Use of aggressor profiling in cyber security risk assessments for industrial control systems

Anders Dahlen Lauvsnes, Håkon Dahl-Olsen, Craig Aaen-Stockdale and Linda Sørensen
Lloyd’s Register Consulting AS

ABSTRACT:

Information security is an ever more important risk factor in safety and automation systems. Risk related to such industrial systems are different from that of accidents caused by human errors or by random system failures, as failures are caused by often malicious actions from insiders or outsiders. As opposed to random failures, for cyber security cases, the probability that a certain scenario will take place may be difficult to quantify. The reason for this is that the underlying mechanism driving the risk is not a random process but rather the result of the aggressor’s motivations, resources and skills. This paper explains how profiling and categorization based on psychological research and how these are included in the assessment of how likely a particular attack is. The article focuses on external aggressors, but outlines the framework for assessment of insider threats. An adoption of the methodology in a recent case is also described and further research is outlined.

1 INTRODUCTION

1.1 *Cyber security*

An increasingly important risk factor in safety and automation systems is information security. This type of risk to an organization's operations is different from that of accidents caused by human errors or by random system failures, as disruptions are caused by deliberate actions from outsiders, and in some cases insiders. Industrial Control Systems (ICS) are computerized systems that control industrial operations, typically on a one to one basis with one communication channel per remote station. One type of ICS is Supervisory Control and Data Acquisition systems (SCADA). SCADA distinguishes itself from other ICS by being used for large scale process and often for remote operation. A SCADA system typically consists of several components ranging from the Human-Machine Interface, through programmable controllers to remote terminals. Also software and telemetry for data acquisition, transfer and analysis is comprised within SCADA. Traditionally safety critical systems and production critical control systems have been largely isolated from external networks, as well as from the enterprise IT systems. Due to market demand driven by convenience and cost focus, industrial automation systems are increasingly relying on commercial off-the-shelf hardware and software, and are to a larger extent than before integrated in the business network. This has several operational advantages over the previous isolated proprietary systems used for control of production plants, where security was mostly ensured by lack of connection to the outside world (NIST, 2011). A number of cyber incidents reported in media over the last decade show that this is no longer sufficient, and that information security challenges related to industrial automation systems must be taken seriously. One may advocate that information security must be seen as an integral part of process safety and that information security issues should be taken into account when designing barriers against major accident hazards. An important question in the context of the subject discussed in this paper is if SCADA is different from other IT systems to such an extent that it requires a separate approach to risk assessment and management. As pointed out in

NIST (2011) the operational and risk differences between ICS and other IT systems with respect to performance requirements and availability suggest that this an important path of investigation. In addition, control system vulnerabilities may lead to a situation where an intruder may cause physical harm, such as destruction of equipment or even larger accidents such as fires and explosions. The reason for this is that ICS acts directly on the physical assets.

As the risk structure is different from that of random failures and normal human errors, traditional risk assessment methods may not be directly applicable to cyber security risk assessments. Risk is typically seen as the convolution of severity with probability for a certain scenario. For cyber security cases, the likelihood that a certain scenario will take place is difficult to quantify. The reason for this is that the underlying mechanism driving the risk is not a random process but rather the result of the aggressor's motivations, resources and skills. Understanding the varying motivations and skills behind an aggressor's malicious behavior thus seems to be important to have a correct risk picture. The likelihood of a cyber-incident is strongly linked to both motivation and resources of an aggressor, who is typically unknown to the asset owner. In order to assess scenario likelihoods it is thus necessary to have a process that links the relevant and available traits of potential attackers to likelihoods of the identified hazards.

Profiling in itself falls within the scope of forensic psychology and consists of creating a relatively short profile that remains dynamic enough to describe and understand the traits and behavior of an unknown criminal, in this case an aggressor.

The work described here explains how profiling based on psychological research can be used to establish aggressor profiles and categories that can be contribute to the assessment of how likely a particular attack vector is. Based on a structured methodology, which includes analysis of the victim and threat environment, so-called victimology, informed decisions can be made about layers of defense against cyber threats.

There are already some normative approaches to categorization within cyber security profiling. In a report from the Government Accountability Office (Protection, 2005) reference is made to data showing what was perceived to be the primary intruders in to information infrastructures in general. Based on data from the FBI, CIA and other US agencies this report proposes some categories of intruders. It is interesting to note however that these categories referenced by the GAO in a general IT security context, are referenced by the National Institute of Standards and Technology (NIST) in their guide to ICS and SCADA security seemingly without any specific consideration to their applicability to SCADA and ICS.

2 THEORETICAL MODELS OF HUMAN INTRUDERS AND RISK

2.1 *Profiling*

According to Likiewicz (2011) psychological profiling may give a valuable contribution in the investigation of computer crimes. The assumption is that there is a clear relationship between the traits of the aggressor and the acts committed. By looking at the modus operandi and traces one should be able to infer the psychosocial characteristics of the cyber intruder. Rogers (2006)) suggest that cyber criminal's count on the relative anonymity of the internet, but typically this does not necessarily affect their modus operandi, motivation and any signatures that they leave. Thus the aim of profiling in this context is to establish a profile that encompasses the unique aspects of the typical cyber-criminal committing a given type of cyber-crime when compared to the general population. There is some mention of the concept 'trait' in profiling in this paper; this is not to be confused with personality traits (e.g. Big-5) as the research linking cyber-crimes in the context of critical infrastructures and personality traits remains virtually non-existent.

Warikoo (2014) proposes a comprehensive methodology for cyber-criminal profiling that suggests attention is given to traditional cyber-crimes such as fraud and information theft this approach does not claim to be specific to general IT or ICS. They differentiate between inductive and deductive profiling. Inductive profiling entails looking at data from previous crimes of a certain type, correlating them to characteristics of the crime in hand and establishes a profile of the typical offender of a certain crime. Deductive profiling means deducing the characteristics of the criminal from the evidence gathered: characteristics of the victim and the victimology (Warikoo, 2014). When doing profiling work it is important to establish a database with valid and comparable data from previous attacks. This will in turn inform the understanding and risk assessment of future cases (Likiewicz, 2011). The evidence and research is to some extent of anecdotal nature in the context of cyber-crimes against SCADA networks. Establishment of such databases of incidents can for certain systems be challenging, due to lack of sharing of data on attacks on critical infrastructure. Some case studies are available from different Computer Emergency response teams (CERT), such as

NorCert and also the Federal Information Security management Act (FISMA). There are also ongoing changes to that situation in many countries with the establishment of sector specific CERT organizations, such as KraftCERT in Norway (Nilsen, 2014). The joint efforts of such institutions will hopefully contribute to the establishment of a database that enables even more empirical investigation in to hacker behavior.

2.1.1 *Categorization of cyber intruders*

The term hacker is the term that is commonly used to refer to cyber intruders and criminals. One typical stereotypical intruder, as perceived by the general population, may be the teenage, computer-addicted that spends the nights gaining access to business computer systems and exploiting the information for social status, political or financial purposes. Such stereotypes are misleading and create ignorance and erroneous risk perception on the side of society and information security decision makers (Rogers (2006). Profiling and categorization in the context of SCADA systems is one way to surpass stereotypes and make more reliable and informed decisions both for forensic purposes and as in the case of this paper as a contribution in risk analysis. It should be noted of course that just as much as profiling serves as a way to broaden our scope of possible perpetrators and their principals, our purpose is to limit the scope and focus on the most likely intruders and their inference on risk quotients.

2.1.2 *External intruders*

Supposing that all intrusions are malicious a general risk model may be proposed based on the categories or profiles of external intruders. In its most basic form, this classification may be based on skill and motivation of the possible intruders (Rogers, 2006). This need however to be broken further down for the purpose of risk analysis. In this paper we suggest a breakdown methodology that allows for structured assessment of the strength of a particular identified threat to an information asset.

An important part of the intruder categorization is to look at the cost/benefit evaluation from the point of view of the intruder, as this represent the intruder's motivation and risk function. This approach requires that the intruder behavior is based on rational decision making. The intruder risk function, as described in this paper as a function of the actions of the aggressor and the defender may give some indication about motivation and skill level (Li, Rickert & Silva, 2013; Moayedi & Azgomi, 2012).

Influenced by Li, Rickert & Silva we propose that an attack typically consists of several actions taken by the intruder, this forms a behavioral pattern. On the opposite side of the ring the security measures taken by the defender forms a pattern that consists of a number of security measures. This is significant as the motivation of the attacker typically varies with the risk he or she takes and the effort that has to be made. So for each action that needs to be taken and each security measure that has to be overcome the risk and effort increases. Risk from the intruder's point of view may be expressed as a function of both the attack pattern and the security measures (Li, Rickert & Silva, 2013).

Further, the consequences of the attack are often observable. They are often of financial, symbolic or political value to the attacker, or whomever he serves as a proxy for. For simplicity, we suppose that the intruder is rational when it comes to relative cost and benefit and do not want to be apprehended by law enforcement and in most cases do not wish to be stopped or sanctioned. However unlikely such prosecution has shown to be. Bearing in mind of course, that for some intruders the attention and symbolical value may be greatest when discovered.

Based on this, the cost/benefit for the attacker is a function of cost and perceived risk in terms of security and complexity as described above. In order to understand and predict intrusion and intruder behavior during an attack we need to understand what type of consequences and the relative value of these that the intruder expects. Also, the number and type of actions that are needed in order to overcome the security measures is of analytical value. Empirical investigation suggest that cyber security will have 'critical mass' at a certain point when the intruders cost effectiveness evaluation gives negative marginal growth (Li, Rickert & Silva, 2013).

Further to the cost/benefit assessment game theory may widen our understanding of intruder behavior and decision-making. According to Roy et al. (2010) game theory gives a framework for studying how multiple players with competing interests interact. As an example, a network administrator and an intruder can be viewed as two competing players participating in a game. It can propose a quantitative framework for calculating risk and modeling network security problems. Game theory also enables us to examine a large number of possible scenarios before taking the best action; hence, it informs the decisions of the security professionals to a large extent. As Moayedi and Azgomi (2012) we presume that we can consider the categories or classes of aggressors to be internally similar in their cost/benefit estimations, strategic decisions and their decisions to act. This is supposing that they have access to the same information. Game theory may enable quantitative and qualitative approaches to scenario modelling.

2.1.3 *Cyber intruder traits*

The research literature within information science, security, psychology and forensics has described several possible categorizations and stereotypes of cyber intruders. As mentioned in the introduction these categories should not be confused with personality traits such as the Big-5 framework.

These features are only to some extent observable, thus making us reliant on empirical knowledge. According to Hald and Pedersen (2014), the cyber intruder categories relevant to SCADA vary based on the following traits and examples:

Motivation

– Risk tolerance and perception of getting caught

– Goals are related to consequences of intrusion such as inflicting damage, becoming famous or achieving financial gains

– The relationship between payoffs (consequences) and goals

Triggers

– External events of political or public interest nature

– Day 0 possibilities

– Opportunism

Skills

– Competence in IT and education level

Resources

– Access to network

– Persistence financially to carry on for an amount of time

Methods

– Speed

– Ability to hide traces, avoid detection

– Social engineering

2.1.4 *Skills and resources*

The relationship between the skills and resources of the intruders and the defenders is important. Their level of skills may be classified as novice, intermediate and expert.

Typically, higher skills and resources in the intruder decrease the risk if the defenders skills and resources are constant and vice versa (Al Mannai, W. I., and T. G. Lewis, 2008; Li, Rickert & Silva, 2013).

2.2 *Insider risk*

Although not a main focus of this paper the issue of insiders warrants mention. People working in the host organization represent both barriers against intrusion and possible threats in themselves. The mixing of these to insider categories may limit the understanding of risk and mitigating actions (Crossler & al., 2013). As Shaw (2006) describes there are significant amounts of research on the traits of the malicious insiders. These include a history of negative and social personal experiences, lack of social skills and propensity for social isolation, a sense of entitlement and ethical flexibility. These traits have to some extent been validated in empirical studies for malicious insider behavior. It is important however to point out that this refers to cyber-crime performed by insiders in general, not SCADA specifically.

2.2.1 *Insider reliability*

The opportunities for the external intruder are influenced by the human reliability of the insiders who functions as barrier elements that activate and guard information security measures. In this case the behavior of the insiders is not malicious. Unintended acts of omissions and commissions when managing the network safeguards is to be expected and influences the risk. Human error probabilities are typically influenced by a range of performance influencing factors such as competence, fatigue, stress, intoxication and user interfaces. The relative contribution of these may be determined in a Human reliability analysis which would include identification of security requirements with consequences influenced by human actions. The types of actions include necessary and desired actions to directly provide a critical function, backup actions to failed automatic responses, and anticipated procedure-guided recovery actions (Spurgin (2010). A typically human reliability analysis consists of the following steps:

- Scenario modelling, where human actions are described.
- A qualitative assessment of error modes, recovery and task characteristics relevant to Performance shaping factors scoring
- A quantitative analysis, where human error probabilities (HEP) are determined for the overall scenario and subtasks

The knowledge about Human Error probabilities may also inform prioritization of competency based risk mitigation (Spurgin, 2010).

2.2.2 *Social engineering*

The way that non-malicious insiders are used to facilitate intrusion is typically by social engineering. This comprises different psychological techniques used to influence and manipulate individuals into performing actions that compromises system security or divulging confidential information (Luo, Brody, Seazzu & Burd, 2011; Mitnick & Simon, 2002). Typically, social engineering consists of four steps. First information gathering about the 'victim' or target individual in order to be able to develop a relationship where the victim exploited or persuaded to provide confidential information or neutralize safety barriers. Social engineering intruders employ a range of different techniques of persuasion and manipulation. Often these are put in to a pretexting scenario that seems plausible to the victims.

Social engineering efforts may be restrained by offering structured training for employees on the recognition and tackling of such attempts, so called Social Engineering Awareness Training (SEAT) (Luo, Robert & al., 2013).

2.3 *Methodology for risk assessment*

Here the general process of SCADA intrusion risk assessment is described and specifically how profiling and categorization of cyber aggressors may contribute to determining the risk in combination with the identified hazards and threats.

2.3.1 *Categorization of intruders related to SCADA intrusion*

For the purpose of SCADA intrusion the database is limited and to some extent of anecdotal nature. The following categories are based on a summary of available research by Hald and Pedersen (2014). It is important to note also that for some types of hackers there may be a continuum between malicious and benign activities, these include so called white, grey and black hats (Crossler & al., 2013). Based on the research of Hald and Pedersen (2014) we identified 9 different categories of agressors and their different motivations, triggers, resources and methods of intrusion. These were combined with the three levels of skills proposed by Li, Rickert and Silva (2013).

The 9 categories that were proposed by Hald & Pedersen (2014) as relevant to SCADA include:

- Script kiddies
- Cyber punks
- Insiders
- Petty Thieves
- Grey hats
- Professional criminals
- Hacktivists
- Nation states
- Terrorists

2.4 *Risk assessment*

The following threat modelling approach is based on a 3-stage process.

First a Cyber HAZID is performed. An HAZID (Hazard Identification) is a method commonly used to evaluate a system or operation to identify any challenges to safety and security (Rausand & Utne, 2009). In the cyber HAZID, each critical item of the inventory list will be reviewed. For each item, the worst case consequences of loss of availability, integrity, confidentiality and traceability will be assessed and categorized according to the project risk matrix. Potential threat descriptions will then be considered, and the feasibility of attacks will be assessed.

The Cyber HAZID should be conducted in a workshop including key personnel, identified prior to the Cyber HAZID. In order to correctly assess severities, it is imperative that the relevant competence is available. This can be particularly challenging with respect to financial and environmental consequences. When concluded the Cyber HAZID will have identified the main hazards facing the organization.

Secondly, a session looking at the victimology and the threat environment, the characteristics of the necessary attacks should be performed to match to relevant cyber-criminal category. The necessary attacks are identified during HAZID.

Thirdly, based on the categorization of potential aggressors a final risk analysis is carried out, using the relevant aggressor profiles as a determinants for the risk categorizations.

The risks together with suggestions for mitigating actions are treated in an ALARP session. ALARP means 'As low as reasonably practicable' and means finding risk mitigations that decreases risk to a reasonable level both for integrity and financially. This is done by putting in place security protocols for organizational/managerial, operational and technical issues.

Finally, the categorizations derived from the victimology, threat assessment and the identified risks from the Cyber HAZID are used to calibrate the likelihood and consequence classification in the risk matrix. This is also described in the case example below before and after mitigations.

3 CASE

A production company has newly invested in a new plant. This plant has a central control room located at significant distance from the physical asset, which handles highly flammable and toxic chemicals. The control room location has been selected to maximize the safety of the process operators in case of an accident in the process plant. The central control room is state-of-the-art and communicates with field equipment of over data networks. The control system is a proprietary software running on a Microsoft Windows server. Real-time information from the plant is made available through a firewall to the enterprise network for the purpose of production optimization and dynamic supply-chain adaptation. All write-back to the control system is performed manually by the process operators in the case of set-point changes, or by a process engineering case of larger changes to the way the control system manages the plant, including set point changes or overrides. The control system of the plant consists of three subsystems; basic process control system (BPCS), the safety instrumented system (SIS) and local unit control panels (UCP's) with internal logics, providing only feedback to the main control system, the BPCS. All communication within this system occurs primarily over a digital bus, physically arranged as a fibre-optic ring structure to allow full availability in the case of a single cable break. The safety signals are in addition to this transmitted over a redundant analog electrical signal from field equipment to the logic card, and further to all actuators.

The operator has long experience from its branch of industry, but not at this specific location. This means that the operator has to establish the operating organization concurrently with engineering and construction activities. Because the organization is new, it cannot be assumed that a consistent and high degree of security awareness exists as part of the culture. The company therefore initiates a security awareness program for all new employees, based on perceived best practices from other locations where they have operations ongoing.

In preparations for the cyber HAZID of this system, a system breakdown was performed and a criticality assessment was performed in accordance with the method described in the Norwegian Oil and Gas Association's Guideline No. 123 (NOROG, 2009). Based on this, all systems considered critical to safety and availability of the production asset was subject to a cyber HAZID. In the cyber HAZID, the systems were separated into nodes based on functionality. Examples of such nodes are "organization and people", "shared services for industrial control systems", "industrial control applications", "computer networks" and "remote access", among other things. Each of these systems were analyzed based on a set of guidewords to identify risks. Examples of guidewords for "organization and people" would be "dissipation" for identification of when personnel are used for other tasks or purposes than they are competent, trained and authorized to do, "overload" for identification of strain on personnel and organizations. Examples applicable to networks would be "passive interception" for network snooping (packet inspection), or "modification" for identification of attacks capable of changing key data during transmission from source to recipient. Based on such data, risks were identified for each of the nodes.

Based on the cyber HAZID, risks were subject to risk ranking. A severity analysis was performed to assess the potential worst-case consequences of loss of confidentiality, integrity, availability and traceability. Based on the identified risks, a probability assessment is performed based on the representative hacker categories and typical traits they are assumed to possess. This is used again to classify the risks in terms of a company specific risk matrix, where all risks were classified as belonging to one of the main classes "unacceptable", "acceptable if risk is as low as reasonably practicable (ALARP)", or "risk is generally acceptable". For all risks falling into the categories "unacceptable" or "acceptable if ALARP", the risk is put forward to risk mitigation planning. This way, a structured analysis could be applied to identify and prioritize risks to critical information systems in process control and safety management. This has made it possible to achieve a proactive approach to information security risk management, focusing energy on those risk factors that are contributing to the largest degree to the overall risk level of the plant.

4 DISCUSSION

Our aim in this paper has been to explain the initial background for an approach to profiling in the context of SCADA cyber intrusion risk management. As the empirical data remains fairly scarce there are of course some limitations as well as areas that merit further investigation both with respect to the profiling and categorization itself, the input to risk assessment and finally risk management. As applied to our case we believe that this source of information is useful and contributes to increased validity of the risk assessment and risk management process.

The cost benefit and game theoretical considerations give some insights in to the mental models of intruders, always with an assumption of rational decision making. However they are not exhaustive as to the cognitive models of intruders. Hacking is in this context considered a cognitive activity that requires exceptional technical and reasoning abilities. In this domain, a mental model can be thought of as a hacker's internal representation of the components and operating rules of an extremely complex software and hardware system and the process or game of controlling or breaking in to it. Mental models help hackers describe, explain, and predict system attributes and behaviors. As explained by Summers & al. (2013) the further understanding of these mental models may inform security measures. This may be investigated in both laboratory and to some extent, naturalistic settings.

The game theoretical approach to cost/benefit estimation may also be challenging as the different actors will not have access to all information about the other actors actions, skills and motivations. This will be an issue of technical, analytical and organizational nature for the defenders. This will probably change during the course of an attack or in multiple attacks as the actors will learn, this learning is crucial for information security specialists abased on organizational learning research it is unlikely that this will occur and to be generalized to the next attack without some structured intervention form the organization.

The categorizations of aggressors involves actors such as nation states and terrorists. We believe that this does not represent a methodological problem in this context as the focus on internal psychological traits is fairly limited and that skills, resources and cost/benefit analysis is applicable both to individual and institutional aggressors.

In addition the state of the system that is being attacked may influence the threat environment and the actor profiles relevant for the risk analysis. Looking at the rates of transitions between the different system states is important for risk analysis.

As part of their risk management work, organizations and authorities will typically put in place governing systems that comprises operational, managerial/organizational and technical security controls. In the context of looking at the insider human impact on risk, looking at the adherence to this will inform the risk assessment. This may also inform investigations and learning from cyber incidents. Although not reported specifically in this paper our proposed methodology extends into such risk management.

In further research, the distinction between insider errors and malicious behavior and the associated profiles is interesting and may be accessed in survey studies. Looking at the cross-cultural applicability of categories should be considered.

The further work will help increase the granularity of profiles; extend our knowledge on hacker mental models and game theoretical approaches. We conclude that the use of profiling is an important step in both risk calibration and risk mitigation. And that this approach will benefit further from the establishment and consolidation of data from an increasing number of government and industry actors.

REFERENCES

Al Mannai, W. I., and T. G. Lewis. 2008. A general defender-attacker risk model for networks. *The Journal of Risk Finance* 9.3: 244-261.

Crossler, Robert E., et al. 2013. Future directions for behavioral information security research." *Computers & security* 32: 90-101.

Hald, S.L.N. and Jens Myrup Pedersen. 2014. "The Threat of Digital Hacker Sabotage to Critical Infrastructures." *Image Processing and Communications Challenges 5*. Springer International Publishing: 379-390.

Jahankhani, H., and Ameer Al-Nemrat. 2012. "Examination of cyber-criminal behaviour." *International Journal of Information Science and Management (IJISM)*: 41-48.

Kshetri, Nir. 2009. "Positive externality, increasing returns, and the rise in cybercrimes." *Communications of the ACM* 52.12: 141-144.

Li, Si, Ryan Rickert, and Amy Sliva. 2013. Risk-Based models of attacker behavior in cybersecurity. *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer Berlin Heidelberg: 523-532.

Luo, Xin Robert, et al. 2013. "Social Engineering: The Neglected Human Factor for. *Managing Information Resources and Technology: Emerging Applications and Theories: Emerging Applications and Theories*: 151.

Mitnick, Kevin D., and William L. Simon. 2011. *The art of deception: Controlling the human element of security*. John Wiley & Sons.

Moayedi, Behzad Zare, and Mohammad Abdollahi Azgomi. 2012. "A game theoretic framework for evaluation of the impacts of hackers diversity on security measures." *Reliability Engineering & System Safety* 99: 45-54.

Nilsen, J. 2014. Kraftbransjen etablerer eget sikkerhetsselskap, Accessed via http://www.tu.no/kraft/2014/09/01/kraftbransjen-etablerer-eget-sikkerhetsselskap on April 30th 2015.

Norwegian Oil and Gas Association. 2009. Guideline 123: Classification of process control, safety and support ICT systems based on criticality.

Protection, C. I. 2005. *Department of Homeland Security Faces Challenges in Fulfilling Cybersecurity Responsibilities*. GAO-05-434, May

Rausand, M., & Utne, I. B. 2011. *Risikoanalyse: teori og metoder*. Trondheim: Tapir Akademisk Forlag.

Rogers, Marcus K. 2006. A two-dimensional circumplex approach to the development of a hacker taxonomy." *Digital investigation* 3.2, pp. 97-102.

Roy, Sankardas, et al. 2010. A survey of game theory as applied to network security. *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*. IEEE.

Shaw, Eric D. 2006. The role of behavioral research and profiling in malicious cyber insider investigations." *Digital investigation* 3.1: 20-31.

Summers, Timothy C., et al. 2013. "How Hackers Think: A Study of Cybersecurity Experts and Their Mental Models." *Third Annual International Conference on Engaged Management Scholarship, Atlanta, Georgia*.