EXPLORATORY SEQUENTIAL DATA ANALYSIS OF A CYBER DEFENCE EXERCISE

Dennis Andersson[(a)], Magdalena Granåsen[(a)], Thomas Sundmark[(a)], Hannes Holm[(b)], Jonas Hallberg[(a)]

[(a)]Swedish Defence Research Agency, Box 1175, SE-581 11 Linköping, Sweden
[(b)]KTH Royal Institute of Technology, Kungl Tekniska Högskolan, SE-100 44 Stockholm, Sweden

[(a)]{dennis.andersson,magdalena.granasen,thomas.sundmark,jonas.hallberg}@foi.se, [(b)]hannes.holm@ics.kth.se

ABSTRACT
Baltic Cyber Shield 2010 (BCS), a multi-national civil-military cyber defence exercise (CDX), aimed to improve the capability of performing a CDX and investigate how IT attacks and defence of critical infrastructure can be studied. The exercise resulted in a massive dataset to be analyzed and many lessons learned in planning and executing a large-scale multi-national CDX. A reconstruction & exploration (R&E) approach was used to capture incidents such as attacks and defensive counter-measures during the exercise. This paper introduces the usage of R&E combined with exploratory sequential data analysis (ESDA) and discusses benefits and limitations of using these methods for analyzing multi-national cyber defence exercises.

Using ESDA we were able to generate statistical data on attacks from BCS, such as number of reported attacks by the attackers and the defenders on different type of services. Initial results from these explorations will be analyzed and discussed.

Keywords: cyber defence exercise, data analysis, experimentation, reconstruction and exploration

1. INTRODUCTION

Most organizations and services are critically dependent on reliable and secure information systems. Thereby, cyber warfare and terrorism is becoming a significant threat to recognize in today's society. Incidents such as the cyber attacks on Estonia in 2007 and the attacks on U.K., U.S., German and French resources in 2005 (Greenemeier 2007) are frequently cited and evidences of that the threat is real. However, the amount of publicly available data from such incidents is limited, which makes it difficult to study the associated phenomena. Hence, there is a need for data that conceptualize the phenomena of cyber warfare and terrorism, which thereby motivates cyber defense exercises (CDX) simulating such attacks and training teams in how to defend critical information systems.

In May 2010, the Cooperative Cyber Defense Centre of Excellence and the Swedish National Defense College hosted the Baltic Cyber Shield (BCS) international cyber defense exercise (CDX). For two days, six Blue Teams from northern European government, military and academic institutions defended simulated power generation companies against a Red Team of 20 computer hackers. The scenario described a volatile geopolitical environment in which a hired-gun Rapid Response Team of network security personnel defended Critical Information Infrastructure (CII) from cyber attacks sponsored by a non-state terrorist group. (Geers 2010)

The technical infrastructure was designed and implemented in a computer cluster located at, and hosted by, the Swedish Defense Research Agency (FOI). Each blue team network consisted of a number of virtual computers on the cluster, containing vulnerabilities to be exploited by the red team. The network connections were established through Virtual Private Networks (VPNs) enabling the teams to be physically distributed. Moreover, the networks were connected to the Programmable Logic Controllers (PLCs) of a power infrastructure model, including steam engines, solar panels, a simulated distribution network and factories with butane flames that could be detonated by the red team. Thus, a mixed-reality supervisory control and data acquisition (SCADA) network was created. (Hammervik, Andersson and Hallberg 2010)

ISBN 978-88-903724-3-8

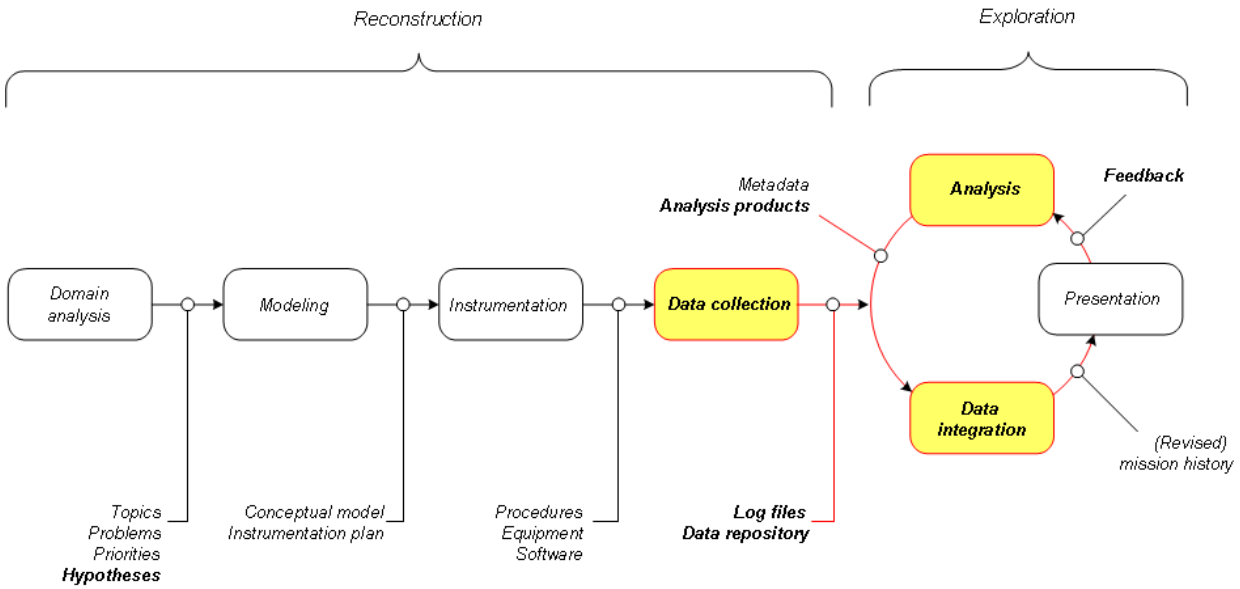

Figure 1: The Process of Reconstruction & Exploration (Andersson 2009)

The BCS CDX had three main goals: training of the Blue Teams, highlighting the international aspects of cyber defense, and improving the knowledge on how to perform CDXs (Geers 2010). To accomplish those goals, the outcome of the CDX needs to be carefully studied in terms of teamwork, collaboration models, scenario validity, attack patterns and C2 structures.

As the teams were given a large degree of freedom in how to organize and perform their work – there was little a priori knowledge from the experiment team on how the events would unfold and what phenomenon to monitor. A consequence of this is that there are too many unknown variables to adequately model the teams and their processes as would be desired. Instead, we chose to collect a massive heterogeneous dataset, containing both qualitative and quantitative data.

The resulting model has shown great potential for creating an understanding, or *situation awareness*, when studying the course of events after the exercise, and therefore enabling discovery experimentation (Alberts and Hayes 2002) using the exercise data. The objective of this paper is to describe the actual data collection, the analysis process, and discuss initial findings.

2. METHOD

During planning of the exercise, it was soon recognized that a structured way of organizing data collection was needed to be able to handle the multitude of available data sources and enable the analysis required to fulfill the goals of the BCS CDX. The Reconstruction & Exploration approach (R&E) (Figure 1) was selected due to its capacity to deal with large and complex data sets as well as being well-known by the analyst team (Andersson 2009). R&E was originally designed for use

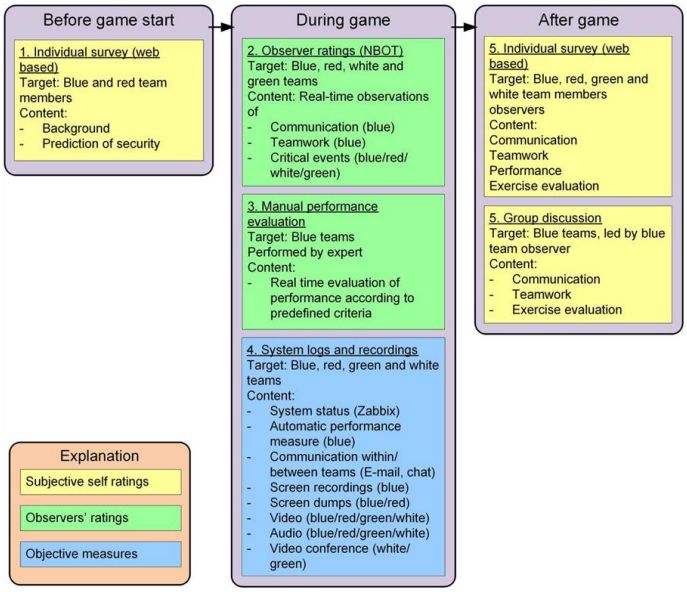

Figure 2: Prioritized Data Collection Nodes

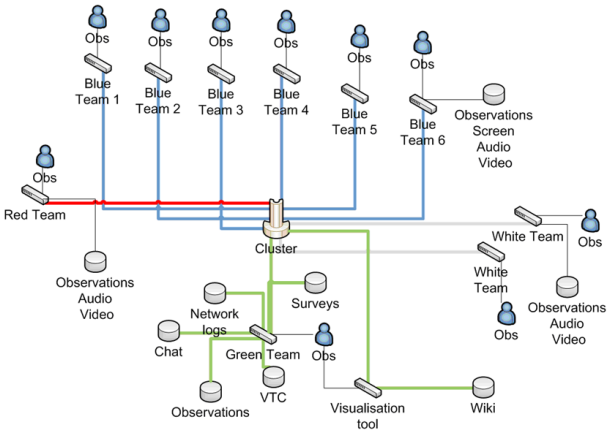

Figure 3: Logical Distribution of Teams, Observers and Data Collection Nodes during BCS 2010 (Hammervik, Andersson and Hallberg 2010)

ISBN 978-88-903724-3-8

with distributed tactical operations (DTOs) like military or crisis management operations (Pilemalm, Andersson and Hallberg 2008) and as such had never before been applied in its entirety to the IT security domain.

R&E consists of 7 steps, domain analysis, modeling, instrumentation, data collection, data integration, presentation and analysis. The output of the minimalistic domain analysis and modeling showed the need for collection of both quantitative and qualitative data (Figure 2) to enable reconstruction of situation awareness for the analyst, with the main focus on data from the blue teams. However, data collection also included red team activities as a reference for understanding blue team actions. Objective data in terms of system logs was assumed to provide results on the teams' activity in the system, but in order to understand why the teams chose the actions they did, there was a need to collect also the participants' views of what was happening and the reasoning within the teams. Therefore, it was decided that observers would be placed within each team, and that questionnaires would be used as a means to collect the subjective estimations of what they were experiencing.

It was decided that video cameras, audio recorders, screen capture tools and human observers should be placed in each team and surveys were to be distributed among the training audience (Figure 3) to try to capture the behavioral aspects of the teams. Observers were equipped with Network-Based Observation Tool (NBOT) (Thorstensson 2008) to enable quick and intuitive reporting of interesting events. Data collection for the objective measures included e-mails, chat sessions, keyboard interactions, network traffic and utilization of memory, processors and hard disk space on each node in the virtual network. In order to capture screen video and keyboard interactions, custom made scripts had to be installed on every machine used in the network. Because it was decided that some of the teams should use their own computers, the analyst team had to rely on participants' willingness to cooperate and install these scripts on their respective machines. For the teams that were supplied workstations by the exercise organizers, however, it was easier to setup and control this logging. For the supplied Windows computers a custom-made screen capture program was used, while on Linux the participants were recommended to use xvidcap, but any other appropriate application was allowed. To capture the terminal I/O a script to be executed by the participants was supplied as part of the team packages. Some data, such as e-mails, video feeds and NBOT reports were also available in real-time for the exercise judges (the White team, WT) who used that information to score blue team performance (Geers 2010).

Data was collected throughout the whole two-day exercise, in total 3 TB of data was collected. F-REX (Andersson 2009) was used for the exploration part of R&E. F-REX is a completely configurable tool

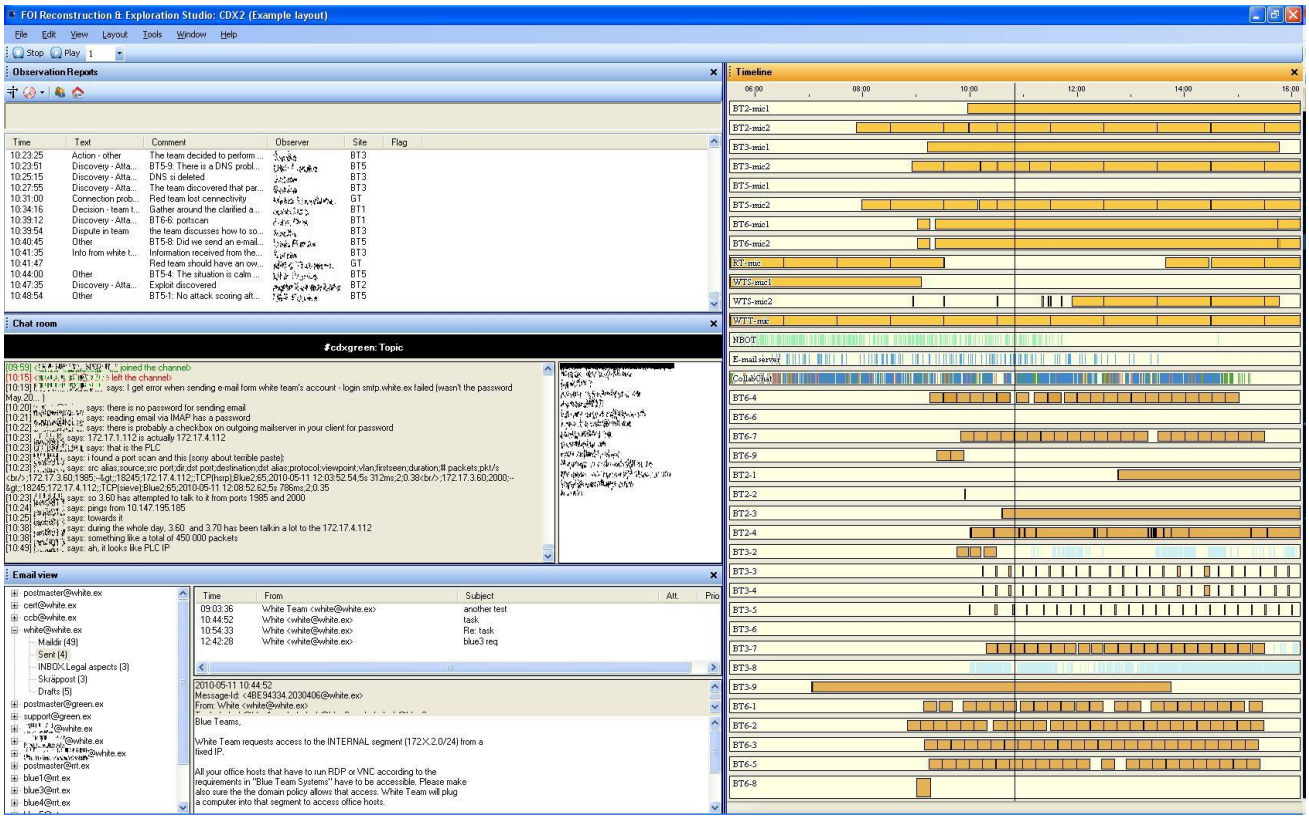

Figure 4: F-REX screenshot showing analysis in the first cycle. The layout shows observer reports, chat room log, e-mail to the left and a timeline of events currently in the mission history, separated by source, to the right. (Note: names are scrambled to preserve anonymity)

ISBN 978-88-903724-3-8

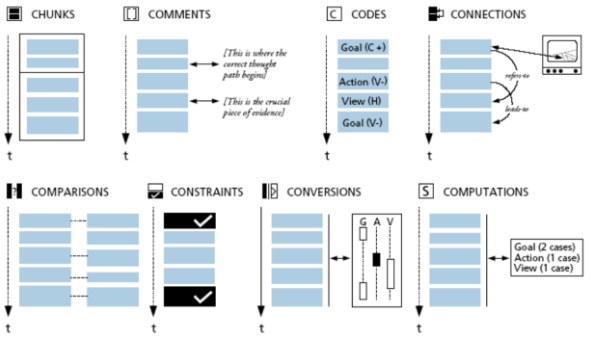

Figure 5: 8C's (Sanderson 1991)

allowing users to view a large and heterogeneous data set in a uniform and synchronized manner, much like playing a DVD back and forth (Figure 4). Its features include quick and easy timeline-based navigation in data based on timestamps from the data collection tools. At any time the analyst can shift focus in F-REX by applying a new layout with any views he or she prefers.

The captured data could be imported to an F-REX project, known in R&E as a mission history, for further synchronization, presentation and analysis according to the Exploratory Sequential Data Analysis (ESDA) method (Sanderson 1991). ESDA is an empirical exploratory approach, as opposed to confirmatory, in which the analyst uses temporal ordering of data to try to make sense of a dataset. Sanderson describes ESDA as a family of observational methodologies that are used when the objective is to observe what people do over time. Sanderson & Fisher (Sanderson and Fisher 1994) outlines the 8 C's (Figure 5) as the main operations needed in ESDA, the reader is advised to read their article for a thorough explanation of the 8 C's. They do not claim that all of the 8 steps must be performed at every study, nor do they have to be performed in a particular order. However, the way that they present them seems to create a fairly logical work flow that can easily be followed using F-REX and the exploration part of R&E.

In R&E, the exploration phase is cyclic, with analysis results and presentation comments being fed back into the model to create revised mission histories. In this study, the first reconstruction cycle used only chat logs, e-mail communication and observer reports - simply because they were estimated to generate the most value for least effort into the analysis. It is easy to assume that neglecting a large portion of the data set as this will impact results. This is, however, merely a case of "data guilt" as Fisher & Sanderson so accurately defines "unless there is a formal commitment to analyzing all the data to meet sampling assumptions, it may not be necessary. Thoroughly analyzing a subset of the data may be more informative". (Fisher and Sanderson 1996)

3. RESULTS

A CDX differs from the type of operations that are typically analyzed by the team in the sense that almost all data is in the virtual domain, and there is little real action to observe and analyze in the physical world. As such, the R&E approach had much to prove. Setting up the data capture was indeed a journey into uncharted waters as, to the authors' knowledge, this type of comprehensive data capture had never before been tried in a CDX.

Since most of the action happens in the virtual domain, in a CDX data capture is mostly a matter of running software that log system-system and human-system interactions. It turned out that it is not always easy to capture these interactions in a easily quantifiable format, as many programs, protocols and data formats being used are proprietary. The chosen fallback solution was to capture screen videos, keyboard interactions and network traffic for systems that could not be tapped in to in any other way. This data capture is not optimal because it is very crude and hard to interpret as needed to do the chunks, connections and codes. Still, the work proved possible, although very time consuming. The process became especially cumbersome since some of the blue teams were allowed to use their private laptops, and the data capturing was therefore dependent on their willingness to install and run special software and scripts to capture these interactions. Getting them to do so proved difficult, probably because of lack of understanding of the importance of the evaluation process. As a result the data set is missing some interactions that could potentially be vital for the detailed analysis of the work that was going on in these teams, and as mentioned before this introduces too many unknown variables to enable a satisfactory analysis of these interactions.

We found however, that the data discussed above can be used by analysts to acquire an increased understanding of the *what*, *when*, *where*, *who* and *how* (Whetten 1989) of the events that unfolded in the exercise. The remaining question, and the most fruitful one, *why*, is less straight-forward, and we found that there are too many unknowns within the setup of the exercise to fully answer why the teams took certain actions with only that data.

Perhaps the answer to the *why* lies in the multimedia data that was captured, i.e. video cameras and audio recorders in each team that were employed to capture human-human interactions. Analysis of these interactions is typically very time consuming, but resembles more traditional R&E work (and ESDA for that matter). This analysis has yet to be performed, but is essential not only to understand the *why*, but also to answer questions such as how human-human interaction affects the team collaboration or performance in a CDX.

ISBN 978-88-903724-3-8

Table 1: Compromised Services as Reported by Attacking vs. Defending Teams

Service	# reports by attacking team (s_a)	# reports by defending teams (s_d)	s_d/s_a
Operator	2	1	0.500
Fileserver	5	1	0.200
External firewall	4	3	0.750
Historian	8	3	0.375
Mail server	6	9	1.500
News server	4	5	1.250
DNS/NTP	1	3	3.000
Database	3	3	1.000
Intranet	3	2	0.667
Public web server	11	12	1.091
Portal	6	7	1.167
Other	7	13	1.857

The third type of data capturing that took place during the BCS CDX was surveying. The background survey showed that the teams consisted of highly experienced personnel on both a technical and strategic level, most of which worked with IT security on a daily basis. Having this in mind, the participants perceived scenario complexity and realism as perfectly sufficient and were highly motivated throughout the exercise. Teamwork was experienced as smooth, probably due to that most team members were familiar to each other. In some teams, the members reported lacking technical competencies within fields experienced as crucial, which could be a possible explanation to differences in performance between the teams. Another aspect which was captured using surveys was the participants' prior assumptions regarding the probability of successful compromise of hosts with specific properties. These data were not included in the mission history, but are expected to be useful for separate studies, such as comparing experts' expectations with actual results to measure the accuracy of expert assessment as a metric for IT security.

As mentioned before there was not enough control within the exercise to a priori generate variables to measure. Instead the analyst team put together data from the exercise to try to make sense of what actually happened from both a technical and teamwork point of view. The first version of the mission history enabled finding an initial classification of the targets for all discovered compromises, as reported by the red and blue teams respectively (Table 1). The table does not yield any strong interpretations, however it hints that the most frequently attacked services during the BCS CDX were the historian, the public web server and the customer portal. The defending teams seem to have reported most of the incidents on the public web servers and the customer portals, while the attacks on the historians would be more likely to have passed undetected.

Our experience from this work is that ESDA is a very useful complement to R&E when analyzing massive multimedia-heavy datasets such as the one collected during the BCS CDX. While one can argue that any analysis made with the assistance of R&E could be categorized as ESDA, it is the structured way of working through the data set, as outlined by the 8C's, that makes ESDA so powerful. From our experience, the 8C's should be considered as guidelines that help structuring the analysis process.

4. CONCLUSION

This study has shown the successful use of R&E and F-REX for analyzing cyber defense exercises (CDXs). In order to perform the actual analysis, Exploratory Sequential Data Analysis was applied in the exploration phase. R&E with ESDA has shown great potential for analyzing CDXs.

It can be argued that any analysis with R&E is automatically ESDA and that would indeed be the case according to the definition of Fisher & Sanderson, since they do not enforce usage of all C's or enforce a certain ordering between the steps. Being aware of ESDA and the 8C's when performing the analysis helps with structuring the analysis and as such ESDA should be regarded as a useful technique to know for R&E analysts.

Capturing human-human interactions in a CDX is not very different from any DTO, although it is reasonable to assume that more of the communication will use digital foras, as opposed to a DTO which typically uses radio as the primary means of communication. A CDX does however, focus more on human-system interactions, which are not always easy to capture. To successfully do so, the analysts must carefully plan their instrumentation. Moreover, it is important to work closely together with the exercise organizers to make sure they understand the need for capturing the necessary data.

For the data collection part it could be concluded that having an observer tool with predefined coding schemas was very helpful for the observers and the analysts, but that the coding schema needs to be tested and verified in advance to avoid having to change schema during the exercise. The observer reports and the different teams' self-reporting via e-mail seem to be the most valuable resources for analyzing the data. From the reports it seems that the historians, the portals and the public web servers were the most frequently attacked targets during the BCS CDX.

Although the CDX does not primarily serve as an experiment, we have shown that the data set acquired from it can be used for discovery experimentation

ISBN 978-88-903724-3-8

(Alberts and Hayes 2002). To enable more detailed studies on specific research questions, e.g. relating to IT security or teamwork, a higher level of experiment control than was used in BCS CDX 2010 is desired.

ACKNOWLEDGMENTS

This paper was originally published in proceedings of the 16th International Command and Control Research and Technology Symposium (ICCRTS) in Québec City, Canada (Andersson, Granåsen, Sundmark, Holm and Hallberg 2011). The authors want to thank the organizers of the Cyberspace management track, the reviewers and the audience of ICCRTS who contributed with valuable comments on the work. This version of the paper contains minor modifications from the original version, with the objective of addressing some of those comments. In addition, the comments from the anonymous reviewers of DHSS 2011 have been taken into account to further improve the contribution of the paper.

REFERENCES

Alberts, D. and Hayes, R., 2002. *Code of Best Practice for Experimentation*, Washington: CCRP Publication Series.

Andersson, D., 2009. F-REX: Event-Driven Synchronized Multimedia Model Visualization, *Proceedings of the 15th International Conference on Distributed Multimedia Systems*, pp. 140-145, September 10-12, Redwood City (California, USA).

Andersson, D., Granåsen, M., Sundmark, T., Holm, H. and Hallberg, J., 2011. Analysis of a Cyber Defense Exercise using Exploratory Sequential Data Analysis, *Proceedings of the 16th International Command and Control Research and Technology Symposium*, June 21-23, Québec City (Québec, Canada).

Fisher, C. and Sanderson, P., 1996. Exploratory sequential data analysis: exploring continuous observational data, *Interactions*, 3(2), pp. 25-34, New York (New York, USA).

Geers, K., 2010. Live Fire Exercise: Preparing for Cyber War, *Journal of Homeland Security and Emergency Management*, 7(1), article 74.

Greenemeier, L., 2007. China's Cyber Attacks Signal New Battlefield Is Online, *Scientific American*, September 18.

Hammervik, M., Andersson, D. and Hallberg, J., 2010. Capturing a Cyber Defence Exercise (extended abstract), *Proceedings of the 1st National symposium on Technology and Methodology for Security and Crisis Management*, pp. 30, October 27-28, Linköping (Sweden). (Note: Hammervik, M. is now publishing under Granåsen, M.)

Pilemalm, S., Andersson, D., and Hallberg, N., 2008. Reconstruction and exploration of large-scale distributed operations—Multimedia tools for evaluation of emergency management response, *Journal of Emergency Management*, 6(4), pp. 31-47.

Sanderson, P. 1991. Exploratory sequential data analysis, *Engineering Psychology Research Laboratory Technical Report EPRL-91-04*, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana (Illinois, USA).

Sanderson, P. and Fisher, C., 1994. Exploratory Sequential Data-Analysis: Foundations, *Human-computer interaction*, 9(3-4), pp. 251-317.

Thorstensson, M., 2008. *Using Observers for Model Based Data Collection in Distributed Tactical Operations*, Linköping Studies in Science and Technology, Thesis No. 1386, Linköpings Universitet, Linköping, (Sweden).

Whetten, D., 1989. What Constitutes a Theoretical Contribution?, *Academy of Management Review*, 14(4), pp. 490-495.

AUTHORS BIOGRAPHIES

Mr. **Dennis Andersson** is a researcher at the Swedish Defence Research Agency, Division of Information Systems. He holds a master in Computer Science at Linköping University, where he is now also a PhD student in Information Science. His research is focused on capturing and sharing experiences obtained from complex exercises and operations.

Ms. **Magdalena Granåsen** is a researcher at the Swedish Defence Research Agency, Division of Information Systems. She holds a master in Cognitive Science at Linköping University, where she is also a PhD student in Information Science. Her research is focused team cognition and social media.

Mr. **Thomas Sundmark** is a software developer at the Swedish Defence Research Agency, Division of Information Systems. He holds a master in Computer Science at Linköping University.

Mr. **Hannes Holm** is a PhD student at the Royal Institute of Technology, Department of Industrial Information and Control Systems, in Stockholm, Sweden. His research is focused on methods for measuring and assessing security in computer networks.

Dr. **Jonas Hallberg** is a senior researcher at the Swedish Defence Research Agency, Division of Information Systems. He holds a PhD in Computer Systems at Linköping University. His research is focused on methods for assessment of IT security.

ISBN 978-88-903724-3-8