21, rue d'Artois, F-75008 Paris
http://www.cigre.org

B5-105

Session 2004

IT Security for Utility Automation Systems

Bernhard Deck*
ABB Schweiz AG
Switzerland

Martin Naedele
ABB Schweiz AG
Switzerland

The influence of utility automation systems pervades many aspects of everyday life in most parts of the world. In the shape of factory and process control systems they enable high productivity in industrial production, and in the shape of electric power, gas, and water utility systems they provide the backbone of technical civilization. The sensitivity of electric power systems is very high for misleading functionality because it can turn easily into a local or mayor black out. Those mostly have impact to daily business and life. Furthermore if a blackout would be caused from an external intruder to the automation system of the utility the impact would be even bigger. This is because, as seen from outsiders, that the basic infrastructure can be damaged from externals means also from other parts of the world in a worst-case scenario.

Up to now, most of these systems are isolated, but for the last couple of years, due to market pressures and novel technology capabilities, a new trend has been rising to interconnect automation systems to achieve faster reaction times, to optimize decisions, and to collaborate between plants, enterprises and industry sectors. Initially, such interconnections were based on obscure, specialized, and proprietary communication means and protocols. Now more and more open and standardized Internet technologies are used for that purpose.

In security terminology, a risk exists if there is a vulnerability, that is, an opportunity to cause damage, together with a threat, that is, the fact that someone will try to find and exploit a vulnerability in order to init damage.

The importance of utility automation network systems for the functioning of modern society together with market pressure and competition on the one hand and geopolitical tensions on the other hand let the existence of security threats from terrorism, business competitor sabotage, and other criminal activity appear likely.

The pervasiveness of utility automation systems that are nowadays accessible from anywhere in the world via communications and information technologies for which there are thousands of experts worldwide and which have a large number of well-known security issues creates many IT security vulnerabilities. In consequence, there are good reasons to investigate and invest into how to reduce the IT security vulnerabilities of utility automation systems, and thus the resulting risks of large financial damage, deteriorated quality of life, and potentially physical harm to humans. This chapter presents an overview of state-of-the-art best practices to that respect, and an outlook into future opportunities.

The scope of utility automation systems considered in this chapter ranges from embedded devices,

* bernhard.deck@ch.abb.com

potentially in isolated locations, via plant control systems to plant and enterprise level supervisory control and coordination system, both in the distributed control system (DCS) flavor, more common in factory automation, and the supervisory control and data acquisition (SCADA) flavor, widespread in utility systems [5].

In the associated types of applications, in contrast to commercial and administrative data processing, often not typical data security issues (e.g. confidentiality, integrity) as such are the most important goal, but IT security is one component of the safety and fault-tolerance strategy and architecture for the utility system.

1 Security objectives

IT security has a number of different facets, which are to some extent independent of each other. When defining the security requirements for a system, these facets, on which risk analysis and in turn design of counter-measures are based, can be expressed in terms of the eight security objectives explained in the following. These security objectives must be carefully looked at which are maybe in contradiction to have still the needed performance of the installed system and the fast action of an authorized person.

1.1 Confidentiality

The confidentiality objective refers to preventing disclosure of information to unauthorized persons or systems. For automation systems this is relevant both with respect to domain specific information, such as product recipes or plant performance and planning data, and to the secrets specific to the security mechanisms themselves, such as passwords and encryption keys.

1.2 Integrity

The integrity objective refers to preventing modification of information by unauthorized persons or systems. For automation systems this applies to information coming from and going to the plant, such as product recipes, sensor values, or control commands, and information exchanged inside the plant control network. This objective includes defense against information modification via message injection, message replay, and message delay on the network. Violation of integrity may cause safety issues, that is, equipment or people may be harmed.

1.3 Availability

Availability refers to ensuring that unauthorized persons or systems cannot deny access/use to authorized users. For utility automation systems this refers to all the IT elements of the application, like control systems, protection systems, operator workplaces as well as the communications systems between these elements and to the outside world. Violation of availability may cause safety issues, as operators may lose the ability to monitor and control the process.

1.4 Authorization

The authorization objective, also known as access control, is concerned with preventing access to or use of the system or parts by persons or systems without permission to do so. In the wider sense authorization refers to the mechanism that distinguishes between legitimate and illegitimate users for all other security objectives, e.g. confidentiality, integrity, etc. In the narrower sense of access control it refers to restricting the ability to issue commands to the control system. Violation of authorization may cause safety issues.

1.5 Authentication

Authentication is concerned with determination of the true identity of a system user (e.g. by means of user-supplied credentials such as username/password combination) and mapping of this identity to a system-internal principal (e.g. valid user account) under which this user is known to the system. Authentication is the process of determining who the person is that tries to interact with the system, and whether he really is who he claims to be. Most other security objectives, most notably authorization,

distinguish between authorized and unauthorized users. The base for making this distinction is to associate the interacting user by means of authentication with an internal representation of his permissions used for access control.

1.6 Non-repudiability

The non-repudiability objective refers to being able to provide irrefutable proof to a third party of who initiated a certain action in the system. This security objective is mostly relevant to establish accountability and liability with respect to fulfillment of contractual obligations or compensation for damages caused.

In the context of automation systems this is most important with regard to regulatory requirements, e.g. FDA approval. Violation of this security objective has typically legal/commercial consequences, but no safety implications.

1.7 Auditability

Auditability is concerned with being able to reconstruct the complete behavioral history of the system from historical records of all (relevant) actions executed on it. While in this case it might very well be of interest to record also who initiated an action, the difference between the auditability security objective and non-reputability is the ability of proving the actor identity to a third party, even if the actor concerned is not cooperating. This security objective is mostly relevant to discover and find reasons for malfunctions in the system after the fact, and to establish the scope of the malfunction or the consequences of a security incident. In the context of automation systems this is most important in the context of regulatory requirements, e.g. FDA approval. Note that auditability without authentication may serve diagnostic purposes but does not provide accountability.

1.8 Third party protection

The third party protection objective refers to averting damage done to third parties directly via the IT system, that is, damage that does not involve safety hazards of the controlled plant. The risk to third parties through possible safety relevant failures of the plant arising out of attacks against the plant automation system is covered by other security objectives, most notably the authorization/access control objective. However, there is a different kind of damage only involving IT systems: The successfully attacked and subverted automation system could be used for various attacks on the IT systems or data or users of external third parties, e.g. via distributed-denial-of-service (DDOS) or worm attacks. Consequences could reach from a damaged reputation of the automation system owner up to legal liability for the damages of the third party. There is also a certain probability that the attacked third party may retaliate against the subverted automation system causing access control and availability issues.

This type of counter attack may even be legal in certain jurisdictions.

2 Differences to conventional IT security

As the security objectives in the previous section are generally valid, many security issues are the same for utility automation systems and conventional, office-type IT systems, and many tools can be used successfully in both domains. However, there are also major differences between these two domains with respect to requirements and operating environment, characteristics, and constraints, which make some security issues easier and others more difficult to address. In the following some of these differences are explained.

2.1 Requirements

While office IT security requirements center around confidentiality and privacy issues, for any protection and process control system the foremost operational requirement is safety, the avoidance of injury to humans. Second after that is availability: The application and the automation system have to be up and running continuously over extended periods of time, with hard real-time response requirements in the millisecond range. This in many cases precludes standard IT system administration practices of

system rebooting for fixing problems, and makes the installation of up-to-date SW patches, e.g. addressing security problems in the running application or the underlying operating system difficult if not impossible. On the other hand, in contrast to e-commerce applications, connectivity to outside networks including the company intranet is normally not mandatory for the automation system, and although extended periods of disconnection are inconvenient, they will not have severe consequences - after all, many automation systems nowadays still run completely isolated.

2.2 Operational environment

The configuration, both of HW and SW, of the protection and control system part, which contains the safety critical automation and control devices, is comparatively static. Therefore all involved devices and their normal, legitimate communication patterns (regarding communication partners, frequency, message size, message interaction patterns, etc) are known at configuration time, so that protection and detection mechanisms can be tailored to the system. Modifications of the system are rare enough to tolerate a certain additional engineering effort for reconfiguring the security settings and thus being able to trade in the convenience of dynamic, administration-free protocols like DHCP against higher determinism and, in consequence, security of e.g. statically setting up tables with communication partners/addresses in all devices. Static structure and behavioral patterns also make the process of anomaly discovery for intrusion detection easier.

The hosts and devices in the automation system zone are not used for general purpose computing, preventing the risks created by mainstream applications like email, instant messaging, office application macro viruses, etc. Often, they are even specialized embedded devices dedicated to the automation functionality, such as power line protection in substation automation. All appropriate technical and administrative means are taken to ensure that only authorized and trustworthy personnel have physical access to the automation equipment. An automation system personnel is accustomed to a higher level of care and inconvenience when operating computer systems, than office staff. This increases the acceptance and likeliness of correct execution of security relevant operating procedures even if they are not absolutely straightforward and convenient. In many plants additional non-networked (out-of-band) safety and fault tolerance mechanisms are available to mitigate the consequences of failure of one or multiple components of the automation system. One problem with security mechanisms is that they cannot prevent all attacks directly themselves, but produce outputs like alerts and log these entries which a human needs to review to decide on the criticality of an event and to initiate appropriate responses. This is often neglected, as the expert IT staff is not around the clock available to monitor the system. A dedicated staff, in contrast, usually continuously monitors automated plants. Defense architecture could make use of this fact, even though these operators do not have IT or even IT security expertise.

2.3 Challenges

On the other hand, the characteristics of automation systems and devices create some additional security challenges:

Automation devices often have lower processing performance compared to desktop computers, which limits the applicability of mainstream cryptographic protocols.

The operating systems of such devices in many cases do not provide authentication, access control, fine-granular file system protection, and memory isolation between processes - or these features are optional and not used due to the above-mentioned limited processing power.

Especially in remote monitoring applications (e.g. SCADA) communication channels with small to very small bandwidth like telephone, mobile phone or even satellite phone lines are used, which makes it imperative to reduce communication overhead and thus collides with certain security protocols.

Automation systems tend to have very long lifetimes. This has consequences both for the currently operative systems and for newly implemented systems:

Those currently operative "legacy" automation systems, as far as IT security was given a thought at all, were designed based on a philosophy of "security by obscurity", assuming that the system would be

isolated and only operable by a small, very trustworthy group of people. This kind of thinking even persists until today, as can be seen from the 2002 IEEE 1588 standard on precision time synchronization for automation systems for which it is explicitly stated as design rationale that security functionality is neglected as all relevant systems can be assumed to be secure. Another consequence of longevity is that automation system installations tend to be very heterogeneous with respect to both subsystem vendors and subsystem technology generations.

For newly built automation systems the long expected lifetime means that the data communication and authentication/access control functionality must be designed so that it will be able to interoperate with reasonable effort with systems and protocols to appear on the market 10 or 20 years later.

Last, but not least, automation systems are operated by technicians and process operators. Due to there training background they have a very different attitude towards IT system operation and security than corporate IT staff, and frequently a mutual lack of trust has to be overcome to implement effective security architecture.

3 Building secure automation systems

Building secure systems is difficult, as it is necessary to spread effort and budget so that a wide variety of attacks are efficiently and effectively prevented. For automation systems the challenges mentioned in Section 2.3 create additional difficulties. In the following, two common approaches to secure systems are explained and their effectiveness is assessed. By looking to that two examples it is even more important to think thoroughly about security.

3.1 Hard perimeter

A popular doctrine for defense, be it of cities or IT systems, is the notion of the hard perimeter. The idea is to have one impenetrable wall around the system and to neglect all security issues inside. In general, however, this approach does not work, for a variety of reasons. The hard perimeter approach does not make use of reaction capabilities: At the time of detection of a successful attack, the attacker has already broken through the single wall and the whole system is open to him. In consequence, this means that the wall would need to be infinitely strong and think, because it needs to resist infinitely long [14]. Also, monoculture is dangerous: The wall is based on one principle or product. If that principle or product fails for some reason to resist the attack, the whole defense is ineffective. The wall must have doors to be usable, which opens it to both technical and non-technical (social engineering) security risks. Once the attacker has managed to sneak inside, the system is without defense - the risk of the proverbial Trojan horse. A hard perimeter is also, by definition, ineffective against insider attacks. Progressing technology gives the attacker continuously better wall-penetration capabilities. Last, but not least, humans make mistakes:

It is illusory to assume that we can design a wall that is without weak spots either in design, implementation, or operation - various border walls in history serve as example.

3.2 Defense-in-depth

The alternative approach is defense-in-depth. Here several zones/shells are placed around the object, which is to be protected. Different types of mechanisms are used concurrently around and inside each zone to defend it. The outer zones contain less valuable targets, the most precious goods, in this case the (safety critical) automation system, are in the innermost zone. In addition to defense mechanisms there are also detection mechanisms, which allow the automation system operators to detect attacks, and reactive mechanisms and processes to actively defend against them. Each zone also buys time to detect and fend off the attacker. In the spirit of Schwartau's time-based security [14] this allows to live with the fact of imperfect protection mechanisms, as only a security architecture strength of $P>=D+R$ has to be achieved, where P is the time during which the protection offered by the security system resists the attacker, D is the delay until the ongoing attack is detected, and R is the time until a defensive reaction on the attack has been completed.

Conclusion: There are two basic approaches for securing systems commonly used today, but only one, defense-in-depth, will result in a secure system, provided it is properly implemented.

4 Elements of a security architecture

In this section the most important technical elements of security architecture for utility systems are surveyed. Note however, that a system cannot be secured purely using technical means. Appropriate user behavior is essential to ensure the effectiveness of any technical means. Acceptable and required user behavior should be clearly documented in a set of policies, which are strictly and visibly enforced by plant management. Such policies should address among other things user account provisioning, password selection, virus checking, private use, logging and auditing, etc.

The topic of policies and user behavior will not be further discussed here. Example documents are available from various government agencies, IT security organizations, as well as in a number of books. According to their physical and logical location in the architecture - in the spirit of a defense-in-depth -, security mechanisms can be classified as belonging to one or multiple of the following categories. These categories are orthogonal to the security objectives of Section 1.

Deterrence Means of pointing out to the potential attacker that his personal pain in case of getting caught does not make the attack worthwhile. However, in most threat scenarios for safety-critical and infrastructure systems the deterrence component, especially the threat of legal action, is ineffective. A warning sign at the location should appear and state clearly that unauthorized access is prohibited and legal action may be the consequence.

Connection authorization Means to decide whether the host trying to initiate a communication is at all permitted to talk to the protected system, and to prevent such connections in case of a negative decision. Here mechanisms have to be installed like firewalls, routers and switches supporting IP Security (IP-Sec), call back communication systems.

User authorization Means to decide whether and with which level of privileges a user or application is permitted to interact with the protected system and to prevent such interaction in case of a negative decision. Different login mechanism has to be chosen to have a clear user authentication. Selection of a suitable password in case a fixed password is used, or an one-time password will be generated from the system itself. Other methods are the detection of biometrics like fingerprints or iris recognition. Such systems are rather complicate and expensive. In any case, if authentication fails, the user has to initiate a new session to continue communication with the protected system, e.g. for another login attempt.

Action authorization Means to decide whether a user or application is permitted to initiate specific actions and action sequences on the protected system or application, and to prevent such interaction in case of a negative decision. Action authorization is an additional barrier assuming a preceding positive user authorization decision. This can be either a role based authorization or a restriction of the interaction from an outside system. Dual authorization can be used for emergency commands where in case always a confirmation from a second authorized user is required.

Intrusion detection Means to detect whether an attacker has managed to get past the authorization mechanisms. Most intrusion detection systems are based on monitoring and detecting whether anything "unusual" is going on in the system. One possibility is to have all actions of the authentications devices logged, and these logs are manually or automatically screened for unusual occurrences or patterns, e.g. that an authorized user is suddenly accessing the system outside his normal work hours or access time. A network-based intrusion detection system (NIDS) tries to discover attacks based on known attack profiles and/or unusual system behavior from communication traffic seen on a network segment (type, content, frequency, path of the transmitted messages). Authentication failure alerts and log analysis are further tools to detect intruders.

As is shown in [8], network-based electronic attacks originating from malicious devices in an automation system e.g. in a power substation can be categorized as either message injection, message modification, or message suppression. Using a suitable communication protocol for detection of invalid messages, one can reduce these three categories to message suppression, which can in many cases be regarded as a system failure that conventional fault-tolerance and fault-response mechanisms such as redundant devices and emergency shutdown sequences can handle.

Response Means to remove an attacker and the damage done by him from the system. Means to lessen the negative impact of the attack on the system and its environment. Means to prevent a future

recurrence of the same type of attack.

By isolating the connections between the compromised subsystem, e.g. the outer part of the security zone at the interface between automation system and other networks, and other, more important parts of the automation system, are closed to avoid further spreading of the attack. Depending on system/remote access functionality and importance, and on whether delaying the attacker and collection of further evidence, or quick restoration of operation is of higher importance, it is an option to shut-down all remote connections, both for the affected and the not-yet affected systems, until the effects of the attack are removed. Like electric power grids, the automation system should already be architected and designed such that it can be partitioned into zones, which can be isolated with minimum disturbance of the whole system.

Further activities could be the activation of dedicated safety mechanisms, generating new passwords randomly or switch over to back up system with the minimum functionality.

Mechanism protection Means to protect the mechanisms for the above listed categories against subversion. This refers, for example, to not sending passwords in clear text over public networks and hardening operating systems and applications by fixing well-known bugs and vulnerabilities, etc. A dedicated physical network or virtual private network (VPN) are scenarios which addressing it.

5 Further reading

A large number of technical and research publications exist on the issue of IT security for home and office information systems. [13] Gives a good general introduction into the topic, and [12] is the reference on cryptographic algorithms and protocols. [9] is a comprehensive resource on the practical issues of securing a computer network, while [1] and [15] address the issue of engineering secure (software) systems from a larger perspective.

On the other hand, apart from some vendor whitepapers, there is almost no literature on the specific security needs and capabilities of industrial automation systems. [11] investigates remote access to automation systems, specifically home automation systems, with potentially malicious devices and proposes the use of smartcards as trusted processors to achieve end-to-end security between each device and its legitimate communication partners. In [7] IT security mechanisms applicable for automation systems are presented according to which conceptual zone they defend - remote access, operator workstations, or automation devices. [3] investigates networking and network-level security issues for Ethernet networks on the plant floor. [5] motivates and describes efforts to create a protection profile for process control systems according to the Common Criteria security evaluation standard. [6] reports on a survey about IT security conducted among automation system users. [4] promotes the use of a Public key Infrastructure (PKI) for process control systems, and [2] suggests a lightweight PKI for power utility SCADA systems. [10] applies standard IT security mechanisms to power utility control systems, with a special emphasis on password management.

6 Conclusion

The security situation grew with the complexity and openness of the automation systems. In the same way the complexity of the security possibilities are growing as showed in this report. It allows today the implementation of a true defense–in-depth application, provided it is properly implemented. It is illusory to assume that system just protected by its surrounded building and a basic authorization can be seen as secure, compared to the today's possibilities to break in a system. A true defense-in-depth with its elements of a secure architecture, will result in a secure system.

It should be noted that the vendor got some additional responsibility issues in security of their systems. It will be a challenge for the vendor of power utility SCADA systems to implement them. More regular standards would be positive, so that the vendors have to follow them, with the result, that security systems are properly implemented and installed.

7 Research issues

As can be seen from the scarcity of published work, automation system IT security is a comparatively new field. Much more research will be necessary until both security requirements and opportunities specific for utility automation systems have been explored to the current level of home and business system IT security.

These are just some of the topics to be addressed in future:

Considering that plant control systems have a life-time of 20 to 30 years, how can effective defense-in-depth security mechanisms cost-efficiently be retrofitted onto them?

What security mechanisms can and should be required for automation systems? This topic is being currently addressed by various industry standard organizations, such as the ISA SP99 working group.

What criteria should be used for assessing and auditing the security mechanisms in automation devices?

How to ensure appropriate levels of access control, data integrity and data confidentiality for automation devices with low computing power [16]?

How can a plant operator, who is not an IT security expert, effectively contribute to plant security?

8 References

[1] Ross Anderson. Security Engineering. Wiley, 2001.

[2] Cheryl Beaver, Donald Gallup, William Neumann, and Mark Torgerson. Key management for SCADA. Technical Report SAND2001-3252, Cryptography and Information Systems Security Department, Sandia National Laboratories, March 2002.

[3] Eric Byres. Designing secure networks for process control. IEEE Industry Applications Magazine, 6(5):33-39, Sep/Oct 2000.

[4] Ferdinand J. Dafelmair. Improvements in process control dependability through internet security technology. In Proceedings Safecomp 2000, volume 1943 of LNCS, pages 321-332. Springer, 2000.

[5] Joe Falco, Keith Stouffer, AlbertWavering, and Frederick Proctor. IT Security for Industrial Control Systems. Technical report, Intelligent Systems Division, (US) National Institute of Standards and Technology (NIST), 2002.

[6] Bill Moore, Dick Slansky, and Dick Hill. Security strategies for plant automation networks. Technical report, ARC Advisory Group, July 2002.

[7] Martin Naedele. IT Security for Automation Systems - Motivations and Mechanisms. atp, 45(5), May 2003.

[8] Martin Naedele, Dacfey Dzung, and Michael Stanimirov. Network security for substation automation systems. In Udo Voges, editor, Computer Safety, Reliability and Security (Proceedings Safecomp 2001), volume 2187 of LNCS, 2001.

[9] Stephen Northcutt, Lenny Zeltser, Scott Winters, Karen Fredrick, and Ronald W. Ritchey. Inside Network Perimeter Security: The Definitive Guide to Firewalls, Virtual Private Networks (VPNs), Routers, and Intrusion Detection Systems. Que, 2002.

[10] Paul Oman, Edmund Schweitzer, and Deborah Frincke. Concerns about intrusions into remotely accessible substation controllers and SCADA systems. Technical report, Schweitzer Engineering Laboratories, 2000.

[11] Peter Palensky and Thilo Sauter. Security considerations for FAN-Internet connections. In proceedings 2000 IEEE International Workshop on Factory Communication Systems, 2000.

[12] Bruce Schneier. Applied Cryptography. Wiley, 2nd edition, 1996.

[13] Bruce Schneier. Secrets and Lies - Digital Security in a Networked World.Wiley, 2000.

[14] Winn Schwartau. Time based Security. Interpact Press, 1999.

[15] John Viega and Gary McGraw. Building Secure Software. Addison-Wesley, 2001.

[16] Thomas P. von Hoff, Mario Crevatin. HTTP Digest Authentication in Embedded Automation Systems, 9th IEEE International Conference on Emerging Technologies and Factory Automation, Lisbon, Portugal, 2003, accepted for publication.