. Although it remains to be thoroughly analysed to what extent the listed measurement techniques according to Salmon et al. [46] can be used for measuring situation awareness in the cyber domain, it is our belief that it may be fruitful to assemble components from several of the existing techniques in order to create a feasible measurement solution for the cyber domain.

Endsley's proposed situation awareness definition, i.e., a person's "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" [8] may have to be carefully reconsidered because both "time" and "space" can be viewed differently in the cyber domain than in other domains, and both of these aspects are judged to be of importance in the situation awareness

model construct. Temporal aspects of situation awareness are mentioned [12] and further elaborated on [13] by Endsley, where she notes that (1) the perception of time, (2) the temporal dynamics associated with events, and (3) the dynamic aspect of real-world situations, are aspects that may be considered. Spatial aspects of SA are also mentioned by Endsley [12] who points out that, in order to gain situation awareness, an operator needs to take the subsets of the environment that are relevant to tasks and goals into account.

As derived from the U.S. Army Field Manual mentioned above [56], the spatial properties of cyberspace is plainly that cyberspace is global, which makes the task of determining the outer geographical boundaries of a situation according to the situation awareness model problematic if not "everything everywhere" should be included. As the other delimiting boundary, the location of one's own system or network along with its externally facing connection point/points may be suitable.

Regarding the relevant temporal aspects to be considered in the cyber domain, we feel that it is of essence to keep several parallel time scales in mind, namely those that may be labelled near real-time, mid-term, and long-term. The near real-time perspective pertains to the time for signals to traverse through various communication systems to and from one's own system or network, and the processing time of those signals in electronic circuitry, which typically takes place during fractions of a second. The mid-term perspective may constitute the interval between minutes, e.g., updating software or applying a patch, and months, e.g., the increased user security awareness with regard to social engineering attacks. This is the timeframe in which different additional effects, other than the near instantaneous, of (cyber) actions will surface and be understood. The long-term perspective may stretch from months to years, and involves relevant aspects of the evolution of the domain itself, e.g., introduction of new (technical) protocols or changes in the governance of the internet.

In Table 1 the discussed cyber domain characteristics with regard to time and space are contrasted relative to other domains that are commonly discussed within the SA literature.

Table 1. Domain comparison with regard to geographical and temporal boundaries for situational awareness.

Domain/context	Geographical boundaries	Temporal boundaries
Tactical flight operations	The aircraft vs. the immediate vicinity of the aircraft	Start of flight mission vs. end of flight mission
Nuclear power plant process control	The power plant	Arbitrary starting point vs. continuous/infinite time
Military command and control	Own position vs. area of operations	Arbitrary starting point vs. mission/campaign time
Cyber defence	Own network vs. globally interconnected computers	Near real-time vs. continuous/infinite time

Endsley originally asserted [12, p. 50] that information reaches the operator from two sources, the real world and through an interface of a system, but later refined her assertion to include a third source [13, p. 7], the communication with team members and others, without, as far as we know, revising her SA model. Consequently, how information reaches the operator is another factor that may differentiate the cyber domain from other domains. In the cyber domain no direct observations, e.g., looking out the window, of the external physical world are feasible. All information about the state of the external environment comes mediated to the operator through artefacts or direct interpersonal communication, e.g., the status of a remote industrial control system is conveyed via sensors, a telecommunication system and displays. Details about a cyber threat may be learned through a conversation.

Drawing from another U.S. military publication, the Joint Publication 3–12 entitled "Cyberspace Operations" [55], we obtain another, functional, view of cyberspace, in terms of three layers:

1. the **physical** network layer in which the physical network components reside in the geography,
2. the **logical** network layer where nodes are interconnected, sometimes without a straightforward mapping to the other network layers, and
3. the **cyber-persona** layer, which takes advantage of the rules that apply in the logical network layer to "develop a digital representation" of an individual or entity identity in cyberspace.

We submit that all three layers have to be treated in a holistic way, but that the logical network layer is the layer that distinguishes the cyber domain from other domains the most. According to the mentioned Joint Publication 3–12, the "logical network layer consists of those elements of the network that are related to one another in a way that is abstracted from the physical network, i.e., the form or relationships are not tied to an individual, specific path, or node" [55]. Hence, the logical layer is an intangible abstraction that exists in computer memory only, that provides the cohesion between the physical hardware and the humans in the other two tangible layers.

We argue that, if made known to operators, a well-performed situation awareness requirements analysis, perhaps using the above mentioned viewpoints as a basis, may be viewed as an educational effort. The resulting hierarchical goal structure, subsequently, can also be used to inform the operator and drive her or his data collection. However, it has recently been shown that availability of an increased volume of additional task-relevant information does not result in significant effects with regard to mission performance [33]. Therefore it is rather the right information than the amount of information that counts. Endsley [13] concur based on her statement that more data does not equal more (relevant) information. Besides being a component in SA measurement, the information can also provide input to system designers who can design better systems, thus contributing indirectly to greater situation awareness and better mission outcomes.

To further expand the scope, yet other dimensions that constitute the cyber domain, besides a cyber security perspective, may be added. Recently Rid and Buchanan proposed a framework, named the Q Model, that covers multiple dimensions and levels of presumably attainable knowledge in an article that discussed cyber attack attribution [42]. We assert that the proposed framework is also useful for identifying required elements for CSA. The quite extensive and coherent model contains three functional, not necessarily hierarchical[4], levels: the tactical/technical level that mainly deals with the questions "What?" and "How?," the operational level that mainly deals with the question "Who?," and the strategic level dealing with the "Why?" The model proposes several specific questions for each level concerning, e.g., technical modus operandi, attacker characteristics, and involved organisations, but also some questions related to predictive knowledge concerning, e.g., attacker intent, second-order effects, and so forth. In short, we notice that the Q Model levels have a striking resemblance to Endsley's three-level model as discussed throughout this paper, in that they also show an ascending complexity using three levels. We feel that the proposed Q Model bears promise to be used to further the understanding of the cyber domain, and specifically contribute to the development of CSA and its associated measurement techniques.

Concerning measurement design for cyber, however, it should for the purpose of this paper suffice to mention that many more or less elaborate and valid methods to measure CSA can be developed using, e.g., SAGAT that can be adapted to suit different domains, along with other awareness measurement techniques such as QUASA, as a basis. In general, we propose the development and use of an SA measurement technique that is constructed specifically for the cyber domain, taking into account relevant elements, as mentioned above, combined with the measurement of bottom-line mission performance. We are well aware of that it is questionable if performance measures, mainly external measures, contribute to the measurement of situation awareness per se, but assert that it is indeed useful to measure performance as related to the mission goals, which is the ultimate rationale for having (good) situation awareness in the first place. The correlation between the level of CSA and the overall mission performance can also be used to gain second order insights.

4 Experiment Design

From a cyber threat perspective it is not easy to "know your enemy." Attackers typically possess a number of varying skills, have complex motives, might be organised in teams, etc. Moreover, the defending organisation's computer infrastructure is often complex and distributed, which makes knowing one's own environment a nontrivial task. It is in this context a cyber threat management solution needs to be evaluated, and this assessment needs to take the actual

[4] In military theory, the hierarchical war levels consist of the (lowest) tactical, operational, strategic, and political (highest) levels.

understanding of the cyber threat into consideration rather than solely evaluating the extent of being able to successfully make use of physical protective measures.

As a basis for measurement, the previously mentioned awareness levels proposed by Endsley serve as a baseline. That is, for any cyber management solution there is an underlying bigger picture that can be more or less understood, and for tackling, e.g., CC and/or cyberterrorism (CT) strategically it will be beneficial to have an understanding that to the greatest extent possible makes it feasible to understand the cyber threat not only in terms of mere perception of attacks but also in terms of working knowledge regarding the ulterior motives of the attack, additional attacker profiles, how to predict future attacks, how to devise new forms of training, etc.

Depending on the nature of the cyber threat of interest and the chosen measurement scheme according to Sect. 3, questionnaires or simpler simulations might suffice for situational awareness measurement in some situations whilst in other cases a more elaborate solution that can account for a higher degree of realism is required. In the following we elaborate on and suggest the use of CDXs that are adapted to accommodate possibilities for performing measurement, thereby testing the level of developed CSA.

4.1 Cyber Defence Exercises

CDXs are today being undertaken at regular intervals with relevant personnel participating in an environment that provides for a good level of realism. As an example, during the "cyber defence exercise" in the U.S., the participating schools are tasked to design and implement a computer environment providing a number of services which the participants are later supposed to defend from cyber attacks that are initiated by the "red force" of hackers which are in reality provided by the NSA [38]. The "Baltic Cyber Shield" exercise provides a similar example where six teams from across northern Europe were tasked to defend critical infrastructure networks from a group of professional penetration testers [26].

As indicated, a CDX provides an environment which can be tailored to resemble a relevant cyber threat arena, which can be further used for obtaining additional insight regarding true hacker motives. For a CDX to provide relevant higher-level data concerning a cyber threat, the CDX needs to be designed in a way that puts the cyber threat in focus and lends itself to observing the relevant aspects. The remainder of Sect. 4 discusses possible CDX setups, and the way to gain CSA insight through using both qualitative and quantitative observations. The main idea is to carefully insert suitable activities within the CDX in order to bring about a behaviour that can be observed and that makes the CDX participants engage in the cyber activity that the cyber threat management solution focuses on. As an example, setting up a honeypot of a suitable kind might attract certain types of attackers. The attacker behaviour can then be observed and used for determining the user's characteristics. In the long run, a number of such observations can turn, e.g., a stereotypical "script kiddie profile" into a

more well-informed understanding of the attacker that can later play an integral role for analysing the overall organisational threat and the strategic measures that ought to be undertaken according to a higher CSA level.

4.2 Games

It is known that forensic psychology can be of great assistance to CC investigation [30], which assumes realistic hacker profiles and personality characteristics to be an important means for cyber defence and, hence, for informing CSA. Whilst many theories regarding hacker motives indeed abound, these are seldom based on actual empirical data and it is unclear whether the current knowledge is at all representative. Notable exceptions exist, though, with the "honeynet project"[5] being an interesting initiative where honeypots are placed on the internet to allure hackers in order to learn about their methods. The knowledge gained from the honeypots is used for raising awareness through issuing "know your enemy papers" where people can gain insight regarding the development of cyber threats and the measures that ought to be undertaken. From a pedagogical viewpoint, some insight regarding hacker behaviour has been gained through hands-on training within specifically designed isolated computer labs which the students are able to use as a playground for trying out various security related tools in a secure fashion. Although a number of successful initiatives have been reported on [3,24,28,43], these still remain fairly small-scale and are typically dependent on specific individuals. Full-scale exercises in terms of CDXs provide for more realism, and better chances of gaining insight that can be considered to be more relevant from a CSA perspective.

It is important, however, to consider both the limitations and the strengths of this claim. Following Raser [41], we distinguish between four criteria for the validity of gaming as a research tool: psychological reality, structural validity, process validity, and predictive validity.

For some cyber threats, these criteria are relatively easy to meet. If the objective is to find the success rate of remote code execution attacks as described by Holm et al. [27], then the exercise environment can be set up accordingly, and whenever a remote code execution attack is performed by the red team, the simulation environment ensures structural validity (operating systems, communication protocols, etc., all work just like in reality), process validity (finding vulnerabilities, using exploits, obtaining privilege escalation, etc., all work just like in reality), and predictive validity (what works in the simulated environment works in reality—if the real systems are configured just like the simulated ones). As for psychological reality, this cyber threat requires only that participants, once in a while, actually attempt to perform a remote code execution attack.

For other kinds of threats, however, the criteria are much more demanding. As noted by Sommestad and Hallberg [51], "the incentives that real attackers or defenders act upon" appear difficult to assess in exercises or competitions. The requirement for psychological reality now becomes prohibitive, as it more or less

[5] https://www.honeynet.org/.

requires the participants to actually *be*, say, ideologically or financially motivated. Indeed, not even economic incentives for the participants are certain to make them financially motivated since they "may make competitive choices not because they want to maximise their point totals, but because they want to beat the other person" [49]. There is, however, a middle ground. Even if questions regarding the psychology of attackers are beyond our reach, questions about their actions *given their incentives* are not. And the incentives of the game can be set to reflect motivation structures found in the IT security literature, gained from questionnaires, inspired by expert assessments, etc.

In the following, we consider a few examples of possible game setups, constructed to measure various aspects of CSA. Each game assumes an ongoing CDX with at least two opposing teams:

Benefits from eavesdropping. The team under attack (blue team) is given access to the communication channel(s), e.g., IRC, of the attacking team (red team). In the basic setup, the blue team has to manually read all the information in person, and take appropriate defensive measures. In more advanced setups, traffic is either pre-processed to highlight terms of interest or fused with other information sources. All of these setups can either be real-time, or lagged by a number of minutes. These setups can be compared to a baseline of no IRC access. In this way, the relative benefits of eavesdropping on the opponent can be measured. If enough trials are conducted, quantitative measures such as time-to-compromise or probability-of-compromise can be elicited. *This scenario measures the value of CSA for defence.*

Targeting with social network analysis. One team is given the ability to partially disrupt the IRC communications of the opponent. In one setup, the team can inhibit the IRC communications of a random member of the opposing team. In a more advanced game, the team has a software tool that displays the social network of the opposing team along with the centrality of each member. The team can then make a more informed decision regarding which IRC communications to disrupt. *This scenario measures the value of CSA for attack.*

Information overload. In this game, the blue team is attacked and is fed with accurate information about this attack, but is also simultaneously fed with a significant amount of irrelevant information. Variants include overloads aimed at single decision-makers, or overloads crafted to make several people in the team all slow down at a time. Quantitative measurements from this scenario include delays in decision-making, delegation of decisions and shutting down certain inputs (measures taken from Libicki [31]). *This scenario measures the extent to which competent information management and fusion tools offer remedies to information overload.*

Insider threat. In this game, the team is subject to an attack from one of their own. The individual is covertly given this task as part of the exercise setup. As noted by many authors, the insider cyber attack is a significant threat. In one setup, there is no system dedicated to detecting insiders. In another setup, an insider detector such as ELICIT [32] is employed. Additional setups would

fuse ELICIT with information from other sensors. *This scenario measures the value of CSA for insider detection.*

Value of honeypots. The team under attack is allowed to configure a honeypot within their network, in order to learn from red team attacks on it. In one setup, the honeypot is monitored in real-time. In another setup, historical data from previous exercises is used instead. *This scenario measures the relative value of historical attack data vs. honeypot data for CSA.*

Automatic hypothesis monitoring. Computer network defence is not only about real-time operational measures, but also about risk analysis and planning beforehand. In this game, the team is allowed to identify high-level attack plans against their own systems before the exercise starts. They also build a threat assessment model with indicators (detectable with sensors at their disposal) allowing the model to provide a continuous threat assessment throughout the exercise. *This scenario measures the value of model-based threat reasoning for CSA.*

Service level agreements. Situational awareness is important not only during IT service operation, but also in the procurement and planning phases. In this game, the team does not fully control all of its IT infrastructure. Rather, some services are "bought" from a service provider, and the team must procure service level agreements regarding guaranteed restore times (e.g., service X is always restored within five minutes for $1,000 or within one hour for $100) before the actual exercise starts. With a limited budget, they must prioritise—some services must be deemed more important than others. In the baseline setup, no historical information is available. In subsequent setups, historical data from previous exercises is made available to the team, allowing more informed decisions. With the advent of cloud services and the notion of SOA, such decision scenarios are rapidly becoming increasingly relevant, but recent research suggests that decision-makers do not always make rational choices in SLA decision-making [21]. *This scenario measures the value of CSA regarding the past when making management decisions for the future.*

Aggressor identification. Four different teams at different locations participate in the exercise. One of the teams is secretively selected to be the aggressor and will during the exercise attack a randomly selected team, possibly hijacking resources from the other teams for the purpose. The task of the attacked team is to identify the aggressor using cyber information fusion techniques, optionally including help from the other teams. *This scenario measures the value of CSA for attribution.*

These examples have shed light on the interplay between specific cyber threat management scenarios and CDXs. The cyber threat specifics is required for proper incentive structures in exercises to be set up. The exercises can then serve to evaluate the level of CSA with respect to a specific cyber threat solution through conducting exercises where relevant and realistic courses of action for different attacker types are operationalised through using appropriate exercise incentives. Such behavioural information can be both qualitative, e.g., common

modi operandi for espionage, and quantitative, e.g., the relative detection rates of ideological attackers compared to insiders.

4.3 Principles for Cyber Situational Awareness Measurement

In this section we discuss the differences between SA measurement experiments for the cyber domain and other domains, and highlight some important aspects to take into account for measurement of CSA. As an experiment platform, the cyber range not only enables the simulation—its computers and networks, real or simulated, are also an integral part of the system that includes the *subject* for training, experiment or measure in the cyber domain. For other domains the computers and networks are used as instruments of the simulation, but for cyber purposes the computers and the networks are at the same time the tools that are used by the operators.

It must be remembered that SA is measured on the operator, even if complex CDXs are used as a backdrop. The operator, or operators, work in an environment with all available means that we have at our disposal to execute the (cyber) mission, e.g., specific arrangements of hardware and software (a technical setup). The operators perform work in work processes. They may also have different degrees of organisation. We call this socio-technical system the *cyber solution*. By measuring the SA of the operators we ultimately aim at improving the cyber solution, be it with new and faster computers, novel pieces of software, new configurations of the software, improved visualisation techniques, or better work processes. Depending on the need, the measurement experiment can be conducted through, e.g., small-scale exercises or full-scale CDXs using exercise designs where it is possible to perform relevant training whilst at the same time evaluating to what extent the cyber solution has resulted in individual understanding of the overall cyber situation.

As discussed, information reaches the cyber operator in two ways, through artefacts via telecommunication systems, and via direct communication. Therefore, the cyber solution is of utmost importance. The cyber solution determines to what degree the operator *can* perceive, and consequently comprehend and predict future events.

Given the above we assume that the performance of the cyber solution is dependent on, and will vary with, at least three different factors: (1) how information is presented to the operators, e.g., how the technological portions of the cyber solution is configured which in turn will affect the operator's CSA, (2) the work processes, and (3) the properties of the operators themselves (including knowledge, experience, cognitive abilities etc.) We assume that, in all cases and experiments, these are the factors that affect the CSA of the operators and the levels of performance relative to the mission. We therefore assume that if we change one or more of the factors, the technical setup, e.g., the configurations of firewalls and intrusion prevention systems, etc., or the work processes, e.g., the order of which tasks are carried out, or the operators, e.g., novice or expert operators, the CSA and the performance will vary. (Alas, as noted in Sect. 3, the

relation between these factors and the resulting CSA is not perfect, but subject to both random and systematic errors, making measurement more challenging.)

Noticing that SA measurement is highly context dependent according to the previous discussion, we emphasise the distinct properties of the cyber domain with regards to the missions and the cyber solutions as well as the importance of testing relevant measures in a carefully crafted game (e.g., a cyber range simulation) to be integral parts of the experiment design. Accordingly, we propose the following elements and associated criteria to be used for guiding CSA experiment design:

Mission. Existence of a clearly defined cyber mission that is realistic and attainable. Its expected outcome has to be measurable. If applicable, spatial and temporal boundaries are to be specified.

Cyber solution. Arrangements of hardware and software (a technical setup), the operators, and their associated work processes.

Metrics. Relevant metrics for (1) SA (given by an SA requirements analysis), and (2) performance (implicit and explicit "bottom-line").

Game. Simulation with a realistic scenario, planned sequence of events, and injects that provide a controlled environment.

In addition we propose using several suitable measurement techniques that are adapted to the cyber domain, e.g., domain-specific SAGAT and QUASA techniques, and both explicit and implicit measures of performance.

To make this more concrete, consider the following example from the banking domain. Nowadays most banks offer online services, e.g., internet access to their product portfolio of financial services, to customers. According to press reports the HSBC bank was struck by a distributed denial of service, DDoS, attack against their web services in January 2016[6]. These kinds of attacks, which are often carried out with the aim to intimidate or damage the reputation of its target organisations, may cause disruptions to online services for legitimate users. In other words they affect the availability of information. According to the same source, HSBC has been hit several times in the past as well, including the end of 2012. Now, expanding the view of this incident, we may add that during the approximate same time period, in the winter of 2012 and spring of 2013, other web sites belonging to other large financial institutions were also attacked by DDoS attacks, including Bank of America, Chase, Citigroup, JP Morgan, Wells Fargo, and others[7]. Furthermore, other types of malicious activity were also detected in conjunction with some of the DDoS attacks. More precisely, attempts to gain unauthorised access and carry out unauthorised transactions that are likely precursors and indicators of fraudulent wire transfers were detected. Data breach and information manipulation of this kind is an attack on the confidentiality and integrity of information.

[6] http://www.telegraph.co.uk/finance/personalfinance/bank-accounts/12129786/HSBC-online-banking-fails-again-after-succumbing-to-cyber-attack.html.

[7] http://www.cnet.com/news/cybercrooks-use-ddos-attacks-to-mask-theft-of-banks-millions/.

Some time before the incidents mentioned above, in September 2012, the U.S. Federal Bureau of Investigation, FBI, issued a warning of a new *modus operandi* for cyber criminals: that "DDoS attacks were likely used as a distraction for bank personnel to prevent them from immediately identifying a fraudulent transaction, which in most cases is necessary to stop the wire transfer" [18]. In other words, the FBI warned that they had observed DDoS attacks being used as diversion manoeuvres by criminals to cloak other more severe types of CC.

As an interpretation of these events in terms of the concept discussed within this paper, we assert that the level 1 understanding (perception) of the situation, according to Endsley, is about detecting the existence of malicious activity in the network. Level 2 understanding (comprehension) is about drawing conclusions about the types of attacks (DDoS and unauthorised access) and their immediate implications. Level 3, i.e., the higher-order understanding (projection), would be to draw conclusions about the specific *modus operandi*, i.e., the use of DDoS as a diversion manoeuvre for the purpose of hiding other attacks. In concrete terms, such insight can contribute to the prioritisation of the work of the IT (security) department to primarily focus on preventing unauthorised access attempts (even if drowned in a simultaneous DDoS attack), and not divert critical manpower to mitigate the effects of the DDoS (a less critical mission goal). For a bank we assume that the protection of the confidentiality and integrity of customer data takes precedence over the goal of protecting the availability of services (though both are important).

Our hypothesis is that it is indeed possible to defend the network (cyber mission) with only a first or second level appreciation of the cyber situation, but that it is possible to do it *even better* with additional third level insights.

4.4 Sample Cyber Situational Awareness Experiment Setup

As elaborated on throughout this section, a good way to perform measurements of CSA is within the context of CDXs. By convention the active (trained) participants of a CDX are named the blue team and the red team. The blue team, normally the primary training audience, is assigned for defensive tasks, while the red team is assigned to be the offensive attacking team. The best way to perform CSA measurement of a blue team, is by controlling the activities of the red team to the fullest extent possible in order to provide uniform conditions in several consecutive experiments, i.e., to rigidly script the attacks with regards to sequencing and timing. In this way it is possible to isolate the measured variable reasonably well. In such a case, however, the training effects for the red team are close to non-existent. Furthermore, if the activities of the attackers are fully scripted there is a risk that the blue team questions the psychological reality of the simulation [41] and that the exercise becomes static and deterministic (see Sect. 3), and is experienced as artificial.

Instead we suggest performing CSA measurements during the regular execution of CDXs (e.g., for training purposes). By giving red teams a certain degree of autonomy, a more dynamic interplay with the blue team(s) can be achieved. Through managing the red teams using a combination of loosely formulated

tasks and an incentive structure (as mentioned in Sect. 4.2), possibly combined with direct instructions, both training effects and good conditions for measurements can be achieved for both blue and red teams. Cyber ranges generally have excellent data collection capabilities that enable extensive post-action analysis.

Using the banking CC case mentioned in Sect. 4.3 as a an example, we propose and discuss a possible CSA measurement experiment setup according to the principles in Sect. 4.3 as follows:

Mission. We would have one red team, and four blue teams. The cyber mission is to detect and prevent CC by protecting the information assets of the bank with regards to confidentiality, integrity, and availability. Sub-goals and subtasks include, e.g., continuous monitoring of network perimeter, matching of known malware parameters with incoming traffic, detecting suspicious network activities, logging and analysing activities on the internal network, stopping ongoing access attempts, etc.

The **cyber solution** is the computers and networks, hardware and software, that the bank has globally. The cyber solution includes the IT departments with their IT security functions and, specifically, the organisation, the personnel and the associated work processes that govern these functions. The mission has to be carried out continuously.

Metrics for availability is uptime/downtime of services. Other metrics, for confidentiality and integrity, are hard to define and measure directly. Implicit metrics can include, e.g., number of detected scans, number of refused connections, as well as quantifications of other kinds of attempts.

Game. As part of the game the red team would be given an incentive structure that awards high scores for fraudulent wire transfers. The red team would also be directly instructed to perform a DDoS attack as a diversion prior to a subsequent attempt to gain authorised access for the purpose of doing the wire transfer.

In this case it would be interesting to investigate, e.g., what changes in the cyber solution that would be required to enable the blue teams to focus on detecting and ultimately deflecting the attempts to gain unauthorised access, whilst under a distracting DDoS attack.

To gain a baseline we would instruct the red team to carry out the DDoS and the illicit transfer attacks as described. We would stop the simulation intermittently and ask the blue teams' questions according to the SAGAT and QUASA protocols. Level 1 questions would include, e.g., "What activity did you observe in the network?" Level 2 questions would include: "Which activities are hostile?," "What are the characteristics of those hostile activities?," and "How are the attacks carried out?" Level 3 questions would include, e.g., "Why are we attacked?," and "What will happen next?," for all four blue teams. At the same time we would record up-time of services (explicit performance) as well as successes or failures of the illicit transfers from customer accounts.

Next, we would test *changes* in the cyber solution to determine what might, and what might not, affect CSA and performance. A plethora of possible experiments can then be undertaken to test any number of ideas, such as, e.g.,

changes in firewall rules, changes in intrusion prevention system (IPS) calibration, changes in hardware, changes in software configuration, changes in information presentation, giving additional information to operators (e.g., FBI warnings, introducing bi-hourly briefings for operators for the purpose of information sharing, etc.) The changes would then be introduced to two of the teams and the simulation resumed. In further measurements the differences in CSA and performance, if any, between the teams can be used to draw conclusions about the effects of the implemented changes.

5 Conclusions

Based on the notion of situational awareness and its use for determining the level of cyber insight in terms of so-called cyber situational awareness (CSA), this paper has served to provide the foundation for developing suitable measurement techniques to be used for testing to what extent a person or a team has been able to acquire and/or maintain CSA. Being able to perform such measurement is critical for making it possible to test, e.g., to what extent training goals have been met, if a technical solution provides the sought for insight, whether a security process is capable of providing strategic insight, etc.

Although the notion of situational awareness and its role as a unique phenomenon has gained acceptance during the years, the way to measure situational awareness has been widely debated and many views exist. Also, measurement is naturally dependent on the domain, which by necessity requires that tailor-made protocols are being developed for the respective applications of interest. Hence, the development of the principles for CSA measurement that have been presented and exemplified in this paper have been based on (1) an overview of a few current situational awareness measurement techniques, in relation to (2) an analysis of the cyber domain and its similarities and differences in contrast to other domains.

It is vital to take the experiment design into account at an early stage in order for CSA testing to provide results that are relevant and applicable to the cyber aspect of interest. Albeit simpler methods requiring less resources, such as questionnaires, could sometimes be used, more elaborate simulations will most often be required for being able to providing sufficient realism and the associated measurement validity. As a result, the basis for constructing more elaborate testing mechanisms utilising cyber defence exercises (CDXs) has been provided in the article. The obvious next step and plan for future work is to develop these principles further and to validate them during the execution of a relevant CDX.

References

1. Artman, H.: Team situation assessment and information distribution. Ergonomics **43**(8), 1111–1128 (2000)
2. Bedny, G., Meister, D.: Theory of activity and situation awareness. Int. J. Cogn. Ergon. **3**(1), 63–72 (1999)
3. Brynielsson, J.: An information assurance curriculum for commanding officers using hands-on experiments. ACM SIGCSE Bull. **41**(1), 236–240 (2009)
4. Carroll, L.A.: Desperately seeking SA. TAC Attack **32**(3), 5–6 (1992)
5. Dekker, S.W.A., Hummerdal, D.H., Smith, K.: Situation awareness: some remaining questions. Theor. Issues Ergon. Sci. **11**(1–2), 131–135 (2010)
6. Dennehy, K.: Cranfield situation awareness scale: users manual. Technical report 9702, Applied Psychology Unit, College of Aeronautics, Cranfield University, Bedford, United Kingdom, January 1997
7. Durso, F.T., Hackworth, C.A., Truitt, T.R., Crutchfield, J., Nikolic, D., Manning, C.A.: Situation awareness as a predictor of performance in en route air traffic controllers. Technical report DOT/FAA/AM-99/3, Office of Aviation Medicine, Federal Aviation Administration, U.S. Department of Transportation, Washington, District of Columbia, January 1999
8. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society 32nd Annual Meeting, Anaheim, California, pp. 97–101, October 1988
9. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference (NAECON 1988), Dayton, Ohio, pp. 789–795, May 1988
10. Endsley, M.R.: A survey of situation awareness requirements in air-to-air combat fighters. Int. J. Aviat. Psychol. **3**(2), 157–168 (1993)
11. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors **37**(1), 65–84 (1995)
12. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995)
13. Endsley, M.R.: Theoretical underpinnings of situation awareness: a critical review. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness Analysis and Measurement, pp. 3–32. Lawrence Erlbaum Associates Inc., Mahwah (2000)
14. Endsley, M.R.: Situation awareness misconceptions and misunderstandings. J. Cogn. Eng. Decis. Making **9**(1), 4–32 (2015)
15. Endsley, M.R., Rodgers, M.D.: Situation awareness information requirements for en route air traffic control. Technical report DOT/FAA/AM-94/27, Office of Aviation Medicine, Federal Aviation Administration, U.S. Department of Transportation, Washington, District of Columbia, December 1994
16. Endsley, M.R., Selcon, S.J., Hardiman, T.D., Croft, D.G.: A comparative analysis of SAGAT and SART for evaluations of situation awareness. In: Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, Illinois, pp. 82–86, October 1998
17. Europol: Hackers deployed to facilitate drugs smuggling. Intelligence Notification 004-2013, European Cybercrime Centre (EC3), Hague, Netherlands, June 2013. https://www.europol.europa.eu/sites/default/files/publications/cyberbits_04_ocean13.pdf

18. Federal Bureau of Investigation: Fraud alert - cyber criminals targeting financial institution employee credentials to conduct wire transfer fraud. Press release, Financial Services Information Sharing and Analysis Center (FS-ISAC) and the Internet Crime Complaint Center (IC3), September 2012. http://www.ic3.gov/media/2012/fraudalertfinancialinstitutionemployeecredentialstargeted.pdf
19. Flach, J.M.: Situation awareness: proceed with caution. Hum. Factors **37**(1), 149–157 (1995)
20. Franke, U., Brynielsson, J.: Cyber situational awareness - a systematic review of the literature. Comput. Secur. **46**, 18–31 (2014)
21. Franke, U., Buschle, M.: Experimental evidence on decision-making in availability service level agreements. IEEE Trans. Netw. Serv. Manage. **13**(1), 58–70 (2016)
22. Gorman, J.C., Cooke, N.J., Winner, J.L.: Measuring team situation awareness in decentralized command and control environments. Ergonomics **49**(12–13), 1312–1325 (2006)
23. Hauss, Y., Eyferth, K.: Securing future ATM-concepts' safety by measuring situation awareness in ATC. Aerosp. Sci. Technol. **7**(6), 417–427 (2003)
24. Hill, J., Carver, C., Humphries, J., Pooch, U.: Using an isolated network laboratory to teach advanced networks and security. In: Proceedings of the 32nd ACM SIGCSE Technical Symposium on Computer Science Education, Charlotte, North Carolina, pp. 36–40, February 2001
25. Hogg, D.N., Follesø, K., Strand-Volden, F., Torralba, B.: Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. Ergonomics **38**(11), 2394–2413 (1995)
26. Holm, H.: Baltic cyber shield: research from a red team versus blue team exercise. PenTest magazine **2**(5), 80–86 (2012)
27. Holm, H., Sommestad, T., Franke, U., Ekstedt, M.: Success rate of remote code execution attacks: expert assessments and observations. J. Univ. Comput. Sci. **18**(6), 732–749 (2012)
28. Jacobson, D.: Teaching information warfare with lab experiments via the internet. In: Proceedings of the 34th ASEE/IEEE Frontiers in Education Conference, Savannah, Georgia, pp. T3C/7–12, October 2004
29. Jeannot, E., Kelly, C., Thompson, D.: The development of situation awareness measures in ATM systems. Technical report HRS/HSP-005-REP-01, European Organisation for the Safety of Air Navigation (EUROCONTROL), Brussels, Belgium, June 2003
30. Kirwan, G., Power, A.: Cybercrime: The Psychology of Online Offenders. Cambridge University Press, Cambridge (2013)
31. Libicki, M.C.: Conquest in Cyberspace: National Security and Information Warfare. Cambridge University Press, Cambridge (2007)
32. Maloof, M.A., Stephens, G.D.: ELICIT: a system for detecting insiders who violate need-to-know. In: Kruegel, C., Lippmann, R., Clark, A. (eds.) RAID 2007. LNCS, vol. 4637, pp. 146–166. Springer, Heidelberg (2007)
33. Marusich, L.R., Bakdash, J.Z., Onal, E., Yu, M.S., Schaffer, J., O'Donovan, J., Höllerer, T., Buchler, N., Gonzalez, C.: Effects of information availability on command-and-control decision making: performance, trust, and situation awareness. Hum. Factors **58**(2), 301–321 (2016)
34. Matthews, M.D., Beal, S.A.: Assessing situation awareness in field training exercises. Research report 1795, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, Virginia, September 2002
35. Matthews, M.D., Strater, L.D., Endsley, M.R.: Situation awareness requirements for infantry platoon leaders. Mil. Psychol. **16**(3), 149–161 (2004)

36. McGuinness, B.: Quantitative analysis of situational awareness (QUASA): applying signal detection theory to true/false probes and self-ratings. In: Proceedings of the 2004 Command and Control Research and Technology Symposium (CCRTS), San Diego, California, June 2004
37. McGuinness, B., Foy, L.: A subjective measure of SA: the crew awareness rating scale (CARS). In: Proceedings of the First Human Performance. Situation Awareness and Automation Conference, Savannah, Georgia, pp. 286–291, October 2000
38. Mullins, B.E., Lacey, T.H., Mills, R.F., Trechter, J.M., Bass, S.D.: How the cyber defense exercise shaped an information-assurance curriculum. IEEE Secur. Priv. **5**(5), 40–49 (2007)
39. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J. Cogn. Eng. Decis. Making **2**(2), 140–160 (2008)
40. Patrick, J., Morgan, P.L.: Approaches to understanding, analysing and developing situation awareness. Theor. Issues Ergon. Sci. **11**(1–2), 41–57 (2010)
41. Raser, J.R.: Simulation and Society: An Exploration of Scientific Gaming. Allyn and Bacon Inc., Boston (1969)
42. Rid, T., Buchanan, B.: Attributing cyber attacks. J. Strateg. Stud. **38**(1–2), 4–37 (2015)
43. Romney, G.W., Higby, C., Stevenson, B.R., Blackham, N.: A teaching prototype for educating IT security engineers in emerging environments. In: Proceedings of the Fifth IEEE International Conference on Information Technology Based Higher Education and Training, Istanbul, Turkey, pp. 662–667, May-Jun 2004
44. Salas, E., Prince, C., Baker, D.P., Shrestha, L.: Situation awareness in team performace: implications for measurement and training. Hum. Factors **37**(1), 123–136 (1995)
45. Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., Young, M.S.: What really is going on? Review of situation awareness models for individuals and teams. Theor. Issues Ergon. Sci. **9**(4), 297–323 (2008)
46. Salmon, P.M., Stanton, N.A., Walker, G.H., Green, D.: Situation awareness measurement: a review of applicability for C4i environments. Appl. Ergon. **37**(2), 225–238 (2006)
47. Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L., Young, M.: Measuring situation awareness in complex systems: comparison of measures study. Int. J. Ind. Ergon. **39**(3), 490–500 (2009)
48. Sarter, N.B., Woods, D.D.: Situation awareness: a critical but ill-defined phenomenon. Int. J. Aviat. Psychol. **1**(1), 45–57 (1991)
49. Schlenker, B.R., Bonoma, T.V.: Fun and games: the validity of games for the study of conflict. J. Conflict Resolut. **22**(1), 7–38 (1978)
50. Smith, K., Hancock, P.A.: Situation awareness is adaptive, externally-directed consciousness. In: Gilson, R.D., Garland, D.J., Koonce, J.M. (eds.) Situational Awareness in Complex Systems. Aviation Human Factors Series, pp. 59–68. Embry-Riddle Aeronautical University Press, Daytona Beach, Florida (1994)
51. Sommestad, T., Hallberg, J.: Cyber security exercises and competitions as a platform for cyber security experiments. In: Jøsang, A., Carlsson, B. (eds.) NordSec 2012. LNCS, vol. 7617, pp. 47–60. Springer, Heidelberg (2012)
52. Stanton, N.A., Chambers, P.R.G., Piggott, J.: Situational awareness and safety. Saf. Sci. **39**(3), 189–204 (2001)
53. Stevens, S.S.: Measurement, statistics, and the schemapiric view. Science **161**(3844), 849–856 (1968)

54. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: AGARD Conference Proceedings No. 178: Situational Awareness in Aerospace Operations, pp. 3/1–17, April 1990
55. U.S. Department of Defense: Cyberspace operations. Joint Publication 3–12(R), Joint Chiefs of Staff, Washington, District of Columbia, February 2013
56. U.S. Department of Defense: Cyber electromagnetic activities. Field Manual 3–38, Headquarters, Department of the Army, Washington, District of Columbia, February 2014
57. Vidulich, M.A., Hughes, E.R.: Testing a subjective metric of situation awareness. In: Proceedings of the Human Factors Society 35th Annual Meeting, San Francisco, California, pp. 1307–1311, September 1991
58. Waag, W.L., Houck, M.R.: Tools for assessing situational awareness in an operational fighter environment. Aviat. Space Environ. Med. **65**(5), A13–A19 (1994)