The Protection of Substation Communications

Shailendra Fuloria, Ross Anderson
Computer Lab, Cambridge University
Shailendra.Fuloria@cl.cam.ac.uk, Ross.Anderson@cl.cam.ac.uk

Kevin McGrath, Kai Hansen, Fernando Alvarez
ABB Corporate Research
kevin.mcgrath@no.abb.com, kai.hansen@no.abb.com, fernando.alvarez@ch.abb.com

Abstract

The last few years have seen considerable investment in the security of industrial control systems. In the power sector, there has been a focus on operational measures directed by NERC, while technical solutions have been proposed through the IEC and the ISA. Many of these proposals use cryptography to secure communications, so that some of the defensive effort can be moved from the perimeter to the end systems. However the security mechanisms need to take into account implementation costs as well as operational challenges. In this paper, we discuss these challenges in the context of protecting communications within substations. Proposals to use digital signatures are impractical because of both performance and cost. We analyse the complexities of these solutions both from security economics as well as engineering perspectives and argue that the right technology for this application is shared-key, namely message authentication codes: it gives the necessary performance at a fraction of the cost.

1 Introduction

The protection of critical national infrastructure has become a higher priority over the past decade, and electricity generation and distribution is arguably the most critical infrastructure of all; when it stops, so, in short order, does everything else. For example, a six-week failure of the power supply to Auckland, New Zealand in 1996 led to almost 90% of the population relocating; most businesses relocated or closed temporarily [8]. Many shorter regional blackouts, from Brazil in 1999 through the Northeastern USA in 2003 to the damage caused by Hurricane Katrina, have reinforced this message. The military too is aware of the strategic importance of taking down power systems; the US has used this tactic against Iraq as well as Serbia.

Like everything else in a modern society, power systems are becoming computerised, and the control systems are moving from closed proprietary networks to IP. This can save a lot of money, but it can open up new vulnerabilities. Over the last few years, an active control system security research community has evolved, and regulators have decided to intervene. The North American Electricity Reliability Council (NERC) published a set of standards called NERC-CIP [15] aimed at protecting critical assets in the power sector; utilities that fail to comply with these standards face heavy fines. NERC's chain of command goes to the Federal Energy Regulatory Commission (FERC) and finally to the Department of Homeland Security (DHS). NERC is not alone: the National Institute of Standards and Technology (NIST) has also proposed several regulations or guidelines, the most recent being draft standards on Smart Grid Interoperability [19] and Smart Grid Security [20]. NIST also has SP 800-82 [18] on industrial control system security and SP

800-53 [17] on controls for federal information systems and organisations. The potential impact of smart metering on communications networks has also prompted interest from the FCC [7].

As well as governments, international trade and technical associations are getting involved. The ISA 99 (IEC 62443) working group is an ambitious effort to deal with security concerns of the wider industrial automation community [14]. Other standards include IEEE 1686 [12] on substation IED cyber security capabilities; IEEE P1711 [13], a trial-use standard for SCADA serial-link cryptographic modules; IEEE P1689, a trial-use standard for cyber security of serial SCADA links and IED remote access; IEEE H13 on understanding requirements and applications of the substation cyber security standards; and the AMI-SEC [21] or Advanced Metering Infrastructure system security requirements. These multiple standardisation efforts can create confusion leading to standards that conflict, or that let problems fall between the cracks.

This paper focusses on the protection of communications within a substation. The most relevant standard is the IEC 62351 (part 1–8) series of technical specifications for data and communication security [9].

The rest of the paper is organised as follows. In section 2, we discuss IEC 62351, and focus on how IEC 62351-6 [11] sees communication security within substations. In section 3, we point out the areas where this standardisation effort was misaligned with real security objectives and discuss how it is not practicable for equipment vendors to implement the specification in its current form. In sections 4 and 5, we present a first cut of an alternative using mechanisms that can meet the performance goals for such networks.

2 Recommendations of the IEC 62351 Draft

Communication protocols define the architecture of control systems – they determine how a system can gather information from field devices and send them control signals. Yet, despite their criticality, current protocols have almost no inbuilt protection: they evolved in a world of closed proprietary networks that were isolated from the Internet. This assumption, of a strong perimeter defence, is under increasing pressure as operators move to IP networks for reasons of cost and flexibility. Consequently anyone who can talk to a sensor can receive messages and anyone who can talk to an actuator can operate it. This was recognised as a serious vulnerability some ten years ago [4]. Considerable work has been done since then on re-perimeterisation: on ensuring that firewalls and other boundary control devices prevent random people on the Internet from interacting with control systems.

But perimeter defences can be expensive to maintain, particularly in complex networks whose configurations change constantly and which may contain devices or subnetworks managed by multiple firms. It would be desirable to be able to protect individual devices too. This already happens to some extent, in that remote field devices may be connected to a control centre using a VPN to encrypt communications and tunnel them over the Internet. However, rather than using commercial products in an ad-hoc way, it would be preferable to have a proper framework of appropriate standards that work well with control system communications, and that can be built into devices rather than bolted on later. Hence the IEC 62351 series not only aims to tackle system-level security of the information infrastructure, but also communication security.

IEC 62351-1 provides an overview of the scope and objectives of the standard series. It also describes two classes of threats to the information infrastructure: inadvertent threats like equipment failure, carelessness and natural disasters; and deliberate threats from disgruntled employees, industrial espionage, vandalism, 'hackers', malware and terrorism. It further identifies some constraints with protecting power systems: many communication channels are narrowband and most field equipment is constrained in processing power and memory. Moreover the impact of denial-of-service attacks could be much more severe in power systems than in typical Internet transactions. For example, denying authorised personnel access to the emergency power system control in a substation could be far more serious than denying a customer access to her bank account.

IEC 62351-3 to IEC 62351-6 deal with the security of communication protocols. They have been designed for backward compatibility and allow phased implementation. IEC 62351-4 [10] provides mechanisms to protect ISO 9506 – Manufacturing Message Specification (MMS) – and the communications between a control center and its substations. The specification suggests the use of TLS and recommends several cipher suites. IEC 62351-6 takes communication security to the next step – protecting communications within the boundaries of a substation.

2.1 IEC 62351-6

IEC 62351-6 specifies mechanisms for protecting IEC 61850 (specifically IEC 61850-8-1, IEC 61850-9-2 and IEC 61850-6). Here we describe the specification's perspective on the substation communication security – the threats and possible countermeasures.

IEC 62351-6 partly acknowledges the operational challenge of implementing cryptography in field equipment with severely constrained memory and processing power. In section 4.1, it suggests that encryption should not be used for data confidentiality in applications which use GOOSE/SMV and require response times of 4ms, support multicast configurations and must have low processor overhead. The specification argues that since GOOSE/SMV communication is restricted to a logical LAN, the communication path selection process would ensure data confidentiality. These devices however, would still ensure data integrity. All other devices which don't have the above mentioned constraints would implement both data integrity as well as data confidentiality measures.

Section 4.2 suggests that message integrity would be ensured by "message level authentication of the messages." Section 4.3 states that the attack methods countered as a result of employing these security primitives are "man-in-the middle", "tamper detection/message integrity" and "replay". Section 7.2.2.1 suggests that a message authentication code (MAC) would be used for ensuring message integrity. SHA256 would be the computational algorithm to generate the MAC.

However, immediately after this, the specification recommends "the value of the hash shall then be digitally signed" by using RSASA-PSS[1]. To prevent replay attacks, the specification suggest that the clients would keep track of their current times, and a GOOSE message whose timestamp exceeds a 2 minute skew would not be processed. As we will discuss in what follows, the use of digital signatures is not only expensive, but it is also unviable. They simply take too long to compute and to verify.

3 Challenges with IEC 62351-6

There has been much discussion about what the specification actually requires. The ambiguity arises mainly from the fact that it implies using MACs to authenticate messages and also orders the use of digital signatures to sign these messages. It appears that the specification indeed started out by mandating MACs to ensure message authenticity, and later switched over to digital signatures without the document being edited thoroughly. The general view finally adopted by the vendors was that digital signatures would be used to authenticate messages. We now discuss the problems this would cause.

[1] This should have really been RSASSA-PSS (RSA Signature Scheme with Appendix – Probabilistic Signature Scheme).

3.1 Resource-constrained equipment

As noted, IEC 62351-6 provides relief from encryption to applications that use GOOSE/SMV for multicast messaging and that must meet latency requirements of 4ms for computing a signature on one device and verifying it on a separate device. It was thought that encrypting such messages would impose an overhead on resource-constrained processors which might adversely affect overall system dependability. (This isn't actually true; encryption with AES, once keys are set up, takes about three cycles per bit on a 32-bit core up to a few dozen on an 8-bit, and is thus essentially free for the data quantities in question.) However, it appears that this logic was ignored while proposing digital signatures to authenticate messages. These are much more computationally intensive processes than straightfoward encryption and are simply not feasible with any sensible latency in slow processors. For example, low-cost smartcards which lack high-performance cryptoprocessors use variants of banking protocols that use message authentication codes rather than digital signatures.

3.2 The 4ms time-line

Meeting the 4ms latency requirement is the real show-stopper. Even high-end processors such as the 32-bit cores from Intel and ARM cannot in general compute and verify a digital signature using RSA with 1024-bit keys within 4ms. For example, an RSA 1024-bit private key signature operation takes 7–8ms on a single 1.7GHz Intel core using the OpenSSL library, and the computation can't usefully be parallelised. Measurements of performance on a range of PCs and handheld devices give even worse figures [2]. Given that RSA keys with about 700 bits can be broken, and that hardware and mathematics continue to improve, 1024 bit keys are the shortest that a prudent vendor might deploy. (In fact, risk-averse vendors might well prefer 2048-bit keys for which the 4ms target is way out of reach.)

We investigated the options currently available for doing high-speed digital signature computation and verification, whether using the RSA algorithm specified in the standard with 1024-bit keys, or the significantly faster elliptic curve DSA signature algorithm. Considering that the typical processor in substation equipment is significantly less capable than a top-end ARM or Intel core, the best that could be done would be to either replace it with a custom processor (e.g. an ARM with an added crypto accelerator core) or add a separate crypto chip to the board. Either option would significantly increase the cost of the equipment, of course, but could it even be done if cost were no object?

3.2.1 Using a crypto core

The first option is to use an RSA (or elliptic curve) accelerator core. We surveyed the performance claims of some of the commercial IP core vendors; a brief summary is shown in Table 1. The executive summary is that only elliptic curve signatures come remotely close.

Since there are only these two companies to choose from, equipment vendors would be concerned about competition, pricing and lock-in, even if the standard changed from RSA to elliptic curve signatures. (That would cause further problems because elliptic curve cryptography is new enough to have significant patent encumbrance.)

3.2.2 Using crypto-chips

The other option is to use cryptographic accelerator chips. A summary of the performance claims for some commercially available chips is shown in Table 2. Only one chip claims to be able to compute and verify 1024-bit RSA signatures under 4ms – SafeNet's SafeXcel 1840. But reliance on a single product would exacerbate concerns about competition, pricing and lock-in.

IP Core	*Target*	*Clock Frequency (MHz)*	*Density*	*Execution time for one 1024-bit RSA operation*
Helion ModExp STD 256	Xilinx Virtex 5-3	259	642 Slices, 1 RAMB18	24.94
Helion ModExp TINY32	Xilinx Virtex 5-3	148	155 Slices, 1 RAMB18	212.70
Talika	Xilinx Virtex 4LX160	100	N/A	11.97
Intopix IPX-RSA	Xilinx Virtex 4	320	520 Slices, 2x18 kb bRAMs	151.50
Elliptic CLP-23 config. 3	ASIC	200	45K gates	17.12
Elliptic CLP-23 config. 6	ASIC	200	536K gates	0.83
Cadence T-CS-EN-0006-100	ASIC	50	23K gates	62.50
SafeXcel EIP-28b-PE-4	ASIC	250	60K gates	6.30
Certicom Suite B	ASIC	200	60K gates	0.16

Table 1: RSA, EC performance on IP cores

This part is also strictly export-controlled, which would create difficulties for vendors selling outside the OECD countries. And the chip purchasing cost is just a part of the total; the engineering effort required to integrate these chips into existing field equipment would make the overall exercise prohibitively expensive. What's more, as the hash function SHA1 suffers from shortcut collision attacks, it is considered almost broken by crypto researchers. We would have to wait for a chip from Safenet or elsewhere implementing SHA256 or the new SHA3 standard once it's adopted.

Device	*Supported Algorithms*	*Performance*
Atmel AT98SC016CU	SHA-1, SHA-256, RSA, RSA-CRT, RSA-DS, ECDSA, FIPS-140 Random number generator	RSA 2048-bit signature generation in less than 360ms RSA 2048-bit signature verification in less than 60ms
Broadcom BCM 5823	HMAC-SHA-1, RSA	550 1024-bit RSA transactions per second
SafeXcel 1741	HMAC-SHA-1, RSA	RSA 1024-bit sign in 8.4ms RSA 1024-bit verify in 0.85ms
SafeXcel 1840	HMAC-SHA-1, RSA	RSA 1024-bit sign in 0.82ms RSA 1024-bit verify in 0.26ms

Table 2: RSA performance on crypto chips

4 What to Protect, and Why

As it is not technically feasible to use digital signatures to protect messaging within substations, alternate strategies need to be explored. Also, prior to finalising the specification, a vulnerability analysis should be done. IEC 62351-1 emphasises the importance of 'Security risk assessment' in Section 5.5 and recommends that all assets should be assessed for their need and degree of security. However, no specific risk assessment or threat modeling appears to have been done for substation communication. IEC 62351-1 does describe a few threats, but most of them are generic scenarios ('hackers', espionage, terrorism) and not specific threats to communication within the boundaries of a substation. In hindsight, it appears that the specification was technology-driven rather than requirements-driven; it seems to have started with the assumption that public key mechanisms would be used.

The foundation of any solution should have been a threat model followed by security design requirements. Given these, the technical design options should have been compared for cost, performance and other factors such as export control.

We now set out a first cut at a threat model and security policy for substation control systems.

4.1 Threat model

There are perhaps four types of cyber-threat to power grids; these are briefly described in IEC 62351.

1. The first is a deliberate attack by a hostile state or by substate groups; there have been comments, for example, by US government officials about Chinese attempts to map out US critical infrastructure [5, 22]. The worry is that a capable opponent might discover systemic vulnerabilities that allow large parts of the grid to be attacked, causing widespread blackouts at a time of national tension, or that a terrorist might cause blackouts in order to cause alarm and provoke an overreaction.

2. The second is that an insider might sabotage systems. It does happen from time to time that disgruntled employees damage assets, and cyber-vulnerabilities give extra options in addition to more traditional physical damage.

3. The third is that systems might be damaged as a side-effect of malicious activity directed at other targets. For example, a worm infection might cause a server on a monitoring network to send out large volumes of packet traffic, jamming the LAN. If the LAN were shared with control systems, this might make an asset uncontrollable, and even if not the loss of visibility resulting from the compromise of the monitoring system might lead to a precautionary asset shutdown.

4. The fourth is that, as systems become more complex and eventually develop into regional smart grids with millions of participants – including not just traditional generators, distributors and users but large numbers of microgeneration operators and third-party companies, damage may result from the participants' strategic behaviour. This is behaviour not aimed at causing damage maliciously, but merely at maximising profits – for example by misreporting sensor data, or just not bothering to faithfully implement parts of protocols that appear to give them no business benefit.

4.2 Security policy

The next step is to tie down a security policy for the core of the network. By a security policy, we mean a succinct statement of the protection goals, usually in the form of information flow constraints. For a discussion of security policies and examples, see [1]. For smart grids, this is a work in progress. The draft NIST IR 7628 [20] touches on several such policies including multilevel confidentiality (p 77 of the pdf pagination) and dual control (p 106) but implies that security requirements are still to be specified (p 11).

We believe the appropriate security policy at the core of the network is multilevel integrity, also known as the Biba model [3]. Just as typical government systems allow information to flow upwards only from Unclassified to Confidential to Secret to Top Secret, and with various compartments at Secret and Top Secret, so control systems also have multiple levels. However, here the information flows downwards only, from the safety system (the level of highest assurance) to the control system, to the monitoring system, to the enterprise system and finally to the outside world. There is also some compartmentation at the control level (e.g. separate LANs for different parts of a large asset) and even more compartmentation at the top at the safety level (where systems often protect a specific machine or other resource).

In the substation environment there will be typically three levels.

1. First, the safety systems will typically trip on overvoltage and operate reclosers or circuit breakers. After operation, the devices in question notify the control system. Their functionality should never be compromised by anything that happens in the control system; in other words, the information flow is from safety to control but not vice versa.

2. Second, the control system will typically consist of a controller (plus possibly a backup) that communicates to a number of intelligent electronic devices (IEDs). By default this network has a star topology, in that the IEDs communicate only with the controller. In recent years vendors have started using virtual LANs and in some cases it's possible for an IED to broadcast a message to other IEDs (for example, of a protective trip). In older equipment, and with WirelessHART, a broadcast message must be sent to the controller which multicasts it to the other devices. The controller will also typically communicate to the outside world via a firewall or other boundary device, which may be implemented in the controller or a separate device, and which will secure wide-area communications using TLS as per IEC 62351-4. The firewall should further control information flows to the outside world to prevent the IP addresses of internal IEDs becoming externally reachable or visible, and to prevent attacks on external systems causing service-denial or worse to the control system.

3. Third, there may be attached external monitoring systems that take a data feed from the control system but that report externally, for example to give the distribution operator data for consolidated event reporting, quality of supply monitoring, stability monitoring, line fault location, asset condition and grid code compliance. Such systems should not pass data directly up to the control system.

Thus we have a multilevel system in which we want to prevent information flows from monitoring up to control and from control up to safety.

The implementation of this policy at present largely depends on boundary control. Older systems have air gaps, bolstered by proprietary protocols in closed networks. In more recent systems, and in older systems to which modern monitoring systems are being retrofitted, boundary control depends on firewalls of various kinds.

4.3 The function of authentication

However it is increasingly felt that relying on boundary control devices alone is not sustainable in the medium term. There are several reasons for this.

1. It is prudent to assume that state-level actors, such as the national intelligence agency or an electronic warfare unit of a major power, would be able to defeat firewalls by deploying zero-day exploits against its underlying operating system.
2. Even in the absence of a capable motivated opponent, managing firewalls is hard. If configuration is left to local staff, they will typically not touch it after deployment, so some firewalls at least will be misconfigured. If, on the other hand, the company centralises their management, then some companies will have all their firewalls misconfigured.
3. Firewalls are of limited use against malicious insiders.
4. Maintaining a tight perimeter between private and public IP networks is a lot harder than it might seem. People are forever deploying little workarounds in order to get their jobs done, even if these are technically in contravention of company policy; so IP networks that are thought to be private often aren't really. The leakage may take the form of network connections of which security managers are unaware, or even radio access devices that give technicians access to substation equipment from the comfort of their trucks. The former can open up a private network to arbitrary remote access; the latter to an attacker who simply parks close to an asset.

These are known problems for private networks in general. In the specific case of substations, the required outcome is for engineers to be able to configure a substation so that IEDs can talk only to the controller, and perhaps to a backup controller, regardless of any perimeter failures. It should not be possible for an outsider, who acquires the ability to send packets to the IP address of an IED and receive packets back in return, to either read sensors or operate equipment.

In order to enforce this property, IEC 62351 sets out to specify mechanisms for message authentication which can ensure that an IED, or controller, that receives a message has assurance of the authenticity of the device that sent the message. Unfortunately the mechanisms proposed in the current draft of IEC 62351-6 are unusable because of poor performance and high cost. We now turn to the question of what message authentication mechanisms might be appropriate.

5 Authentication Protocol Design

It turns out that there are at least two other applications where designers have faced broadly similar authentication challenges; where they initially preferred a public-key architecture; but where they found that for performance, cost and other reasons a shared-key approach was preferable.

5.1 The European Railway Traffic Management System

Our first case study is the European Railway Traffic Management System (ERTMS) [6], where the challenge was to ensure rail interoperability with a robust yet simple cryptographic mechanism that would span all the major countries in Europe. The approach they take is broadly modelled on the Kerberos approach to authentication [1]. Europe's rail transport system is divided into several Key Validity Areas (KVA) on the basis of geographical boundaries. Each KVA is controlled by a Key Distribution Center (KDC). There are, within the geographical boundaries of each KVA, several Radio Block Centres (RBCs) which interface with local sensors to obtain track occupation and route status. They also interact with the trains' on-board equipment to send them relevant details about the track status etc. This communication between trains and RBCs is sent in the clear, but message authentication codes (MACs) are used to ensure integrity of the data.

To explain the key management principle in brief: Each KDC generates a unique triple-DES transport key (KTRANS) for each RBC under its control and for each train engine that belongs to its zone. These keys are used for initial bootstrapping of trust and must be installed manually. Once this initial trust is established, the KDC sends unique KMAC keys to each pair of (train, RBC). This KMAC is used by the train to authenticate its radio communications and is valid only in the zone governed by that KDC. If a train has to travel to another zone, more KMAC keys are needed. Each new set of KMAC keys is generated by the KDC of the visited zone and passed to the KDC of the home zone which passes them in turn to the train. The basic idea, borrowed from Kerberos, is that a train or RBC need interact only with their KDC; this aids both manageability and scalability. In fact, once all the KTRANS and KMAC keys are distributed for a particular time period, trains and RBCs only need to interact with the KDC in exceptional circumstances.

This mechanism has proved both efficient and robust; it has been adopted by 18 countries in Europe and another nine countries outside the EU, including India and China. There are very strong analogies with power systems: both work on the principle of 'safety first', while information integrity is vital and message confidentiality is of negligible importance.

5.2 Homeplug

Our second case history comes from the security management mechanism of the HomePlug AV domestic power-line communication system [16]. HomePlug supports 155Mb/s communications over domestic mains power lines; a typical use is to provide transparent network bridges between DSL modems and wifi hubs. The challenge here was to design a usable and robust system that would prevent appliances mating by accident with adjacent, interfering networks, while ensuring that no rogue device could add itself to the network.

Some similar systems, such as Bluetooth and IEEE 802.11, had used public-key mechanisms; Bluetooth 2.1 SSP uses a proprietary protocol while for IEEE 802.11, PEAP, WPA EAP-TLS and EAP-TTLS use TLS. However public-key mechanisms impose significant implementation costs, as already noted.

Furthermore, in a hostile environment, it's not enough to leave the authentication to an automatic public-key mechanism. Consider a sensitive user – say an attorney working at home with highly sensitive patent applications. A private eye performs an attack by parking nearby in a van and jamming her wifi hub. The attorney goes to her network management application and sees the message: "Netgear wifi hub wishes to join network: certificate hash a427 b34f 01ff 63bc 9087 bc23 1524 af20. Admit? (y/n)" She assumes a transient network failure and admits the device – and her network is compromised, as unknown to her this is the certificate of a Trojanned device that the private eye has attached to the mains cable outside her house. The moral is that it's not enough to get a user to compare a certificate hash with a displayed value (e.g. the value displayed on her wifi hub) – you have to get her to actually type the value in. And if the users are going to do that, then the public-key mathematics isn't doing any real work. You might as well get the user to type in an AES key that's printed on the device, or on its packaging. That is what HomePlug does, in its Secure mode of operation. (There is a further mode, Simply Connect, where the network controller simply sends a raw AES key to the device; sending a clear key over the network leaves a vulnerability if there's an attacker listening right then, but for most domestic users this isn't even remotely a concern.)

The design of HomePlug was driven above all else by a concern for usability. With high-volume low-margin consumer electronics gear, a device that doesn't work first time (or that fails subsequently) is likely to be returned for a refund, which vendors are extremely anxious to minimise. Avoiding the extra cost of building in public-key capabilities into the chipset was a secondary consideration.

Substation communications are remarkably similar. They require safe usability and high dependability; crypto keys should be installed once by engineers and must thereafter just work. The last thing a utility needs is to have to replace public-key certificates that expire every two years in tends of thousands of IEDs: repair visits are just as much to be avoided as are returns in consumer markets. And in the threat models we have considered here, attacks by insiders at the time of installation can be largely discounted.

5.3 The design space for an updated IEC 62351-6

When protecting communications within a substation, there is no real need for non-repudiation, as the equipment on the virtual LAN all belongs to the same operator. (In fact, using the non-repudiation properties of digital signatures is fraught with practical and legal difficulties even in multi-stakeholder e-commerce systems.) Thus the choice between shared-key and public-key authentication mechanisms is fundamentally one of key management.

In wide-area networks, the choice between using public-key mechanisms with a public-key infrastructure (PKI), as with TLS, versus using shared-key mechanisms with a key distribution centre, as with ERTMS, is an empirical design question. In the limit of a large number n of users, each has a per-user cost of k log n; and which will have the larger value of k depends on the design detail. The choice of TLS for IEC 62351-4 commits the industry to the construction of a PKI for substation controllers, but it does not follow that this PKI can be economically or usefully extended from one certificate per substation to one for every IED within each substation.

We must start from the basic protection requirement: that an IED, or controller, that receives a message has assurance of the authenticity of the device that sent the message. For this it is sufficient that each IED share a key with its controller so that the two of them can compute, and verify, authentication codes on messages between them. In the common case of a star topology, there is no requirement for authenticated communication between the IED and devices outside the substation, and it follows that a MAC is sufficient.

Thus the simplest possible mechanism is that each IED comes with a factory-installed 128-bit key, which is also printed on its packaging, perhaps as 16 hex digits. The engineer who installs it enters these values into the controller. The controller then exchanges a message pair with the IED to verify that the key works. Thereafter each device D uses its device key, KD, to authenticate messages to and from the controller. In standard protocol notation:

$$D \longrightarrow C : M_1, HMAC(KD; M_1)$$

$$C \longrightarrow D : M_2, HMAC(KD; M_2)$$

The second case we have to consider is where equipment supports broadcast; for example, GOOSE broadcast messages sent by an IED to every device on the virtual LAN to notify the whole substation of a trip. The canonical mechanism is for the controller to send each IED a shared network broadcast key KBT regularly (once per time period). The network key would be sent to the device, encrypted under its device key, and the encrypted key packet should contain a timestamp and a time-to-live in order to prevent replay:

$$C \longrightarrow D : \{T, L, KBT\}_{KD}$$

The network broadcast key is then used to authenticate broadcast messages:

$$D_1 \longrightarrow D_2, D_3, ..., D_k : M_B, HMAC(KBT; M_B)$$

A further refinement is needed at this point. It is poor crypto design practice to use keys for more than one purpose; so if we're going to use KD to encrypt KBT, and KBT to authenticate messages from the IED to the LAN, we had better also set up a separate device authentication key KDT, encrypted under the initial device key KD, using a similar mechanism:

$$C \longrightarrow D : \{D, T, L, KDT\}_{KD}$$

We believe that this is as simple as the authentication can safely be made. Simplicity is important, first for robustness, and second because there has grown up a thicket of patents on authentication protocols in the past 10–15 years. A prudent designer will if possible use simple mechanisms that were within the scope of someone skilled in the art of 20 years ago. Fortunately, in our application, this is possible (as in HomePlug, which uses essentially similar mechanisms).

When it comes to the choice between the basic HMAC mechanism, and the more sophisticated mechanism with KBT and KDT, there are two minor matters to consider: hash function standards and export control.

NIST is currently running a competition to replace SHA-256 with a more modern hash function, SHA3. There is no particular reason to believe that HMACs computed using SHA256 will become vulnerable any time soon, but given the nature of the standards business there will be pressure to move to SHA3 once it is selected and promulgated. For this reason, vendors might prefer to use AES-CBC MACs rather than HMACs.

Second, the USA, the EU and the other countries party to the Wassennaar Arrangement have agreed that devices supporting encryption with key lengths in excess of 56 bits require licenses to be exported from member states. Thus if a system simply implements MACs using the HMAC mechanism, it will fall outside the export licensing regime. If on the other hand AES mechanisms are used, some kind of license is likely to be necessary; and although it is likely that the industry can secure open general export licenses (OGELs) for IEDs and other equipment, this may still take some time. It may be prudent for the industry to start talking to the export licensing authorities (particularly in the USA) before finalising the standard.

5.4 Wider area key management

As the Smart Grid programme continues to drive investment in, and increased complexity of, power grid control systems, it is not impossible that at some point it will be felt restrictive to have authenticated communication only between IEDs and the substation controller on the one hand, and between the controller and the outside world on the other. There may be a demand for direct, authenticated communications between external systems and particular IEDs.

Such a change would challenge the core network security policy, of multilevel integrity, and its implementation using boundary control devices. It would therefore involve significant effort to work out what access controls and information flow controls would be needed to support the security policy, and in turn what protocol-level mechanisms would be needed to support these controls. So it would be a significant security engineering design effort, and not to be undertaken lightly. Should it eventually be demanded by operators and permitted by regulators, the key management aspects should be easy enough in principle, with two options for provisioning additional key material to IEDs. The first would be to extend the PKI that will be built to support IEC 62351-4, and the second is to build a network of KDCs along the lines of ERTMS.

However we would like to suggest that it would be premature to start designing complex protocol mechanisms for which a need is not yet apparent. We suggest instead that the IEC standardise simple authentication mechanisms, as described here, in the next part of IEC 62351 and leave open the design of wide-area device-level authentication mechanisms to a further part, if and when that should ever become necessary.

6 Conclusions

We have described how the digital signature mechanisms proposed in the current draft of IEC 62351-6 to support the authentication of broadcast GOOSE messages within electricity substations are not practical. It simply takes too long for one device to compute a digital signature and another to verify it.

We have proposed instead a simple authentication approach based on message authentication codes, and explored the options. We sketched a set of protocols to support communication where broadcast is required, and an even simpler set for where broadcast is not required.

Acknowledgement

The first author's research is funded by ABB. The contents of this article do not necessarily express the views of ABB.

References

[1] R Anderson, *'Security Engineering – A Guide to building Dependable Distributed Systems'*, Second edition, Wiley 2008

[2] D Berbecaru, 'On Measuring SSL-based Secure Data Transfer with Handheld Devices', Politecnico di Torino, Dip. di Automatica e Informatica at `http://ala.isti.cnr.it/atti/ISWCS05/iswcs05/pdf/WP1-33.pdf`

[3] K Biba, *'Integrity Considerations for Secure Computer Systems'*, Mitre Corporation MTR-3153 (1975)

[4] EJ Byres, 'Network Secures Process Control', *Tech Magazine*, Instrumentation Systems and Automation Society (Oct 1998)

[5] *'2008 REPORT TO CONGRESS of the U.S.-CHINA ECONOMIC AND SECURITY REVIEW COMMISSION'*, 110th Congress, Nov 2008, at `http://www.uscc.gov/annual_report/2008/annual_report_full_08.pdf`

[6] The European Railway Traffic Management System at `http://www.ertms.com/`

[7] Federal Communications Commission, *'Comments Sought on the Implementation of Smart grid Technology'*, public notice, Sep 2009 at `http://www.fcc.gov/Daily_Releases/Daily_Business/2009/db0904/DA-09-2017A1.txt`

[8] P Gutmann, *'Auckland's Power Outage, or Auckland – Your Y2K Test Site'*, at `www.cs.auckland.ac.nz/~pgut001/misc/mercury.txt`

[9] IEC, *'Power systems management and associated information exchange - Data and communications security'*

[10] IEC, *'Power systems management and associated information exchange Data and communications security. Part 4: Profiles including MMS'*

[11] IEC, *'Power systems management and associated information exchange Data and communications security. Part 6: Security for IEC 61850'*

[12] IEEE, *'IEEE Standard for Substation Intelligent Electronic Devices (IEDs) Cyber Security Capabilities'* at `http://standards.ieee.org/standardswire/archives/sw_apr08_email.html`

[13] IEEE, *'IEEE approves substation data file standard, begins standard for substation cybersecurity'* at `http://standards.ieee.org/announcements/PR_IEEEDataFileStandard.html`

[14] ISA, *'Manufacturing and Control Systems Security'* at `http://www.isa-99.com/`

[15] NERC, 'Critical Infrastructure Protection (CIP)' at `http://www.nerc.com/page.php?cid=2|20`

[16] R Newman, S Gavette, L Yonge, R Anderson, *'Protecting Domestic Power-line Communications'* presented at *Symposium on Usable Privacy and Security (SOUPS)*, July 2006

[17] NIST, *'Recommended Security Controls for Federal Information Systems and Organizations'*, Aug 2008 at `http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final-errata.pdf`

[18] NIST, *'Guide to Industrial Control Systems (ICS) Security'*, Sep 2008 at `http://csrc.nist.gov/publications/drafts/800-82/draft_sp800-82-fpd.pdf`

[19] NIST, *'Framework and Roadmap for Smart Grid Interoperability Standards' Release 1.0 (Draft)*, Sep 2009 at `http://www.nist.gov/public_affairs/releases/smartgrid_interoperability.pdf`

[20] NIST, *'Smart Grid Cyber Security Strategy and Requirements'*, Sep 2009 at `http://csrc.nist.gov/publications/drafts/nistir-7628/draft-nistir-7628.pdf`

[21] OpenSG Users Group, *'Advanced Metering Infrastructure Security'* at `http://osgug.ucaiug.org/utilisec/amisec/default.aspx`

[22] TL Thomas, *'Dragon Bytes: Chinese Information-War Theory and Practice'*, Foreign Military Studies Office, Fort Leavenworth, Kansas, 2004