24th International Command and Control Research & Technology Symposium

Topic 9: Experimentation, Analysis, Assessment and Metrics

Data Collection and Research in CDXs

- Command and Control, Cyber Situational Awareness and Intelligence Perspectives on Cyber Defense

Magdalena Granåsen[a], Gazmend Huskaj[c,e], Stefan Varga[b,d]

[a] FOI Swedish Defence Research Agency, SE-164 90 Stockholm, Sweden
[b] KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
[c] Swedish Defence University, Box 278 05, SE-115 93, Stockholm, Sweden
[d] Swedish Armed Forces Headquarters, SE-107 85 Stockholm, Sweden
[e] University of Skövde, Box 408 05, SE-541 28, Skövde, Sweden

Abstract

The annual cyber defense exercise Locked Shields is the world's largest unclassified defensive exercise. The exercise participants form "blue teams" that are tasked to defend their critical infrastructure against an attacking "red team." The blue teams are scored based on how well they keep their essential system functions running and the extent to which they manage to assess and report what they are exposed to. During Locked Shields 2019, 24 blue teams from 30 countries participated in a two-day exercise. The case study presented in this paper focuses on one of the blue teams. The team consisted of around 60 people from governmental institutions as well as private companies. The objective of this paper is to explore the possibilities to collect meaningful data for research on Command and Control, C^2, Cyber Situational Awareness, CSA, and Intelligence in conjunction with an inter-organizational cyber defense team during a cyber defense exercise. During preparations preceding the exercise, the research team observed the development of strategy, coordination structures and organization in the temporarily formed team as it prepared to solve the highly challenging exercise tasks. During the exercise, data collection included questionnaires, observations, team communication logs, reporting from the blue to the white team and performance scores. The data collection sought to satisfy needs within three research themes - 1) command and control, C^2, 2) cyber situational awareness, and 3) intelligence. A review of the dataset showed that the data is well suited for further analysis. The paper presents initial results as well as an outline of how the different types of data collected contribute to research within the three research themes.

1. Introduction

The complexity of the cyber domain and the society's increasing dependency on information and communication technology place extensive demands on cyber defense teams [56]. While the technical skills and output of cyber defense teams are in focus for a lot of cyber research, it is important to embrace that the cyber defense teams also operate in a highly socio-technical context, that requires skills such as decision making, critical thinking, communication and the ability to make sense of the situation [21].

There are several challenges when researchers want to examine cyber defense teams. First, it may prove difficult to access fully manned and operational teams in their original environments, due to security concerns, etc. Second, the nature of cyber security work is such that it is not always necessary for the teams to engage in interpersonal communication to be successful [8]. Third, it may be extremely difficult for external observers to capture what is actually going on, due to the speedy execution of work processes and complex nature of the work, particularly when combined with the absence of externally observable events [19]. Finally, much of the work, e.g. reasoning and decision making, can be expected to be carried out by individual team members solely in the cognitive domain, which in itself is not directly observable [9].

Cyber defense exercises, CDXs, have proved to be useful for the training of cyber security personnel. The main purpose, arguably, for CDXs are (to provide) "interactive learning opportunities in realistic scenarios" [29]. Even if the main goal of a CDX is to train technical aspects, such exercises might also enable training for other work areas, such as legal, ethical, and forensic work [22]. There are a number of advantages with CDXs. Foremost, they provide a controlled environment where it is possible to observe attacks in some detail, but also to positively attribute them to attackers. CDXs further permit training of soft skills, such as the improvement of teamwork and other interpersonal skills [29]. CDXs can also produce scientifically valuable data. One example is labeled datasets [40, 45] that reflect network activities. While fully artificially manufactured data are conceivable, it has been shown that they still may have flaws [28]. Data collected from CDXs is at least created by real people who perform attacks, even if some of them are carried out using simple standard tools [40].

1.1. Objective and scope

The objective of this paper is to explore the possibilities to collect meaningful data for research on Command and Control (C^2) Cyber Situational Awareness (CSA), and Intelligence in conjunction with an inter-organizational cyber defense team during a cyber defense exercise.

Data was collected from the Locked Shields 2019 (LS19) with focus on the Swedish Blue Team. The collected data aimed to address the three research themes mentioned above. The present paper primarily focuses on the data collection and discusses to what extent the data is valid for research.

1.2. The Cyber Domain

In military settings, the cyber domain constitutes the fifth domain alongside land, air, sea, and space. This fifth domain is the only one that is fully man-made. It consists of hardware and software that are connected in networks through information and communications technologies. Six characteristics makes it unique compared to other traditional domains. The first is *speed*; latency, "the time it takes to get a response to information sent" [43], is 50 ms in 4G [25]. In the emerging 5G telecommunications network, latency is expected to be 1 ms or less [37], and in a future 6G, latency is speculatively expected to diminish even further, to 0.1 ms [50]. The second is *depth*; in 1921, Douhet stated that modern air power could hit a target deep behind enemy lines some distance away [13]. Cyber attacks can hit a target in an office, living room, and even a smartphone anywhere on earth [1]. The third is *obstacles*; obstacles in the cyber domain are generally man-made and consist of security controls such as authentication, firewalls, and logical and physical separation. Bypassing these requires exploiting flaws in the security

controls, or physically bypassing logical and physical separation. The fourth is *weapons*; cyber-weapons are computer programs, programmed by developers for the purpose of exploiting one or several vulnerabilities in a target. These programs, or capabilities, have to be tested before they are deployed. Their development requires a lot of preparation time and operations security, OPSEC, due to the risk of vulnerabilities being patched. Furthermore, cyber-weapons differ from traditional weapons due to the risk of cascading effects, making it difficult to assess the risk of spreading unwanted damage beyond the intended target. The fifth is *proxies*; the cyber domain is global and attackers can hide their tracks by exploiting several weakly protected information systems to make it appear as if an attack is coming from a different country than their own. This makes attribution difficult, but not impossible to make. The sixth challenge is *confidence-, and security-building measures* (CSBM). Countries "exchange information on their armed forces, military organization, manpower and major weapon and equipment systems [that ensure] military stability, predictability and transparency" [35], [33]. The challenge with confidence-building measures in the cyber domain is that they are voluntary [34]. These six characteristics make the cyber domain unique and complex. Given the characteristics above, there are many challenges for conducting successful operations in the domain, but there are also plenty of opportunities.

1.3. Cyber Defense Exercises

CDXs are typically constructed to develop defensive skills, while offensive (attack) skills are practiced in capture-the-flag style exercises. There are different types of CDXs with regard to scope, time frames and layout [22]. A CDX that aims to train technical personnel typically involves computers that are interconnected and run a number of services that are to be secured and remain available to legitimate users. Such an infrastructure, sometimes referred to as a *cyber range*, may be set up with either virtual computers and network equipment, dedicated hardware, or both. Medium-, and large-scale exercises often connect participants at different geographical locations through secure virtual private networks.

Exercise participants are typically divided into teams that, by convention, are designated by different colors. Normally, at least one blue team is assigned to defend resources, a red team is tasked to attack, degrade or destroy the resources, while a white team is the exercise management, which may include observers and judges [46, 26, 53]. There may be other teams with additional colours as well.

However useful a CDX may be, there are also a number of drawbacks. The training environment is still an artificial milieu with a disproportionate amount of malicious activities compared to the everyday situation [40]. The duration of a CDX is limited, which in turn increases the workload per time unit of the participants. Furthermore, the absence of background internet noise [40] or internet background radiation [36], such as scans, diverse automated attacks, malformed packets, flooding backscatter, etc., may contribute to distorting both the psychological reality of the exercise and the validity when exercise data is used to, e.g., tune intrusion detection systems [44]. Finally, psychological factors such as workload may be hard to recreate in exercises. However, the disproportionate amount of malicious activities in conjunction with a competitive element created by visualizing the performance score contributes to raising stress levels [41].

To summarize, it appears that CDXs are suitable for the training of cyber security specialists, but also for the collection of data sets that can be used for research.

1.4. Locked Shields

The annual CDX Locked Shields is the world's largest unclassified defensive CDX. It has been conducted since 2010 and is hosted by the NATO Cooperative Cyber Defence Centre of Excellence, CCDCOE [2].

The training audience consists of several blue teams that are tasked to defend an assigned critical infrastructure against an attacking red team, which in this exercise is a part of the exercise control. The blue teams are scored based on how well they keep their essential system

Figure 1: Swedish Blue Team participants.

functions running and how well they assess and report what they are exposed to. The exercise has continuously grown more popular, and thus expanded to include more and larger teams. The scenario has matured and the technical infrastructure has been developed over the years, which now makes the CDX quite complex. There are no formal restrictions on team size, and each team is responsible for staffing the team with relevant competencies as they see fit. The live phase of the exercise lasts for two days. There are several scheduled preparatory activities before the exercise.

The scenario includes a fictious island nation, Berylia. Berylia has traditionally experienced political and military tensions with the neighbouring island state Crimsonia. Crimsonia possesses advanced offensive cyber capabilities (the red team). Berylia's government has asked for support from other nations. The blue teams enter the picture as rapid reaction teams that have agreed to support Berylia in restoring their vital information systems. The roles of the Berylian government are played by members of a white team. The basic scenario has remained the same over the years, even if injects, technical infrastructure and parts of the scenario have evolved in order to create new challenges for the participants.

During LS19, 24 blue teams from 30 countries participated in the two-day exercise. During the two months before the exercise, CCDCOE distributed the scenario, held webinars and provided textual instructions to the blue teams, mainly through a wiki page. On two occasions - one week before the exercise and the day before the exercise - the blue teams gained limited access to the exercise environment, the *gamenet*, in order to familiarize themselves with the infrastructure they were supposed to protect. Based on the overall timeline, each team was allowed to structure their own preparatory work. The composition and preparation phase of the Swedish blue team, which was the subject of interest for the data collection, is further described in section 3.

The remaining parts of the paper is structured as follows: Section 2, Background, provides an overview of various research fields that use data collected during CDXs. The research fields of interest are presented, as well as different data collection methods that are used in CDXs. Section 3, Method, describes the method used for the current study. Section 4, Results, details the review of the collected data, and some initial results are presented. These are followed by a discussion in section 5. The paper concludes with a summary and conclusions in section 6, and some notes on planned future work in section 7.

2. Background

This section presents research activities that are conducted in conjunction with CDXs in general as well as our own research fields of interest. The section is concluded by an overview of data collection methods used in CDXs.

2.1. Research during Cyber Defense Exercises

In order to acquire an overview of research conducted at CDXs, a search was performed in the scientific publication database Scopus[1]. The enquiry yielded 140 articles, of which 99 were deemed to be relevant. The results were manually clustered into ten categories, that are listed in Table 1.

Table 1: The characteristics of the 99 articles.

Category	Amount
Tools	35
Education	16
Training	13
General	12
Cognition	5
Human Factors	5
Techniques	5
Performance	3
Network	3
Simulation	2

The papers in the *tools* category discussed tools, testbeds, emulation, frameworks, how red teams conduct attacks, and monitoring of information flows.

The *education* category presented papers of lessons learned form various educational efforts, curricula, course designs, etc. The *training* and hands-on-exercises category involves papers that discussed the improvement of sense-making, understanding, and remote code execution, but also battle space situational awareness.

Next, the *a) cognition, b) human factors*, and *c) performance* categories discussed a) cognitive agility and evaluation of human performance; b) the importance of understanding how adversaries adapt at various attack trajectories for anticipatory defensive measures; and c) how the performance of human teams in CDXs is affected based on whether systems and procedures for data collection exist.

Finally, the publications in the *techniques* category described how techniques that facilitate developing attacker profiles enable effective countermeasures and could potentially enhance situational awareness.

The overview leads to the conclusion that virtually no one has examined C^2 issues. Only a few papers mentioned intelligence aspects, but several papers concerned cyber situational awareness.

[1]Search term: cyber* AND defence AND exercise*

Our research fields of interest are command and control (C^2), Cyber Situational Awareness and Intelligence. These three research themes are worthwhile to study as it is hypothesized that they most certainly affect the effectiveness of cyber defense teams. The fields are inter-related insofar that timely intelligence provide operators' with increased cyber situational awareness, which in turn facilitates command and control that is fit for purpose [54].

2.2. Command and Control

Organized human activity encompasses the division of labour and the achievement of coordination of these tasks [32]. In the military domain, this is generally denoted as command and control, C^2, which is defined as "the exercise of authority and direction by a properly designated commander over assigned forces in the accomplishment of the mission" [12]. There have been attempts within military research to replace C^2 with terms less associated with traditional hierarchical approaches, such as focus and convergence [3], and direction and coordination [6].

In other domains, such as business and crisis response, management or administration are more commonly used terms. Crisis management is characterized by pragmatic decision making under time and resource constraints, as well as increased needs for coordination and reorganization [52]. Inter-organizational crisis management is to a large extent based on emergent networks of actors - *adhocracies* [30]. Mintzberg defined adhocracy as one of five different archetypical structural configurations that organizations can adopt[32]. Each of these structures is associated with certain coordinating mechanisms, type of centralization/decentralization, and focuses on different key parts of the organization. Adhocracy is the least structured configuration of the five. An adhocracy is loosely structured and coordinated through mutual adjustment, allowing for creativity and initiative, efficient use of resources and rapid adaptation to changes in the environment [27].

The CDX environment is an environment that typically demands creativity and initiative, efficient use of resources and rapid adaptation to changes. CDX teams solve highly complex problems under time pressure and are only partly familiar with the infrastructure they are to protect. The cyber defense team studied for this paper was a temporarily formed team with no previously established coordination structure. Thus, it may be categorized as an adhocracy. It seems that this type of organization may be suitable for the CDX environment. This remains to be investigated, and will consequently be a topic for further research within the C^2 research theme connected to this case study.

Mission command is another concept of interest to explore for the type of work conducted in a blue team during CDX. Mission command denotes decentralized leadership and requires and facilitates initiative on all levels of command [42]. It thus takes advantage of each individual's decision-making capability. Initiatives are encouraged down to the individual level. Applying mission command requires trust, intent focus, initiative, common ground, and risk acceptance [16], [20]. In order to investigate C^2 from a mission command perspective, it is thus relevant to investigate whether the prerequisites for applying mission command are in place.

As stated in section 2.1, it appears that research on C^2/management in relation to cyber defence teams is scarce. The authors of the few related papers that were found share this view [8]. Buchler et al. (2018) used an observational protocol assessing certain aspects of teamwork and leadership. They found that adopting different collaborative and leadership approaches and varying face-to-face interactions according to the nature of task at hand was beneficial to team performance. Using observations in cyber defense exercises is not trivial as communication may occur through technical systems and thus be difficult to observe [19]. A combination of data collection methods may be an option, employing a pluralist perspective in which various types of data may contribute in different ways [31].

2.3. Cyber Situational Awareness

Situational awareness, SA, is a concept concerning an individual's ability to appreciate a situation. There are several theoretical constructs that seek to frame the concept [38]. A widely

used model is Mika Endsley's three-tier model that describes ascending levels of understanding; level (1) comprises the ability to perceive elements in the environment within a volume of time and space, level (2) adheres to the comprehension of their meaning, and (3) the capability to project their trajectory or status in the immediate future [14]. Other theories seek to explain SA for teams. Shared SA concerns the degree to which team members possess similar SA on shared requirements [15]. Furthermore, Distributed Situational Awareness (DSA), challenges and expands the original and shared SA concepts, using the socio-technical system as the unit of analysis rather than the individual mind, thus assuming that artefacts as well as humans may possess SA [49]. A concept related to SA is the Common Operational Picture (COP). According to Wolbers and Boersma [55] there are two distinct prevalent perspectives on COPs in the literature. First, that it can be seen as an artefact that accumulates and conveys allegedly useful pieces of information, e.g. the "information warehouse" [11]. Second, it can also be seen as a process, in which the meaning of the information for various users is actively negotiated in social interactions, e.g., the "trading zone" [55]. When teams rather than individuals are examined, the "trading zone" perspective with regards to SA becomes important. Cyber situational awareness (CSA) which is a special case of general SA, is a heterogeneous academic field [17] of which several different aspects are studied in the literature. CSA is about having situational awareness of different aspects of the cyber domain.

Barford et al. [5] suggested that "Situation Awareness (SA) for cyber defense consists of at least seven aspects" (pp. 3-4), with requirements that need to be fulfilled:

1. awareness of the current situation (which may include network security and the wider cyber influence),
2. awareness of the impact of the attack,
3. awareness of how situations evolve,
4. awareness of adversary behavior,
5. awareness of why and how the current situation is caused,
6. awareness of the quality and trustworthiness of the situational awareness information, and
7. assessment of plausible futures of the current situation.

All three levels of Endsley's three-tier model are accounted for in the seven SA requirements.

A computer-based CDX provides a platform to perform CSA measurements, e.g., SAGAT, SART, QUASA (For a review of SA measurement techniques see Stanton et al. [39]), and opportunities to observe all relevant aspects of participant performance related to CSA. Furthermore, Brynielsson et al. [7] have proposed different ways of constructing CDXs with specific scenarios in which CSA can be measured. In sum, CDXs provide excellent opportunities for scientists to examine how individuals and teams acquire CSA.

2.4. Intelligence in Support of Cyber Operations

The purpose of intelligence is to provide decision-makers with information collected, processed and exploited from numerous sources. This can be information about the threat actor capabilities, intent and opportunity, but also about one's own information systems that are critical for supporting the mission. In other words, intelligence in cyberspace operations consists of two types; intelligence for supporting cyberspace operations, and intelligence for cyberspace operations to support other domains [23]. Therefore, as in the other domains, the first step is to generate an understanding of the operational environment, which in this case is cyberspace. It consists of three layers: the physical network layer, the logical network layer and the cyber-persona layer [48].

Intelligence preparation of the operational environment, IPOE, is a methodology that considers all "conditions, circumstances, and influence that affect the employment of capabilities and bear on the decisions of the commander" [47, p. I-1]. All conditions and circumstances

that influence the employment of capabilities include the threat's capabilities, intent and opportunity, and strengths and weaknesses of own information systems supporting the mission. The supporting intelligence theory for this research is based on Gill's (2009) definition as "targeting, collection, analysis, dissemination and action - as collectively constituting an intelligence 'cycle' or 'process'" (p.219).

According to Joint Chiefs of Staff (2018), the *physical network layer* is comprised of the physical hardware, or information technology, that enables network communication. The *logical network layer* consists of the different protocols, such as TCP/IP and DNS, that enable information to be transferred from one place to another. Finally, the *cyber-persona layer* consists of the virtualized self of physical people in the form of user accounts, whether they be for e-mailing, administrative purposes, or social media purposes. Therefore, a target in cyberspace could be the physical location of an information system, a virtual machine acting as a virtual information system, and a particular administrator managing either of the two. Each of these three layers would have to be considered as part of the operational environment, including how the adversary exploits them.

From a defense perspective, the mission of a defensive operation (DCO) is different. Williams (2014) noted that defensive cyberspace operations "provide the ability to discover, detect, analyze, and mitigate threats, to include insider threats" ([54] p.15). This has implications on the intelligence cycle, because instead of identifying adversarial targets the first step in a DCO would be to identify which information systems are key for a successful mission. The next step would be to identify the vulnerabilities or software flaws, in those information systems. Finally, the intent and capabilities of the threat need to be taken into consideration [54].

Therefore, the intelligence cycle for a DCO would be to identify key information systems, their vulnerabilities, as well as the threats, followed by analysis, dissemination and action. In LS19, the Swedish blue team already knew their key information systems. The team was also aware that they would come under attack by a threat actor known as the red team. However, the team was unaware of all the vulnerabilities in their information systems and the type of capabilities that the opposing red team had at their disposal. The role of intelligence would then be to support the team with actions to patch vulnerabilities, manage misconfigurations, and to discover, detect, analyze and mitigate the threats. These are the focus areas of the data collection to assess how intelligence is generated to support the aforementioned actions, a field that has been neglected academically ([24]; [51] as cited in [4]).

2.5. Data Collection Methods in Cyber Defense Exercises

As mentioned in the Introduction (Section 1), one of the main advantages of performing training activities in CDXs, is the excellent opportunities to collect data. Twelve papers from the review of existing literature (Section 2.1) specifically mentioned data collection methods in CDXs. The paper titles along with their data collection methods are displayed in Table 2.

The table shows that the preferred data collection method is network traffic (papers 3, 5, 6, 10, 11), followed by observations (papers 1, 4, 7, 10). Collection methods that are referred to in only one paper include data from the red team (paper 10); from injects (paper 10); interviews (paper 10), network vulnerabilities (paper 6), observer logs (paper 6), red team reporting (paper 10), self-assessment questions (paper 7), self-reporting application (paper 8), wearable social sensor (paper 1), and yellow team reporting (paper 10). It should be noted, however, that the naming conventions for the different data sources vary between the papers, meaning that one data collection method may have different names in the table.

2.6. Data Collection during CDXs

This subsection lists available means of data collection. Collection can be carried out prior to the live phase of a CDX, as well as during and after it. Before the CDX, planned scenarios, scoring system construction and other contextual factors are of interest. It is also possible to

Table 2: Data collection methods in CDXs.

Authors	1	2	3	4	5	6	7	8	9	10	11	12
Year	2017	2003	2013	2015	2016	2012	2011	2018	2016	2017	2011	2012
Activity logs				x	x							
Attacker logs				x		x						
audio			x								x	
E-mail			x								x	
Event logs				x	x							
IDS-logs		x		x	x							
Info from RT										x		
Injects										x		
Interviews										x		
Machine logs			x	x							x	
Network traffic			x		x	x				x	x	
Network vulns						x						
Observations	x			x			x			x		
Observer logs						x						
Post-event survey				x	x				x			
Pre-event survey				x	x				x			
RT Reporting										x		
SA			x								x	
Scores			x							x	x	
Self-assessment Questionnaire							x					
Self—reporting Application								x				
Survey				x						x		
Team comms					x					x		
User feedback									x			x
Video			x								x	
Wearable Social Sensor	x											
YT reporting										x		
ToT Data collection methods	2	1	7	9	7	4	2	1	3	10	7	1

Note: The papers are:
1. Buchler N., Rajivan P., Marusich L.R., Lightner L., Gonzalez C. (2017) "Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition"
2. Dodge Jr. R.C., Wilson T. (2003) "Network traffic analysis from the cyber defense exercise"
3. Fink G., Best D., Manz D., Popovsky V., Endicott-Popovsky B. (2013) "Gamification for measuring cyber security situational awareness" collected data in the same manner as Malviya et al. (2011).
4. Granåsen M., Andersson D. (2015) "Measuring team effectiveness in cyber-defense exercises: a cross-disciplinary case study"
5. Henshel D.S., Deckard G.M., Lufkin B., Buchler N., Hoffman B., Rajivan P., Collman S. (2016) "Predicting proficiency in cyber defense team exercises"
6. Holm H., Ekstedt M., Andersson D. (2012) "Empirical analysis of system-level vulnerability metrics through actual attacks"
7. Holm H., Sommestad T., Franke U., Ekstedt M. (2011) "Expert assessment on the probability of successful remote code execution attacks"
8. Jøsok Ø., Hedberg M., Knox B.J., Helkala K., Sütterlin S., Lugo R.G. (2018) "Development and application of the hybrid space app for measuring cognitive focus in hybrid contexts"
9. Li Y., Xie M. (2016) "Platoon: A virtual platform for team-oriented cybersecurity training and exercises"
10. Maennel K., Ottis R., Maennel O. (2017) "Improving and measuring learning effectiveness at cyber defense exercises"
11. Malviya A., Fink G.A., Sego L., Endicott-Popovsky B. (2011) "Situational awareness as a measure of performance in cyber security collaborative work"
12. Mullins B.E. (2012) "Developing cyber warriors from computer engineers et al"

follow the formation of a functioning team. Sometimes the participating teams are allowed to collect information about the training environment, e.g. the networks to be defended, prior to the exercise. During the live phase there is an opportunity to monitor participant activities, including the deliverables that they produce. After the live phase, after-action reviews, so called lessons learned activities and other processes are often of interest.

In order to extract users' inherent knowledge, for example their intentions, a plethora of techniques, such as *questionnaires*, *surveys, interviews*, etc. are feasible. **Knowledge elicitation** may contribute with, for example, *demographic background data, perceived levels of stress*, etc. For extensive details of knowledge elicitation techniques, see Cooke (1994) [10]. Data derived from individuals include *self-assessments*.

Performance metrics include two main types. The users produce direct tangible products, e.g. various forms of reports as mandated by the upper echelons of the in-game leadership (such as *threat-, situation-, and adversary reports*), but they also leave a trail of indirect traces of their behavior, such as different types of *artefacts or logs* (IP-traffic at the packet level, Netflow-data, intrusion detection/prevention systems etc.). A third type of performance metric that emanates from user performance are the *scores* that are calculated or awarded by exercise control.

A third data type involves data created due to **user actions**. Users may engage in teamwork/collaborational activities, strategy/tactics discussions and other meetings. Questions related to leadership and C^2 can also be investigated. User actions can be captured with *video* where video cameras can be used to record interesting sequences. *Audio* can be captured with microphones at the premises of the exercise, or by recording phone conversations (including VoIP traffic). Other types of activities, such as the *use of equipment*, can be caught by the logging of, e.g., 1) *screen captures*, 2) *keyloggers*, 3) *mouse movements*, and 4) *running programs/processes*

in the computers.

A fourth data type is **user communication**. Here, *oral communication*, as well as other manifestations of communication can be captured, e.g., *chat-logs, conversations in forums and e-mail*. In the physical domain, *images of whiteboards* with scribbled information as well as *paper (post-it) notes* may contain useful information.

An important data collection method is *observations* by observers. Information that cannot be recorded or documented in any other way can be collected by observers who document and log events. Examples include user behavior in meetings and face-to-face conversations (see oral communication above). Observers can also take complementary notes on incidents, performance, teamwork/collaboration, strategy/tactics and leadership/communications, etc.

Finally, it is valuable to collect the **reporting from other teams** than the one that is subject to the study, as is information submitted by the exercise control, e.g., injects and other types of information.

3. Method

The current section describes the participants and the data collected for the case study of the Swedish blue team during LS19. In order to answer research questions within the three research themes of interest in this exercise, a number of data collection methods were employed. These are described in the current section.

3.1. Participants

The Swedish team was composed of around 60 cyber security experts who participated for the duration of the exercise. A majority of these also participated during the preparatory meetings. The number of participants is somewhat inexact due to last-minute recruitments as well as last-minute drop-outs of anticipated participants. At an initial meeting six months before the exercise, the team leader was appointed, and a handful of organizations from the security sector agreed to participate, forming the primary basis for recruitment of the team members. For anonymity reasons, these organizations are not explicitly described in this paper. The team further contained personnel outside the core organizations who were recruited based on personal knowledge. There was no central funding of participants, meaning that it was up to each organization to enable exercise participation for their personnel as well as in the preparatory activities. Several participants reported that they had voluntarily spent considerable time outside normal working hours for preparations in addition to the scheduled activities during the last weeks before the exercise. These efforts were based on personal motivation and ambition to maximize performance during the exercise.

Participant background information was mainly collected from the pre-exercise survey, which had a response rate of 88%. Among the team members, 70% had their ordinary employment within governmental institutions, while around 30% worked in private companies. Mean age was 39 years. Mean time at current workplace was 1.9 years, however, the total experience of working with cyber security-related issues was 4.7 years (ranging between 0-22 years).

The main part of the team participated in the preparations as well as in the two-day exercise. During the preparation days, preliminary subteams were formed. They were structured according to competencies, e.g. a Windows team, a Linux team, an Industrial Control Systems team, and a team responsible for reporting.

3.2. Data Collection during Locked Shields 19

The data collection served to gather all relevant data associated with the CDX that could be of use for investigating the research areas. Various information channels were used to capture the data. Unless otherwise specified, the data collection concerned only the Swedish blue team.

Observations were collected by the research team during the game days as well as during the blue team's preparatory meetings. During five preparation days preceding the exercise, the

research team observed the development of strategy, coordination structures and organization in the temporarily formed team, as they prepared for the solving of the highly challenging task. During the game, the observers/research team attended the main meetings and observed the work in the team. The observers had access to the the Swedish blue team's wiki and chat channels, and could thus keep track of within-team conversation, score and reporting during the game. The observers further attended the hot washup.

A pre-exercise survey handed out to all participants in the Swedish blue team on the first day of the exercise collected information about the participants' background, motivation and expectations for the exercise.*A post-exercise survey* to all participants in the Swedish blue team during the last day of the exercise collected information about the participants' perception of whether the team was composed of sufficient competencies, team performance, within-team collaboration, situational awareness, strategy and learning.

The performance score was designed and administered by the central white team. During the exercise, the scoreboard was displayed in real-time to all teams on the wiki page. This meant that the teams could keep track of their own and other teams' score at all times. The research team collected the score by taking a screenshot of the scoreboard approximately every hour. This means that the score was collected for all teams.

The score was a composite measure. The scoring components were 1) Attack - A negative score based on successful red team attacks on the blue team's infrastructure. 2) Availability - Service uptime, which was automatically measured by scoring agents, 3) Usability - a measure of the extent to which certain services were available to users. A user simulation team (part of the white team) manually checked access to services and complained to the blue teams when services were not available. The score dropped dramatically if the blue teams refused to open the service. 4) Forensics - Score based on a specific forensic challenge, 5) Injects - tasks assigned to the blue teams during the game, divided into legal, media, or scenario themes. These were communicated via e-mail, 6) Adversary assessment - manual assessment of the extent to which the teams managed to compile the adversary assessment report according to the instructions by the white team, as well as the quality of the content, 7) Threat and situation reports - manual assessment of the extent to which the teams managed to compile reports according to the instructions by the white team, as well as the quality of the content 8) System revert - a negative score was applied if the blue team damaged part of their infrastructure in a way so that the white team needed to revert it, 9) Special score - supplementary adjustment of scores based on mistakes in the initial scoring. The blue teams could make complaints if they found mistakes in the scoring, which was analyzed and corrected by the white team. A log of corrected scores was displayed on the wiki page.

Reports were compiled by the blue teams in accordance with the white team's instructions, which included timings as well as report formats. Most reports adhered to one of three report types: 1)Threat reports - compiled and reported as soon as the team noticed a threat of some dignity to their critical infrastructure, 2) Situation reports - a summary of threat reports and key events, compiled twice a day, 3) Adversary assessments, which were to be compiled at fixed time intervals. In total, 10 adversary reports were published. All threat reports, situation reports and adversary assessments were reported through the blue team's wiki page. They were accessible during as well as after the game. Each report type had a specified format and template. For the threat reports, the white team displayed an overview of the number of threat reports per team, which was updated continuously. The research team took a screenshot of the threat report board on nine occasions during the exercise. This enabled analysis of the progress of the threat reporting. In addition to the three aforementioned reports, the blue teams received additional tasks during the exercise, which were mainly reported through e-mail, such as legal assessments. The e-mail communication between the blue team and the white team was exported after game stop.

Within- and between-team communication, which consisted of chat and e-mail communication,

was exported immediately after the exercise. The Swedish blue team communicated internally mainly through a chat-tool, using 22 different chat channels. Sub-teams reported discovered threats to the reporting sub-team, which then produced and published the threat reports. Based on input from the research team, the within-team communication for threat reports used the threat report including a template relating to Endsley's three levels of situational awareness. Communication between the blue teams and the white team occurred mainly through e-mail.

Availability of services was displayed to the blue teams on their wiki pages in real-time. The Swedish blue team denominated this screen their operational picture (Figure 2). This information was not collected on the first exercise day. However, on the second day, a script was utilized, which produced a screenshot of the availability of services board of the Swedish blue team every minute.

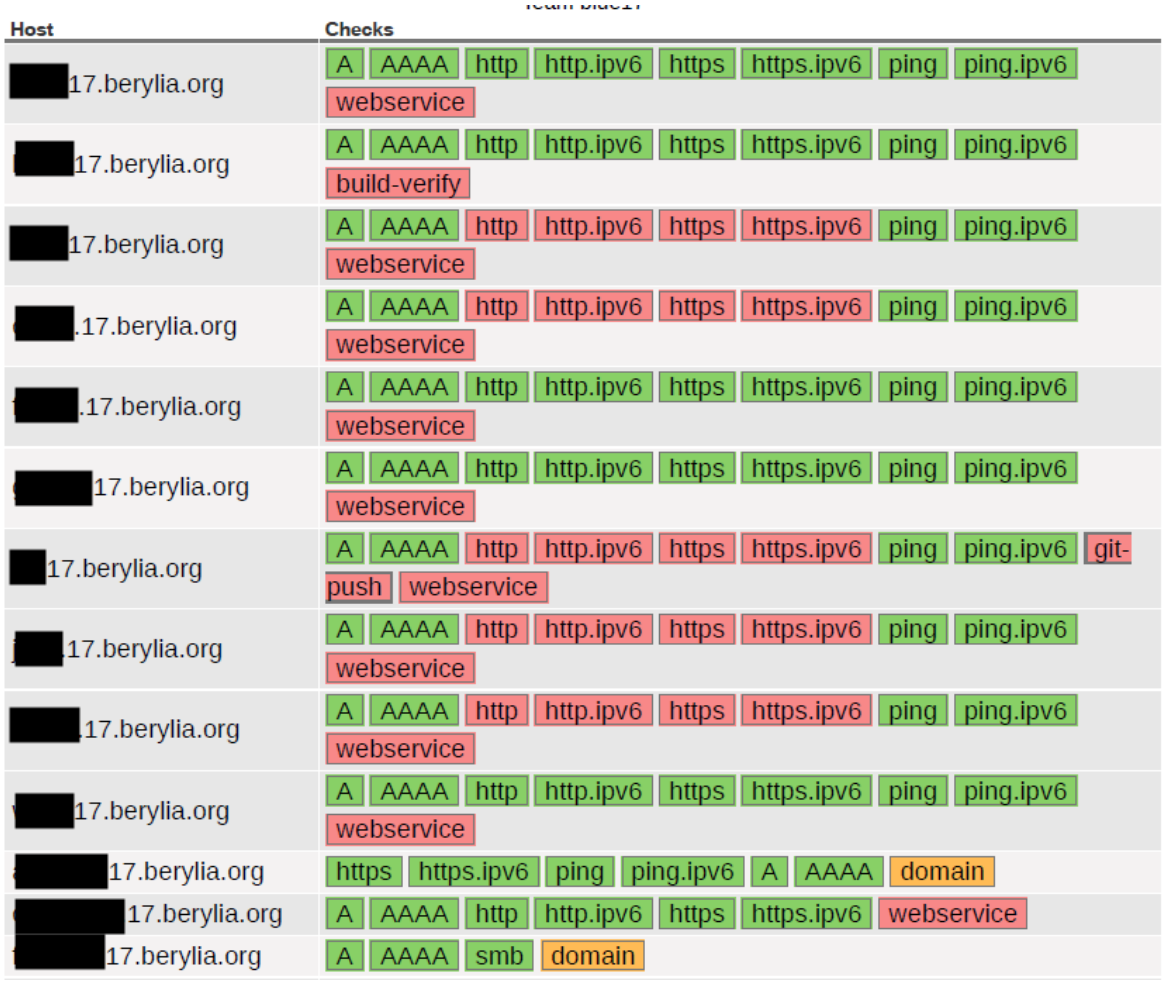

Figure 2: Availability of services for the Swedish team at one point of time during LS19.

4. Results

As this paper mainly has a descriptive and methodological focus, the results and discussion sections primarily aims to respond to the quality of the data collected, and how it will be used for the thorough analyzes relating to the research themes. Still, some initial analyses have been conducted, why initial results are presented here.

The response rate for the surveys was high, bearing in mind the fact that a few of the 60 participants in the Swedish blue team did not attend for the full duration of the exercise and were thus not available for survey completion. The pre-exercise survey was completed by 53

participants (88%) and 51 (85%) responded to the post-exercise survey. The survey respondents decided on their own identification code which they used for both surveys. In this way, the pre- and post-exercise surveys could be paired afterwards without violating participant anonymity. After pairing the surveys, it was found that 49 participants (82%) responded to both surveys, while six (10%) responded to only one of the surveys. Only 8% did not respond to any survey. There were no major differences in response rate between the sub-teams. Two types of data loss were found: The six cases where participants only responded to one of the two surveys, and cases where participants had skipped specific questions in otherwise completed surveys. Additional preparation of the survey data is necessary in order to handle the data loss. To allow comparisons when there is data loss, missing data may be replaced by the mean of the whole team for the particular question, or the mean of the sub-team. The consequences needs to be analyzed, why for this paper, no comparisons between questions were conducted.

The surveys provided a general overview of the participants' perception of their participation in the exercise. *Initial results* from the *pre-exercise survey* showed that motivation among participants was very high (M=4.5), and that they expected to learn a lot from the event (M=4.7).

Regarding assessment of preparations of technology (T), strategy (S), organization (O) and ways of working (W), the participants perceived that they were neither prepared nor unprepared for these issues (M(T)=3.24, M(S)=2.9, M(O)=3.1 and M(W)=3.3). Very similar results were obtained when they assessed the preparations in their own sub-team.

As for expected position in the competition, the majority of the participants guessed that they would end up somewhere between the 10^{th} and 15^{th} place out of the 24 participating teams. The mean was 9^{th} place. The mean result for expected position was somewhat biased, as handful of the participants submitted their pre-exercise survey after the exercise, thus already being aware of the team's actual result - third place - and that at least two of these put the final position as their expectation in the survey. This means that reaching the third place in the competition was quite unexpected to most participants, which was reflected in a somewhat joyful celebration at the Swedish National Defence University pub after the en of the exercise.

In the *post-exercise survey*, the participants rated the overall team performance as very high (M=4.7). They further rated the difficulty (D) and workload (WL) as high (M(D)=4.3, M(WL)=3.9). Collaboration and information sharing seems to have worked sufficiently (M=3.34), and the collaboration within the sub-team was rated as exceptionally good (M=4.5).

The surveys will be further analyzed within the research themes C^2 (questions about strategy, information sharing and collaboration), and situational awareness (questions about situational awareness and access to information).

Observations during the preparations provided an understanding of how the Swedish blue team formed sub-teams and appointed sub-team leaders, established coordination structures and formed an initial strategy. During the game, the observers noted changes in structures and strategy, primarily during discussions and meetings between the team- and sub-teamleaders. Times-points for observations were noted. Observations will constitute the primary source for analysis of decision making within the team. As there were only three persons in the research team, the observers could not pin-point specific sub-teams over time. Beside an observational role, the research team designed and distributed the surveys and conducted the required data exports, which at times impaired the ability to observe team activity. It is assessed that the research team managed to note most of the essential decisions made during the meetings in which team- and sub-team leaders met. However, decisions made in the sub-teams that were not explicitly mentioned in meetings remained undetected.

Initial results from the observations in relation to the *C^2 theme* revealed that team coordination in the Swedish blue team occurred mainly in accordance with the adhocratic structure [32], although there was a loose hierarchy in terms of the division into sub-teams and a management team. Furthermore, a mission command philosophy was clearly influencing the way in which the management team approached the subteams. The subteams were given extensive mandates to

work on their issues and coordinate the work with other subteams, while the focus of the small management team was to obtain an overview of the situation and team as a whole. Only on a few occasions did the management team interfere with how the sub-teams chose to prioritize their tasks. During the second day of the exercise, coordination meetings with the sub-teamleaders were conducted every second hour, as it was experienced on the first day that some issues required more work between sub-teams. These coordination meetings were generally kept to less than ten minutes. During the second day, a temporary team was formed focusing on solving a specific problem.

Analysis of various reports will be the primary task for the exploration of the *CSA research theme*. The content of the Swedish team's threat-, and situation reports was collected as well as data about how many threat reports that were produced by each team was retrieved. An initial analysis of the number of threat reports produced in relation to the total score of the team showed no correlation between performance and the number of threat reports produced. Thus, detecting and reporting threats is not a good indicator of performance. One team participant commented on this, stating that the teams that are less effective in protecting their infrastructure may detect and report a larger amount of threats than those who protect their infrastructure well. It cannot be confirmed to what extent the teams reported on the threats that they detected, or to what extent detected threats correlated with actual threats. The second data set of high importance for the CSA track is the communication between the sub-teams and individual participants. Threat reports were submitted internally (within the team) by a chat-channel. The researchers had added two additional research questions to the threat report template. The response rate of this approach was 100%. Overall, the CSA side of the research effort acquired good volumes of high quality data.

Knowing the flows of information and the content of reporting facilitated obtaining an overview of how *intelligence* supported the defensive cyberspace operation, i.e., the blue team's actions. Therefore, observations, e-mails and chat-logs will be used to assess information flows in the blue team. The information flows can indicate how information about red team activities were shared. In addition, information flows can indicate how the blue team's own assets were defended. Furthermore, analysis of reports, particularly adversary reports, enables the exploration of red and blue team activities and their capabilities. In addition, the reports provide information about targets and attack-vectors. However, further investigation is required to identify any potential courses of action to plan defensive actions.

To obtain an understanding of events and processes, the observations need to be analyzed in conjunction with the logged communication (chat, e-mail) and the published threat and situation reports. This will be useful for all three research areas (command and control, situational awareness and intelligence).

Analysis of chat communication will need to be combined with other data sources in order to understand the content in the chat. The chat conversations used a very sparse and technical language, e.g., "rm -rf", meaning that an asset was destroyed. Without the context in which the chat message was written, it is difficult to understand the content. A comment from the sub-team responsible for producing the threat and situation reports was that chat communication worked very well for the communication between the other sub-teams and the reporting team in order to extract information for threat reports. However, it was noted that information that was shared through chat communication but not qualifying for a threat report, such as a vague indication, were in a few cases omitted due to that no one being responsible for taking care of that type of information in the chat channels. Suddenly a small issue became a much more severe problem. Thus, someone had shared information, but no one had payed attention to it. This might be a drawback with chat communication compared to more structured approaches to information sharing.

5. Discussion

The results showed that the data collected for the case study of the Swedish blue team during LS19 is meaningful and of sufficient quality for further analysis within the three research themes. Some issues concerning the quality of data were raised in the results section. These and other issues are further discussed in the current section.

Concerning data quality, it can be concluded that data loss in the collected data will not impair the ability to conduct the analyses needed in order to explore the research themes C^2, CSA and intelligence. Analysis of different data sources in conjunction will be necessary in order to make sense of the data. The language in chat logs was sparse, and although the researcher is acquainted with the terminology used in these logs, it does not make sense until correlated with other information such as observations or published threat reports.

As the exercise was organized by the CCDCOE, the research team only had part control of the data. This is the reason why data on the progress of threat reporting, score and network overview (available services) was collected through screenshots. It will be a time-consuming task to manually transfer these screenshots to a format that can be further analyzed. Still, the research team assessed that a screenshot of the information was better than not having access to this information at all. Using a well-established CDX such as Locked Shields for research is cost-efficient in that the scenario and scenario injects (including a 70-person red team), set-up, gamenet (including support personnel), communication infrastructure, participant invitation, scoring system, exercise control personnel and a lot of the administration already are in place. Furthermore, Locked Shields has gained a reputation that attracts very proficient cyber security experts, while it is likely that these people would not feel equally attracted to participate in a pure research experiment. From a Swedish perspective, one of the greatest benefits of participation in LS19 was the ability to meet and work with people from other organizations.

Research within an international CDX clearly entails lack of experimental control and reduced access to data for the researchers, as well as continuous adjustments to fit the data collection to the exercise schedule. However, it also entails highly motivated, skilled and competitive participants, and a far more complex exercise environment than a normal research project would be able to finance.

The initial results from the surveys indicated that the exercise as a whole seemed to have progressed smoothly, with positive ratings of teamwork, information sharing and strategy, as well as cyber situational awareness. The high performance score further emphasizes the success of the team. It is therefore clear that the case studied is well suited for analysis of how a successful team develops its strategy and what functional structures for coordination in a cyber defence team may look like.

The participation of the research team can also be viewed as a success story. The team was engaged from the very beginning of the preparatory phase of the exercise. In this way, the researchers were able to learn contextual factors, such as the team leadership view on issues such as strategy and coordination, which may have affected the final outcome of the exercise, that were not necessarily written down or documented in any other way. The researchers were given the opportunity to describe their initial planning, which included the aim to be as non-intrusive as possible. This means that the data collection did not include any "freezes" in the exercise in order to collect data, neither were there any plans to collect excessive unnecessary amounts of data (video recordings, key-logging or screenshots). The research team explained their measures to preserve the participants' anonymity (if this was requested). The nature of the meetings, e.g., personal interactions between exercise participants and the research team contributed to mutual trust and confidence building. The cordial relationship between the researchers and the participants probably contributed to the high survey response rates and a benevolent attitude to answering the questions. Having three persons in the research team meant a high workload, and an appreciation for that every team activity would not be observed. It should be noted that there could be a limit as to how many researchers are feasible for data collection to reduce the

risk of collecting too much data, duplicates of data, or disturbing the observed team.

The inclusion of CSA-related questions in the team's ordinary reporting procedure provided a less intrusive method for collecting SA-related data than if a traditional probing measure such as SAGAT or SART had been used. There was no indication of that the inference of SA-related questions in the internal reporting template disturbed the sub-teams. The internal reporting template was designed and administered by the reporting team, and the sub-teams were not aware of that some of the questions in the internal reporting template were induced by the research team. The research team may have positively affected the SA and even the performance in the team, in that they directed the team's attention towards issues useful to include in the threat reports as well as for the understanding of the threats. This is, however, a risk with all data collection. For instance, the presence of observers may affect the behavior of the team, and the questions in the start-up survey may direct the team's focus towards the issues mentioned in the survey.

For integrity reasons as well as risk of data overload, the research team made a deliberate decision to not to capture network activity. It is assessed that network activity logs would produce a lot of data of which only a small part would be useful. Subsequently, a great deal of analysis would be required, and there would probably not be a great deal of added value for the research themes.

For the current paper, a pluralist methodology [31] was employed, where different types of data were collected in order to address the research themes from several perspectives. It needs to be acknowledged that some data will almost always suffer from some data loss (for instance surveys), while other data is more complete, but requiring quite a lot of analytic effort and needing to be combined with other data in order to interpret it (i.e. communication logs). Some data is based on subjective assessment (surveys, observations). However who is best suited to assess the workload, if not those who are experiencing it? The pluralist approach finds answers in combinations and comparisons of data. The dataset will suit the analysis within all three research themes, and the same data items may be combined and analyzed in different ways to correspond to research questions within different themes.

6. Summary and Conclusions

The objective of this paper was to explore the possibilities to collect meaningful data for research on Command and Control (C^2) Cyber Situational Awareness (CSA), and Intelligence in conjunction with an interorganizational cyber defense team during a cyber defense exercise. The purpose of the data collection was to explore the possibilities to conduct reserach within these research areas. A variety of carefully selected data collection methods specifically tailored to suit the needs for relevant data for our research fields of interest was employed. Our research has shown that a CDX indeed provided a good opportunity to collect pertinent data to analyze Command and Control, Cyber Situational Awareness and Intelligence related research questions.

7. Future work

The trove of collected data will be used to write papers according to our research themes. The working titles of some future papers in the works are: i) Exploring the development of coordination structures and strategy in a cyber defense team, ii) Acquiring Cyber Situational Awareness in a cyber defense team, and iii) Intelligence in support of a defensive cyberspace operation.

The results obtained in this paper is further expected to be useful for refining data collection methods for CDXs.

Acknowledgements

This work was partially funded by the Swedish Armed Forces and partially by the Swedish Civil Contingencies Agency. We want to thank Per Wikberg and Sofia McGarvey at the Swedish Defence Research Agency for valuable comments on the draft manuscript. We are further grateful to the Swedish blue team for accepting the presence of the research team without hesitation. Last but not least, we would like to thank the exercise organization at CCDCOE.

References

[1] Nsa tapped german chancellery for decades, wikileaks claims. *The Guardian*, 2015. URL https://www.theguardian.com/us-news/2015/jul/08/nsa-tapped-german-chancellery-decades-wikileaks-claims-merkel.

[2] Locked shields, 06 2019. URL https://ccdcoe.org/exercises/locked-shields/.

[3] D. S. Alberts. Agility, Focus, and Convergence: The Future of Command and Control. *The International C2 Journal*, 1(1):1–30, 2007.

[4] M. Bang. *Military Intelligence Analysis: Institutional Influence*. PhD thesis, 2017. URL http://www.doria.fi/bitstream/handle/10024/144017/Bangdissertation{%}28web-copyright{%}29.pdf?sequence=1{&}isAllowed=y.

[5] P. Barford, M. Dacier, T. G. Dietterich, M. Fredrikson, J. Giffin, S. Jajodia, S. Jha, J. Li, P. Liu, P. Ning, X. Ou, D. Song, L. Strater, V. Swarup, G. Tadda, C. Wang, and J. Yen. Cyber SA: Situational awareness for cyber defense. In S. Jajodia, P. Liu, V. Swarup, and C. Wang, editors, *Cyber Situational Awareness: Issues and Research*, volume 46 of *Advances in Information Security*, chapter 1, pages 3–14. Springer, Boston, MA, 2010. doi: 10.1007/978-1-4419-0140-8_1.

[6] B. Brehmer. The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control. In *Proceedings of the10th International Command and Control Research and Technology Symposium*, Newport, RI, 2005. International Command and Control Institute. URL http://www.dodccrp.org/events/10th{_}ICCRTS/CD/papers/365.pdf.

[7] J. Brynielsson, U. Franke, and S. Varga. Cyber situational awareness testing. In B. Akhgar and B. Brewster, editors, *Combatting Cybercrime and Cyberterrorism: Challenges, Trends and Priorities*, chapter 12, pages 209–233. Springer International Publishing, 2016. doi: 10.1007/978-3-319-38930-1_12.

[8] N. Buchler, P. Rajivan, L. R. Marusich, L. Lightner, and C. Gonzalez. Sociometrics and observational assessment of teaming and leadership in a cyber security defense competition. *Computers & Security*, 73: 114–136, mar 2018. doi: 10.1016/j.cose.2017.10.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0167404817302298.

[9] M. A. Champion, P. Rajivan, N. J. Cooke, and S. Jariwala. Team-based cyber defence analysis. In *Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, pages 218–221, 2012. ISBN 9781467301831. doi: 10.1109/SSP.2012.6319793. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4049859{&}queryText=substrate+mode{&}newsearch=true{&}searchField=Search{_}All{%}5Cnhttp://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1443594{%}5Cnhttp://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnum.

[10] N. J. Cooke. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41(6):801–849, 1994. ISSN 1071-5819. doi: https://doi.org/10.1006/ijhc.1994.1083. URL http://www.sciencedirect.com/science/article/pii/S1071581984710834.

[11] J. Copeland. Emergency response: Unity of effort through a common operational picture. Strategy research project, U.S. Army War College, Carlisle, PA, Mar. 2008.

[12] Department of Defense. Command and control, 2007. URL http://www.military-dictionary.org/DOD-Military-Terms/.

[13] G. Douhet. *The Command of the Air*. Air Force History and Museums Program, Washington D.C., 1998.

[14] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

[15] M. R. Endsley. The role of situation awareness in naturalistic decision making. In C. Zsambok and G. Klein, editors, *Naturalistic decision making*, pages 269–283. LEA, Mahwah, New Jersey, 1997.

[16] M. Flynn and C. Schrankel. Applying Mission Command through the Operations Process. *Military Review*, 93(2):25–32, 2013. URL https://www.armyupress.army.mil/Portals/7/military-review/Archives/English/MilitaryReview{_}20130430{_}art006.pdf.

[17] U. Franke and J. Brynielsson. Cyber situational awareness – A systematic review of the literature. *Computers & Security*, 46:18–31, Oct. 2014. doi: 10.1016/j.cose.2014.06.008.

[18] P. Gill, S. Marrin, and M. Phythian. *Intelligence Theory: Key Questions and Debates*. Studies in Intelligence Series. Routledge, 2009. ISBN 9780415553377. URL https://books.google.se/books?id=Aq14PgAACAAJ.

[19] M. Granåsen and D. Andersson. Measuring team effectiveness in cyber-defense exercises: a cross-disciplinary case study. *Cognition, Technology and Work*, 18(1):121–143, 2016. ISSN 14355566 14355558. doi: 10.1007/s10111-015-0350-2.

[20] M. Granåsen, P. Barius, N. Hallberg, and A. Josefsson. Exploring Mission Command in a Concept for Future Command and Control. In *Proceedings of the 23rd International Command and Control Research and Technology Symposium*. International Command and Control Institute, 2018. URL `https://www.diva-portal.org/smash/get/diva2:1299251/FULLTEXT01.pdf`.

[21] K. Helkala, B. J. Knox, Ø. Jøsok, R. G. Lugo, S. Sütterlin, G. O. Dyrkolbotn, and N. K. Svendsen. Supporting the human in cyber defence. In S. K. Katsikas, F. Cuppens, N. Cuppens, C. Lambrinoudakis, C. Kalloniatis, J. Mylopoulos, A. Antón, and S. Gritzalis, editors, *Computer Security*, pages 147–162, Cham, 2018. Springer International Publishing. ISBN 978-3-319-72817-9. doi: 10.1007/978-3-319-72817-9_10.

[22] L. J. Hoffman, T. Rosenberg, R. Dodge, and D. Ragsdale. Exploring a national cybersecurity exercise for universities. *IEEE Security & Privacy*, 3(5):27–33, Sept.–Oct. 2005. doi: 10.1109/MSP.2005.120.

[23] M. Hurley, Matthew. For and from cyberspace - conceptualizing cyber intelligence, surveillance, and reconnaissance. *Air & Space Power Journal*, 26(6):12–33, Nov.–Dec. 2012. URL `https://www.airuniversity.af.edu/Portals/10/ASPJ/journals/Volume-26_Issue-6/F-Hurley.pdf`.

[24] K. Johnson, Loch. *The development of intelligence studies*. Routledge, 2014.

[25] S. Kavanagh. 5g vs 4g: No contest, 2018. URL `https://5g.co.uk/guides/4g-versus-5g-what-will-the-next-generation-bring/`. Last accessed 5 August 2019.

[26] J. Kick. Cyber exercise playbook. Technical Report MP140714, MITRE Corporation, Wiesbaden, Germany, Nov. 2014.

[27] F. C. Lunenburg. Organizational Structure: Mintzberg's Framework. *International Journal of Scholarly, Academic, Intellectual Diversity*, 14(1), 2012. URL `http://www.nationalforum.com/ElectronicJournalVolumes/Lunenburg,FredC.OrganizationalStructureMintzbergFrameworkIJSAIDV14N12012.pdf`.

[28] M. V. Mahoney and P. K. Chan. An analysis of the 1999 DARPA/Lincoln Laboratory evaluation data for network anomaly detection. In *Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection (RAID 2003)*, pages 220–237, Pittsburgh, Pennsylvania, Sept. 2003. doi: 10.1007/978-3-540-45248-5_13.

[29] B. Mauer, W. Stackpole, and D. Johnson. Developing small team-based cyber security exercises. In *Proceedings of the 2012 International Conference on Security and Management (SAM'12)*, pages 213–217, Las Vegas, Nevada, July 2012.

[30] D. Mendonca, T. Jefferson, and J. R. Harrald. Collaborative Adhocracies and Mix-and-Match Technologies in Emergency Management - Using the Emergent Interoperability Approach to Address Unanticipated Contingencies During Emergency Response. *Communications of the ACM*, 50(3):45–49, 2007. URL `https://www.researchgate.net/publication/278361293`.

[31] J. Mingers. Combining IS Research Methods: Towards a Pluralist Methodology. Technical Report 3, 2001. URL `http://gkmc.utah.edu/7910F/papers/ISRcombiningISresearchmethods.pdf`.

[32] H. Mintzberg. *Structure in fives: Designing effective organizations*. Prentice-Hall, Inc, 1993.

[33] OSCE. Vienna document 2011: achievements and prospects for further updates. Technical report, Organisation for Security and Co-operation in Europe, n.d.. URL `https://www.osce.org/fsc/103978`.

[34] OSCE. Permanent council decision no. 1202. Technical report, Organisation for Security and Co-operation in Europe, n.d.. URL `https://www.osce.org/pc/227281`.

[35] OSCE. Ensuring military transparency – the vienna document. Technical report, Organisation for Security and Co-operation in Europe, n.d.. URL `https://www.osce.org/fsc/74528`.

[36] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of internet background radiation. In *Proceedings of the 2004 ACM SIGCOMM Internet Measurement Conference (IMC 2004)*, pages 27–40, Taormina, Sicily, Italy, Oct. 2004. doi: 10.1145/1028788.1028794.

[37] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai. A survey on low latency towards 5g: Ran, core network and caching solutions. *IEEE Communications Surveys Tutorials*, 20(4):3098–3130, Fourthquarter 2018. ISSN 1553-877X. doi: 10.1109/COMST.2018.2841349.

[38] P. M. Salmon, N. A. Stanton, G. H. Walker, C. Baber, D. P. Jenkins, R. McMaster, and M. S. Young. What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, 9(4):297–323, 2008. doi: 10.1080/14639220701561775.

[39] P. M. Salmon, N. A. Stanton, G. H. Walker, D. Jenkins, D. Ladva, L. Rafferty, and M. Young. Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3):490–500, May 2009.

[40] B. Sangster, T. J. O'Connor, T. Cook, R. Fanelli, E. Dean, W. J. Adams, C. Morrell, and G. Conti. Toward instrumenting network warfare competitions to generate labeled datasets. In *Proceedings of the 2nd Workshop on Cyber Security Experimentation and Test (CSET'09)*, Montreal, Canada, Aug. 2009.

[41] E. Seker and H. H. Ozbenli. The concept of cyber defence exercises (cdx): Planning, execution, evaluation. In *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–9, June 2018. doi: 10.1109/CyberSecPODS.2018.8560673.

[42] E. Shamir. *Transforming command : the pursuit of mission command in the U.S., British, and Israeli armies*. Stanford University Press, Stanford, 2011. ISBN 0804772037.

[43] S. Shankland. How 5g aims to end network latency, 2018. URL `https://www.cnet.com/news/how-5g-aims-to-end-network-latency-response-time/`. Last accessed 5 August 2019.

[44] T. Sommestad and U. Franke. A test of intrusion alert filtering based on network information. *Security and Communication Networks*, 8(13):2291–2301, Sept. 2015. doi: 10.1002/sec.1173.

[45] T. Sommestad and J. Hallberg. Cyber security exercises and competitions as a platform for cyber security experiments. In *Proceedings of the 17th Nordic Conference on Secure IT Systems (NordSec 2012)*, pages 47–60, Karlskrona, Sweden, Oct.–Nov. 2012. doi: 10.1007/978-3-642-34210-3_4.

[46] P. Sroufe, S. Tate, R. Dantu, and E. Çankaya Celikel. Experiences during a collegiate cyber defense competition. *Journal of Applied Security Research*, 5(3):382–396, 2010. doi: 10.1080/19361611003601280.

[47] J. C. o. Staff. Joint intelligence preparation of the operational environment (jp 2-01.3). Technical report, Joint Chiefs of Staff, 2014. URL `https://www.airuniversity.af.edu/Portals/10/ASPJ/journals/Volume-26_Issue-6/F-Hurley.pdf`.

[48] J. C. o. Staff. Cyberspace operations (jp 3-12). Technical report, Joint Chiefs of Staff, 2018. URL `https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf`.

[49] N. A. Stanton. Distributed situation awareness. *Theoretical Issues in Ergonomics Science*, 17(1):1–7, jan 2016. doi: 10.1080/1463922X.2015.1106615. URL `http://www.tandfonline.com/doi/full/10.1080/1463922X.2015.1106615`.

[50] F. Tariq et al. A speculative study on 6g. 2019. doi: arXiv:1902.06700. URL `https://arxiv.org/pdf/1902.06700`.

[51] D. Thomas. *U.S Military Intelligence Analysis: old and New Challenges*. Georgetown University Press, 2008.

[52] M. Van Wart and N. Kapucu. Crisis Management Competencies. *Crisis Management Competencies, Public Management Review*, 13(4):489–511, 2011. doi: 10.1080/14719037.2010.525034. URL `https://doi.org/10.1080/14719037.2010.525034`.

[53] N. Wilhelmson and T. Svensson. *Handbook for planning, running and evaluating information technology and cyber security exercises*. National Defence College, Stockholm, Sweden, 2014.

[54] B. T. Williams. The joint force commander's guide to cyberspace operations. *Joint Force Quarterly*, 73(2):12–19, 2014. URL `http://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-73/jfq-73_12-19_Williams.pdf`.

[55] J. Wolbers and K. Boersma. The common operational picture as collective sensemaking. *Journal of Contingencies and Crisis Management*, 21(4):186–199, Dec. 2013. doi: 10.1111/1468-5973.12027.

[56] World Economic Forum. The Global Risks Report 2019. 14th Edition. Technical report, World Economic Forum, Geneva, Switzerland, 2019. URL `http://wef.ch/risks2019`.