Security for Industrial Communication Systems

DACFEY DZUNG, MEMBER, IEEE, MARTIN NAEDELE, THOMAS P. VON HOFF, AND MARIO CREVATIN, MEMBER, IEEE

Invited Paper

Modern industrial communication networks are increasingly based on open protocols and platforms that are also used in the office IT and Internet environment. This reuse facilitates development and deployment of highly connected systems, but also makes the communication system vulnerable to electronic attacks. This paper gives an overview of IT security issues in industrial automation systems which are based on open communication systems. First, security objectives, electronic attack methods, and the available countermeasures for general IT systems are described. General security objectives and best practices are listed. Particularly for the TCP/IP protocol suite, a wide range of cryptography-based secure communication protocols is available. The paper describes their principles and scope of application. Next, we focus on industrial communication systems, which have a number of security-relevant characteristics distinct from the office IT systems. Confidentiality of transmitted data may not be required; however, data and user authentication, as well as access control are crucial for the mission critical and safety critical operation of the automation system. As a result, modern industrial automation systems, if they include security measures at all, emphasize various forms of access control. The paper describes the status of relevant specifications and implementations for a number of standardized automation protocols. Finally, we illustrate the application of security concepts and tools by brief case studies describing security issues in the configuration and operation of substations, plants, or for remote access.

Keywords*—*Cryptography, embedded systems, industrial automation, industrial communication systems, remote access, security objectives, security protocols, security standards.

I. INTRODUCTION

A. Motivation

Industrial automation systems are used in application domains like discrete, batch, and process manufacturing; electric power generation and distribution; gas and water supply; and transportation. Depending on the type and purpose of the automation system, its components are distributed on a local, wide-area, or even global scale. Communication links and facilities, as described in the other papers of this issue, are an important element of such distributed control systems (DCSs). In the past, automation systems were not linked to each other and were not connected to public networks like the Internet. Today, the market puts pressure on companies to make fast and cost effective decisions. For this purpose, accurate and up-to-date information about the plant and the process status have to be available not only on the plant floor, but also at the management level in the enterprise and even for supply chain partners. This results in increasing interconnection between different automation systems as well as between automation and office systems. Initially, such interconnections were based on specialized, proprietary communication mechanisms and protocols. Today, open and standardized Internet technologies as well as the Internet itself are increasingly used for that purpose.

In the terminology of information system security, a risk exists if there is a vulnerability and a threat. A vulnerability is the opportunity to cause damage. A vulnerability of an information system may be caused by a logical design flaw (e.g., a badly designed protocol), an implementation flaw (e.g., a buffer overflow), or a fundamental weakness (e.g., passwords and cryptographic keys that can be guessed). A threat arises from an attacker trying to find and exploit the vulnerability in order to inflict damage. Damage may also be caused by an incidental, nonintentional exploitation of a vulnerability.

The industrial communication systems of today are to a large extent based on commercial operating systems, protocol implementations, and communication applications which are known to have vulnerabilities. By connecting to the Internet, or other public networks, these vulnerabilities are exposed to potential attackers. Attackers need expertise and motivation. With open Internet technologies, the expertise is easily available. Motivations for attacking industrial communication systems may be political or economical:

> Water, electricity, [...] and other critical functions are directed by computer control [...]. The threat is that in a future crisis a criminal cartel, terrorist group, or hostile nation will seek to inflict economic damage [...]

Manuscript received March 25, 2004; revised February 4, 2004.

The authors are with ABB Corporate Research, Baden CH-5405, Switzerland.

Digital Object Identifier 10.1109/JPROC.2005.849714

by attacking those critical networks. [...] The threat is very real [1].

Deregulation and competition among power utilities has created what IEEE-USA calls "financial incentives for malicious intrusion into computers and communication systems of the electric power industry and marketplace participants." [2]

In summary, large and interconnected networks of industrial communication and automation systems are vulnerable to electronics attacks, and threats exist. Therefore, security has become an important issue.

B. Reported Incidents

Actual occurrences of both targeted and untargeted security incidents at critical industrial and infrastructural plants demonstrate that the threat is real [3]:

In August 2003, a worm infected the communication system of the U.S. railway company CSX Transportation. The dispatching and signaling systems were affected and all passenger and freight traffic, including morning commuter traffic in the Washington, DC, area, had to be stopped for about half a day [4].

In January 2003, the "Slammer" worm disabled the computerized safety monitoring system at the Davis-Besse nuclear power plant in Ohio, which was shut down for repair at that time. The responsible managers considered the plant "secure," as its outside network connection was protected by a firewall. The worm entered the plant network via a contractor's infected computer connected via telephone dial-up directly to the plant network, thus bypassing the firewall [5], [6].

In March 2000, a former consultant to a waste water plant in Maroochy Shire, Queensland, Australia, accessed the control system of the plant and released up to 1 million L of sewage into the surrounding waterways [7].

The Internet Engineering Lab of the British Columbia Institute of Technology has set up an industrial control system security incident tracking database which in spring 2004 contained around 41 entries, with some additional investigations pending [8].

These examples show that security vulnerabilities in industrial automation and communication systems pose the risk of financial damage for the plant owner, as well as harm to humans and the environment. Industrial communication systems share some security-relevant characteristics with information and communication systems in the office and Internet domain, but they also exhibit major differences (see Section III-A), which create both obstacles and advantages for securing them.

C. Outline

This paper presents an overview of the state-of-the-art security technologies and best practices for industrial communication system security, and a look into standardization in this area. In Section II, general aspects of IT security are reviewed. This covers the general security objectives, typical attacks against them and countermeasures such as cryptography-based security in communication protocols, as well as design principles for secure system architecture and operation. Section III presents security-relevant characteristics of industrial communication systems and analyzes the main types of industrial and utility communication network topologies and protocols with respect to their security features. Section IV surveys general IT and industrial-communication-specific security standards. Section V illustrates the application of the concepts and tools discussed in the previous sections in case studies covering substation automation, plant automation, and remote access to a stand-alone embedded device.

II. Communication Security Fundamentals

A. Security Objectives

When looking for a definition of security for a certain system, which shall serve as a base for a security policy or a system specification, we can distinguish three main perspectives.

The first perspective deals with the question how the attacker reaches the target information system: in person, in which case the countermeasures are collectively called physical security, or via the electronic network, in which case the countermeasures are network and system security. This paper is not concerned with physical security.

The second perspective looks at what fundamental types of threats the system is to be secured against, or in other words, the security objectives that are to be achieved. The eight security objectives explained in the following offer a framework for categorizing and comparing the security mechanisms of various systems:

- **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.
- **Integrity:** The integrity objective refers to preventing undetected modification of information by unauthorized persons or systems. For automation systems, this applies to information such as product recipes, sensor values, or control commands. 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.
- **Availability:** Availability refers to ensuring that unauthorized persons or systems cannot deny access or use to authorized users. For automation systems, this refers to all the IT elements of the plant, like control systems, safety systems, operator workstations, engineering workstations, manufacturing execution systems, as well as the communication systems between these elements and to the outside world. Violation of availability, also known as denial-of-service (DoS), may

not only cause economic damages but may also affect safety issues as operators may lose the ability to monitor and control the process.

- **Authentication:** Authentication is concerned with determination of the true identity of a system user and mapping of this identity to a system-internal principal (e.g., valid user account) by which this user is known to the system. Most other security objectives, most notably authorization, distinguish between legitimate and illegitimate users based on authentication.
- **Authorization:** The authorization objective, also known as access control, is concerned with preventing access to the system 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 plant control system. Violation of authorization may cause safety issues.
- **Auditability:** Auditability is concerned with being able to reconstruct the complete history of the system behavior from historical records of all (relevant) actions executed on it. 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. Note that auditability without authentication may serve diagnostic purposes, but does not provide accountability.
- **Nonrepudiability:** The nonrepudiability objective refers to being able to provide irrefutable proof to a third party of who initiated a certain action in the system, even if this actor is not cooperating. This security objective is relevant to establish accountability and liability. In the context of automation systems, this is most important with regard to regulatory requirements, e.g., U.S. Food and Drug Administration (FDA) approval. Violation of this security objective has typically legal/commercial consequences, but no safety implications.
- **Third-party protection:** The third-party protection objective refers to averting damage done to third parties via the IT system, that is, damage that does not involve safety hazards of the controlled plant itself: 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 DoS (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. (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.)

Some of these security objectives are to a certain extent independent of each other, and for most systems only a subset of the security objectives will have a high priority or be applicable at all.

The third perspective is concerned with the level of confidence one can have in the means that are in a particular system used to achieve the relevant security objectives. The level of confidence can be defined in terms of formal certification results, like the Common Criteria (CC) evaluation (see Section IV-A1), or in terms of standards and generally accepted best practices (see Section II-E). Typically, the level of confidence in the security mechanisms of a system decreases over time, as new attacks and vulnerabilities become known, so the security architecture and implementation need to be reviewed regularly and updated as necessary.

B. Types of Attacks

Depending on its specific function and environment, each industrial communication system has to satisfy a subset of the security objectives described in Section II-A. An intentional violation of a security objective is called an attack. Attacks may either be initiated by persons outside the plant or by insiders. We distinguish between targeted and untargeted attacks. Targeted attacks intend to harm a specific communication system or type of system, e.g., for purposes of industrial espionage, warfare, or terrorism. Untargeted attacks victimize any vulnerable system they discover. Targeted attacks are typically preceded by a phase of gathering information about the target, e.g., using online and offline available references, as well as dedicated tools for discovering vulnerable systems on a network [9]. Some common types of attacks are the following.

- **DoS:** The goal of the attacker is to decrease the *availability* of the system for its intended purpose.
- **Eavesdropping:**The goal of the attacker is to violate the *confidentiality* of the communication, e.g., by sniffing packets on the LAN or by intercepting wireless transmissions.
- **Man-in-the-middle:** In a man-in-the-middle attack, the attacker acts toward both end points of the communication as if the attacker were the expected, legitimate partner. In addition to *confidentiality* violations, this also allows modifying the exchanged messages (*integrity*). Via man-in-the-middle attacks, weaknesses in the implementation, or usage of certain key exchange and authentication protocols, can be exploited to gain control even over encrypted sessions [10].
- **Breaking into a system:** Through violation of the *authentication* and *access control* objectives, the attacker obtains the ability to control aspects of the behavior of the communication system and the connected plant at his will, including the ability to overcome *confidentiality* and *integrity* objectives. A break-in usually involves the consecutive penetration of multiple subsystems and the step-wise elevation of the privileges of the attacker.

- **Virus:** A virus-based attack manipulates a legitimate user to bypass *authentication* and *access control* mechanisms in order to execute the malicious code injected by the attacker. In practice, virus attacks are often untargeted and spread among vulnerable systems and users. Virus attacks often directly or indirectly decrease the *availability* of infected systems by consuming excessive amounts of processing power or network bandwidth.
- **Trojan:** A Trojan is a virus where the malicious functionality is hidden behind functionality that is desired and used by the user. Trojans are typically employed to circumvent *confidentiality* or *access control* objectives.
- **Worm:** A worm is malicious code whose propagation mechanisms rely on automatic exploration and exploitation of vulnerabilities in the targeted system, without involvement of any user. Worm infections are untargeted and usually create *availability* problems for the affected systems or even the Internet as a whole [11]. In addition, the worm may carry malicious code to launch a distributed, targeted attack from all the infected hosts.

C. Cryptographic Methods

Among the security objectives mentioned in Section II-A confidentiality, integrity, authentication, and nonrepudiability are achieved by cryptographic methods. Cryptographic algorithms are employed for secure data storage and for secure transmission. For secure data transmission involving more than one party, the *algorithms* must be embedded in cryptographic *protocols* which define the sequence of steps to be undertaken by the participating parties. For comprehensive overviews of cryptographic algorithms and protocols, see [12] and [13].

1) Cryptographic Algorithms: Before the invention of modern cryptographic techniques, cryptography relied on keeping the algorithm secret. However, modern cryptography follows Kerckhoffs' principle, which states that only the *key* must be kept secret, while the encryption algorithm may be public [14]. The longer an algorithm has been published with no vulnerabilities discovered, the more the confidence in its robustness against attacks increases. However, it cannot be excluded that unpublished attacks exist or that new cryptanalysis methods will be found to break the algorithm. In the long term, cryptographic algorithms with given key lengths become weaker against brute-force attacks (exhaustive testing of all possible keys), as the computational power available to attackers keeps growing.

Symmetric encryption algorithms are characterized by the fact that the decryption key is identical to the encryption key. The key must be exchanged in advance between sender and receiver in a secure manner and must be kept secret. There are stream ciphers and block ciphers: stream ciphers combine input data bit- or byte-wise with a key stream, while block ciphers transform input data block-wise in a key-dependent way. There exists a wide range of symmetric algorithms with different characteristics [12], [13]. Typical symmetric algorithms are the following.

- RC4 is a stream cipher with byte-wise processing [15]. This makes it attractive for implementation on 8-bit processors.
- DES is a block cipher defined as U.S. standard for encryption in 1977 [16]. Due to the relatively short key length (56 bits), DES in its original form is not recommended for new systems. Key length can, however, be easily doubled by applying DES three times in a process known as Triple DES [17].
- Advanced Encryption Standard (AES) is the current U.S. standard for encryption [18]. By design, it is at least as secure as Triple DES, but much faster and suitable for different processor word lengths. The AES processes 128-bit blocks and its key size is 128, 192, or 256 bit. A key size of 128 bit is considered strong enough today.

The block algorithms mentioned above encrypt only a fixed-size block. There are different modes to encrypt a message of arbitrary length by processing a sequence of blocks. The simplest one, the electronic code book (ECB), encrypts block by block separately. However, this procedure has severe cryptographic weaknesses if the plain text contains repeated patterns. To avoid this problem, other modes are defined where information of previously encrypted blocks influence the encryption of the current block. These are cipher block chaining mode (CBC), cipher feedback mode (CFB), output feedback mode (OFB), and the counter mode (CTR). A detailed description can be found in [19].

With symmetric encryption, an individual secret key is required for each communication link. Hence, a system with n participants needs $n(n-1)/2$ keys. The distribution of the keys becomes difficult to scale for increasing n.

Public-key algorithms [20] have different encryption and decryption keys and the latter cannot be derived from the former by any efficient algorithm. The encryption key is made public. Any prospective sender can encrypt the message using the public key of the receiver, but only the receiver owning the corresponding private key can decrypt the message. Hence, n participants need only n key pairs. However, compared to symmetric algorithms, public-key algorithms are slower by an order of magnitude and require longer keys to achieve the same level of security [21]. The best known public-key encryption algorithms are the following.

- RSA [22]: Its security is based on the difficulty of factoring large numbers into their prime factors, i.e., decryption, given the public key, is believed to be equivalent to solving the factorization problem. For large keys (large integers), no efficient factorization algorithms are known today.
- ElGamal [23]: Its security is based on the difficulty of computing the discrete logarithm in a finite field.

Elliptic curves over finite fields allow the implementation of faster public-key cryptosystems with a smaller key size (see [24]).

One-way hash functions are important building blocks for various cryptographic protocols. They are used to generate

message digests (also called cryptographic checksums, or fingerprints). A one-way function is characterized by following properties.

- A finite-length output (hash) is calculated from an arbitrary-length input.
- It is easy to calculate the output.
- Given an output, it is hard to find a corresponding input.
- Modification of one bit in the input message leads to modification of about half of the bits in the output message.

One-way hash functions are mainly used for integrity protection and authentication. The most widely used hash functions are MD5 and SHA. Many collisions (messages with the same hash) are found with MD5 [25], so its use is no longer recommended. The Secure Hash Standard (SHS) [26] defines a family of hash algorithms. The algorithms differ in the number of generated hash bits. It is recommended to use a high number of hash bits whenever possible.

2) Message Authentication: There are three levels of protecting integrity and authenticity of a message.

Integrity protection: The integrity can be protected by a cryptographic checksum (hash) and secure storing or transmitting the checksum together with the message for later verification.

Message Authentication: If the message and its checksum are transmitted between two entities, the checksum is calculated over the message concatenated with a secret shared by the two entities [27]. This message authentication code can then only be verified by entities which share the secret.

Digital Signature: The purpose of a digital signature is to prove message integrity and origin. Digital signatures are realized using public-key cryptography. The source of a message uses its private key to generate a message signature by encrypting the message or its hash. Any receiver can verify the validity of this signature by decrypting it using the originator's public key and comparing it with the original message or its hash, respectively. Any public-key encryption scheme can be used as signature scheme. A Digital Signature Standard (DSS) was defined in [28]. The predominant algorithm for digital signatures is RSA as defined in [29].

3) Hybrid Encryption: As mentioned, symmetric encryption requires a much larger number of keys than public-key encryption to protect communication between a given number of participants. However, symmetric algorithms are much faster than public-key algorithms. Hybrid encryption combines the advantages of both types. First, a public-key based *key exchange* (e.g., by the Diffie–Hellman key exchange protocol [20]) takes place to agree on a session key. Then the session key is used for actual data encryption using a symmetric encryption algorithm.

4) Key Distribution: A prerequisite to secure communication is the secure distribution of the keys. Regardless of the method used to distribute a key from A to B, B has to authenticate the key purportedly received from A. In addition, the distribution of secret keys requires confidentiality.

- A *key server* is a trusted entity that hands out all keys. To use the service of this system, a secret key shared with the key server is required for each participating entity. When A wants to communicate securely with B, A asks the key server to generate a secret key to share between A and B. The key server sends this key to A in two versions, each encrypted with the key shared with A and B, respectively. The most popular key management system based on key servers is Kerberos [30]–[32].
- A *public-key infrastructure* (PKI) supports distribution and authentication of public keys. The central role is played by a certificate authority (CA). An entity generates its private/public key pair and presents the public key to the CA for certification. If the CA has identified the entity successfully, it digitally signs a document containing the public key, information about the key owner, the CA, and the expiration date of the key. These documents are called certificates. The most common standard for their format is X.509 [33], and [34] profiles its use in the Internet. Any receiver with the public key of the CA can now authenticate the owner's public key in the certificate. [35] describes a smartcard-supported PKI for a process control environment.
- *Distributed key management* operates without a central institution. Every participant generates its own public/private key pair and may sign the public key of others. If A wants to communicate with B, but does not know B's public key, A requests it from a trusted participant C whose public key is known to A and who knows B's public key. Once A has obtained B's public key, A can use it for secure communication with B, and A can sign it as well. Thus, the users rely on, and build up, a nonhierarchical trust relationship (*web of trust*). Such webs of trust are used by PGP (see Section II-D4).

5) Entity Authentication: In entity authentication protocols, an entity proves a claimed identity to another entity [36]. There are two categories of authentication protocols, differing in whether a trusted third party is involved or not. In the case where no third party is involved, the two entities must have previously established a shared secret.

- The weakest form of authentication is based on passwords. Here, the authenticating entity asks the other entity to send his username and password in order to prove its identity and checks whether this pair belongs to the list of authorized users. The weakness of this procedure is its vulnerability to eavesdropping, since username and password are transmitted as clear text.
- The *challenge/response mechanism* avoids sending the password in clear text. Instead, the authenticating entity A challenges the entity B by sending a random nonce (number used once). B calculates a response from the shared secret and the challenge using some cryptographic operation, e.g., a hash calculation. A makes the same calculation and compares its result

Table 1
Network Layers and Common Security Protocols

Layer	Protocol	Security Protocol	Confidentiality	Integrity	Authentication	to be secured
Applications	SOAP	WS-Security	yes	yes	data origin	document parts
	SMTP	PGP/GnuPG	yes	yes	message	mail content
		S/MIME	yes	yes	message	
	HTTP	HTTP Digest Authentication	no	no[a]	user[b]	user-to-server
Transport layer	TCP	SSH Transport Layer Protocol	yes	yes	server[c]	client(user)-to-server
		SSL/TLS	yes	yes	server[d]	client-to-server
Internet layer	IP	IPSec	yes[e]	yes	host[f]	host-to-host
Link layer	PPP	CHAP/PAP	no	no	client	end-point of link
	Bluetooth	Bluetooth Security	yes	yes	device	air interface
	WLAN IEEE802.11	WEP/WPA/802.1X	yes	yes	device	

[a] optional, but usually not implemented
[b] server (or mutual) authentication optional
[c] user authentication provided by SSH User Authentication Protocol
[d] client authentication optional
[e] optional, only in Encapsulated Security Payload (ESP)-mode
[f] data origin authentication optional, only in Authentication Header (AH)-mode

with the response from B. If they coincide, the B is authenticated to A. The randomness of the nonce prevents attacks with precomputed or replayed responses. Hence, the authenticator A can verify the freshness of B's response. Alternatively, public-key cryptography can also be applied: B's response is its signature of the nonce and can be verified by A using B's public key.

In authentication procedures with *trusted third-party* support, two entities A and B delegate their mutual authentication to a third party. This third party is usually the key server, which also performs authenticated key distribution, as described above. The best known, trusted third-party authentication system is Kerberos [30]–[32], which is based in part on the Needham–Schroeder protocol [37].

D. Security in Communication Protocols and Networks

In communication networks, security objectives are achieved by security protocols at different layers of the communication network [32]. Table 1 contains a selection of the most common security protocols and the services they provide. The emphasis of these security protocols is to protect against network-based attacks on the communication links. This can be provided largely independent of the application, so that industrial communication systems may rely on these protocols for *network security*, as described in this section. In contrast, *access control* issues are mostly specific to the requirements of the application. Hence, Section III-C on security in industrial automation protocols emphasizes access control features.

The security services provided by a certain layer secure the link between the end points pertaining to that layer, and should be transparent to the layers above. The security protocols on the network and the transport layer, IPSec and SSL, respectively, are currently most widely deployed.

1) Link Layer Security: As extensions to the Point-to-Point Protocol (PPP) [38], the Password Authentication Protocol (PAP) and the stronger Challenge Handshake Authentication Protocol (CHAP) [39] provide authentication. While the former is password-based, the latter uses a challenge/response mechanism (see Section II-C). The Extensible Authentication Protocol (EAP) [40] defines a general framework for authentication.

Security for short range wireless links: Wireless short range links such as Bluetooth or IEEE 802.11 wireless LANs (WLANs) are particularly vulnerable. DoS attacks by radio jamming can be done easily. Ultimately, there are no means to protect against such attacks, but at least they are easily detected. More subtle attackers may eavesdrop on the traffic or even break into the automation system. Protection measures against eavesdropping and active attacks exist, however, for modern short range wireless communication systems:

Bluetooth: Bluetooth authentication and encryption rely on a shared "link key" between each pair of devices [41]. To establish a link key between the Bluetooth-enabled automation controller and e.g., some handheld configuration device, the two devices must be "paired" the first time they connect to each other, by entering an identical personal identification number (PIN) on both devices. This pairing is critical and should be performed in an environment which is secure against eavesdroppers. As an alternative to such a link layer specific solution, some of the higher layer security protocols described below, in particular IPSec, could be applied. However, this would leave link layer management messages unprotected and vulnerable to spoofing attacks.

IEEE 802.11 Wireless LAN: WLAN systems have higher transmission power, i.e., higher range, than Bluetooth systems, and are therefore more vulnerable to eavesdropping and active attacks. The IEEE 802.11 optional security standard, Wireless Equivalent Privacy (WEP), defines an encryption service based on the RC4 stream cipher (symmetric encryption). However, its problematic usage of initialization vectors has allowed WEP to be cryptographically broken if the attacker can capture sufficient data [42]. The IEEE 802.11i

task group has corrected these problems by increasing the key length and improving the handling of the initialization vector in the Wireless Protected Access (WPA) RC4-based solution. The preferred long-term IEEE 802.11i solution replaces the encryption algorithm by AES, but this requires a hardware modification. WEP also assumes a network-wide preshared secret to derive the encryption key for all nodes. Since automatic key management is not included, WEP cannot be scaled to a high number of terminals. If, as is often the case in industrial automation applications, only a limited number of preconfigurable users must be supported, it may be sufficient to deploy encryption with static keys using WEP or its improved successors, together with MAC-address filtering. However, in more flexible networks with high security requirements, client and network must be securely authenticated and keys must be distributed automatically. The IEEE 802.1X protocol has been designed for port-based network access control in LANs; IEEE 802.11i specifies its use for WLANs [43]. With IEEE 802.1X, the client node and the WLAN access point first run the EAP [40]. The access point allows network access to the client node only after it has successfully authenticated itself to the network. The network stores user account information typically on a remote access dial-in user server (RADIUS) [44]. A useful addition, or alternative to secure WLAN operation, is the deployment of IPSec/VPN solutions described below.

2) Internet Layer Security: Security at the Internet layer is provided by Internet Protocol Security (IPSec) [45]. IPSec is an optional extension to version 4 of the Internet Protocol (IPv4), but is a mandatory part of IPv6, and is supported by most modern operating systems. As protection is implemented at the IP layer, IPSec provides a single means of protection for all UDP and TCP applications. It is transparent to the upper layers and addresses the following security objectives for host-to-host communication:

- confidentiality (only in encapsulating security payload (ESP) mode [46]);
- data origin authentication (only in authentication header (AH) mode [47]);
- data integrity (in ESP and AH modes);
- access control, with respect to individual hosts (in ESP and AH modes).

ESP and AH are modes of IPSec and are described in [46] and [47]. Both modes can be applied together or alone. Additionally, the IPSec key exchange (IKE) [48] provides a framework for key management and negotiation of security associations (containing, e.g., the session key for symmetric encryption). IPSec provides machine-to-machine security, but does not perform authentication of individual users. Therefore, IPSec is predominantly deployed to establish virtual private networks (VPNs), but can also be used to provide security between end hosts. (Reference [10] describes a man-in-the-middle attack on VPN sessions which exploited implementation weaknesses.) Since IPSec's security associations are identified by the source and destination IP addresses, IPSec configurations may fail if some intervening network address translation (NAT) modifies IP addresses and ports. This is addressed in [49].

3) Transport Layer Security: Applications running on top of TCP (but not UDP) can be secured by the secure sockets layer (SSL) [50], or its extension, the transport layer security (TLS) standard [51]. Typical applications are HTTP and FTP, whose usage of SSL is indicated by the URL prefixes https and ftps, respectively. The main features of SSL are:

- session key management and negotiation of cryptographic algorithms;
- server authentication using certificates;
- confidentiality;
- data integrity protection.

SSL includes *client* authentication as an option, but this is rarely implemented, as it requires client certificates. Instead, client authentication is often implemented separately at the application level, under the confidentiality protection provided by SSL encryption.

The secure shell (SSH) transport layer protocol (SSH-TRANS) [52] provides confidentiality, server authentication, session key exchange, and integrity protection for remote login applications. SSH-TRANS supports the SSH Authentication Protocol (SSH-AUTH) and the SSH Connection Protocol (SSH-CONN), both running on top of it. While SSH-AUTH provides user authentication, SSH-CONN offers a number of network services, to mention secure remote login and secure TCP port forwardings, which cover functionalities of telnet and the Berkeley r-tools, e.g., rlogin [53].

4) Application Layer Security: Not all security objectives can be dealt with by the lower layer security protocols mentioned above. In particular, user and application authentication must be executed by the application layer.

Typical examples for application layer security are the following.

- *HTTP digest authentication (DA)* for user authentication following a challenge/response procedure [54]. By definition, it is foreseen that DA provides user authentication and integrity protection. However, the latter is not necessarily included in implementations. If an application does not require confidentiality and SSL and IPSec cannot be applied for memory and performance reasons, DA can be deployed instead [55].
- *Pretty Good Privacy (PGP)*. PGP [56], or its open-source derivative Gnu Privacy Guard (GnuPG), are used to encrypt and sign files for asynchronous data transfer (e.g., via e-mail) and for storage.
- *XML Web services*: The human-readable, self-documenting eXtensible Markup Language (XML) is increasingly used to exchange information and invoke services in a platform and application independent format. Various specifications are being developed to address the security objectives of fine granular confidentiality, integrity, authentication and authorization for XML documents or XML elements, both during transport and in permanent storage [57].

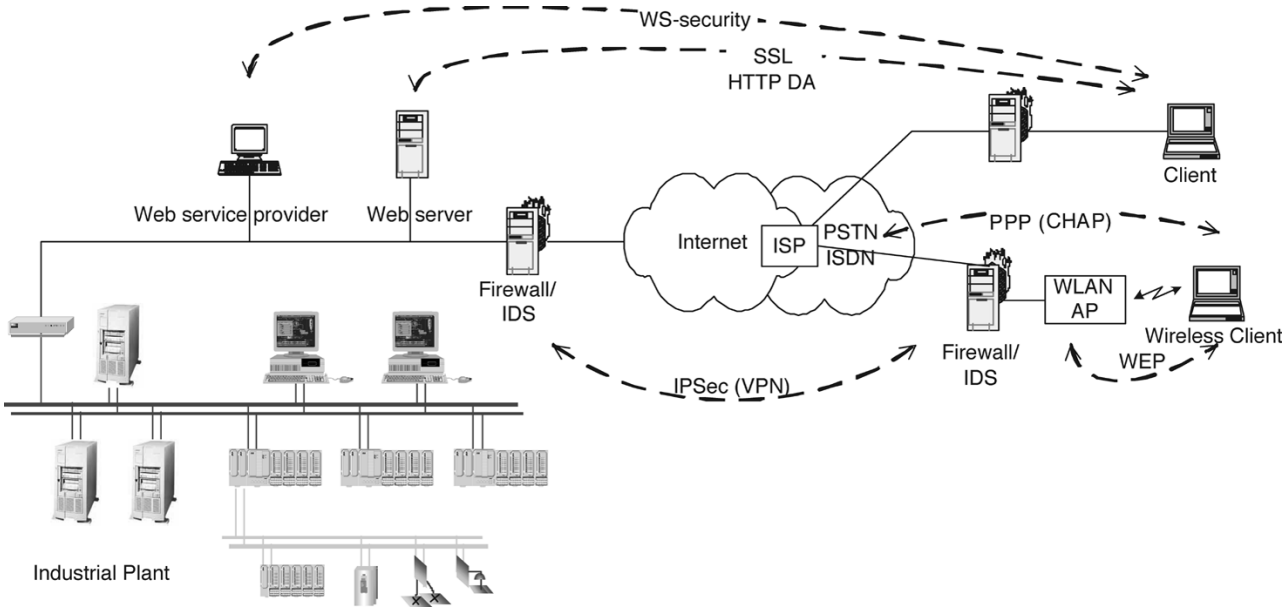

Fig. 1. Typical deployment areas of the different security protocols.

5) Firewalls: A firewall is a network equipment or software module deployed at the border of the network or network zone to protect it against unauthorized access from neighboring zones (e.g., the Internet) [58], [59]. It basically examines the protocol specific header fields of all incoming data packets and makes its admission decision based on those. Firewall types and products vary based on the protocol layers they inspect and the evaluation logic. The basic firewall types are the following.

- Filtering routers let packets pass or not dependent on the packet's IP source and destination address, as well as the protocol type.
- Packet filter firewalls also use TCP and UDP header fields like ports and flags for their admission decision.
- Stateful firewalls remember session state in order to decide whether a packet belongs to a previously established session, for instance.
- Application layer gateways relay requests on the application layer between the external and internal networks (using separate connections), after checking for their admissibility based on application layer data items and criteria.

6) Intrusion Detection Systems: An intrusion detection system (IDS) tries to discover attacks based on known attack profiles and/or unusual system behavior [59]. A network-based IDS (NIDS) obtains its information from the traffic observed on the network segment. (This information includes type, content, frequency, and path of the transmitted messages).

A host-based intrusion detection (HIDS) tries to discover attacks from information seen locally on the host on which it is running. A host-based IDS obtains its information from file system integrity checkers for example, which monitor whether important system files change without operational reason. The most important and security-relevant hosts such as firewalls and network-based IDS servers should be secured by a host-based IDS.

Fig. 1 shows a typical deployment of the different security protocols and network security equipment.

E. Security Architecture Best Practices

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. In this section, several design principles for secure systems will be surveyed. Some of these are also treated in [60].

1) Security Policy: The security policy states who has access to what and when, who is accountable for what, and what threats must be countered. It reflects the results of a risk evaluation, i.e., the security mechanisms must be in a reasonable relation to the mitigated risk. Without an explicit security policy, it is likely that effort is wasted on securing system aspects that do not need protection from the business point of view, and at the same time important threats are not mitigated. A security policy must thus be in place before one can start to design the security architecture, the body of technical and nontechnical security mechanisms, for the system.

2) "Security Is a Process, Not a Product": Due to changes in the operating environment and the availability of new attacks, no security system will be able to fulfill its purpose forever without maintenance. Maintenance starts with a regular review of rules, for instance for access control. This review should verify that the rules still correspond to the purpose of the system as stated in an up-to-date security policy. It should also establish that those rules reflect the current configuration of devices, users, and applications/services in the system. Other maintenance operations involve installation of patches for newly discovered application and operating system vulnerabilities, or even redesign or extension of parts of the system to cope with new threats or risks associated with new services and protocols. Examples for such new protocols are automation protocols based on HTTP and XML. The TCP/IP level filtering rules of conventional firewalls are to a large extent ineffective against such protocols.

3) Importance of the Weakest Link: The effort spent on the various security objectives required by a system has to be balanced so that all mechanisms facing an attacker are of comparable strength. Choosing a very strong encryption algorithm with a large key length, but then restricting the user password, from which this key is generated, to six-digit uppercase letters, or even transmitting this weak password in clear text as part of the authentication procedure, would only steer an attacker to break the password.

4) "Security by Obscurity" is Not Reliable Anymore: Automation systems used to be considered secure against electronic attacks because the vendors and plant owners assumed that no attacker would have the detailed technical knowledge necessary to exploit the proprietary protocols and systems for an attack [61]. If this were ever true, it is not true anymore. Due to the proliferation of plant automation, there is an increasing number of experts with knowledge of these proprietary systems. With the trend toward "open" and "standards" based systems, nowadays many attackers can use their knowledge of the office information and communication systems also against the systems used in industrial automation.

5) Least Privilege: One nonnegligible threat is insiders, i.e., authorized users who execute actions beyond their authorization. The principle of least privilege says that every user's authorization should be defined in a way that he/she has only the minimal permissions necessary to do his/her job. This requirement also requires technical security mechanisms being able to support such fine-grained authorization rules. For example, in embedded systems, where a single password is usually the maximum of protection for a device, this is often not the case.

6) Protect Your Secrets: Many automation systems contain information that should not be disclosed to the end user or be modifiable by him. Examples include proprietary algorithms or data that the vendor would like to keep secret from its customers, as well as authentication information like passwords that the customer defines. Ensuring that the information is really protected against more than the casual observer in a hostile environment is difficult, if not practically impossible [62]. There are basically five different approaches, presented here in order of increasing strength.

- Hiding/obfuscation: The secret is, perhaps after subjecting it to a transformation, placed in a nonobvious place in storage (memory, hard disk, or other storage media) [63]. If we assume that the attacker has technical means to debug software execution on hardware and software level, the attacker will be able to find the location of the secret or catch it while it is being transported for use in the CPU.
- Encryption: If an appropriately strong algorithm is chosen (see Section II-C), encryption protects information satisfactorily against a direct attack. However, in order to use the information, it must be decrypted in the system. The decryption makes the system susceptible to two types of attack: the attacker catches the information after it has been decrypted, or the attacker finds the decryption key that is stored somewhere in the system. Finding the key is harder, but not impossible, if the encryption key is not read from storage but is dynamically composed from (physical) characteristics of the system, such as a processor identification number.
- One-way functions: If the secret itself is not needed during program execution, but only a Boolean statement whether a certain data item is identical to the secret (typically the case in authentication) then a rather secure approach is to subject the secret to an irreversible transformation (see Section II-C) before storing it on the system. During execution, the secret is never exposed, as the data item is compared to the secret in the transformed form.
- Secure hardware subsystem: In some environments dedicated tamper-resistant hardware, such as smartcards [64] or the trusted platform module (TPM) [65], is available to store and protect certain secrets, e.g., cryptographic keys [62].
- Secure server: In some situations, it might be feasible to store the secret throughout the system lifetime, not in the system deployed in a hostile environment, but on a secure external server under control of the owner of the secret. In this case, all operations involving the secret are executed in the protected environment, and the low-level hardware/software debugging approach to attack will not work. This is the architecture used in e-business applications, for instance.

7) End-to-End Security: A security architecture has to take into account that the transmission phase is not the only target for attacks. In many cases, it is easier for an attacker to subvert the communication end points or intermediate nodes where the data are decrypted. Therefore, it is necessary to design security mechanisms to provide end-to-end protection from the data source application to the destination application.

8) Defense-in-Depth: There are two basic approaches for securing both physical and information systems commonly used today, namely hard perimeter and defense-in-depth.

The basic idea of the hard perimeter approach is to put a single impenetrable wall around the system and to disregard all security issues inside. There are several problems with this approach, such as lack of reaction capability resulting in the unrealistic requirement for a perfect defense, no diversity in mechanisms, and no protection against the malicious insider.

In the defense-in-depth approach, several zones, like 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, while the most critical automation systems are in the innermost zone. In addition to defense mechanisms, defense-in-depth also requires detection mechanisms which allow the automation system operators to detect attacks, as well as reactive mechanisms and processes to actively defend against attacks [66]. Each zone buys time to detect and fend off the attacker. With this approach, no defense mechanism has to delay the attacker for an indefinite length of time. The protection mechanisms have to resist the attacker only until the ongoing attack is detected and until a defensive reaction on the attack is completed. This approach to securing a system is called time-based security [67].

9) Standard Cryptographic Algorithms: Developing and implementing cryptographic algorithms without conceptual or implementational flaws is very hard. As the theoretical proof of the quality of an encryption algorithm is often impossible, expert evaluation for known and theoretical attacks, and "proven in use" experience are the main methods of evaluation. To conduct such a thorough evaluation that results in a satisfactory confidence in the quality of the algorithm takes a lot of time, effort, and expertise. Many very good cryptographic algorithms exist for a variety of usage scenarios and constraints (see Section II-C and [12], [13]). Therefore, there is hardly any reason to implement new and proprietary algorithms with unproven security characteristics. Instead, one of the standard algorithms should be used, preferably in the form of the implementation in a proven cryptographic library.

III. Security for Specific Industrial Protocols

A. Security-Relevant Characteristics of Industrial Communication Systems

1) Requirements: While office IT security requirements center around confidentiality and integrity issues, the operational requirements are different for automation and process control systems [68]. The most important requirement is safety; the absence of catastrophic consequences for humans and environment. Security violations may endanger plant safety. The second requirement, in priority order, is availability; the plant and the automation system have to be safe-operational over extended periods of time. This requirement, in many cases, precludes using standard IT system administration practices such as system rebooting for fixing problems. It also makes the installation of up-to-date software patches, for addressing security problems in the running application or the underlying operating system, difficult if not impossible.

The focus of automation system security is on the automation devices such as programmable logic controllers (PLCs), typically the edge devices of the automation network. These devices may be mission and safety critical and must be protected from electronic attacks, e.g., by authentication of *clients* attempting to address the device. This is in contrast to office and e-commerce applications, where protection of the central servers is considered of paramount importance and *server* authentication is used to verify that the server is not compromised. Connectivity to outside networks including the company intranet is often not mandatory for the automation system, and although extended periods of disconnection are inconvenient, they may not have severe consequences.

2) Operational Environment: Automation systems are operated by a team of specialized plant operators, technicians, and process engineers. This is typically a small and restricted user group, whose members have well-defined roles. Adding or removing members, or changing their roles and permissions, is an infrequent task which can be performed manually by a system administrator. Compared to large office IT or Web environments, scalability and flexibility of administration procedures are less important so that manual procedures may be acceptable and preferable to automatic, but, possibly, vulnerable mechanisms.

Also, the topology and configuration, both of hardware and software, of the automation system part, which contains the safety-critical automation and control devices, is comparatively static. Therefore, all involved devices and their normal and legitimate communication patterns (regarding communication partners, frequency, message size, message interaction patterns, etc.) are known at the configuration time. As a result, protection and detection mechanisms can be tailored to the system. Modifications of the communication system are rare enough to afford to tolerate a certain additional engineering effort for reconfiguring the security settings. Thus, one is able to trade off the convenience of administration-free protocols like the Dynamic Host Configuration Protocol (DHCP) for the higher security of statically setting up tables with communication partners/addresses in all devices. Static structure and patterns of behavior also reduce the number of false positive alerts produced by intrusion detection algorithms based on anomaly discovery.

The hosts and devices in the automation system are not used for general purpose computing, precluding the risks created by mainstream applications like e-mail, instant messaging, office application macro viruses, etc. Often, they are specialized embedded devices dedicated to the automation functionality, such as power line protection in substation automation or robot control in discrete manufacturing plants.

In many plants additional safety and fault-tolerance mechanisms are available which are independent of network based communication. They use, for instance, direct electric signaling. Such independent and isolated safety mechanisms can help to mitigate the consequences of failure of one or multiple components of the automation system. This reduces the risk to the environment that is caused by the threat of attacks. Many security mechanisms do not directly stop an attack, but just produce security information and rely on human intervention to respond appropriately. If specialized security staff for monitoring such security notifications is not available in the enterprise, response occurs late or never at all. However, automated plants are usually continuously monitored by staff dedicated to its core operation. A defense architecture could make use of these operators to shorten response times, even though plant operators may not have IT security expertise [69].

3) Challenges: The characteristics of automation systems and devices create some additional security challenges:

Automation devices often have less CPU processing power compared to desktop computers, but may have to satisfy hard real-time response requirements, frequently in the millisecond range. This limits the applicability of the mainstream cryptographic protocols described in Sections II-C and II-D. Communications channels with small bandwidth such as telephone, mobile phone, or even satellite phone lines, typically used in telemonitoring applications, make it imperative to reduce communication overhead. This conflicts with the requirements of certain security protocols.

The operating systems of automation devices in many cases do not provide authentication, access control,

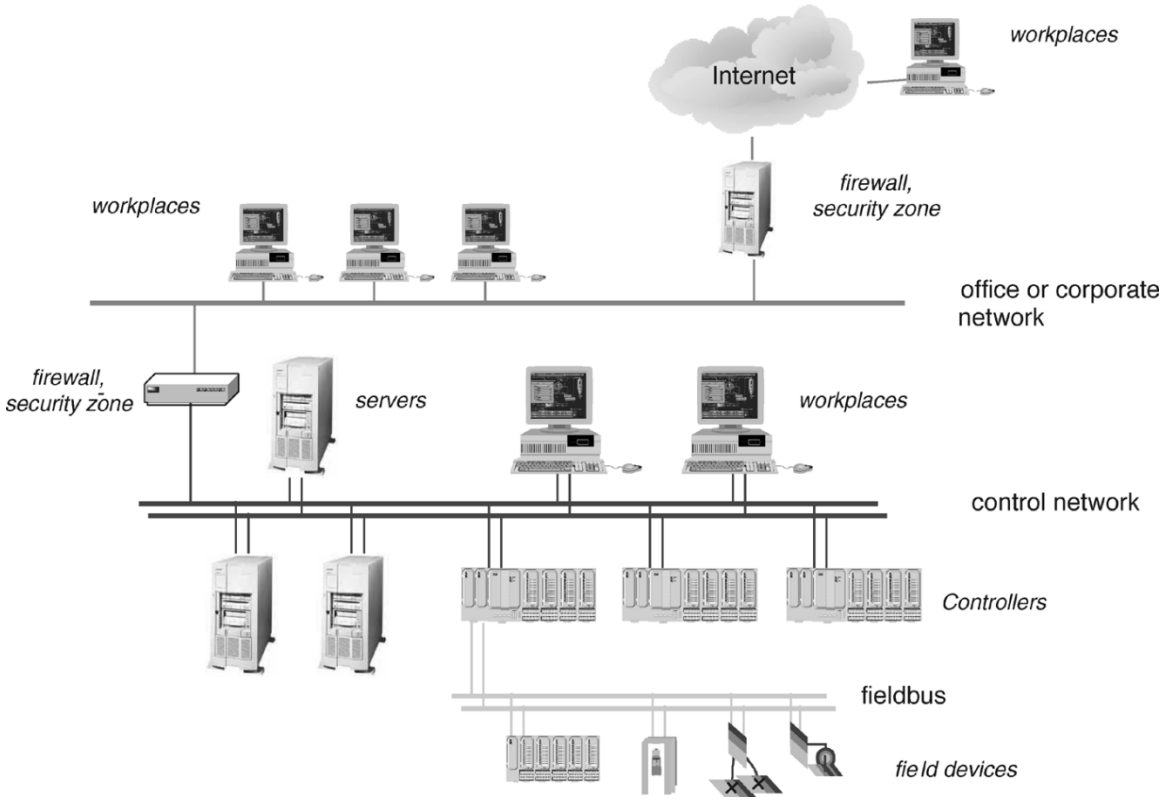

Fig. 2. Network architecture for industrial plant automation.

fine-grained file system protection, or memory isolation between processes—or these features are optional and not used due to limited processing power. However, [70] presents a concept for extending an industrial controller with hard real-time requirements to support role-based access control while maintaining deterministic timing. The protection against malicious code and privilege elevation is addressed in [71] via hardware and operating system mechanisms, using a Harvard architecture to separate data and executable code. While these radical mechanisms are hard to establish on the office PC market, they may be applicable for the development of industrial field devices with custom hardware and software.

Automation systems tend to have a very long lifetime. This has consequences both for systems currently in operation and for those newly implemented. The "legacy" automation systems currently in operation were designed based on the assumption that the system is isolated, and relied on "security by obscurity." No active security measures were taken. 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 installed industrial communication systems, the expected long lifetime means that the data communication and authentication/access control functionality must be designed to be able to interoperate with reasonable effort with systems and protocols to appear on the market 10 or 20 years from now.

Automation systems are often supported by multiple third-party contractors who request to install their own communication infrastructure for remote maintenance, e.g., dial-in modems. If suitable policies and processes are not in place and enforced to manage the proper configuration and operation of such access links, there is a high level of risk that they will expose the plant to additional vulnerabilities.

Last but not least, automation system operators and plant technicians have, due to their professional background and their core tasks in the plant, a very different attitude toward IT system operation and security than corporate IT staff. Frequently, a mutual lack of trust has to be overcome to implement an effective security architecture.

B. Network Architectures

This section describes the main types of industrial and utility communication network topologies and protocols, in preparation for the discussion of specific security issues in the later sections. Communication networks for industrial automation are typically built in hierarchical fashion, with hierarchy levels ranging from sensor and actuator wires at the bottom, plant LANs, and possibly WANs at the top. The introduction of hierarchy levels is motivated by the requirement of handling large amounts of data, not all of which is relevant at all levels. Hierarchy levels are separated by gateways and servers. From the security viewpoint, the hierarchical structuring with multiple network levels (or zones) provides a basis for implementing a defense-in-depth security architecture [66].

1) Plant Local Area Networks: A network for a DCS in an industrial plant is depicted in Fig. 2. On a given site, such as an industrial plant, factory, or utility substation, the top communication level may be a plantwide local area network linking office workplaces used, for example, for manu-

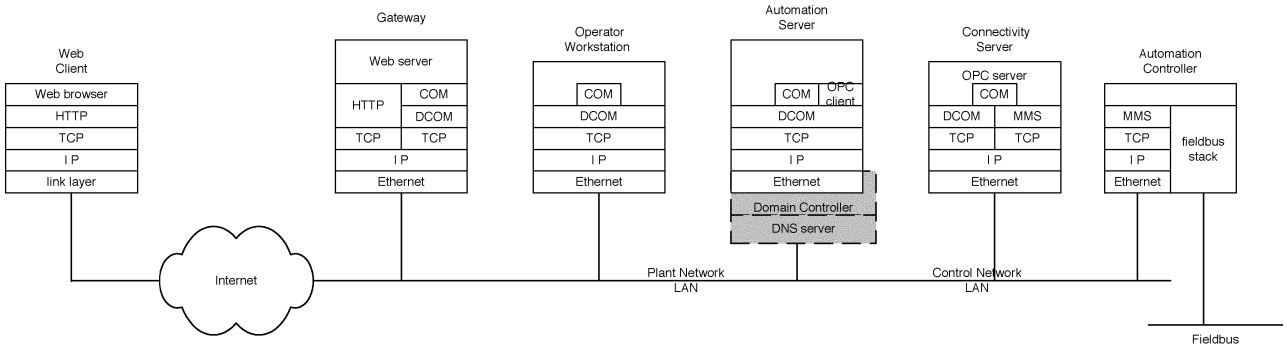

Fig. 3. Protocol stacks for industrial plant automation.

facturing execution and enterprise management applications. This office network is connected to the middle level, the control network that links workplaces used for process supervision and control. The control network carries real-time data between process controllers (automation servers) and operator workstations (clients). The control network may be divided into a number of separate network segments. For example, in a substation each segment would carry the communication for and within a switch bay. The process controllers are in turn connected to the bottom level, the field or process level. Here field buses or dedicated wiring provide the link to field devices such as sensors and actuators. Whether all levels are distinct as described here, or whether some or all levels are split or merged, depends on the plant size and the data traffic load. Going from lower to higher hierarchy levels, data is filtered and aggregated at the servers and gateways located between the levels. Data delay and availability requirements are more stringent toward the bottom of the hierarchy, so that networks may have to be segmented to distribute the traffic load.

A number of communication protocol stacks are in use at the various hierarchy levels. Fig. 3 shows a possible usage of protocols. Today's office networks are predominantly based on Ethernet and the TCP/IP protocol suite, and these are also increasingly applied to the higher and less time-critical levels in the industrial automation hierarchy. However, for the time-critical connections to the actual automation devices, field buses or dedicated wiring still dominates, although Ethernet is entering this field as well. Field buses have their own specific protocols. Gateways must be introduced for protocol conversion and to provide a common interface to the higher levels. Well known industry standards for such interfaces are the Manufacturing Messaging Specification (MMS) [72] and the standards defined by the Open Process Control (OPC) Foundation [73]. The interfaces provided by middleware according to such standards hide the details of field bus protocols, and thus allow efficient design and implementation of automation applications. Most common implementations of both MMS and OPC are built on top of the TCP protocol, thus permitting the use of widely available implementations of the lower layer communication stack. The security features of these protocols are discussed in Section III-C.

Most LANs are based on the TCP/IP and Ethernet protocol suite and rely on a number of specialized communication devices running network protocols, in particular for the addressing and routing; see [74]. A host may own a static IP address, or it may be assigned a dynamic IP address by a Dynamic Host Configuration Protocol (DHCP [75]) server. The domain name system (DNS [76]) server provides other hosts with the IP address of hosts identified by their host names. Routers perform routing (packet forwarding) between subnets (Ethernet segments) based on IP addresses. Switches perform packet forwarding within and between Ethernet segments, based on the Ethernet MAC addresses. The Address Resolution Protocol (ARP [77]) in turn maps IP addresses to Ethernet Medium Access Control (MAC) addresses, as required for Ethernet switching.

Reliable operation of these mechanisms and protocols are critical to network security. [78] discusses the security relevance of bridges, switches, gateways, and routers in TCP/IP plant networks. The flexibility of network management delivered by the DHCP, DNS, ARP, and similar protocols, unfortunately make the network vulnerable to attacks. For example, routers and switches may adaptively learn their routing and forwarding tables based on the source addresses of received packets. Attackers may insert malicious messages with spoofed IP or MAC addresses to mis-configure the routing tables, thus causing erroneous or unreliable network behavior. Routers and switches may allow remote management over the network, typically using the Simple Network Management Protocol (SNMP [79]). No secure access control is included in the commonly used version 1 of SNMP, so that switches and routers are vulnerable to electronic attacks. See [80] for a discussion of weaknesses of SNMP and [81] on efforts to improve security with SNMPv3.

Many other open protocols of the TCP/IP/Ethernet suite may be deployed in industrial automation LANs, and their operation may introduce further vulnerabilities. For example, [82] analyzes the case where a firewall must be opened to pass Simple Network Time Distribution Protocol (SNTP [83]) messages, thus opening a potential security hole.

Virtual LAN (VLAN): Normal (non-VLAN) Ethernet switches forward node-to-node packets only to the port attached to the destination node, but must forward broadcast packets to all ports. This broadcast traffic, for network housekeeping and for distribution of process status and

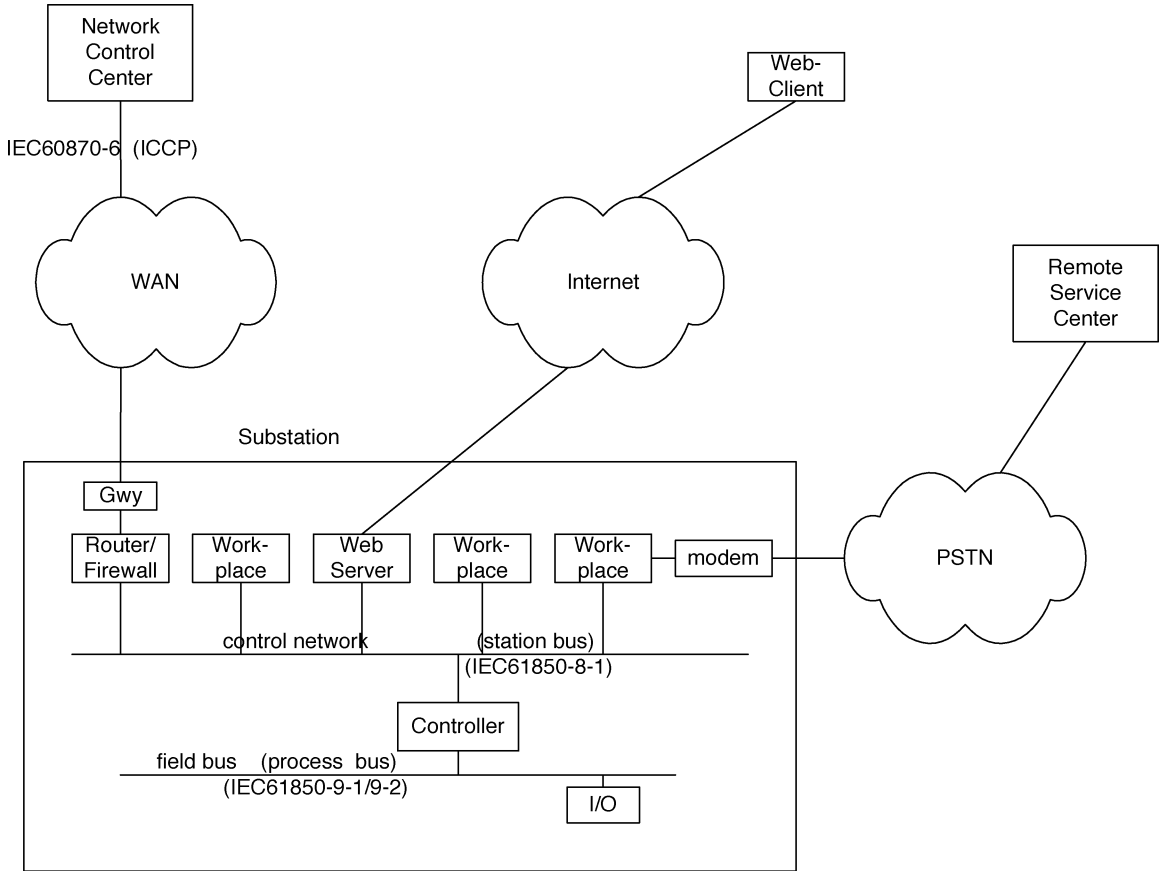

Fig. 4. Substation automation with wide area connections (without security features).

measurement values, may result in a high traffic load. With VLANs, the node or the first receiving switch adds a tag to each Ethernet frame indicating the VLAN membership of the source node. Packet forwarding by the switches is then based on the tag, thus decoupling the broadcast domains from the physical network structure, in a faster manner than via IP routers. IEEE 802.1Q specifies a tag with a 12-bit VLAN id. This scheme is used for the fast and efficient distribution of sampled measurement values over the substation Ethernet according to the IEC 61850 standard (see Section III-C3). The VLAN concept also allows to hide nodes by blocking identity information broadcast by the nodes. However, this is not a strong security measure, since the segmenting can be disabled by attacking the VLAN switches [84].

2) WANs for Automation: Energy and water utilities maintain large, geographically distributed networks for the transport of the primary goods such as electrical power, gas, and water. Supervision and control of such systems requires corresponding wide area communication networks. Fig. 4 shows an example of the communication links for an electricity substation. Utilities typically lease communication capacity from a telecommunication operator or maintain their own communication networks. In the latter case, separate physical cabling is laid or fiber capacity is rented from a communications provider. Electricity utilities may even use the power transmission network itself to carry automation data.

The largest automation networks are run by the electric power transmission operators for supervisory control and data acquisition (SCADA) of the power generation and transmission systems and for energy management tasks. These large networks connect power plants and substations of various power voltage levels to network control centers. The electricity industry has established open standards such as the Inter Center Control Protocol (ICCP) [85]. These are typically also built on top of MMS and the TCP/IP protocol stack.

Communication networks with smaller geographical coverage are required for water and electrical distribution networks. Applications involve water and electric power distribution automation, load management, and automatic meter reading (AMR). These networks must provide communication to various small automation devices such as valves, switches, and meters, which are often installed at isolated locations. Radio networks are often the appropriate solution for such applications, either as a utility-operated private radio network, or using the data services of public mobile cellular or even satellite radio systems. The choice is determined by coverage, traffic load, and cost. Many automation devices attached to such networks are small and implemented on embedded controllers with a small communication module and a simple serial line interface. This is sufficient for the occasional transfer of status and control information between devices and a data concentrator. Where time-critical data transfer is required, application

specific protocols with low overhead may have to be used. Communication standards are available for the distribution automation in electrical distribution networks [86].

3) Remote Point-to-Point Links for Remote Operation: Remote operation refers here to the possibility to access the automation system or some automation device from a geographically separate location. Often, the communication link does not need to operate continuously. This is preferable for security reasons. An example is an operator on emergency duty accessing devices in an unmanned substation, possibly on receipt of an alarm message. Remote access then occurs by dialing into the system using modem connections through the public switched telecommunication network (PSTN), or through the Internet. A device may also be able to spontaneously connect itself toward some predefined server to emit alarm or other status messages.

Larger plants may have a dedicated communication gateway handling such communications. Operator workstations in the plant may offer communication ports to support remote access services such as remote login, thus allowing the remote operator to take full control of the automation workplace. It is evident that such setups are highly critical in terms of security and thus need protection of the communication links and strong access control.

4) Public WANs: Industrial and utility wide area links and remote point-to-point links may use a public WAN to carry all or part of their traffic. There are two public WANs, the circuit-switched telephone network and the packet-switched Internet. These networks are conceptually distinct and have different security and availability properties. In practice however, a combination of both networks is often employed.

PSTN: The PSTN is the most widely available telecommunication network. Subscriber terminals are connected over individual access lines to the local exchanges, which are connected globally via a hierarchy of switches and transmission (trunk) lines. Connections in the PSTN are circuit-switched, i.e., a communication path is set up initially and its transmission capacity is exclusively assigned to the connection for the duration of a call. Control of the circuits is handled by the telephone system switching centers using the Signaling System No. 7 (SS7) network and SS7 protocol suite [87]. For data connections which transmit only intermittently, circuit-switching is inefficient and costly. However, as the coverage and availability of the global PSTN is unsurpassed, in many applications, PSTN access lines are still employed to connect the data terminals via some gateway to a data backbone network such as the Internet. PSTN connections are also used for direct modem connections over dial-up or leased lines for remote access to automation systems. The PSTN is considered to be much less vulnerable to electronic attacks than the Internet. PSTN access and transmission lines are typically laid in specialized cables and ducts, and telecommunication switches are housed in physically protected buildings and operated by specialized and trusted staff. Also, it is difficult to interface to the SS7 network to launch electronic attacks. Pure modem-to-modem data connections over the PSTN are thus considered to be more secure than connections involving the Internet, although some attacks are conceivable (Section V).

Internet: The Internet is a global data network based on the Internet Protocol (IP) and the IP addressing scheme. IP routers forward each data packet individually, according to the IP destination address carried in each packet. Packet transmission is multiplexed on the transmission lines connecting the routers. This is in general more efficient than circuit switching, but leads to load dependent behavior. Internet backbone transmission lines may be shared with or leased from the PSTN operators. However, routing is performed in a distributed manner by routers belonging to a variety of Internet operators and Internet service providers. The large number of organizations involved in operating the Internet, and the large number of powerful computers connected to it, increases the number of potential sources of electronic attacks to the network and to the devices attached to it. IT security has thus become a dominating concern in Internet applications and has led to the introduction of various security protocols, as described in Section II-D. Similar concerns are relevant with the use of Internet technology for safety-critical communications; see [88].

C. Security in Industrial Automation Protocols on LAN/WAN Level

Where protocol standards for industrial automation do include security issues, these deal mainly with *access control* for the objects defined in the respective standard. The automation system must offer appropriate administration tools to manage the security settings. For access control, user groups and roles must be defined, and permissible operations must be classified. It must be possible to configure access control lists within the required granularity. For some general discussion of the role-based access control and multilevel security, see [62]. The following description surveys the security functions defined by some common standards for industrial communication protocols.

1) OPC: OPC is an application layer specification for the communication between software components for industrial automation [73]. Specifically, it defines standardized interfaces through which OPC clients access the objects in OPC servers. Typically, operator workstations are OPC clients, which access OPC servers acting as gateways to the automation field devices. The OPC standard interface hides the complexities of the device-dependent communication protocols, hence OPC's claim that "OPC is open connectivity [and interoperability] in industrial automation and the enterprise systems that support industry." OPC builds upon Microsoft's component object model (COM) [89], so that OPC implementations are in practice restricted to platforms supporting COM.

The OPC data access (DA) specification defines standardized read operations to transfer real-time process data, together with time stamp and status information, from automation devices such as controllers to higher level applications such as process supervision and manufacturing execution systems. Other OPC specifications define alarms and event notifications, data exchange for

server-to-server communication across Ethernet networks, access to stored historical data, and other functions [73]. More recent OPC specifications apply XML to provide rules and formats for representing and publishing plant floor data. The use of the XML standard promises to make OPC platform-independent.

The OPC security specification [90] defines optional security interfaces of OPC objects. The specification deals with access control and is based on the Microsoft Windows security model. *Principals* (human users, computer processes), on accessing a *security object*, must present their *credentials* (access token) over a protected communication channel. A *reference monitor* checks the credentials against the information in the *access control list* (ACL) associated with the object, and grants or denies access accordingly. Where available, i.e., in purely Windows networks, the OPC servers implementing security will use Windows/Kerberos credentials and the reference monitor service of the underlying COM middleware. The credentials must be generated for authenticated users and processes and stored (cached) in a protected manner [91]. User authentication is left in practice to the Windows log-on procedure. Where Windows credentials are not available, for instance, if the OPC server is accessed from a different domain or from the Internet, OPC servers must specify and manage their own "private" credentials, such as server specific combinations of user IDs and passwords. It is up to the OPC server to use mechanisms that prevent the compromise of these credentials, for example not to store and transmit passwords in clear text.

An OPC server may implement one of three levels of security.

- Disabled security: No security is enforced.
- DCOM security: Launch and access permissions to the OPC server are limited to selected clients. This is the default security level provided by distributed COM (DCOM; see below), and is typically administered with the DCOM declarative security configuration tool.
- OPC security: The OPC server serves as a reference monitor to control access to vendor-specific security objects that are exposed by the OPC server. The implementation uses DCOM's programmatic security. The standard does not prescribe or suggest which objects should be secured. If OPC security is implemented, DCOM security must be configured to grant access to the server interfaces.

The OPC security specification covers only server/object access control, but is not concerned with confidentiality and integrity during transmission. In distributed configurations where OPC clients and servers reside on different hosts, DCOM is normally used in place of COM. Network security measures are relegated to DCOM. OPC implementations may choose to use COM, but implement their own networking middleware in place of DCOM. Some implementations extend DCOM services regarding fault-tolerance and redundancy in particular. Any such alternative communication basis for OPC should offer similar services as the DCOM services described below, in particular with regards to network security. As mentioned above, the new OPC-XML specifications replace the DCOM services by the simpler and platform independent SOAP transport protocol. SOAP typically runs over HTTP, in which case it may rely on a transport layer security mechanism like SSL to protect the communication.

DCOM: DCOM supports distributed COM applications. DCOM is not specific to industrial automation, but many distributed OPC implementations are based on DCOM. COM objects, identified by a globally unique class identifier (CLSID), provide services through interfaces. Clients address the object by specifying the CLSID (in which case DCOM retrieves the location/server of the object from the fixed, configured registry), or by specifying both the CLSID and the server machine name (if the latter is dynamically selected by the client). Typically, the object method is invoked by a remote procedure call (RPC) sent via TCP connections or UDP datagrams over the network. DCOM performs marshaling (arranging object data in a representation suitable for serial transfer over the network), call-back services from server to client, and a number of optional security services. The latter comprises access control, including impersonation (the transfer of credentials for cascaded calls), and the communication security measures described below. Access control and launch permissions can be set for individual DCOM objects, groups of objects, or all objects on the server.

DCOM *communication security* distinguishes between connection security, per call security, and per packet security. With connection security, client hosts authenticate themselves on system start-up toward the server using the Microsoft Windows NT LAN manager (NTLM) challenge/response protocol. There is no server authentication and no further message authenticity or privacy (encryption) services are invoked on subsequent RPCs. This is the default security level. DCOM's definitions of per packet authentication and encryption is too fine grained for practical use and seems to be superseded by the usage of the Windows IPSec implementation. IPSec protects either all or none of the communication between two ports of a pair of computers (see Section II-D). DCOM's usage of a wide range of port numbers may, however, conflict with restrictive IPSec and firewall configurations. This makes DCOM-based OPC unsuitable as a protocol for communication between enterprise systems and control systems across the control network perimeter, which will likely prohibit passing DCOM messages.

DCOM offers two ways for the setting of security parameters. With *declarative security*, administrators employ a separate tool to configure the settings, while with *programmatic security* the settings are specified directly in the program code. The latter allows a finer granularity of security settings and is used for the OPC security level.

2) MMS: MMS [72] is an application-level standard for the messaging communication to and from field devices or PLCs in a computer integrated manufacturing environment. MMS defines only generic services and objects (variables)

and their addressing. Separate companion standards are required for the definition of the application specific objects. Today, MMS is widely used in automotive manufacturing, and is also a basis for the IEC 61850 standard suite (see Section III-C3). MMS was defined in the 1980s and thus predates OPC. MMS not only covers client/server communication, but also peer-to-peer communication in distributed networks.

MMS defines some access control features based on simple password authentication. Neither confidentiality nor nonrepudiation facilities are provided. The initial MMS association request sets up an MMS environment between client and server, i.e., opens semipermanent communication connections and exchanges information on the capabilities of the server. The association request has an optional authentication parameter (password) to be supplied by the client. This authentication value is used to control access to MMS objects. Objects contain access control lists (ACL) specifying the conditions under which services requested from the object are permitted. Conditions are user (application) identity, and/or submission of a password, while services include create, read, write, execute, delete, or modify. Conditions may be specified separately for individual objects or for all objects.

MMS assumes that an Ethernet LAN is the underlying communication network between clients and servers. The original specification described a protocol stack according to ISO, but most implementations today are based on the TCP/IP protocol stack. For TCP/IP, the mapping of MMS onto TCP follows RFC 1006 [92] (ISO transport service on top of TCP), which specifies data formats, frame assembly/disassembly, and the port number. If communication security services are required, these may then be provided by deployment of IPSec or SSL.

3) IEC 61850: IEC 61850 specifies the software entities, data models, services, protocols and data formats for automation of substations of the electric power network [93]. Of special relevance to IT security are the access control options and the communication protocol mappings, which determine the network security characteristics. IEC 61850 stipulates that nodes should provide access control based on node identification (for machine-to-machine communication) and on user authentication and system access control (for HMI users). Users must be authenticated as operators, administrators, etc., and obtain corresponding access privileges to create/view/execute etc.

On initial association, the client should send authentication parameters (= user id + view + password) to the server. The access control to be performed by the server restricts the views (the accessible set of objects) based on the client id. On operator control actions such as commands sent to substation switches, the object should check the authorization of the client before performing the selected operation. The IEC 61850 association is directly mapped to the MMS association request. As described above, MMS is further mapped to TCP/IP/Ethernet.

Substation automation devices run algorithms for the protection of the electric power network, where very low latency data transmission is a crucial requirement. For the fast transmission of time-critical data for generic substation events (GSEs), such as activation of power network protection switches, or for the fast distribution of sampled measured values (SMVs) of currents and voltages, IEC 61850 defines a direct mapping of simple data packets onto Ethernet packets. Peer-to-peer data distribution and support of the publisher/subscriber paradigm uses Ethernet layer broadcast/multicast. IP layer routing is bypassed for latency reasons. This is appropriate on the field or control networks for the fast raw data transfer from the field devices to substation controllers. To segment traffic load and to ensure fast transfer, the IEEE802.1Q VLAN and priority schemes may be applied.

IEC 61850 does not address communication security, but refers instead to work done by the IEC working group on security [94]. Where IP is part of the communication protocol stack, the security services of IPSec may be employed. However, no security standards exist on the Ethernet layer, so that there are no standard tools available to protect GSE and SMV data transmission. Receivers identify the data using the source MAC addresses, but this provides only weak authentication, since MAC addresses are easy to spoof. As noted in Section III-B, VLANs may be used to confine the distribution of data to restricted network segments, but this is not a strong security measure.

4) ICCP: The ICCP of the IEC [85] is a standard for the wide-area communication between centers of the electric power transmission network, such as power plants and network control centers and substations. It is similar in scope to OPC, but is not tied to a particular operating system. The standard originally did not address security; however, a report issued by a security working group of IEC discusses vulnerabilities and countermeasures of ICCP systems in detail [95]. As noted above, most implementations of ICCP run over MMS and TCP, and the access control functions of MMS should be implemented. To protect the communication between centers, [95] recommends the deployment of the SSL protocol and specifies the mandatory cipher suites to be supported. SSL provides security between clients and servers and is compatible with any network address translations in the WAN, in contrast to IPSec. In order to make use of SSL even in non-TCP implementations of ICCP according to the ISO protocol stack, [95] also defines the encapsulation of SSL messages to such non-TCP (ISO) packets.

Security in the industrial automation protocols, as reviewed in this section, emphasizes *access control*. The assignment of access permissions is a function of the roles of the automation system users, and hence belongs to the automation application layer. In contrast, *network security* can be provided by the lower communication layers using the general purpose security protocols reviewed in Section II-D. With a proper combination of these security features offered by the various protocol layers, secure systems can be built.

D. Security on the Field Bus and Device Level

As described in Section III-B, Fig. 2, industrial communication networks involve a number of levels. The lowest level is closest to the application specific devices such as sensors,

meters, and actuators. A large number of specialized and partly proprietary communication systems, media, and protocols can be found on this level. Most were developed at a time when security issues were of lesser concern than today, and when no practical security measures were available.

1) Field Buses: Field buses are employed where a large number of automation devices must be connected to controllers. Transmission media are copper wires, coaxial cables, or optical fibers. Field bus protocols are optimized to provide fast and deterministic access to a large number of devices. In order to ensure high availability and reliability, many forms of redundancy are introduced to provide fault tolerance against random errors and equipment failures, but none of the traditional field buses offers security features against intentional attacks. It was considered unlikely that attackers could perform electronic attacks such as eavesdropping or message tampering on the field bus lines, as they would need physical access to the lines. Also, for the specialized field bus protocols the "security by obscurity" principle was often considered to be good enough. Recent studies have now begun to address the issue of field bus security. [96] looks at fieldbus-to-Internet gateways in general, and [64] surveys security facilities and flaws in the building automation field buses (LonWorks, European Installation Bus, and BACnet). The use of smartcards as cryptographic coprocessors is proposed to provide industrial communication devices with the capability for confidentiality or integrity protection algorithms.

2) Industrial Ethernet: The new generation of Ethernet-based field buses uses Ethernet and TCP/IP protocols and services; see, e.g., [97]–[99] and the IEC 61850 substation automation standard [93] described above. Such networks are more vulnerable to attacks and should thus provide security services. The general communication security tools for the TCP/IP protocol suites described in Section II-D are applicable; see, e.g., [100][101]. As discussed in Section III-B, a crucial component of Ethernet security is the protection of networking equipment such as the Ethernet switches.

For efficient distribution of industrial process values, Ethernet-based field buses employ the publisher/subscriber paradigm, whereby a publisher (data source) sends its data to the IP multicast destination address to which any interested nodes may subscribe. In principle, IPSec can protect all IP traffic. However, IPSec, or more specifically its key exchange protocol IKE, is not designed to handle security associations for the multicast case [45]. Protocols to extend IPSec for multicast security services, including group key management, have been defined; see [102]. Some Ethernet-based industrial protocols bypass the TCP/IP protocol layers and use Ethernet unicast and multicast for the transmission of time-critical data. As already noted in Section III-C3, there are currently no standard security tools available on the Ethernet layer, so that such direct Ethernet transmission cannot easily be protected.

3) Wireless Systems for Automation, Wide Area Wireless Data Networks: Dedicated radio data networks are deployed by energy and water distribution utilities to serve automation equipment located in a wider geographical area. Typically, such systems operate in the UHF/VHF radio bands on narrow-band channels and carry status and commands between an operation center to remote terminal units. Often, such systems are highly specialized requiring dedicated equipment and running proprietary protocols, so that their designs have not included any security features and simply rely on "security by obscurity."

Where traffic load is only moderate, and where remote units are located in areas covered by public cellular networks, it is more efficient to use data services of such public radio networks. Modern public networks offer a variety of data services which all include basic security services such as encryption, at least over the air interface.

Radio networks are vulnerable to radio jamming, for which no effective countermeasure exists. Automation systems using wireless links must be designed to tolerate loss of communication links and to guarantee safety under such circumstances.

Local area wireless sensor networks: Specialized wireless sensor networks are appropriate for data collection from a large number of devices. As with all radio networks, these networks are particularly vulnerable to various DoS attacks [103]. Most such wireless sensors are powered by batteries and thus energy-limited. A subtle attack may hence provoke continuous retransmissions from a device by suppressing the transmission acknowledgment it expects, thus rapidly exhausting the battery of the device. Where such wireless networks support packet forwarding over multiple hops between nodes, as in some proposed wireless sensor networks, attackers may upset the adaptive route discovery by replying with spoofed "route-through-me" messages. This is similar to attacks on ARP and IP-routers in IP networks.

In some applications, such as in factories, it may be appropriate to deploy general purpose wireless communication systems such as Bluetooth or wireless LANs to carry data from field devices. Security issues of such systems are described in Section II-D.

4) Power Line Communication Systems: The cables and transmission lines installed to transport electric energy may also serve as data communication lines, by using appropriate coupling equipment and specialized data transmission terminals. By definition, such power line transmission lines provide coverage in principle to any location relevant for the electricity network. A number of industrial standards exists for utility data transmission on power lines; see [104] for long-distance high voltage transmission lines and [86] for distribution application and electricity meter reading. Reference [105] allocates frequency bands for power line transmission for utility use. However, power lines are not designed for frequencies required for data transmission. Asymmetries and imperfect impedance matching may cause considerable spurious emission of the data communication signal from the power line. Eavesdropping is thus easily done with a simple radio receiver placed in the vicinity of the power line. Some early power line ripple control systems, e.g., for load shedding and tariff switching, use very simple broadcast communication signals without any message authentication. Despite this fact, none of the known

utility power line transmission standards has included any strong security measures, as they have been established at a time where electronic attacks were not considered likely. The power line communication medium is as vulnerable as the radio communication medium to jamming and eavesdropping. For the latter, security issues are well recognized. Modern systems communicating over power lines should therefore deploy similar security measures as developed for modern radio communication systems. Reference [106] is concerned with the specific requirements of industrial communication systems via power line carriers and discusses the use of smart cards to handle cryptographic protocols.

E. Security of Embedded Systems for Industrial Control and Communication

Industrial automation controllers are typically implemented on embedded computers. Such embedded systems have to cope with restrictions on cost, real-time performance, power consumption, and other constraints which are even more demanding than in large workstations. Reference [107] discusses these aspects with the example of a thermostat connected to the Internet.

1) Real-Time Requirements: The primary function of industrial automation devices is measurement and control. The automation tasks are run cyclically with well-defined execution times to meet real-time deadlines. External attackers may use the communication interface to perform DoS attacks: by flooding the device with heavy traffic, the processor receives a high rate of interrupts from the communication interface requesting additional processing. Hence the design of the task and interrupt priority selection in the processor is important to reduce vulnerability to DoS attacks.

2) Memory and Processing Limitations: Limited memory and processing capability are the main characteristics of embedded systems. The selection of security protocols and cipher options to be supported by an embedded device must take these constraints into account. Reference [108] gives an overview on cryptography in embedded systems, and [109], [110], and [55] consider the implementation of the SSL protocol and of HTTP DA, respectively, in such systems. According to [110], public-key encryption based on elliptic-curve cryptography (ECC) have advantages over the better known RSA algorithm in terms of required key lengths and processing times [24]. In embedded devices, SSL should thus select the ECC-based cipher suites.

3) Robustness: Proper error and exception handling is an issue for any software. However, embedded automation controllers must often operate autonomously and must hence be particularly robust. An attacker may insert malformed messages, attempting to crash the system (e.g., by a buffer overflow), and the embedded controller must be capable to withstand such attacks autonomously. Battery-powered embedded systems can be caused to fail by provoking some unnecessary processing (e.g., by fake exception conditions), thus draining the battery. Such attacks can be prevented by careful validation of received messages and commands.

4) Software Implementation Issues: Some simple protocols used by embedded devices transmit passwords, which must be compared against a copy stored in the nonvolatile memory of the device. To prevent password theft, the password should be stored in encrypted form, as recommended in Section II-E.

Cryptographic protocols use random numbers for key generation, challenges and initialization vectors to prevent replay attacks. Care must be taken to generate numbers that cannot be predicted by an attacker. For example, if a network stack implementation generates predictable TCP sequence numbers, then an attacker may spoof TCP segments and hijack TCP connections. The generation of random numbers is particularly difficult in embedded systems; see [60], [111].

5) Software Configuration: Proper configuration of the communication subsystem of the embedded system is crucial. Some examples of security measures for embedded industrial Web servers are the following.

- Disable insecure protocols like PAP for PPP or Basic Authentication for HTTP [55].
- Enable timeouts and renegotiations. CHAP should be renegotiated periodically and HTTP DA sessions should be closed after a given timeout has elapsed.
- Set limits on the number of login attempts per session. In this way, automatic password guessing would be very slow and therefore practically impossible. For example, one may close an HTTP SSL session after the user has tried three times to login without success.
- Enable client and server authentication. For example, mutual authentication is included in the SSL specification [50], but practical deployment often omits client authentication, since it would require complex management and processing of client certificates.

In general, it must be verified that only the specified tasks and services are active, and that no backdoors are left open after installation and commissioning of the embedded device.

IV. Security Recommendations for Industrial Systems

A. Applicable General IT Security Standards

This section presents two general IT security standards that are also used to develop security mechanisms and procedures in the industrial communications area. For cryptographic standards, see Section II-C.

1) CC: The CC standard [112], collectively developed by the United States, Canada, and various European countries, is the successor of early country specific IT security standards, such as the U.S. "orange book" [113]. This standard is concerned with certifying devices and applications (e.g., firewalls or operating systems) according to predefined feature/requirement sets, which are called protection profiles (PPs). The effort that is put into the evaluation and thus the resulting confidence of the certification authority in its verdict, is stated in terms of seven evaluation assurance levels (EALs). This effort can range from superficial usage of the product to verification using formal methods. A CC certification on levels EAL1 to EAL4 is automatically recognized by

participating countries according to the CC recognition arrangement (CCRA). For comprehensive information on CC, see [114].

2) ISO/IEC 17 799: Whereas the CC standard focuses on product features, the standard ISO/IEC 17 799 [115], which started out as a British standard, BS7799, is concerned with company internal processes and controls that need to be established in ten different areas such as business continuity planning, system access control, physical security, computer and operations management, and security policy, in order to have a standard compliant, certifiable security management system. Much like ISO 900x, this standard deals more with procedures and documentation than with actual effectivity and efficiency.

B. Security Standards for Industrial Communication Systems

1) IEEE 1402: The substation security standard IEEE 1402-2000 [116] is mostly concerned with physical security, but it mentions defense against electronic intrusions at various places.

2) Process Control Security Requirements Forum (PCSRF): The PCSRF [117], initiated by the U.S. National Institute of Standards and Technology (NIST), aims to create sets of standard requirements for the procurement of new process control systems. So far, PCSRF has produced an Industrial Control System Security Capabilities Profile (SCP-ICS) [118], which provides background information, and an Industrial Control System (ICS) System Security Profile (SPP-ICS) [119], which defines a set of baseline security requirements for the entire ICS lifecycle, including both functional and operational requirements. This SPP-ICS serves as a starting point for more specific system PPs, e.g., for DCSs or SCADA systems, and for component PPs, e.g., for PLCs or sensor authentication. PPs developed within the PCSRF are intended to serve as specifications of security capabilities that are desired in new products and systems. Use of these PPs for certification and evaluation is independent of its use for requirements definition and is still under consideration by the PCSRF.

3) ISA SP99: The ISA Committee SP99 "Manufacturing and Control Systems Security" intends to create guidance documents on introducing IT security to existing industrial control and automation systems. The scope of this effort includes, but is not restricted to, hardware and software systems such as DCSs, PLCs, SCADA, networked electronic sensing, and monitoring and diagnostic systems. Two guidance documents have been published. The first report [120] is a comprehensive survey of the state of the art in security technologies and mechanisms, and provides comments on their applicability for the plant floor. The second report [121] presents recommendations for a security architecture, and for procedures to achieve and maintain (including auditing) IT security for industrial control systems, based on the mechanisms described in [120].

4) American Gas Association (AGA) 12: The AGA is developing a standard for communication security for SCADA called AGA 12 [122]. This standard will mainly deal with issues of encryption, especially retrofitting encryption mechanisms, that comply with the U.S. Federal Information Processing Standard FIPS-140 [123], on the communication links of gas, water and power SCADA systems. The existing draft [122] gives a very good introduction into the security risks and issues surrounding SCADA systems. Members of the committee working on this standard are also active in the ISA SP99 and PCSRF, which ensures coordination between these related efforts.

5) IEC TC65: The IEC Technical Subcommittee 65C "Digital Communications" started in early 2004 to address security issues for fieldbuses and other industrial communication networks in a new part 4 "Digital data communications for measurement and control—Profiles for secure communications in industrial networks" of the IEC 61 784 standard [124] in its working group WG13 (Cyber Security). In 2003, the IEC Technical Sub-Committee 65A "Industrial Process Measurement and Control—System Aspects" had started to consider complementing its standard IEC 61 508 "Functional safety of electrical/electronic/programmable electronic safety-related systems" [125] with security guidance. In 2004, however, this subcommittee decided to wait for the results of SC65c/WG13 before proceeding further.

6) North American Electric Reliability Council (NERC): NERC passed in August 2003 its Urgent Action Standard 1200 "Cyber Security" [126]. This Urgent Action Standard was extended for one year in August 2004, while work on its permanent successor CIP-002-1 through 009-1 (formerly called NERC 1300) is under way. A first draft of NERC 1300 became available in September 2004. The CIP-002-1 through 009-1 implementation plan calls for the standard to become effective October 1, 2005, and to begin to require compliance in the first quarter of 2006. Drafts are now available for review.

NERC 1200 requires U.S. transmission and distribution utilities to create and maintain documentation about 16 different aspects of IT security, such as access control, physical security, incident handling, training, and policy. The CIP-002-1 through 009-1 series appears to be rather similar in character and content of its prescriptions, but it will extend its scope to power generation companies and also impose penalties for noncompliance.

7) FDA: The U.S. FDA requires a detailed audit trail (who did what on what batch?) as part of the good manufacturing practice (GMP) for products in its areas of supervision. These audit trails can also be created and maintained as electronic documents. Regulation 21 CFR part 11 [127] states requirements on software systems in the area of operator authentication, multiple log-on, and electronic signatures, for such electronic audit trails to be considered equivalent to handwritten records and signatures.

8) Industrial Automation Open Networking Alliance (IAONA)IAONA: A joint technical working group of the IAONA has produced a security guideline for industrial Ethernet networks on the plant floor [101]. The focus is on recommendations on TCP/UDP/IP services which should or should not be used with regard to their known security weaknesses.

9) International Council on Large Electric Systems (CIGRE)CIGRE: CIGRE established a Joint Working Group JWG D2/B3/C2-01 "Security for Information Systems and Intranets in Electric Power Systems" [128] in June 2003. A final report from this working group addressing the IT security concerns of power utilities is scheduled for 2005. The group cooperates with SP99 activities.

V. Case Studies

The sections above have described concepts and elements of IT security for industrial and utility communication systems. This section describes the application of these security concepts and measures to some representative cases. As discussed in Section II-E, best practice dictates that for a given system a *security policy* must be defined first. The security measures to be implemented are then derived from the security policy.

A. Substation Automation

A substation of the electricity transmission and distribution network is responsible for voltage conversion, control of power flow, and protection of power lines. Main primary equipment are power transformers, circuit breakers, and power switches. The automation network of a substation may consist of a LAN and of links to remote nodes; see Fig. 4. The security of utility substations and SCADA systems as well as their communication networks is the focus of [129], [130]. Attacks on the communication system may cause malfunction of the power equipment and may have disastrous effects on its operation.

Using the results of a survey among utility staff, [131] attempts to quantify the threats and risks of network-based attacks on (water-supply) SCADA systems in terms of their probability of occurrence. Reference [132] is an early discussion of risks to the electric power grid from electronic intrusion into automated substations or attacks on the SCADA communication infrastructure. It is pointed out that a worst case scenario would be an attack on the communications infrastructure combined with an attack on the electric power control system. Restoring the power grid would be difficult and dangerous without the availability of communication links for the coordination between the network control center and power generation and transmission units.

1) Threats: IEC 61850-based substation automation systems are built partly on top of MMS, as described in Section III-C3. MMS access control specifies a clear text password to be supplied by the client for authentication. Attackers with access to the relevant LAN segment may thus simply eavesdrop on the association message to obtain the password, and may then introduce rogue control commands to the system. GSE and SMV packets are multicast directly on the Ethernet level over the (V)LAN. These packets are not protected. An attacker can thus easily insert tampered, replayed, or delayed packets.

Unprotected ancillary substation automation services, over protocols such as SNMP/UDP for switch and router configuration, and SNTP/UDP for time transfer, can be attacked by listening to legitimate traffic and inserting tampered or delayed messages, e.g., using TCP session hijacking. For SNTP, however, eavesdropping attacks are of little interest to potential attackers and thus unlikely.

These attacks concern the LAN in the substation. There are also threats from remote attack sources which use the external interfaces of the substation [130]. The substation maintains a permanently operational WAN connection to the network control center (NCC) through a gateway. An attacker who succeeds in subverting the access control system of the NCC system can also attack the substation remotely. Attacks on the WAN, particularly on the link running the ICCP, also make the substation vulnerable.

Another external interface is offered by dial-up modems in the substation, either attached to an individual host or to a remote access server. PSTN lines are considered less vulnerable to eavesdropping and tampering attacks than Internet connections (Section III-B). However, DoS attacks may be launched toward the substation simply by jamming the modem by continuous dialing. Call-back schemes offer some protection from unauthenticated access, but spoofing of calling numbers in the PSTN has been reported, so authentication should not be based solely on the calling number.

On the lowest substation level, the field bus and I/O signal wires are vulnerable to physical destruction (DoS), and physical rerouting (reconnection of wires for message tampering and man-in-the middle attacks) and eavesdropping. Physical destruction should be rapidly detected by the existing fault-handling subtasks of the control system. Intentional reconnection of the wires is unlikely as it is of limited interest to attackers. Visual site inspections are the only way to protect against such attacks.

2) Effect of Attacks on Substation Operation: An attacker may cause the greatest damages if he succeeds in tampering, inserting, or suppression of control messages to the substation primary equipment. Some potential scenarios are the following [133].

Failure to break circuit: In this scenario, a short circuit in a line of the electric grid occurs but the line is not disconnected, as it should happen according to the line protection system. This may lead to a damage in the primary equipment of the transmission/distribution infrastructure and to power outages on the consumption side (unselective switch-off). Only the inhibition of the "switch off" state is considered as an attack here. The actual short circuit in the line, which, together with the attack, leads to damage, may or may not be a malicious and artificially induced event.

Unnecessary disconnect: In this scenario, a line is disconnected even though there exists no reason (e.g., operator command, short circuit) why this should have happened. This may lead to power outages on the consumption side (unselective switch-off).

Operating maintenance switches under load: In this scenario, maintenance switches in the substation (disconnectors, earthers), which may only be operated on disconnected lines, are operated while under load. This may lead to damage in

the primary equipment of the transmission/distribution infrastructure and to power outages on the consumption side (unselective switch-off).

3) Countermeasures: The attacks considered above are network-based attacks on message integrity, including injection and suppression (or possibly delaying) of messages carrying actuator, sensor, or interlock signals. Well-known mechanisms exist that allow the receiver to detect tampered or injected messages (see Sections II-C and II-D). Message suppression or delay are more difficult to detect. A secure substation automation system must also cope with message suppression, e.g., by implementing redundant communication, regular heartbeat messages, and message sequence numbers. Devices must guarantee fail-safe behavior in case of interrupted communication. Reference [134] makes further suggestions for retrofitting security mechanisms in the deployed SCADA systems and their communication links. References [129] and [130] place special emphasis on password security. Based on results of a public-key cryptography benchmark, [135] suggests that in general public-key cryptography is not feasible in time-critical SCADA networks for electric utilities. Instead, a hierarchical scheme is proposed with stored symmetric keys where elliptic curve asymmetric keys (see Section II-C) are used for key exchange. Also mentioned are requirements on securing keys in stand-alone intelligent electronic devices (IEDs) as well as a peer-to-peer communication scheme between substations. It is argued that for the static SCADA systems, a full PKI with key revocation lists is not needed.

Finally, secure physical and electronic access to substations must be implemented [116].

B. Plant Automation

This section gives a general discussion of the range of security issues and recommendations applicable to the network configuration for plant automation. As one of the earliest publications on security issues in industrial automation systems, [136] summarizes the recommendations given by the NAMUR process industry association on information system security in process control systems. The issues depend on the required level of connectivity between the automation network, any intranet, and external networks. An IT security survey conducted among automation system users and vendors in 2002 found that in nearly half of the plants there is a direct connection from external systems to the automation system, not counting additional connections to and via the enterprise intranet [137]. Only a few of the plants used some kind of intrusion detection system and the majority had no regular security audits on the plant floor. The following paragraphs discuss the security issues for three cases with increasing connectivity.

1) Isolated Automation System: With a stand-alone automation system, security is achieved by physically protecting the system and keeping the network isolated. Only authorized personnel have physical access. Installation of new software or updates, temporary connections of computers, e.g., for servicing and maintenance, is only allowed after explicit authorization and after careful virus scanning, etc.

In addition to operator workplaces, such a system may include "office workplaces." As long as these office workplaces are dedicated to functions related to the automation system (i.e., not used for general office purposes such as e-mailing, and not connected to networks other than the automation network, neither directly nor indirectly), they could be part of the automation system domain, and users could be registered as members of that domain. In order to better control which servers can be accessed from which office workplace and to control the network load they impose on the automation system, the office network should be separated from the automation system network by a packet filtering router.

2) Connecting to an On-Site Office Network: In large factories, office workplaces are also used for functions that are not strictly related to the automation system, the network that they are connected to should be regarded as a general-purpose office network. This should be separated from the automation system by means of a security zone (firewall), as illustrated in Fig. 2. The office and automation system networks should constitute separate domains. Workplaces in the automation system should not be used for accessing the Internet or for incoming e-mail. If there is a strong requirement to support incoming e-mail, the security zone should include an e-mail proxy that removes attachments and active contents.

Furthermore, if the office network is connected to the Internet, particular care must be taken to protect it from attacks. See the survey paper [138] for a discussion of attacks and countermeasures for corporate office networks with Internet access. For high security requirements, separating the office network from the Internet is done by means of a so-called demilitarized zone (DMZ), where corporate Web servers and other servers that are to be accessible from the external network are placed. The DMZ is isolated from both the office network and the external network by firewalls. These firewalls are configured to allow access from the Internet only to selected servers. To ensure that any intrusion attempts to the office network are detected as early as possible, the DMZ should include an intrusion detection system (IDS). A separate security management system should be used to supervise the firewalls and intrusion detection system. It should collect logs from the firewalls and intrusion detection system, analyze these, and generate an alarm if it concludes that there is an attempted intrusion occurring.

3) Connecting to a Corporate Network: As the number of users in the office network grows, so do automation system security concerns. A corporate network with thousands or even tens of thousands of users must, from an automation system security perspective, be regarded as potentially as hostile as, for example, the Internet. The automation system is still assumed to be physically protected (i.e., physical access to the system, including network equipment and cables, should be limited to authorized people). For the corporate network typically spanning several sites or even countries, strict physical protection of all involved network equipment is, however, difficult or impossible to implement and maintain. In this case, the automation system should be isolated from the corporate network in the same way

the corporate network is protected from external networks with a full-blown DMZ in which servers that are to be accessible from the corporate network are placed [66], [100]. In cases where there are several separate automation system installations at the same geographical site, these can be connected to the corporate network through the same DMZ. An IDS should be placed in this DMZ to detect any intrusion attempts emanating from the corporate network to the automation networks.

In the DMZ, a reverse proxy server could be placed. This proxy represents the servers in the automation system that shall be accessible from the corporate network. The best practice is to use a separate proxy server for each additional service that is exposed in this way. The firewalls and proxies should be configured to allow access from the corporate network only to selected servers and services in the automation system, and only from selected nodes in the automation system to selected services and servers in the office or corporate network. The security management system supervises the firewalls and intrusion detection systems, and, in particular, isolates (disconnects) the automation system from the external network in the event an intrusion attempt is detected, until secure and safe operation can be restored.

C. Remote Access to Stand-Alone Embedded Systems

Industrial controllers, especially for power system and transportation applications, are often deployed as stand-alone systems in a geographically dispersed area. Maintenance and service costs of stand-alone embedded systems can be reduced when they can be accessed from remote locations. Remote access services range from read-only actions such as monitoring to control and configuration actions requiring write access. Hence, the impact of actions from remote locations may vary, implying different security requirements.

We consider three remote access scenarios for stand-alone embedded servers.

1) Access Over Dial-In PSTN Line: Here, the server telephone number provides some authentication of the server, as it is unlikely that an attacker would succeed in tampering with the PSTN circuit switching. To authenticate the client over the PPP connection, it is recommended that the server uses a dial-back procedure and applies the CHAP protocol (see Section II-D). Using a dial-back procedure allows to restrict the access to a preconfigured set of phone numbers. An attacker who knows the phone number of the embedded server may, however, perform a DoS attack by continuously dialing the number.

A typical example of such remote access is AMR, where communication modules are attached to the energy or water meter. Periodically, or triggered by some polling, the meter is requested to send its collected data to the load management and billing center. The communication module may be connected to a standard PSTN wired phone line or to a public wireless data network, where temporary connections would be briefly set up for data transfer in off-peak traffic hours. In some areas, specialized wireless or power line communication systems may be deployed. Tampering with meters to alter reported consumption data, or tariff setting commands, is of obvious immediate monetary interest [62], so it is important that security measures are included in such systems. At the meter, mechanical tamper protection or detection is important. The communication-related security measures recommended above should be taken. Specifically, separate passwords for dialing in, configuration, and meter reading or resetting should be enforced.

2) Access Over Unprotected On-Site Network: The embedded system is connected to a LAN, and remote access to the system relies on the third party controlling the LAN to which it is connected. For example, an on-site sensor attached to a LAN is remotely accessed from an off-site location via the Internet. Eavesdropping or message tampering may occur at the LAN or the Internet. Protection measures depend on whether the remote access is required for continuous or for batch data transfer. For continuous data transmission, a secure session must be established, e.g., by deploying SSL between the embedded server and the remote client. However, most SSL implementations do not authenticate the client and there might not be enough computational power and memory available on the embedded system. A further drawback is the need for a certificate on the server side. Maintaining certificates implies additional effort. If confidentiality is not required, the lightweight HTTP DA protocol is an alternative for integrity protection and user authentication (see [55]). If batch data transmission is sufficient, where processing delays are acceptable, the data files can be signed and encrypted offline, e.g., by PGP, before being transmitted.

3) Access Over PPP Connection and the Internet: The stand-alone embedded system may be deployed at some physically unprotected location. PSTN land lines or some public mobile data link provide a PPP connection to a local Internet service provider (ISP), which then connects further to the Internet. The unprotected access links and possibly the ISP are vulnerable to attacks. The same security measures described above for the second scenario, e.g., SSL or HTTP authentication, are also applicable here. The use of IPSec in this scenario often requires appropriate tunneling to cope with network address translations [139]. Reference [140] describes remote access to home automation systems via an intermediate server handling the security services using proprietary protocols.

In all scenarios, measures must be taken to avoid damage by a DoS attack on the embedded system, either by reasonable task priority settings (see Section III-E) or by using different hardware devices for real-time applications and communication. Keeping the number of open ports minimal by deactivating unused services decreases the risk of an attack. Services transmitting passwords as plain text (e.g., FTP, telnet) should be avoided unless the data transfer is secured by lower layer protocols (IPSec or SSL). If none of these security protocols can be deployed and the use of these services is indispensable, they should at least be deactivated by default and activated only temporarily (e.g., by a Web page secured over HTTP DA). It is recommended that highly critical commands should be blocked for remote access and

only be allowed over a local communication interface (e.g., for service personnel physically connecting a maintenance device to the embedded system).

VI. Conclusion

The evolution of industrial communication systems toward increasing interconnection with other enterprise networks or even the Internet, together with a continuously stronger reliance on open standards such as the TCP/IP protocol suite, creates an increased exposure of automation systems to network-based attacks. In consequence, information system security for industrial communication systems is growing in importance.

Information system security can be described in terms of security objectives, such as confidentiality, integrity, availability, authentication, access control, auditability, nonrepudiability, and third-party protection, of which availability and integrity often have the highest priority in industrial systems. Attacks on automation systems, both intrusions and DoS, can have severe consequences ranging from monetary loss up to damages to the environment, as well as injury and loss of life. The threat is real—several documented incidents have occurred in recent years.

Even though there are very different types of industrial communication systems, most of them share certain security-relevant characteristics such as the high priority of safety concerns, importance of availability and integrity, strict timing requirements, and static topologies and configurations.

IT security has been an issue for office automation and e-commerce environments for quite some time, and many concepts and tools developed for these applications remain relevant and should be reused in industrial automation. However, a number of specific challenges to the security of industrial automation systems can also be identified.

The security level of a given software application tends to degenerate over time, as new vulnerabilities in the applications or in its underlying platform are discovered. Today, the standard approach to this problem in commercial IT environments is to frequently issue software updates. For industrial systems, the long operational lifetime of such systems, infrequent service time slots in the production schedule, and sometimes low bandwidth of the access link, often make this approach impractical. The long expected system lifetime additionally creates two types of challenges: how to secure legacy systems currently in operation, and how to design systems today that can easily be adapted to the security threats and technologies of the future?

Another challenge for securing industrial communication systems is posed by automation devices that lack basic security functionality, e.g., the capability to define user accounts, or support for secure communication protocols. Most currently used industrial communication protocols have no (or only rudimentary or optional) security functionality. This security functionality is mostly concerned with access control to data items. It is mainly intended to prevent accidental operation errors, but not to stop dedicated attackers.

A wide variety of conventional cryptographic algorithms and protocols in different layers of the ISO/OSI stack is available to help address some of the security objectives, but cryptography does not solve all security issues. Also, standard security protocols may introduce too much overhead if the industrial communication systems has tight constraints on computational complexity, transmission latency, or communication topology (e.g., multicast).

Independent of a particular technological solution, there are certain best practices for both security architecture design and operations that can be applied to secure any communication system: implement a security policy, pay attention to the weakest link, do not rely on security by obscurity, enforce the principle of least privilege, protect secrets, design for end-to-end security, implement defense-in-depth, use only well-proven cryptographic algorithms and protocols, and invest the necessary continuous effort to keep a system secure in operation.

In summary, if one is willing to invest the necessary effort in the implementation and operation of the security architecture, it is well possible to achieve a reasonably high level of security for an industrial communication system, using conventional and recently emerging industrial system specific security mechanisms, some of which are described in this paper.

Currently, several standardization initiatives on security for industrial communication systems are under way. Some initiatives (e.g., ISA SP99 and IEC SC65c WG13) are concerned with documenting existing security best practices. These best practices can also be applied to industrial communication systems in operation today. Other initiatives (e.g., PCSRF) produce catalogs of functional and operational security requirements for new industrial communication systems.

Information-system security for industrial automation and communication systems has only very recently emerged as a new field in academic and industrial research. The importance of this research field will grow in the near future, because security considerations and mechanisms are increasingly required as a part of standards-based best practices and because enterprises recognize the business case for protecting industrial plants against electronic attacks.

References

[1] "Practices for securing critical information assets," U.S. Critical Infrastructure Assurance Office, Jan. 2000.

[2] IEEE-USA. (2000) Position: Information security in electric power, part of the IEEE-USA legislative agenda for the 107th Congress. [Online]. Available: http://www.ieeeusa.org/forum/POSITIONS/infopower.html

[3] E. Levy, "Crossover: Online pests plaguing the offline world," *IEEE Security Privacy*, vol. 1, no. 6, pp. 71–73, Nov./Dec. 2003.

[4] "Computer virus strikes CSX transportation computers—Freight and commuter service affected (press release)," CSX Transportation, Aug. 2003.

[5] K. Poulsen. (2003, Aug.) Slammer worm crashed Ohio nuke plant net. [Online]. Available: http://www.securityfocus.com/news/6767

[6] U.S. Nuclear Regulatory Commission. (2003) NRC Information Notice 2003-14. [Online]. Available: http://www.nrc.gov/reading-rm/doc-collections/gen-comm/info-notices/2003/in200 314.pdf

[7] T. Smith. (2001, Oct.) Hacker jailed for revenge sewage attacks. *The Register* [Online]. Available: http://www.theregister.co.uk/content/4/22 579.html

[8] E. Byres and J. Lowe, "The myths and facts behind cyber security risks for industrial control systems," presented at the VDE Kongress, Berlin, Germany, 2004.

[9] E. Cole, *Hackers Beware*. Indianapolis, IN: New Riders, 2002.

[10] J. Viega and M. Messier, "Security is harder than you think," *ACM Queue*, vol. 2, pp. 60–65, Jul./Aug. 2004.

[11] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Stanford, and N. Weaver, "Inside the stammer worm," *IEEE Security Privacy*, vol. 1, no. 4, pp. 33–39, Jul./Aug. 2003.

[12] B. Schneier, *Applied Cryptography*, 2nd ed. New York: Wiley, 1996.

[13] W. Mao, *Modern Cryptography: Theory and Practice*. Upper Saddle River, NJ: Prentice-Hall, 2003.

[14] A. Kerckhoffs, "La cryptographie militaire," *Journal des Sciences Militaires*, vol. 9, pp. 5–38, Jan. 1883.

[15] R. Rivest, "The RC4 encryption algorithm (proprietary)," RSA Data Security Inc., Mar. 12, 1992.

[16] *Data Encryption Standard*, FIPS Pub. 46, 1977.

[17] *Data Encryption Standard*, FIPS Pub. 46-3, 1977.

[18] *Specification of the Advanced Encryption Standard (AES)*, FIPS Pub. 197, 2001.

[19] M. Dworkin, "Recommendation for block cipher modes of operation—Methods and techniques," U.S. Nat. Inst. Standards Technol. (NIST), SP 800-38A, Dec. 2001.

[20] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, Nov. 1976.

[21] A. K. Lenstra and E. R. Verheul, "Selecting cryptographic key sizes," *J. Cryptol.*, vol. 14, no. 4, pp. 255–293, 2001.

[22] R. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[23] T. ElGamal, "A public-key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. 31, no. 4, pp. 469–472, Jul. 1985.

[24] S. V. D. Hankerson and A. Menezes, *Guide to Elliptic Curve Cryptography*. Berlin, Germany: Springer-Verlag, 2004.

[25] X. Wang, D. Feng, X. Lai, and H. Yu, "Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD,", Cryptology ePrint Archive, Report 2004/199, 2004.

[26] *Secure Hash Standard (SHS)*, 2002.

[27] H. Krawczyk, M. Bellare, and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication," Network Working Group, RFC 2104, Feb. 1997.

[28] *Digital Signature Standard (DSS)*, 1994.

[29] *Digital Signatures Using Reversible Public Key Cryptography for the Financial Services Industry (rDSA)*, ANSI X9.31-1998.

[30] J. Kohl and C. Neuman, "The Kerberos network authentication service (V5)," Network Working Group, RFC 1510, Sep. 1993.

[31] B. Neumann and T. Ts'o, "Kerberos: An authentication service for computer networks," *IEEE Commun. Mag.*, vol. 32, no. 9, pp. 33–38, Sep. 1994.

[32] W. Stallings, *Network Security Essentials: Applications and Standards*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[33] "Rec. X.509v3 the Directory: Authentication Framework," ITU-T, 2000.

[34] R. Housley, W. Polk, W. Ford, and D. Solo, "Internet X.509 public key infrastructure certificate and CRL profile (PKIX)," Network Working Group, RFC 3280, Apr. 2002.

[35] F. J. Dafelmair, "Improvements in process control dependability through Internet security technology," in *Lecture Notes on Computer Science, SAFECOMP 2000*. Heidelberg, Germany: Springer-Verlag, 2000, vol. 1943, pp. 321–332.

[36] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, Dec. 2003.

[37] R. Needham and M. Schroeder, "Using encryption for authentication in large network of computers," *Commun. ACM*, vol. 21, pp. 993–999, Dec. 1978.

[38] W. Simpson, "The Point-to-Point Protocol (PPP)," Network Working Group, RFC 1661, Jul. 1994.

[39] ——, "PPP Challenge Handshake Authentication Protocol (CHAP)," Network Working Group, RFC 1994, Aug. 1996.

[40] L. Blunk and J. Vollbrecht, "PPP Extensible Authentication Protocol (EAP)," Network Working Group, RFC 2284, Mar. 1998.

[41] J. C. Haartsen, "The Bluetooth radio system," *IEEE Pers. Commun.*, vol. 7, no. 1, pp. 28–36, Feb. 2000.

[42] N. Cam-Winget, R. Housley, D. Wagner, and J. Walker, "Security flaws in 802.11 data link protocols," *Commun. ACM*, vol. 46, pp. 35–39, May 2003.

[43] R. Housley and W. Arbaugh, "Security problems in 802.11-based networks," *Commun. ACM*, vol. 46, pp. 31–34, May 2003.

[44] C. Rigney, S. Willens, A. Rubens, and W. Simpson, "Remote authentication dial in user service (RADIUS)," Network Working Group, RFC 2865, Jun. 2000.

[45] S. Kent and R. Atkinson, "Security architecture for the Internet Protocol," Network Working Group, RFC 2401, Nov. 1998.

[46] ——, "IP authentication header (AH)," Network Working Group, RFC 2402, Nov. 1998.

[47] ——, "IP encapsulating security payload (ESP)," Network Working Group, RFC 2406, Nov. 1998.

[48] D. Hankins and D. Carrel, "The Internet key exchange (IKE)," Network Working Group, RFC 2409, Nov. 1998.

[49] B. Aboba and W. Dixon, "IPsec-network address translation (NAT) compatibility requirements," Network Working Group, RFC 3715, Mar. 2004.

[50] A. Freier, P. Karlton, and P. Kocher. (1996, Nov.) The SSL Protocol, Version 3.0. Transport Layer Security Working Group. [Online]. Available: http://home.netscape.com/eng/ssl3/

[51] T. Dierks and C. Allen, "The TLS Protocol, Version 1.0," Network Working Group, RFC 2246, Jan. 1999.

[52] T. Ylonen and C. Lonvick. (2004, Jun.) SSH Protocol Architecture. Internet Draft, Secure Shell Working Group. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-secsh-architecture-16.txt

[53] A. Robbins and D. Gilly, *UNIX in a Nutshell*, 3rd ed. Sebastopol, CA: O'Reilly, 1999.

[54] J. Franks, P. Hallam-Baker, J. Hostetler, S. Lawrence, P. Leach, A. Luotonen, and L. Stewart, "HTTP authentication: Basic and digest access authentication," Network Working Group, RFC 2617, Jun. 1999.

[55] T. von Hoff and M. Crevatin, "HTTP digest authentication in embedded automation systems," in *Proc. IEEE Int. Conf. Emerging Technologies for Factory Automation (ETFA'03)*, vol. 1, pp. 390–397.

[56] P. Zimmermann, *The Official PGP User's Guide*. Boston, MA: MIT Press, 1995.

[57] M. Naedele, "Standards for XML and Web services security," *IEEE Computer*, vol. 36, no. 4, pp. 96–98, Apr. 2003.

[58] D. Chapman and E. Zwicky, *Building Internet Firewalls*. Sebastopol, CA: O'Reilly, 1995.

[59] S. Northcutt, L. Zeltser, S. Winters, K. Fredrick, and R. W. Ritchey, *Inside Network Perimeter Security: The Definitive Guide to Firewalls, Virtual Private Networks (VPN's). Routers, and Intrusion Detection Systems*. Indianapolis, IN: New Riders, 2002.

[60] J. Viega and G. McGraw, *Building Secure Software*. Reading, MA: Addison-Wesley, 2001.

[61] E. Byres. (2002, Sep.) The myth of obscurity. ISA Safety Community Update. [Online]. Available: http://www.isa.org/rd.cfm?id=1216

[62] R. Anderson, *Security Engineering*. New York: Wiley, 2001.

[63] C. Collberg, G. Myles, and A. Huntwork, "Sandmark—a tool for software protection research," *IEEE Security Privacy*, vol. 1, no. 4, pp. 40–49, Jul./Aug. 2003.

[64] C. Schwaiger and A. Treytl, "Smart card based security for fieldbus systems," in *Proc. IEEE Int. Conf. Emerging Technologies for Factory Automation (ETFA'03)*, vol. 1, pp. 398–406.

[65] Trusted Computing Group (TCG). (2003, Nov.) Trusted Platform Module Specification, v1.2. [Online] . Available: https://www.trustedcomputinggroup.org/downloads/tpmwg-mainrev62Part1_Design_Principles.pdf

[66] M. Naedele, "IT security for automation systems—motivations and mechanisms," *Automatisierungstechnische Praxis*, vol. 45, pp. 84–91, May 2003.

[67] W. Schwartau, *Time Based Security*. Seminole, FL: Interpact, 1999.

[68] E. Byres, J. Carter, A. Elramly, and D. Hoffman, "Worlds in collision: Ethernet on the plant floor," presented at the ISA Emerging Technologies Conf. Instrumentation Systems and Automation Soc., Chicago, IL, Oct. 2002.

[69] M. Naedele and O. Biderbost, "Human-assisted intrusion detection for process control systems," in *Proc. 2nd Int. Conf. Applied Cryptography and Network Security*, 2004, pp. 216–225.
[70] J. Kiszka and B. Wagner, "Domain type enforcement for real-time operating systems," in *Proc. IEEE Int. Conf. Emerging Technologies for Factory Automation (ETFA'03)*, vol. 2, pp. 439–446.
[71] R. Fitz and W. Halang, *Sichere Abwehr von Viren*. Frechen, Germany: Datakontext, 2002.
[72] *Industrial Automation Systems—Manufacturing Message Specification (MMS)*, ISO 9506-1:2003, 9506-2:2003, 9506-5:1999, 9506-6:1994, 2003.
[73] F. Iwanitz and J. Lange, *OLE for Process Control*. Heidelberg, Germany: Hüthig, 2001.
[74] A. Tanenbaum, *Computer Networks*. Englewood Cliffs, NJ: Prentice-Hall, 2003.
[75] R. Droms, "Domain Host Configuration Protocol (DHCP)," Network Working Group, RFC 2131, Mar. 1997.
[76] P. Mockapetris, "Domain names—Concepts and facilities (DNS)," Network Working Group, RFC 1034, Nov. 1987.
[77] D. Plummer, "An Ethernet Address Resolution Protocol (ARP)," Network Working Group, RFC 826, Nov. 1982.
[78] E. Byres, "Designing secure networks for process control," *IEEE Ind. Appl. Mag.*, vol. 6, no. 5, pp. 33–39, Sep./Oct. 2000.
[79] J. Case, M. Fedor, M. Schoffstall, and J. Davin, "A Simple Network Management Protocol (SNMP)," Network Working Group, RFC 1157, May 1990.
[80] G. Jiang, "Multiple vulnerabilities in SNMP," *IEEE Computer (Security Privacy Suppl.)*, vol. 35, no. 4, pp. suppl2–suppl4, Apr. 2002.
[81] W. Stallings. (1998, Oct.) SNMPv3: A security enhancement for SNMP. *IEEE Commun. Surv.* [Online]. Available: http://www.comsoc.org/livepubs/surveys/public/4q98issue/stallings.html
[82] M. Naedele, "Security log time synchronization for high-availability systems," in *Proc. IEEE Int. Conf. Industrial Informatics (INDIN'03)*, pp. 199–206.
[83] D. Mills, "Simple Network Time Protocol (SNTP) version 4 for IPv4, IPv6 and OSI," Network Working Group, RFC 2030, Oct. 1996.
[84] R. Farrow, "VLANs: virtually insecure?," *Netw. Mag.*, vol. 18, pp. 62–63, Mar. 2003.
[85] *Inter Control Centre Protocol (ICCP)*, Std. IEC 60 870-6 TASE.2, 1997.
[86] "Power system control and associated communications—Distribution automation using distribution line carrier systems," Int. Electrotech. Comm. TC N.57, Tech. Rep. IEC 61334, 1998.
[87] A. Modarressi and R. Skoog, "Signalling system no. 7: A tutorial," *IEEE Commun. Mag.*, vol. 28, no. 7, pp. 19–35, Jul. 1990.
[88] M. English, "Safety implications of industrial uses of internet technology," U.K. Health and Safety Executive Contract, Res. Rep. 408/2002, 2002.
[89] D. Box, *Essential COM*. Reading, MA: Addison-Wesley, 1998.
[90] "OPC Security Custom Interface," OPC Foundation, Oct. 2000.
[91] K. Brown, *Programming Windows Security*. Reading, MA: Addison-Wesley, 2000.
[92] Y. Pouffary and A. Young, "ISO transport service on top of TCP (TOT)," Network Working Group, RFC 2126, Mar. 1997.
[93] *Communication Networks and Systems in Substations*, Int. Std. IEC 61850, 2003.
[94] "Initial report from TC57 ad hoc WG06 data and communication security," Int. Electrotech. Comm., Tech. Rep., Sep. 17, 1999.
[95] "ICCP (TASE.2) security enhancements," Electric Power Res. Inst., Tech. Rep., Sep. 2003.
[96] P. Palensky and T. Sauter, "Security considerations for FAN-Internet connections?," in *Proc. IEEE Int. Workshop Factory Communication Systems*, 2000, pp. 27–35.
[97] *EtherNet/IP Specification, Release 1.0*, Jun. 2001.
[98] "HSE presence," Fieldbus Foundation, Foundation Specification FF-586, Sep. 2001.
[99] "Real-Time Ethernet PROFINET IO," PROFIBUS Int., Proposal 65C/359/NP to IEC SC65C, Dec. 2004.
[100] D. Miller and M. Franz. (2003, Jun.) Building secure industrial Ethernets. [Online]. Available: http://ethernet.industrial-networking.corn/articles/i15security.asp
[101] IAONA JTWG Network Security. (2004) *The IAONA Handbook for Network Security, Version 1.01*. IAONA e.V. [Online]. Available: http://www.lAONA.org
[102] T. Hardjono and L. Dondeti, *Multicast and Group Security*. Norwood, MA: Artech House, 2003.
[103] A. Wood and J. Stankovic, "Denial of service in sensor networks," *IEEE Computer*, vol. 35, pp. 54–62, Oct. 2002.
[104] "Single Sideband Power-Line Carrier Terminals," Int. Electrotech. Comm., 1993.
[105] "Signalling on Low-Voltage Electrical Installations in the Frequency Range 3 kHz to 148.5 kHz," Eur. Comm. Electrotech. Standardization (CENELEC), 2001.
[106] M. Lobashov, G. Pratl, and T. Sauter, "Implications of power-line communication on distributed data acquisition and control systems," in *Proc. IEEE Int. Conf. Emerging Technologies for Factory Automation (ETFA'03)*, vol. 2, pp. 607–613.
[107] P. Koopman, "Embedded system security," *IEEE Computer*, vol. 37, no. 7, pp. 95–97, Jul. 2004.
[108] T. Wollinger, J. Guajardo, and C. Paar, "Cryptography in embedded systems: An overview," in *Proc. Embedded World 2003*, pp. 735–744.
[109] V. Gupta and S. Gupta, "Securing the wireless Internet," *IEEE Commun. Mag.*, vol. 39, no. 12, pp. 68–74, Dec. 2001.
[110] A. Steffen, "Secure communications in distributed embedded systems," in *Proc. 2nd Mechatronic Systems Int. Conf.*, 2002, p. 111 ff.
[111] D. Cambridge, "Modern methods and applications of random number generation in signal processing," in *Proc. Global Signal Processing Conf.*, Apr. 2003.
[112] *Evaluation Criteria for Information Technology Security, Version 2.11*, Std. ISO/IEC 15 408, Dec. 1999.
[113] *Trusted Computer System Evaluation Criteria*, Std. DOD 5200.28-STD, Dec. 1985.
[114] Official Website of the Common Criteria Project. CESG, Cheltenham, U.K. [Online]. Available: http://www.commoncriteriaportal.org
[115] *ISO/IEC 17799, Code of Practice for Information Security Management*, Dec. 2000.
[116] *IEEE Guide for Electric Power Substation Physical and Electronic Security*, IEEE Std. 1402-2000, Apr. 2000.
[117] J. Falco, K. Stouffer, A. Wavering, and E. Proctor. (2002) IT security for industrial control systems. Nat. Inst. Standards Technol. (NIST). [Online]. Available: http://www.isd.mel.nist.gov/documents/falco/ITSecurityProcess.pdf
[118] (2003, Sep.) Security capabilities profile for industrial control systems. [Online]. Available: http://www.isd.mel.nist.gov/projects/processcontrol/
[119] (2004, Apr.) System protection profile for industrial control systems (SPP-ICS) version 1.0. [Online]. Available: http://www.isd.mel.nist.gov/projects/ processcontrol/
[120] "Security technologies for manufacturing and control systems," Instrum., Syst., Autom. Soc., Tech. Rep. ISA-TR99.00.01-2004, Mar. 2004.
[121] "Integrating electronic security into the manufacturing and control systems environment," Instrum., Syst., Autom. Soc., Tech. Rep. ISA-TR99.00.02-2004, Apr. 2004.
[122] "Cryptographic protection of SCADA communications: Background and policies," Amer. Gas Assoc., AGA Rep. 12, Draft 2, Jan. 2004.
[123] "Security requirements for cryptographic modules," U.S. Nat. Inst. Standards Technol. (NIST), FIBS PUB 140-2, May 2001.
[124] *Digital Data Communications for Measurement and Control—Profiles for Functional Safe and Secure Communications in Industrial Networks*, Std. IEC 61 784, Dec. 1998.
[125] *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems, Part 1–7*, Std. IEC 61 508, Dec. 1998.
[126] Urgent Action Standard 1200—Cyber Security (2003). [Online]. Available: http://www.nerc.com/~filez/standards-cyber.html
[127] U.S. Dept. Health Human Services, Food and Drug Admin. (FDA), "Title 21 of the Code of Federal Regulations, Part 11: Electronic Records; Electronic Signatures (21 CFR Part 11)," *Federal Register*, vol. 62, no. 54, Mar. 20, 1997.
[128] "Security for information systems and intranets in electric power systems—Annual report 2003," CIGRE JWG D2/B3/C2-01, 2003.
[129] P. Oman, E. Schweitzer, and D. Frincke. (2000) Concerns about intrusions into remotely accessible substation controllers and SCADA systems. Schweitzer Eng. Labs. [Online]. Available: http://www.selinc.com/techpprs/6111.pdf

[130] P. Oman, A. Risley, J. Roberts, and E. Schweitzer. (2002) Attack and defend tools for remotely accessible control and protection equipment in electric power systems. Schweitzer Engineering Labs. [Online]. Available: http://www.selinc.com/techpprs/6132.pdf
[131] B. Ezell, "Risks of cyber attack to supervisory control and data acquisition for water supply," M.S. thesis, School Eng. Appl. Sci., Univ. Virginia, Charlottesville, 1998.
[132] (1997, Mar.) Electric power information assurance risk assessment. Nat. Security Telecomm. Advisory Committee, Information Assurance Task Force. [Online]. Available: http://www.securitymanagement.com/library/iatf.html
[133] M. Naedele, D. Dzung, and M. Stanimirov, "Network security for substation automation systems," in *Lecture Notes on Computer Science, Computer Safety, Reliability and Security (Proceedings Safecomp 2001)*, U. Voges, Ed. Heidelberg, Germany: Springer-Verlag, 2001, vol. 2187.
[134] E. Jirak. (2002, Jan.) Security issues of integrating a standalone system into corporate network. [Online]. Available: http://www.sans.org/rr/managed/stand_alone.php
[135] C. Beaver, D. Gallup, W. Neumann, and M. Torgerson, "Key management for SCADA," Cryptog. Information Sys. Security Dept., Sandia Nat. Labs, Tech. Rep. SAND2001-3252, Mar. 2002.
[136] J. Eul, "Data security of distributed control systems," *Automatisierungstechnische Praxis*, vol. 40, no. 12, pp. 49–52, 1998.
[137] B. Moore, D. Slansky, and D. Hill, "Security strategies for plant automation networks," ARC Advisory Group, Jul. 2002.
[138] C. Landwehr and D. Goldschlag, "Security issues in networks with Internet access," *Proc. IEEE*, vol. 85, no. 12, pp. 2034–2051, Dec. 1997.
[139] A. Kara, "Secure remote access from office to home," *IEEE Commun. Mag.*, vol. 39, no. 10, pp. 66–72, Oct. 2001.
[140] P. Bergstrom, K. Driscoll, and J. Kimball, "Making home automation communications secure," *IEEE Computer*, vol. 34, no. 10, pp. 50–56, Oct. 2001.

Dacfey Dzung (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology (ETH) in 1975 and 1981, respectively.

He has since been with Brown-Boveri, Alcatel, Ascom, and is now with ABB Corporate Research, Baden, Switzerland. He has worked on a number of communication systems and standards, including cellular mobile radio systems (GSM, TETRA), wireless sensors, and industrial power line communication systems. His main contributions are in the design of communication protocols and of modem signal processing algorithms. His current professional interest is in industrial communication networks with special emphasis on communication security.

Martin Naedele studied electrical engineering at Ruhr-University, Bochum, Germany, and Purdue University, West Lafayette, IN. He received the Ph.D. degree in computer engineering from the Swiss Federal Institute of Technology (ETH), Zurich, in 2000.

Currently, he is working as Principal Scientist and Project Manager at ABB Corporate Research, Baden, Switzerland. His main interests are innovative software technologies and software architectures for industrial automation systems, with a special focus on information system security. He coordinates ABB's research on automation system security and participates in international standardization.

Dr. Naedele is a certified auditor for system and network security.

Thomas P. von Hoff received the M.Sc. and Ph.D. degrees in electrical engineering from the Swiss Institute of Technology (ETH), Zurich, in 1996 and 2000, respectively.

In 2001, he joined ABB Corporate Research, Baden, Switzerland, where he works as a Principal Scientist and Project Leader. His main interests are security protocols, signal processing for communication, and estimation of parameters in industrial processes.

Mario Crevatin (Member, IEEE) studied electrical engineering at the University of Applied Science, Winterthur, Switzerland.

From 1995 to 1998, he was with Zellweger Luwa. Currently, he is working as a Group Leader with ABB Corporate Research, Baden, Switzerland. His main interests are signal processing for sensors, software engineering and Internet technologies for industrial automation systems.