SEVENTH FRAMEWORK PROGRAMME

Information & Communication Technologies
Trustworthy ICT

NETWORK OF EXCELLENCE

A European Network of Excellence in Managing Threats and Vulnerabilities in the Future Internet: *Europe for the World* [†]

Deliverable D6.3: Advanced Report on Smart Environments

Abstract: This deliverable presents a review of the state of the art in security research related to the smart grid, including an in-depth view on some ongoing projects.

Contractual Date of Delivery	August 2013
Actual Date of Delivery	September 2013
Deliverable Dissemination Level	Public
Editor	Magnus Almgren
Contributors	All *SysSec* partners
Quality Assurance	TUV, FORTH

The *SysSec* consortium consists of:

FORTH-ICS	Coordinator	Greece
Politecnico Di Milano	Principal Contractor	Italy
Vrije Universiteit Amsterdam	Principal Contractor	The Netherlands
Institut Eurécom	Principal Contractor	France
IICT-BAS	Principal Contractor	Bulgaria
Technical University of Vienna	Principal Contractor	Austria
Chalmers University	Principal Contractor	Sweden
TUBITAK-BILGEM	Principal Contractor	Turkey

[†] The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 257007.

Contents

List of Figures

Foreword

In the *Smart Environment* work package in the *SysSec* network of excellence, we especially consider the security of networks and devices that comprise smart environments. In this third deliverable, *Advanced Report on Smart Environments* we study the so-called "smart grid". In Europe and elsewhere, the electrical grid is transitioning into the smart grid to increase flexibility and accommodate large scale energy production from renewable sources. This transition involves, among other steps, the installation of new, advanced equipment, including the replacement of traditional domestic electrical meters with smart meters, and remote communication with devices; for example, allowing remote access to an unsupervised energy production site. In the past, the easiest way to attack the electrical grid would have been to physically access and destroy components. However, with the introduction of the smart grid and its increased dependence on information and communication technologies, the future grid may be vulnerable to pernicious cyber attacks performed remotely.

As emphasized in the *Second Report on Threats on the Future Internet and Research Roadmap* (2012), the smart grid with its SCADA systems and advanced metering infrastructure is a critical infrastructure with such fast cross-disciplinary development, that security issues exist and may become worse unless the research community focuses on these problems and the special requirements for this type of environment.

The objective of this deliverable is to give a broad survey of the ongoing research related to the smart grid. As very large datasets are produced to control the smart grid, we also include sections on consumer privacy and scalable data processing with a focus on building a defense framework. Finally, we highlight research areas from partners of the consortium: for example a new methodology to analyze the firmware of embedded devices, such as smart meters.

Previous deliverables in this series

In the first deliverable, *Report on The State of the Art in Security in Sensor Networks*, we considered low-capability devices such as sensor nodes and their respective networks. Research-wise, we considered the fundamental network-service algorithms for such environments.

In the second deliverable, *Intermediate Report on the Security of The Connected Car*, we considered a specific application area to focus the discussion. The *connected car*, as a research area, is being developed actively both by industry and in academia and with reported security problems.

1 Introduction

1.1 Background

The electrical distribution grid is being transitioned from the traditional grid into the new so-called *smart grid*, partly to become more flexible and to be able to accommodate large energy production from renewable sources. This transition involves, among other steps, the installation of advanced equipment in places where it previously was not found. This includes, but is not limited to *smart meters* replacing the traditional domestic electrical meters.

While the current definition of a smart grid is abstract, it can crudely be summarized as "electricity networks that can intelligently integrate the behavior and actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies" [208]. Simply put, the main purpose is to extend the traditional network so it becomes more flexible by adding new equipment and a management layer; this layer controls the equipment, making the system robust, flexible and easier to administer. This change is necessary to, among other reasons, accommodate the use of more renewable energy sources. Today the primary globally-consumed resource to produce electrical energy is coal, which together with natural gas and oil account for 67% of the total energy produced in 2009 [176]. Nuclear power covers another 13%, while the main source of renewable energy comes from hydroelectric plants (16%). Solar, wind and geothermal energy cover only 3% of the energy production. However, the long-term strategy of many countries is to use more renewable energy. Germany, for example, has announced that all nuclear plants operational from before 1980 will be shut down, and that by 2022 all nuclear power production should be shut down [43]. The plan is to replace the nuclear energy with renewable energy, which by 2020 should count for 35% of the national energy production, i.e. double than what it is

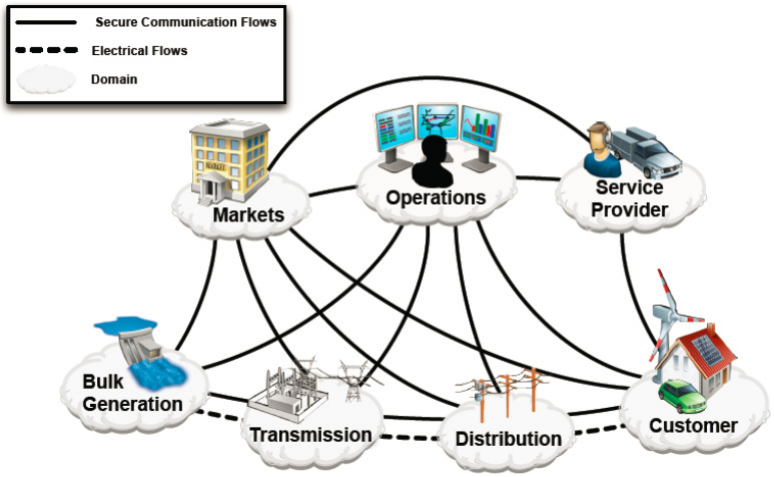

Figure 1.1: Smart grid conceptual model [172]

today, as well as to decrease the total electricity consumption by 10%. The migration to use more renewable energy is one factor driving the adoption of the smart grid, together with the expected wide adoption of hybrid vehicles as well as a better utilization of produced energy, meaning that both the traditional energy transmission and distribution networks are being upgraded.

1.2 Conceptual model for the smart grid

The U.S. National Institute of Standards and Technology (NIST) provided a conceptual model which can be used as a reference for the research and development of a framework to achieve interoperability of smart grid systems and devices [172]. The conceptual model defines the most important domains of the smart grid as shown in Figure 1.1. Here we briefly review the domains. Note that there exist more standards from different areas, countries or organizations, e.g. China [214] or IEEE P2030 [121]. We refer to [193, 227] for a comparison between the different standards.

- *Bulk generation domain*: Energy generators that can use different resources to generate electricity. There the electricity can also be stored for the usage of resource scarcity. The bulk generation domain communicates with the transmission domain and market domain to overcome different quality of service issues such as scarcity and generation failure by routing electricity from different resources.

- *Transmission domain*: Transmission is the bulk transfer of electrical power from generation sources to distribution through multiple sub-

stations. The responsibility of the transmission domain is to maintain the stability of the electric grid by balancing the generation with load across the transmission network. The transmission network is typically monitored and controlled through a SCADA system composed of a communication network, monitoring devices, and control network.

- *Distribution domain*: This domain is the electrical interconnection between the transmission domain and the customer domain. It does not only include the electricity delivery to the customers, but also includes the metering for consumption, distributed storage, and distributed energy generation in order to balance the customer demands and the energy availability. It is within this domain and in the customer domain that you would find the so-called *Advanced Metering Infrastructure* (AMI).

- *Operation domain*: The main responsibility of the operation domain is to analyze and operate power transmission and distribution reliably and efficiently. The operation domain communicates with the transmission and distribution domains to get information of system states, e.g. network connectivity, loading conditions, control device status.

- *Market domain*: In the market domain grid assets are bought and sold. Communications between the market domain, the energy generation domain, and the customer domain are important for efficient matching of energy supply and consumption.

- *Service provider domain*: Service providers in the smart grid manage services like billing and customer account management. They also help the customer for management of energy use and home energy generation. Communication with the operation domain and market domain is important for the service providers, since they need to know the metering values, information about system status and help customers to interact with the market domain.

- *Customer domain*: Customers can consume, store and generate electricity. The customer domain can be divided into sub-domains, including home, commercial buildings, and industrial buildings. In the smart grid system, customers should be enabled to monitor and control their smart devices (further expanded in Chapter 8 about smart homes). It is also necessary that they communicate with distribution, operation, market and service provider domains. Thus the communications within the domain as well as to other domains are important.

Upgrades in parts of the smart grid are going faster than in others. For example, the EU mandates that all the metering devices present in the traditional energy distribution network should be replaced with smart meters

by 2020, in an attempt to better control and monitor the energy consumption. Some countries have already completed the installation of smart meters in the distribution network, such as Sweden, Germany, Italy, and the UK. Smart meters allow remote reading of consumption, hopefully influencing consumer behavior with near-real-time measurements. However, meters will also have other functions such as promptly alerting the distribution company of electrical problems occurring at the site of the customer (such as power outages), and maybe even controlling when consumer devices are allowed to run.

The smart meters, together with other components in the distribution network, form the *Advanced Metering Infrastructure* (AMI). The AMI should be able to provide information in the future of accurate real-time consumption as well as other electric network properties to guide grid operation and also energy production. Other smart grid domains also contain specialized industrial control systems, such as the *Distributed Control System* (DCS) and the *Supervisory Control And Data Acquisition* (SCADA) system.[1]

The AMI and the SCADA system are two very important parts of the information collection done in the smart grid and we will focus on these two systems when discussing protection from attacks in the smart grid.

1.3 The need for security in the smart grid

Many of the components in the smart grid were never designed with security in mind and common communication protocols may lack important security properties; components may simply trust any command sent to them. As long as the networks were localized and isolated, one could argue that it was equally easy to physically attack the grid as it would be to launch a cyber attack as one would have to launch an attack separately for each isolated component. However, nowadays this is no longer true; the individual components and networks of the smart grid, as well as of many other critical infrastructures, are no longer isolated but can be reached through the Internet. The attacker motivation is also clear: the electricity network is a critical infrastructure in society and if it fails, many other systems will in turn cease to function correctly.

The focus of security is thus much stronger in any new component deployed for the smart grid. However, even if many of the security issues in the smart grid are well-known problems in the information and communication technology (ICT) domain, such as buffer overflows in devices and sloppy implementations of cryptographic protocols, the solutions from the more mature ICT domain may not be directly applicable to the smart grid due to resource-constrained devices (smart meters), the life cycle of components

[1] As the functionality of these systems have grown closer together, we will in some parts of this deliverable simplify the presentation by only using the term SCADA system.

(there will always be legacy systems) or the impossibility of immediately shutting down and patching a machine that needs to run 24/7.

There are also challenging new problems originating from the intersection between the electrical engineering and ICT domains, for example where a cyber attack (buffer overflow) in turn affects properties of the electrical grid (power quality), which in turn may propagate back to the ICT domain (vulnerability of control loop). Many of the devices are deployed in the field with little or no physical protection meaning that they run the risk of device tampering.

We discuss attacks against SCADA systems and AMI in further detail in Chapter 2.

1.4 Overview of this deliverable

The smart grid is a complex system of different domains, and to make such a system reliable and secure is challenging. It is important and helpful to study the various characteristics, especially reliability and security requirements of the different domains used in the smart grid. In this report, we focus on research related to the *communication* in the domains or between the domains. We also survey existing detection mechanisms being developed by the research community.

Many of the components now being deployed will create very big datasets, offering both possibilities (in detection and control) as well as challenges (how to analyze the data efficiently). In its current implementation, data processing solutions in smart grids are centralized and will not scale accordingly to the increasing amount of sources and data. To this end, distributed and parallel data processing are needed for better scalability while complying with given time requirements, be it for grid stability or for security monitoring purposes. For that reason, we survey scalable processing of data and provide an outlook on how the so-called *streaming paradigm* can be used both for grid control as well as in different defense frameworks.

Finally, we also describe four research projects connected to *SysSec* partners in greater depth: trusted computing within the smart grid, the possible future of smart homes, improved analysis methodology for embedded devices within the smart grid, and an analysis of how the power quality can be influenced by compromising smart meters.

More specifically, Chapter 2 motivates the need for security in the smart grid and lists known attacks for SCADA systems as well as attacks in the advanced metering infrastructure. Given the sensitive nature of the data collected, Chapter 3 surveys important results in regards to consumer privacy. An overview of the communication used in different parts of the smart grid is given in Chapter 4. We discuss detection frameworks in Chapter 5. We then turn to the need for efficient data processing and give an overview

of the streaming paradigm and how it is used for detection within the smart grid (Chapter 6). In the final four chapters we highlight ongoing research, before concluding the report.

2 Attacks in the smart grid: SCADA systems and AMI

2.1 Introduction

Several key characteristics of smart grids could motivate malicious activities [105]. An attacker might for example be interested in gaining access to a huge communication infrastructure other than the Internet. Such infrastructure would also imply access to millions of devices which, although having constrained resources, could provide a powerful aggregated computational power. Unauthorized access could also be conducted to steal sensitive customer information, which would have high financial value if sold (e.g., to competitor companies) or to illegally monitor household activities (as discussed in [166], behavioral patterns can be easily extracted using basic statistical methods). Finally, access to a smart grid infrastructure could be motivated by the high visibility and impact in case of disruption.

These goals could be achieved following different steps, either "cyber" or "physical". Physical steps are not of particular interest in the context of this deliverable, since they usually require physical countermeasures to protect the infrastructure. As an example, physical tampering of a smart meter or removal of a concentrator unit in charge of collecting data from several smart meters could be avoided by protecting or restricting the access to devices installed in public places.

Among the different steps that can be performed by an attacker, we are particularly interested in the ones that actively depend on the communication channels used by the devices and that can be detected by processing the data exchanged among the latter.

In later chapters we will describe how such attacks could be detected or mitigated in the context of smart grids. However, it is important to grasp the strong connection between the cyber and physical dimensions of such threats in the context of smart grids before proceeding to the actual attack detection descriptions. As discussed in [165], cyber threats launched against

devices in smart grids might result in physical damage, which in turn might result in real-world threats. As an example, a distributed denial of service (DDoS) attack might be conducted instructing thousands of smart meters to flood with messages of a specific power station, which might crash and result in a blackout affecting the energy supply of an hospital. It should be noted that complex DDoS attacks might involve different applications running in the smart grid. As an example, a DDoS attack could be carried out instructing smart meters to report wrong consumption values. If such readings are processed by a demand response application in charge of controlling the supply based on future consumption predictions, a sudden peak in the consumption could lead to a blackout.

Two of the very important control systems of the envisioned smart grid are the SCADA systems and the Advanced Metering Infrastructure (AMI), as discussed in the overview in Chapter 1. SCADA systems can be found in many critical infrastructures, not just in connection to the smart grid. The AMI, on the other hand, is an integrated module in the grid and its deployment may differ depending on in which region of the world it is deployed and the local laws and regulations.

Discussing attacks in conjunction with critical infrastructures is difficult. Many companies do not disclose any details about such events. However, SCADA systems have existed for many years, and some attacks have been documented. For the AMI, most of the suggested attacks are only described in the scientific literature. We give a brief overview of attacks in the two environments here, and return to the topic when we also discuss detection methodologies.

2.2 Attacks against SCADA systems

The first incident on a SCADA system dates back to 1982, when a Trojan supposedly infected the Industrial control system (ICS) that controlled the so-called "Siberian Pipeline" and caused an explosion equivalent to 3 kilotons of TNT [160]. Further exacerbating this scenario, today's SCADA-controlled systems are widespread, given the market traction of smart grids and smart buildings, and thus more appealing to offenders [234, 216, 239]. Although SCADA implementations can vary from vendor to vendor, the specifications of the control protocols (e.g., PLC) are publicly available[1] and the devices can be acquired by anyone given enough funding. In addition, the control software runs on general purpose OSs (e.g., Windows), and were originally deployed in isolated environments where network connectivity was not considered. Needless to say, SCADA software comes with several serious vul-

[1] http://www.modbus.org/docs/Modbus_over_serial_line_V1.pdf
https://www.ashrae.org/resources--publications/bookstore/standard-135

nerabilities[2], most of them caused by buffer-overflow and input validation bugs, which culminated in experts entitling SCADA security as being "laughable"[3]. Unfortunately, these vulnerable ICS are publicly accessible over the Internet. One such center of exploits is called SHODAN[4], a search engine tailored at finding and exposing online embedded devices such as webcams, routers, power plants or even wind turbines. Unsurprisingly, "scada" is the most searched term on SHODAN. How well these exploits perform in real-world scenarios, however, is hard to estimate.

According to the information that CERTS and governments collected, offenders increasingly targeted critical infrastructures of countries: The Industrial Control Systems-Cyber Emergency Response Team (ICS-CERT) responded to 198 incidents against CIs in 2012, 52% more than the previous year. The two most impacted sectors in 2012 are energy (41% of reported incidents) and water (15%)[5].

There are debates among the researcher's community about the accuracy of the answers collected in a recent survey conducted by SANS among industries and organizations that adopt SCADA and process-control systems[6]. Despite such debates, the survey corroborates the anecdotal belief that SCADA and ICS adopters are aware of the security risks. Roughly 50% of the participants reported that they are taking countermeasures that include patching, access control and log analysis. Unfortunately, the PLC layer appears to be a weak spot, where it is often difficult to deploy proper monitoring mechanisms.

For instance, the Stuxnet [84] infection of 2009–2010, which influenced thousands of devices, reached very sensitive targets. A recent report[7] describes that earlier versions of the sophisticated cyber weapon contained other known versions of the malicious code that were reportedly unleashed by the U.S. and Israel several years ago, in an attempt to sabotage Iran's nuclear program.[8] This indicates that Stuxnet was active about two years before the main incident. It also implies that none of the two campaigns of Stuxnet (2007 and 2009–2010) had a serious impact on Iran's nuclear facilities, the avowed main target of the attack. Even though Stuxnet essentially failed, an important fact remains true: Stuxnet was developed (by

[2] http://scadahacker.com/vulndb/ics-vuln-ref-list.html

[3] https://threatpost.com/en_us/blogs/state-scada-security-laughable-researchers-say-020312

[4] http://www.shodanhq.com/

[5] http://resources.infosecinstitute.com/scada-security-of-critical-infrastructures/

[6] https://www.sans.org/reading_room/analysts_program/sans_survey_scada_2013.pdf

[7] http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/stuxnet_0_5_the_missing_link.pdf

[8] http://arstechnica.com/tech-policy/2012/06/confirmed-us-israel-created-stuxnet-lost-control-of-it/

nation states offices, as some experts argue) with careful planning and use of product-specific 0-day vulnerabilities, and it had the potential and the opportunity of causing serious damage on a national level.

The widespread belief that standard protection tools (e.g., VPNs, firewalls, etc.) would suffice to secure network-connected SCADA equipment is just a myth. In fact, Stuxnet reached its targets from an infected USB drive. Then, it used other exploits and local-network probing techniques to find and infect other targets within the production environment. This attack vector is impossible to restrict with network-based access control alone. Instead, a full-blown security infrastructure, including access and account policies would be needed. Something that is not even supported by most SCADA systems and their backbones.

Subsequent milestones were Duqu (2011) and Flame (2012), both designed with intelligence gathering purposes, although Flame is more opportunistic as it spreads also to mobile devices and uses ambient sensors (e.g., microphone) to steal information. These are two examples of the second most important effect of cyber weapons: espionage. Due to the similarity of some code fragments of Duqu, Flame and the variants of Stuxnet, it is not unrealistic to conclude that Duqu was designed to be the precursor of the next Stuxnet [59], to gather intelligence about CI targets.

If Flame will be the precursor of the often predicted "year of cyber attacks (2013)", remains to be seen.

2.3 Attacks in the Advanced Metering Infrastructure (AMI)

Contrary to many other industrial control systems, the smart metering is still in a phase of development and is being rolled out in a number of countries. In Sweden and Italy, the adoption is almost complete and the rest of Europe is following in the same footsteps. Given the novelty of the platform, no known targeted attack has been discovered and discussed openly as with, for example, Stuxnet described above. However, there have been actual attacks and the scientific literature also describes possible paths an attacker may take. The following is a list of attack goals that have been discussed in the scientific literature and in other forums:

- energy fraud
- large scale DoS attack
- destabilization of parts of the grid
- detailed profiling of end clients (issues of privacy)
- targeting specific meters

One of the most commonly discussed goals of an attack would be energy fraud, where a customer would be able to lower his or her electricity bill. Even though not a particularly sophisticated goal, it is one of the few attacks that is claimed to have happened in the wild against the smart meter infrastructure. KrebsonSecurity [139] has reported that the FBI has investigated meter fraud involving many customers in a country, which may have cost the utility several hundreds of millions of dollars.

In a presentation at Black Hat 2009, a researcher showed how a smart meter deployment could be compromised and then used for a DoS attack [70]. As discussed in [105], a DDoS attack against a device deployed in the smart grid (e.g., a Data Collection Unit (DCU)) will usually consist of three steps: 1) installation of malware on the meters, 2) coordination of the attack campaign and 3) flooding of the DCU. Using similar techniques to compromise meters, it has also been shown that the power quality of the grid can be influenced under certain conditions if the attacker has fine-grained control of the remote power switch of the smart meter [65].

These three types of attacks have also been discussed by McLaughlin et al. [158]. They list broad potential attacker goals and use these to create real attack scenarios, covering energy fraud, denial of service, and targeted disconnect of electrical services.

However, from a consumer perspective, the largest concern with smart meters are usually whether the data collection is privacy invasive, especially if the measurements are fine grained. It has been demonstrated that even the television channel can be determined with smart meter data [63]. We devote a separate chapter to privacy (see Chapter 3).

Given the homogeneous nature of a single deployment, a single attack may target many of the meters and thus easily achieve large scale energy fraud or destabilization of parts of the network. However, in some cases an attacker may target specific meters for specific purposes. For example, a targeted attack against a politician may read the energy profiles for possible blackmailing purposes. Single, but important, meters controlling alarm systems may also be targeted in conjunction with other types of attacks (break-ins). Skopik and Ma [207] describe the attacker incentives differently, as financial gain, personal revenge, and chaos.

It should be noted that smart meters are quite rudimentary, with not much protection. Companies in different countries have chosen different types of implementation (further described in Chapter 4 where any of GPRS, powerline communication, or ZigBee networks may be used for communication), so one attack that works in one country might not work in another country. However, within the same deployment the meters are very homogeneous – if it is possible to compromise one meter, it is probably possible to compromise many.

There do not exist many tools to analyze the security of smart meters. In conjunction with Black Hat 2012, a tool was released to probe the optical

port, Termineter [200], now part of Backtrack 5. The need for tools for analysis of embedded systems is further discussed in Chapter 9.

2.4 Summary

An attacker might have many different reasons to attack the smart grid, such as gaining access to a large communication network, espionage, fraud, or to gain visibility by compromising a high profile target. There exists a wide range of possible attacks, and excluding physical attacks, most attacks are targeted either towards the SCADA systems that are used to control the infrastructure, or to the advanced metering infrastructure. A problem is that most companies are unwilling to publicly announce that they have been attacked, but there have been high profile incidents where more details have been made available. SCADA systems usually have very little security and rely on VPNs to protect them from attacks, which is of limited help if the attacker gets physical access to a part of the system. When it comes to attacks to the metering infrastructure the biggest concerns are those of fraud and privacy. An attacker could lower or increase the reported energy consumption to cause disruptions. It is also possible to gain much information about a household simply by monitoring the energy consumption.

3 Data privacy in the advanced metering infrastructure

3.1 Introduction

As a concept, Warren and Brandeis [238] gave in 1890 the definition of "privacy" as the "right to be let alone." Every section of the electrical grid (generation, transmission, distribution) is under a process of transition and modernization towards the new smart grid, where each section will produce data of a higher frequency than before and with more types of information. Thus, each part comes with privacy concerns but in this chapter we will focus on data collected in the Advanced Metering Infrastructure (AMI). These are data most closely tied to an actual person and thus more sensitive than data collected elsewhere. For example, it has been shown that even the television channel can be determined with the metering data of a very high frequency [63]. According to LeMay et al. [145], there can be attackers with different motives trying to get detailed profiles of end clients. In particular, they are *curious eavesdroppers*, who usually try to get the metering information from their neighbors by listening/eavesdropping the surrounding meters; *motivated eavesdroppers*, who use the metering data to get detailed information about the behaviors of the victims, to help them plan crimes, such as theft in a building; *active attackers*, who try to use the detailed profiles to decide how to attack the critical infrastructures in order to maximize the damage.

A *smart meter* is a small embedded device whose main purpose is to automate the energy consumption readings. It supports two-way communication and it is the main actor in the AMI. Before the smart meters, energy consumption readings were made every month, in the best case, usually by a human operator visiting each customer individually. With the help of smart meters, the Electricity Distribution Company (*Distribution System Operator* or shortly *DSO*) can obtain information about the energy consumption almost in real time.

3.2 European smart grid regulation

Data privacy is usually protected by laws, which set up the framework in which personal data are collected and processed by automatic means, the quantity of data that should be collected and also proper specification of the purpose for which data are collected. To the best of our knowledge, there is no specific European Directive which covers smart metering data privacy. Thus, these types of data are covered by the general European Directive, EU Data Protection Directive 95/46/EC [77]. The EU directive sets up the characteristics presented earlier and also sets up the rights of the data subjects to be informed about the data collection process and possible transfer of their data between entities, to be able to verify their collected data, perform correction and to object to the way in which their data may be used (such as direct marketing).

At a national level, the German Federal Office for Information Security released a protection profile for the gateway of a smart metering system,[1] which can be used by electricity companies for the AMI. Closely related to this protection profile, Stegelmann and Kesdogan [217] propose an architecture called *GridPriv* that includes a non-trusted k-anonymity service for pseudonymised meter data.

3.3 Privacy-preserving solutions

Siddiqui et al. [203] make an overview of some of the proposed solutions towards preserving privacy in the smart grid and divide these into the following categories: *anonymous credentials, third party escrow mechanisms, load signature moderation, smart energy gateway, and privacy-preserving authentication.*

Anonymous credentials are based on blind signatures (similar to the ones used in the e-cash payment systems) and have the advantage to offer privacy protection against both DSO and third parties. The disadvantage of this solution is that it does not provide availability of billing data and it can only be used for pre-paid energy.

Third party escrow mechanisms [39, 80, 232] require the presence of a trusted third party entity whose role is to anonymize the data collected from the customers and then present it to the DSO or to aggregate the data and present it in an anonymized form. Efthymiou and Kalogridis [80] present a solution based on separation of data into attributable low-frequency data, collected with a lower frequency and used for billing, and anonymized high-frequency data, collected very often and used for grid operation. Each of these will be reported using a different pseudonym (one public and one

[1] https://www.bsi.bund.de/DE/Themen/SmartMeter/TechnRichtlinie/TR_node.html

private) and only the trusted third party is supposed to know the connection between the anonymous pseudonym and the public one. Their solution offers privacy protection against other third parties and also provides availability for billing data. The open question that remains is if the DSO can later recreate low-frequency data from the high-frequency and match it with the already available low frequency data and so breaking the privacy (see Section 3.5).

Load signature moderation [128, 129] is a good privacy preserving method that can be used by customers. It requires the presence of an energy storage facility at the customer's premises, such as an old battery from an electrical vehicle. The customer can then even out his/her external load signature by drawing energy from the battery in the high-load periods or by charging it during the low consumption periods or when energy is cheaper. This method offers protection both against DSOs and other third parties and also provides availability of billing data, because the Smart Meter will register only the energy used from the electricity network. However, the method has the disadvantage of requiring extra hardware.

The last two categories proposed by Siddiqui et al. [203] are *smart energy gateway* and *privacy-preserving authentication*. In the same way as load signature moderation, these also require the presence at the customer's premises of a dedicated system. In the first case the system is responsible to manage data released from the smart meter on some internal rules based on the data requester, while in the second case its role is to create trusted pseudo-identities that are used in requesting different energy amounts. In the first case privacy protection and availability of billing data can be enforced by setting up proper rules; the second one can only be used in a pre-paid energy scenario.

Hiding in the crowd is another method used to preserve privacy. Borges et al. [42] present a solution based on anonymity networks in which a customer uses two different identities to send his billing data and grid-operational data. While the billing data is directly attributable to the customer, the grid-operational data is forwarded to the DSO through an anonymity network, so the customer cannot be directly identified in a group of customers from the same network.

Focusing on demand-response schemes, Cárdenas et al. [47] present the problem of appropriate sampling intervals in AMI as a trade-off between keeping a good level of customer privacy and gains in the demand-response scheme properties. They also take into consideration the economics behind this problem as a parameter into the proper sampling scheme.

3.4 Data aggregation and homomorphic encryption

Data aggregation can be used as a privacy-preserving solution. Before data is aggregated, one initial step in order to prevent unlawful disclosure of information is to perform mutual authentication [245, 246] between the entities involved in the process. Following this, privacy against DSO's and third parties can be obtained by using homomorphic cryptography [42, 140, 157, 246], or by adding random noise from a known distribution of zero mean [140, 157]. Li et al. [146] also proposed an in-network data aggregation approach which can reduce the network traffic used for reporting the energy consumption data.

Unfortunately aggregating methods do not provide availability of billing data and techniques based on homomorphic cryptography can be expensive on devices with reduced processing power and low resources such as the currently deployed smart meters. Homomorphic encryption is used to encrypt the data for the privacy issue but certain algebraic operations can stilled be performed directly on the ciphertext. In the proposed aggregation approach, there is only additive operations, so they use Paillier cryptosystems [177] which is commonly used for additive homomorphic encryption. A brief description of Paillier cryptosystems is shown below:

Let $E(\cdot)$, m and r be the encryption function, a message and a random number, respectively ($m \in \mathbb{Z}_N$ and $r \in \mathbb{Z}_N^*$). Given m, the ciphertext c, is $c = E(m) = g^m \cdot r^N \ mod \ N^2$, where g is a random number for the base ($g \in \mathbb{Z}_{N^2}^*$), and N is the result of a multiplication of two large prime numbers. Then, the additive homomorphic property is as follows:

$$E(m_1) \cdot E(m_2) = \left(g^{m_1} \cdot r_1^N\right)\left(g^{m_2} \cdot r_2^N\right) mod \ N^2$$

$$= g^{m_1+m_2} \cdot (r_1 r_2)^N \ mod \ N^2$$

$$= E(m_1 + m_2)$$

Seo et al. [201] also use homomorphic encryption to construct a secure and efficient power management mechanism. The proposed mechanism enables to gather the power demands from customers securely and efficiently, and to distribute power to the customer who has valid tickets. Furthermore, each customer can verify whether one's request is correctly delivered to the utility, and each distributor can detect misbehaving customers exceeding their consumption requests. Figure 3.1 illustrates the energy demand aggregation phase in the proposed mechanism.

Instead of using homomorphic encryption to preserve information privacy, code spread scheme is proposed for reporting metering values from the meters to the access point (AP) in the wireless communication infrastructure [147]. When reporting the load, each smart meter uses a random sequence, which is also known to the AP, to spread the signal to a vector.

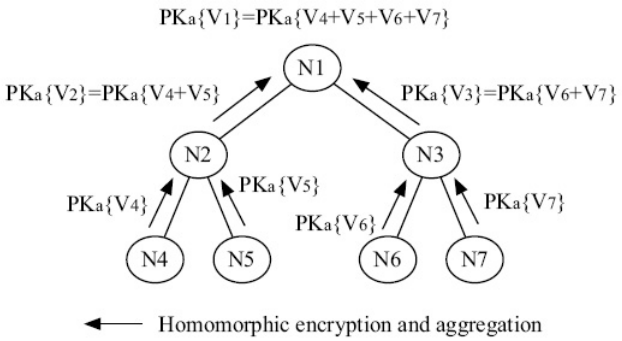

Figure 3.1: Energy demand aggregation [201]. PK_a is used for homomorphic encryption and aggregation of the customer's demand. V_x is the energy demand of node x, where $x \in \{4, 5, 6, 7\}$

.

If the random sequence is unknown to the adversary, the adversary cannot reconstruct the metering values. Rial et al. [189] used a zero-knowledge proof to construct a billing mechanism. In the proposed mechanism, users combine the meter readings with a certified tariff policy, to produce a final bill. The bill is then transmitted to the provider alongside a zero-knowledge proof that ensures the calculation to be correct and leaks no additional information.

3.5 De-anonymization attacks

Privacy enhancing techniques should also be resistant to attacks. Jawurek et al. [126] present the problem of breaking smart meter privacy by using de-pseudonymization. They propose a framework based on machine learning with support vector machines for the analysis of consumption traces and tracking consumption traces across different pseudonyms by using two linking procedures. Linking by Behaviour Anomaly (LA) tries to link a real ID to a consumption trace or two consumption traces together by correlating anomalies that happen in the same time, for example consumption spikes or blackouts. Linking by Behaviour Pattern (LB) tries link to different pseudonyms for one consumer and their method can be applied even if the consumption profiles do not overlap in time.

Buchmann et al. [44] show that identification of individual houses based on their energy-consumption records is possible even by using simple statistical tools such as means and standard deviations on a reduced number of data features. They show that 68% of the records coming from a set of 180 houses can be re-identified by using these simple methods.

3.6 Synthetic datasets

Large datasets of real data are important for research purposes, and because of the privacy concerns, obtaining real (even anonymized) smart metering data is a cumbersome process, and companies are reticent in sharing it. To overcome this obstacle, Tomosada and Sinohara [223] propose a method to generate virtual consumption data from real data. Such virtual data could then be useful in the process of estimating the variation of energy load and energy efficiency. The authors claim that this method keeps the statistical properties of the real data while keeping customer's privacy intact.

3.7 Summary

Smart meters allow electricity distribution companies to monitor the energy consumption of its customers in almost real time. This raises privacy concerns. Currently the gathering of data from smart meters is only covered by the generic EU Data Protection Directive. There do exist proposed solutions to the privacy problem, such as third party anonymization or using batteries to even out the energy consumption, but these need to be evaluated further before they can be put to use. Other possibilities include data aggregation and using homomorphic encryption that allows for algebraic operations on values without revealing their actual numerical value. Regardless of the method used to anonymize the data, it must be resistant to de-anonymization attacks. Even quite simple methods can currently be used to identify individual houses based on their energy consumption.

4 Security threats related to the communication networks

4.1 Introduction

To achieve the envisioned functionality in the smart grid, it is commonly agreed that robust and secure communication networks play important roles in interactivity between devices, applications, energy consumers and grid operators. From the network point of view, the smart grid is a complex system of different networks, and to make such a system reliable and secure is challenging. It is important and helpful to study the various characteristics, especially reliability and security requirements of different communication networks used in the smart grid. Therefore we will survey the reliability and security issues and potential countermeasures related to the communication networks in this chapter.

Take the Advanced Metering Infrastructure (AMI) as an example. The communication can happen in different layers: among meters, from meters to the data concentrators, or from the data concentrators to the data centers. Communication in different layers may have different communication protocols, which can become potential targets of malicious adversaries who want to disturb the energy supply, get private information of energy consumption in order to derive personal behavior of the victim [145], or just reducing the electricity bill, etc.

This chapter is structured as follows: We first give an overview of the components in the smart grid communication architecture in Section 4.2, followed by more thorough investigations of security issues in each component and the proposed solutions in the literature in Section 4.3. In Section 4.4, we discuss and summarize the reliability issues in the smart grid communication network. Parts of the above sections include description of research conducted by *SysSec* partners within related projects; namely the CRISALIS[1] European project (FP7-SEC-285477-CRISALIS) on secure critical

[1] http://www.crisalis-project.eu/

infrastructure environments, and the project "Algorithms for Adaptiveness and Robustness in Electricity Networks" (SN07) within the collaboration framework of Chalmers Energy Area of Advance[2].

4.2 Smart grid components and their communication technologies

An overview of the smart grid and its domain is presented in Chapter 1. Before reviewing the communication technologies, we first review the conceptual reference model of important components and domains in the smart grid in order to keep their interaction patterns in mind. These interaction patterns can help us to understand the significance of security and reliability issues of the corresponding interconnected networks that implement the data communications in the smart grid.

To support interactions among different domains (devices, applications), desirable communication technologies of the smart grid must be designed to satisfy the functionalities and performance requirements of probably heterogeneous subnetworks which in general can be classified as:

- *Wide Area Networks* (WANs), which connect distributed subnetworks that serve power systems in geographically different locations.
- *Field Area Networks* (FANs), which connect devices in the electricity distribution systems, e.g. intelligent electronic devices (IEDs), the electrical sensors on the distribution feeders and transformers.
- *Home Area Networks* (HANs), which consist of customer smart devices and the utility networks within the customer domain.

Due to the various types of networks contained in the smart grid, there is still a lot of discussion on what technologies should be used for smart grid data communications and how they should be implemented [210].

In the rest of this section we briefly review the communication technologies proposed in the literature. We refer the reader to [86, 95, 235] for surveys on communication architectures and technologies in the smart grid.

In general, the proposed communication technologies can be categorized into *wireless technologies* and *wired technologies*. Wireless communication has the advantage of low-cost deployment and can reach the areas that are difficult to reach by communication wires. However, wired communication is more reliable than wireless communication regarding the signal interferences and signal blocking by barriers [107].

[2] http://www.chalmers.se/en/areas-of-advance/energy/society-industry/

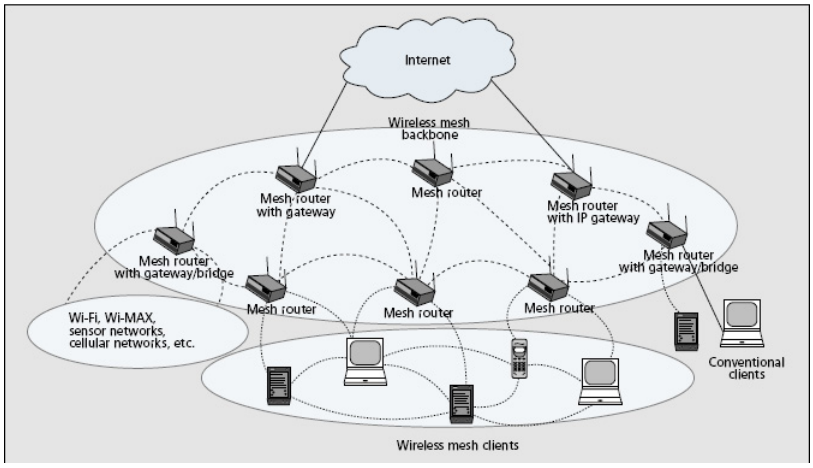

Figure 4.1: Wireless mesh network [22]

4.2.1 Wireless technologies

Below we briefly review the wireless communication technologies proposed in the literature which may be suitable for smart grid usage.

- **Wireless Mesh Networks**: Wireless mesh networks (WMNs) are dynamically self-organized and self-configured, with the nodes in the network automatically establishing an ad hoc network and maintaining the mesh connectivity. WMNs are comprised of two types of nodes: mesh routers and mesh clients. Other than the routing capability for gateway/bridge functions as in a conventional wireless router, a mesh router contains additional routing functions to support mesh networking [22]. As shown in figure 4.1, the mesh routers form a mesh of self-configuring, self-healing links among themselves. Mesh clients can access the network via mesh routers or other mesh clients. Mesh routers can provide connectivity to other networks, such as the Internet, Wi-Fi, WiMAX, cellular.

  The main advantage to use WMNs in the smart grid is the increased communication reliability and the above explained inherent flexibility in the network connectivity. They are crucial for the electric utilities to operate reliably over an extended period of time, even in the presence of a network element failure or network congestion. We refer to [108] for more detailed studies on WMNs used in electric power systems.

- **Zigbee Networks**: Zigbee is a wireless communication technology that is built upon the IEEE 802.15.4 protocol stack [109] to achieve low-complexity, low-cost and low-rate wireless connectivity. Zigbee is an open specification which complements IEEE 802.15.4 standard with network and security layers and application profiles [18]. It

might be one of the most widely used communication technologies in customer home networks (HANs) [86]. Zigbee and Zigbee Smart Energy (SEP) have been defined as one of the communication standards for use in the customer premise network domain of the smart grid by NIST [172]. So far there are many Zigbee certified products (e.g. smart metering devices) from different manufacturers, such as DEVELCO, GE, Itron, TELECOM [17].

- **Cellular Networks** Cellular communication has been a mature technology for data transmission for many years. Due to its established communication infrastructure, a cellular network can be used in smart grid communications without additional time and operational cost to build a dedicated communication infrastructure [107]. Network model and communication methods of using cellular networks for communicating and controlling smart electric devices are proposed in research literature [119, 114]. However, it is also argued that cellular networks (2G,3G) provide good coverage and sufficient throughput to transport metering data, but are insufficient for remote surveillance (e.g. video surveillance), because their uplink capacity is severely constrained and has limited quality of service (QoS) functionality [178]. So there is need for 4G technologies such as WiMAX (Worldwide Interoperability for Microwave Access) and LTE (Long Term Evolution) to help address the issues [178, 151].

- **Satellite Communications** Satellite communication is proposed to be a good alternative communication method for electric system automation in order to reach remote substations where no other communication infrastructure (e.g. cellular network) exists [108]. It can also be a backup communication method for smart grid communications when link failures or network congestions happen in other types of communication networks [81]. However, the disadvantages of satellite communications, such as long communication delay, operation cost and performance dependence on weather conditions and effect of fading [116], should also be taken into consideration.

- **Cognitive Radio** Cognitive radio is an intelligent wireless communication system which can change its operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time according to its surrounding environment. Two primary objectives of cognitive radio are: 1) highly reliable communication whenever and wherever needed and 2) efficient utilization of the radio spectrum [111]. The key challenge in cognitive radio communications is how to make the system intelligent, to be able to achieve the two objectives. Many methods were proposed in the literature for adopting cognitive radio in smart grid communications in order to achieve high volume data

transmission [97], reliable and low-latency routing in large sensor networks [211], and data recovery in the presence of strong wide-band interferences [186].

- **McWiLL Networks** The Multi-carrier Wireless Information Local Loop (McWiLL) is a wireless broadband multimedia trunk system [127]. In paper [127], the advantages for using McWiLL system in the smart grid are discussed, such as fast communication channel setup, multimedia applications support, reliable and efficient radio resources utilization, and IP-sharing.

4.2.2 Wired technologies

Below we list important wired technologies that are used in smart grid communication networks.

- **Powerline Communications** Powerline communication (PLC) is a technology for carrying data on a conductor also used for electric power transmission, so that data and electricity can be transmitted simultaneously without constructing a dedicated communication infrastructure [108]. In the last decades, utility companies around the world have been using PLC for remote metering and load control applications [88], due to its small deployment cost.

  The role of PLC in smart grid communications has broad variations. In [93], the usage of PLC in the electricity distribution grid is discussed. Since traditional substations in the medium voltage distribution grids are not equipped with communication capabilities, using existing powerline infrastructure offers an appealing alternative to the installation of new communication links. In the low voltage distribution grids, PLC can also play an important role in smart metering, communication between electric vehicles and the power grid, and transferring data seamlessly from the smart grid controller to home networks and vice versa [86].

  However, PLC is not as flexible and robust as other communication methods, e.g., wireless communications. If the power distribution network is disconnected, wireless communication connectivity can survive. However, communication of PLC would be degraded and possibly also disconnected in such a case. PLC has the advantage of more accurately predicted message delivery which can avoid network congestion when cooperative schemes are employed [93], and the communication is more power efficient. Data rates on power lines vary from a few hundred bits per second to millions of bits per second, in a reverse proportional relation to the power line distance. Hence, power line communication is mainly used for in-door environments [25] to

provide an alternative broadband networking infrastructure without installing dedicated network wires [235].

- **Wireline Networks** Instead of using power lines to transfer data, dedicated wireline cables can be used to construct data communication networks which can still function well during power line failure. They also have radio interference immunity which wireless communications do not have.

  Networks such as SONET/SDH [13] which use optical fibers can transfer data with rates of hundred-Gbps. It is believed that such networks can play an important role of being communication backbone networks in the smart grid [108] to connect different domains, since they are already deployed as the Internet backbone networks. In the network for home and work places, Ethernet is particularly deployed which can provide data rates up to 10Gbps. The use of Ethernet based networks for substation automation is also proposed [62]. These dedicated networks require extra investment on the cable deployment, but they can offer higher communication capacity and shorter communication delay [108]. Note that, to limit the deployment cost, DSL (Digital Subscriber Line) is proposed as a low-cost and high-bandwidth communication method in the smart grid, since it uses existing wires of the voice telephone network for data transmission [107].

4.3 Communication security issues and countermeasures

As we can see, many communication technologies are used or proposed to be used in the smart grid to establish the automatic interactions between different components. However, integrating communication technologies into the electric grid also creates exploitable vulnerabilities which can be used by attackers of various motivations. In this section, we focus on the attacks aiming at subverting legitimate information transmission. As discussed in [153], the targets of the attacks can be further categorized into *network availability*, *data integrity*, and *information privacy*. Privacy is discussed separately in Chapter 3.

4.3.1 Network availability

The network availability property is fundamental for other network services offered by the communication infrastructure. In the smart grid the network availability is even more important than in the Internet, since the smart grid is more concerned with the message delay than the data throughput

due to the timing constraint of messages transmitted over the power networks [247]. For example, the delay constraint of a generic object oriented substation events (GOOSE) message is 4 ms which is defined in the IEC61850 [124] standard. In general, attacks aiming at network availability can be classified as denial of service (DoS) attacks. In this report, DoS attacks are referred to packets flooding in wired communication networks. In wireless communication, network availability can be affected by *jamming* and *attacks against routing protocols*.

Denial of service attacks: The objective of Flooding-based DoS attacks is to exhaust the victim's network bandwidth or computing resources by means of massive flooding with malicious packets. When the malicious traffic is from distributed sources, the attack becomes a *distributed denial of service* (DDoS) attack. Flooding-based DoS has been one of the major Internet threats for years and it is difficult to mitigate but we list suggested solutions below. When the smart grid communication network is merged into Internet and TCP/IP is adopted as a part of its protocol stacks [124], it becomes a potential target. For example, the adversary can control many end points and flood network or communication links of networking devices and utility business servers.

Several solutions against DoS (DDoS) attacks have been proposed: Some novel ones include *IP-traceback*, *secure overlays* and *network capabilities*. The main idea of IP-traceback is to identify the path through which the attacking traffic is forwarded according to the information in the packets which is attached by the routers along the path, and filter or rate limit the malicious traffic as close to the attacking source as possible [74, 209, 196]. To expand the receiving and filtering capacity of the protected victim, overlays are used to deploy a distributed filter. In overlay-based solutions [135, 26, 215, 92], overlay nodes act as the access points of the potential victim. Any packet that goes to the victim should be checked and forwarded by the overlay nodes. The essential component of the capability-based solutions [244, 248, 179] is a token (capability) which indicates the validity of the packets. Every router along the path may check this capability or some part of it. Only the packets with valid capabilities can reach the protected host. We refer the reader to [181, 161] for a comprehensive survey and taxonomy for DoS (DDoS) attacks and countermeasures.

Jamming attacks: Jamming is a physical layer attack in which the adversary transmits signals over the wireless medium to prevent other nodes from communicating due to the low *signal to noise ratio* [50]. As wireless networks will be widely deployed in the smart grid communication infrastructure, especially in AMI and home-area networks, jamming becomes the primary attack that harms the network availability. In the literature, progress has been made on the development of jamming-resilient commu-

nication schemes [234]. In general the proposed schemes can be classified into coordinated-based schemes and uncoordinated schemes.

Coordinated-based anti-jamming communication schemes can be categorized as *frequency hopping spread spectrum* (FHSS), *direct sequence spread spectrum* (DSSS), and *chirp spread spectrum* (CSS) which are conventional wireless communication techniques against jamming [104, 236]. The main weak point of the coordinated protocols is the secret shared by the communication parties which is used to determine and control the spreading pattern of the signal across the allocated bandwidth, such as the direct sequence in DSSS and the hopping pattern in FHSS. The secret is assumed unknown to the adversary. However, sometimes such an assumption is not valid for open communication standards, such as WiFi and cellular networks. Thus, coordinated protocols are vulnerable to the adversary who has knowledge of the protocol spreading pattern.

Unlike coordinated protocols, uncoordinated protocols [219, 149] do not require the transmitter and the receiver to share a pre-known secret with each other. They randomly generate a secret for each transmission and prevent the adversary from getting sufficient knowledge to disrupt the legitimate communication. For that reason, the uncoordinated protocols are more suitable for secure wireless communication in a distributed environment [234]. Let us take the *Uncoordinated frequency Hopping* (UFH) proposed in [219] as an example. UFH is used for executing a key establishment protocol in the presence of a jamming attack. The established key can then be used to create a secret hopping sequence which can be used by the legitimate communication with conventional frequency hopping. In the key establishment phase, each message is split into multiple parts and then sent with random hopping frequencies chosen from a fixed frequency band. Under the assumption that the adversary cannot jam all the frequency channels at the same time, the sender and the receiver can still communicate through the remaining channels. Initially, the sender and the receiver do not agree on a hopping sequence but instead transmit and listen on random channels. To give the receiver a chance to hear from the sender, the sender hops the channels with a higher rate than the receiver. For instance, if the sender and the receiver hop with 1500 Hz and 100 Hz, respectively, and there are 500 channels, then the receiver will be listening on the frequency where the sender is transmitting in average $\frac{1}{500} \times 15 \times 100 = 3$ times per second. Figure 4.2 illustrates UFH.

Attacks against routing protocols: Another way to prevent legitimate communication is by attacking the routing protocols, since the routing protocols determine the logical connectivity among the network entities. Attacks against routing protocols can happen both in wired networks and wireless networks. Attacks against routing protocols fall into three categories: I) The malicious nodes participate in a route but simply drop some of the data

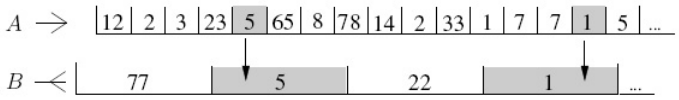

Figure 4.2: Example of UFH [219]. The numbers indicate the frequency channels where sender A is sending and listener B is listening over time. If A and B send and receive simultaneously on the same frequency (5 and 1 in the example) the packet sent on this frequency is successfully transmitted over the undisturbed channel.

packets; II) The malicious nodes modify packets with route information from other nodes, thus harming the routing stability of the network; III) The malicious node advertises itself as having the optimal routes to some destinations, thus attracting a lot of traffic to itself with which it can use for other kinds of attacks, e.g. dropping or analyzing the packets. In wireless scenarios, the first type is called *selective forwarding* and the third type is called *black hole/sink hole* [202].

To fight against the second and the third type of routing attacks, the basic method are digital signatures as described in [185]. Every packet with route information should be signed by the originator for preserving the integrity, thus any modification to the packets can be detected. Furthermore, extra information about the routes should also be included in the messages, e.g. sequence numbers, predecessors and successors in the routes. In such a way the route information can be validated. Most of the research papers in this research area focus on providing efficient cryptographic protocols [32, 125, 134, 117].

Selective forwarding attacks are difficult to deal with, since they are hard to detect. A suggestion is to assign confidence levels to nodes and using routes that provide the highest level of confidence [243]. Bennett et al. [34] also suggest to use a dedicated path between the legitimate communication parties in the Zigbee network of AMI, instead of using the path where the malicious node resides.

Wang et al. [237] proposed a security framework for wireless communication in the smart distribution grid which can be used to limit the effect of routing attacks by malicious nodes with mobility. The core of the framework is a detection and response scheme, called *tracking firewall*. The basic idea is to form a *defense zone* around the detected malicious node. Any node in the defense zone does not send to or receive any of the packets from the malicious node. That way the packets of the malicious node cannot get out of the defense zone and the harmful influence is isolated. Figure 4.3 illustrates the basic behaviors of the nodes to form the tracking firewall.

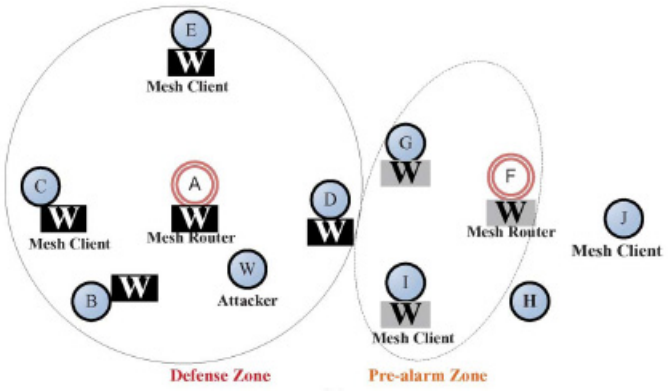

Figure 4.3: Example of tracking firewall [237]. Nodes B,C,D are in the communication area of the malicious node and detect the malicious node. They add the malicious node into the blacklist and generate pre-alarms and broadcast them to their neighbors. Mesh router A receives enough pre-alarms, it adds the malicious node into its blacklist and broadcast this to every node in the cluster. So the cluster of mesh router A forms a defense zone, where any malicious packets cannot get out of the cluster. Node G,I and F receive the pre-alarms from the neighbor cluster, then they form a pre-alarm zone. When the malicious node moves into their communication area, the cluster of mesh router F will become the new defense zone.

4.3.2 Authentication and data integrity

Since correct data or control command transmission is critical for the smart grid, the corresponding communication infrastructure must provide authentication methods to verify whether a traffic flow is correct and is generated by the claimed entity (that can imply a meter, a concentrator or other type of device in the complex system). The authentication mechanisms usually also provide data integrity which is used for verifying that a message has arrived unaltered from its original state [247].

In literature, several works are focused on providing efficient multicast authentication schemes for smart grid applications, since multicast has wide applications in the smart grid, including monitoring, protection, and information dissemination [234]. Using public key based authentication in the multicast authentication scheme is quite straightforward. However, it suffers from a significant computation overhead. Using symmetric key based authentication can achieve computational efficiency, but sharing a single key with a group of nodes may increase the chance that the key is revealed to the adversary if he can compromise some nodes. The research is focused on designing fast and efficient multicast authentication based on asymmetry across receivers. As shown in [55], the proposed methods can be catego-

rized into three categories: *secret information asymmetry*, *time-asymmetry*, and *hybrid-asymmetry*.

In the secret information asymmetry approaches, the sender uses a set of secrets to authenticate a multicast message and gives each receiver a partial view of the secrets used that allow it only to verify authenticity of received messages without being able to generate valid authentication information for messages. The main problem of such approaches is the size of the *message authentication code* (MAC) which is linear with the number of receivers. To deal with this problem, MACs can be truncated across multiple packets thus amortizing the space overhead and being suitable for smart grid applications [221]. Time asymmetry approaches limit the lifetime of keys used to authenticate multicast packets to prevent the adversary to use the key to generate messages with valid MACs. The sender generates the keys periodically to authenticate multicast packets by certain intervals of time. The approaches deal with authenticating the keys themselves and how to generate those keys, so that the lost key can be recovered from the received keys [184]. The main problem of such approaches is the delayed key disclosure and the corresponding packet buffering which may be unacceptable in time-critical applications in the smart grid. The hybrid asymmetry approaches try to eliminate the drawbacks of the other two asymmetry approaches by mixing their underlying asymmetry mechanisms. One-time signatures are proposed to be used in authentication of multicast messages [183], and the "one-time" constraint is also relaxed to "n-time" so that one key can be used to authenticate multiple packets. The method is adopted to construct a multicast authentication protocol for time-critical messages in the smart grid [231].

To achieve secure and efficient multicast data delivery, Kim et al. [137] proposed an overlay-based decentralized data-centric information infrastructure for the smart grid. In the infrastructure, smart devices can join different groups according to the type of data they generate and receive. The multicast group is formed by publisher/subscriber mechanisms [82]. Security control of the infrastructure is implemented in an overlay network of trusted grid hub nodes, as illustrated in Figure 4.4. The hub nodes provide multicast and access control by authenticating and authorizing group joins, providing users with the group keys they are authorized to have. This way, the information generated by a publisher can only be read by the right group of subscribers. Attackers outside the group cannot inject false data into the group.

Replay attack: When the adversary can capture legitimate messages, he can replay them to the destination. These replayed messages may pass the authentication mechanisms, since they are valid packets without any modification. The consequences of successful replay attacks in smart grid communications can be severe, especially if the control commands of critical devices are replayed. To deal with replay attacks, some information may be needed,

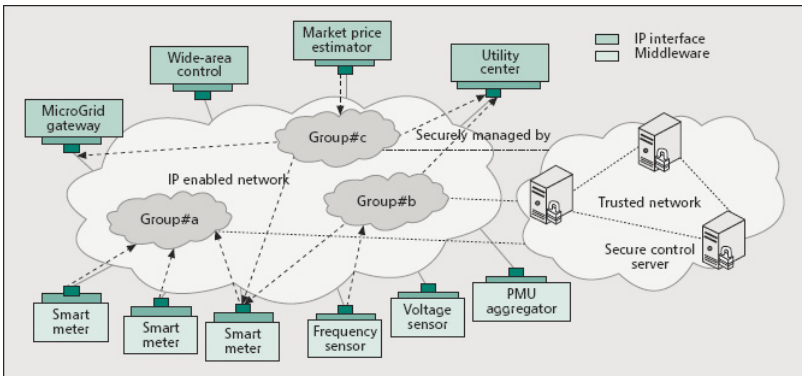

Figure 4.4: Secure grid overlay network [137]

such as timestamps and sequence number. When timestamps are put into the packets, the receiver checks the timestamps with the current time, if the deviation exceeds some threshold, then the packets are dropped. The network-capability based solution is an example of using timestamps to deal with replay attacks [248]. However, using timestamps needs help of clock synchronization which, in turn, can be the target of attacks [155]. Furthermore, network latency also has a big influence on choosing the threshold for timestamp deviations. When sequence numbers are used, the receiver has to save the set of packet sequence numbers it receives; if it receives a packet with a sequence number that is already in the set, the packet is dropped. Saving the sequence numbers increases the space overhead, but it could potentially be done efficiently with bounded space overhead with Bloom filters [164].

4.4 Network reliability in smart grid communications

When using low-power wireless such as 802.15.4 for communication, wireless link dynamics and their impact on network reliability become a key challenge [101, 213, 212]. For example, interference from other wireless technologies operating in the same band or emissions from devices such as microwaves are two main reasons for interference [38]. Both the 900MHz and the 2.4 GHz bands used by 802.15.4 are public bands and, hence, used by multiple wireless technologies. Especially, the 2.4 GHz band that offers the most channels for 802.15.4 and is most common as a result, is shared with many technologies, most notably 802.11. Moreover, even without external interference, the wireless channel between two devices could be influenced by environmental factors such as weather and moving obstacles such as large vehicles between the two devices. Finally, any attacks on the

wireless medium such as jamming attacks add interference to the wireless medium.

Overall, link dynamics are inherent in wireless communication [213, 212]. Moreover, low-power wireless technologies such as 802.15.4 show a stronger sensitivity to these than, for example, 802.11 or cellular technologies due to two key reasons: (1) Their low transmission power increases their susceptibility to interference. (2) Aiming at low cost, small size, and high energy efficiency, 802.15.4 devices use simple radio technologies and thus operate without advanced coding schemes or MIMO capabilities.

As a result, wireless link-dynamics impact the link quality between two devices and potentially lead to connectivity and topology changes. Any wireless communication protocol needs to adapt to these changes to ensure reliable communication. The following mechanisms are common to address link dynamics. They operate on both the link layer and the network layer of the protocol stack.

- **Transmission power management:** Increasing the transmission power allows a device to overpower an interference source. However, this is challenging for most technologies operating in shared bands such as 802.15.4 or 802.11: Their standardization puts strong limitations on the maximum transmission power.

- **Change of coding or frequency:** Another option is to change coding and transmission frequencies to escape a narrow-band interference. While most standards provide a reasonable wide range of channels to operate on, a practical challenge comes in: For example, changing to another channel requires a form of synchronization in the network where the new channel is propagated to all devices. However, when the current channel is interrupted due to interference it is not possible for the system to propagate information about any new channel to use. Thus, devices have to scan the spectrum in order to find the channel to which their neighbors have switched, which is time consuming. Frequent channel hopping (time-synchronized channel-hopping) as utilized by Bluetooth and WirelessHART are an alternative: Here, the network changes channels synchronously every couple of milliseconds following a quasi random pattern known to all devices of the network.

- **Route adaptation:** In dense mesh-networks, where a node has more than one neighbor, nodes can adapt their routing choices and essentially route around an interference source [101, 240]. For this, a node maintains an up-to-date table of link qualities to each neighbor [89, 242]. If the link to one neighbor fails it can choose another forwarder from this neighbor table based on a number of metrics such as link quality and distance to destination, commonly denoted as Expected Transmission Count (ETX) [71].

- **Opportunistic routing:** Opportunistic routing departs from traditional unicast routing as discussed before on one key point: Instead of a sender selecting the next hop, a packet is addressed to a number of potential forwarders [37, 54, 143]. Among the actual receivers, one of the nodes is selected to forward the packet, commonly the one providing the largest routing progress. As a result, opportunistic routing shows a high resilience to node failures and sudden interference, as each packet is addressed to many neighbors at once [143].

4.5 Summary

Threats related to networking can transfer to the smart grid domain, due to network dependencies in the SCADA and the advanced metering infrastructure environments. Clearly, with the necessity of networking, the threat surface of electricity systems changes compared to the classic picture. It is imperative to intensify the research efforts in this direction, both for the sake of the important research challenges, as well as – and more importantly – for the sake of society, due to the criticality of the infrastructure. This chapter provides an overview of the situation and the on-going research in this direction.

5 Intrusion detection systems for the smart grid

This chapter gives an overview of the research regarding Intrusion Detection Systems (*IDS*) adaption to the smart grid. As in Chapter 2, with the overview of attacks, the presentation here also focuses on Supervisory Control and Data Acquisition (*SCADA*) systems and the Advanced Metering Infrastructure (*AMI*). As will be shown we cannot simply take an IDS from the normal Information and Communications Technology (*ICT*) domain and expect it to work well in the smart grid. There are important differences between the ICT domain compared to the smart grid domain that affect the efficacy of a detection system.

The chapter begins with a short overview of intrusion detection systems for the readers who are not familiar with such systems (Section 5.1). Section 5.2 then highlights some of the differences between the ICT domain and systems connected to the electricity network. This is followed by the different requirements posed upon an intrusion detection system deployed to monitor SCADA systems or within the AMI (Section 5.3). Parts of the above include description of actual research conducted by *SysSec* partners on secure critical infrastructure environments. Section 5.4 lists intrusion detection systems proposed in the scientific literature. Challenges facing such research are then discussed in Section 5.5 followed by a summary of the chapter in Section 5.6.

5.1 A brief history of intrusion detection systems

The concept of Intrusion Detection System (*IDS*) for Computer Systems appeared in the early $80's$ [133], with one of the first attributed sources being the report by Anderson [27]. At that time the purpose was to identify misuse of computational systems, often by manual analysis of log files. Automated scripts which could do this job appeared soon, and with them the basis of

what would be called the *host-based IDS*. Later, when the computer systems became interconnected, systems that analysed the network traffic were developed and were thus called *network-based IDS*[1].

Apart from how the events to be analyzed are collected, the intrusion detection systems are often categorized according to the internal models used to detect attacks: signature-based detection versus anomaly-based detection.

The *signature-based detection* is built on a set of signatures of known attacks or improper events for the system or network being monitored. Alarms are raised whenever an event matches one of these signatures. The disadvantage of this method is that it is effective only for the events for which it has the proper signatures. The main advantage is a relatively limited number of false positives (compared to an anomaly-based system).

The *anomaly-based detection* is built on a model of the common behavior of the monitored system or network, and any event that does not fit this behavior is flagged as an anomaly. The advantage here is that this method may discover new anomalies that do not have a proper signature yet, but the disadvantage is a larger number of false positives, compared to the signature-based detection method. A flagged anomaly may not even be an attack but only unusual behavior. As a third option, Hybrid Intrusion Detection Systems use a mix of the above techniques in order to take the best characteristics from each paradigm.

Any type of IDS should be capable of discovering anomalies and misuse; it can then raise an alarm to a human operator or provide input to an Intrusion Prevention System to block an application, turn off a connection or reconfigure the rules of a firewall, among other actions. The efficiency of an Intrusion Detection System is often measured in the percentage of correctly identified events in relation with the number of false positives (the benign events misinterpreted as dangerous) and the number of false negatives (the number of missed detections).

A more detailed classification can be found in Debar et al. [75].

5.2 Differences between the smart grid and the normal ICT domain

Due to the particular nature of the domain, existing IDS solutions for ICT cannot effectively be deployed in their current form in the smart grid. There are differences between the classical ICT systems and the systems in the smart grid that require monitoring [46]. Even though some differences are quite concrete, others are of a more principled nature. For example,

[1] http://www.symantec.com/connect/articles/evolution-intrusion-detection-systems

the smart grid is a critical infrastructure that many other critical infrastructures and services depend upon. A malfunction can cause loss of life in extreme cases, be it either directly or indirectly. Looking at the well-known categorization of security into the *Confidentiality-Integrity-Availability* (CIA) properties, the emphasis is rather Availability-Integrity-Confidentiality in the smart grid. However, *safety* is usually prioritized over the other properties.

Another more concrete difference between these two domains is the lifetime of devices present in each of them. In the ICT domain, the lifetime of devices is usually three to five years, after which they are replaced with a newer generation of devices. During their lifetime, they receive multiple software updates, and the update process is executed every few days or even on a daily basis. In the smart grid the lifetime of devices is usually $15 - 20$ years, sometimes even longer. Many of these devices are also not updated frequently (sometimes not at all after deployment), meaning that the smart grid always has legacy systems with known vulnerabilities.

The types of devices in the smart grid are also quite diverse with a large variety of hardware devices, each with a specific functionality, and these devices can be located tenths or hundreds of kilometers away from the central system. These devices are in turn limited both concerning their computational capability and the available network bandwidth connecting them. Grid stability may depend on the timely reporting of anomalous or malicious events, meaning that the resource limitation may provide challenges for any deployed IDS.

Another difference from classical ICT systems is the number and complexity of protocols used in the smart grid. Together with the classical protocols used in the ICT (TCP/IP, SNTP) there are a number of specific protocols (ModBus, M-bus, Profibus, DLMS/COSEM); some of them are closed source (vendor-locked), while others are based on open standards but with vendor-specific implementations. For an IDS, this may raise difficulties in creating detection semantics.

Even though many of these differences offer challenges to create an effective IDS, there may also be other properties in the domain which makes it easier to develop monitoring solutions. For example, many research efforts focus on the fact that these types of networks seems to be more *regular* than networks in the ICT domain. For that reason, it is believed that anomaly-detection techniques will work quite well (see [30] and [58] among others).

Kush et al. [141] further discuss the challenges faced by IDS.

5.3 SCADA and AMI in the smart grid

As introduced in Chapter 1, the system that monitors and controls the generation and transmission section of the smart grid is called the *Electrical Supervisory Control and Data Acquisition* system (*SCADA*). The complementary

system that is responsible for monitoring a part of the distribution section of the grid is called the Advanced Metering Infrastructure. Figure 5.1 shows the overlay of these two control systems on the electrical network.

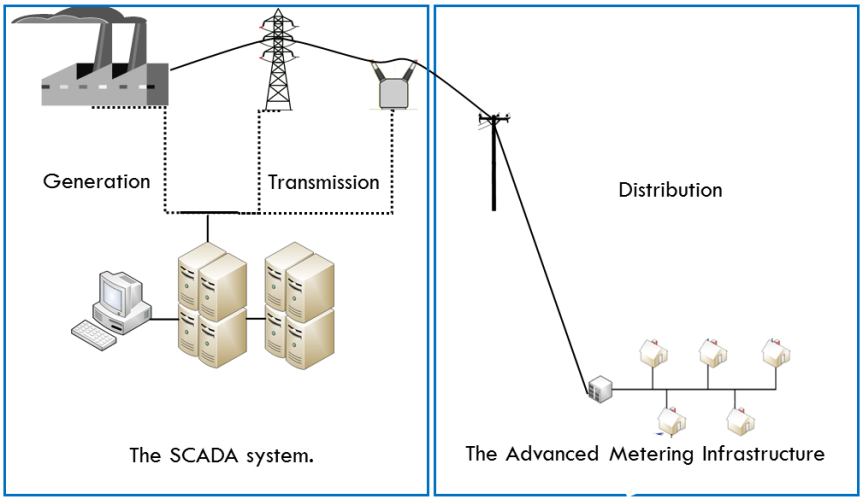

Figure 5.1: Electrical Network - SCADA - AMI

The SCADA system is the core of the electrical control network. It is comprised of a multitude of sensors and controllers, scattered in different parts of the electrical network. The central part receives data from these devices and creates a virtual image of the current state of the electrical network for the personnel that supervises it. The informational flow is bi-directional, status updates are received from the devices in the network and commands are sent back to them if needed.

Figure 5.2 presents a typical architecture of a SCADA network. There are a few types of devices present in such a network. The different switches and valves are controlled by devices called Programmable Logic Controllers (*PLC*) and the sensors are connected to Remote Terminal Units (*RTU*). Data from these two types of devices are presented to the human operator via a third type of device called the Human Machine Interface (*HMI*). Along these devices, the network may also contain a system whose purpose is to log the running process. The computing systems that are used in the SCADA network are nowadays commercial solutions, and most of them use publicly-available commercial operating systems. This makes them vulnerable to many of the known vulnerabilities for commercial systems found in the ICT domain.

The advanced metering infrastructure is an important component of the smart grid. It is mostly localized in the distribution part of the electrical grid, and its main purpose is gathering energy consumption data that is used for the purpose of customer billing. These data can be also used for

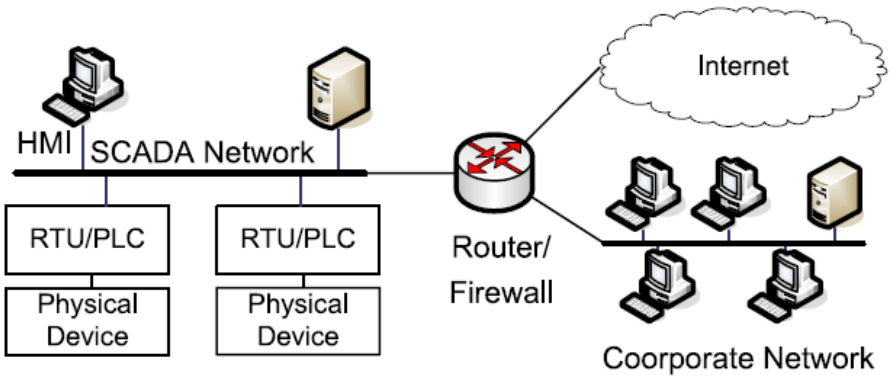

Figure 5.2: SCADA network architecture [30]

grid operation, fraud detection, smart device control and for many other applications. For example, many smart meters have the "remote shut-off" functionality installed, so energy distributors can remotely turn-off energy delivery [28] to customers who forgot to pay the bill, have been relocated from the premises or to regulate the energy consumption in case there are problems in that specific part of the grid. Other functionality that can be implemented on the smart meters is the possibility to remotely control the large energy consumer devices at a customer's premises [195]. By turning off or scheduling these large consumers during periods with high energy demands, one can help reduce the stress on the electrical network.

Even if found in the same domain, these two control systems differ from each other in many characteristics, but mainly in purpose, geographical spread and number of devices they monitor. The *SCADA* system is responsible for monitoring and controlling the electrical energy flow from the point where the energy is produced to the point where it is consumed, while the *AMI* system is present only in the distribution part of the grid to gather energy consumption data used for customer billing. As more energy producing facilities will be installed in the distribution part of the electrical grid, it is likely the AMI will have more functions in the future as detailed above. These two systems are currently not connected to each other, but may become interconnected in the future when new functionalities are added to the AMI system.

When referring to the geographical spread of these two systems, the distances covered by the SCADA system are large; transmission lines could span over tens or hundreds of kilometers. The area covered by the advanced metering infrastructure is moderately large, typically the area occupied by the town or city served by the distribution system operator. One SCADA system may thus cover multiple small areas that are situated at large distances from another, while one AMI system typically covers one large area.

The number of monitored devices also differs between these two systems. SCADA systems employ a number of devices that can vary from hundreds to thousands, while AMI systems can employ tens of thousands of devices (if we consider only the electrical domain, each apartment, house or commercial building must have at least one electrical meter on its premises).

The special characteristics of these two systems and their differences play a big role when designing and developing Intrusion Detection Systems tailored to this domain. For that reason, such systems are either tailored to the SCADA domain or the AMI, seldom for both simultaneously.

5.4 Intrusion detection for the smart grid

Based on the presented background above, let us now have a look at the proposed intrusion detection systems found in scientific literature.

Zhang et al. [251] present a distributed Intrusion Detection System for the smart grid which employs analyzing modules whose role is to provide input to classification algorithms as support vector machines and artificial immune systems models [250] to better classify the events. Their proposed solution covers the whole smart grid (both the SCADA system and the AMI system) by using a hierarchical and distributed Intrusion Detection System called SG-DIDS (Smart Grid Distributed Intrusion Detection System) and it is interesting from a theoretical point of view. Due to the fact that different parts of the grid are managed by different companies, the idea of a general Intrusion Detection System that covers all the parts of the grid is hard to implement in practice.

Kush et al. [141] present an overview of the research concerning Intrusion Detection Systems for smart grid environments, together with a list of nine characteristics of the smart grid environment and nine requirements for an IDS system suitable for this specific environment. They identify the fact that IDS requirements for SCADA networks are different from the ones of AMI IDS and also that the research so far has focused on IDS for SCADA systems and supply-side (transmission) networks and only a small fraction on IDS that can be used in the AMI networks. One explanation for this is that the SCADA systems have been around for a longer period of time than AMI systems, so the interest in them has been greater.

Below we survey the systems suggested in the literature separated based on their use in the SCADA domain or for the AMI.

5.4.1 Intrusion detection systems for the SCADA environment

Zerbst et al. [249] present the different network zones in which the SCADA network can be divided. Different zones contain devices with different functionalities and this division makes applying security levels and defining the

correct data information flows between these zones more precise. Figure 5.3 gives one example of defining these zones for two different companies.

From a theoretical point of view, Liu et al. [150] show that false data may be injected in these sensors that may give false results in the operator's overview. This can lead to the point of making incorrect decisions regarding the stability of the process, which can further lead to unbalancing the process and even shutting it off. It is almost the same with the results from the Intrusion Detection Systems, false negatives and the incorrect interpretation of false positives can have adverse effects in the correct running of the monitored and controlled process.

From a practical point of view, the discovery of Stuxnet[2] showed that SCADA systems are not safe from targeted attacks and specially crafted malware. The damaging economical effects of malware such as Stuxnet are important, and the presence and actions of such malware should be identified early by the Intrusion Detection System.

The SCADA network is sometimes connected to an office network where data from the SCADA network is used for other purposes such as administration and marketing. The data flow between these two networks should be in an ideal case unidirectional, from the SCADA network towards the office network. External actors can also have connections to the SCADA network, for example equipment providers for maintenance and servicing.

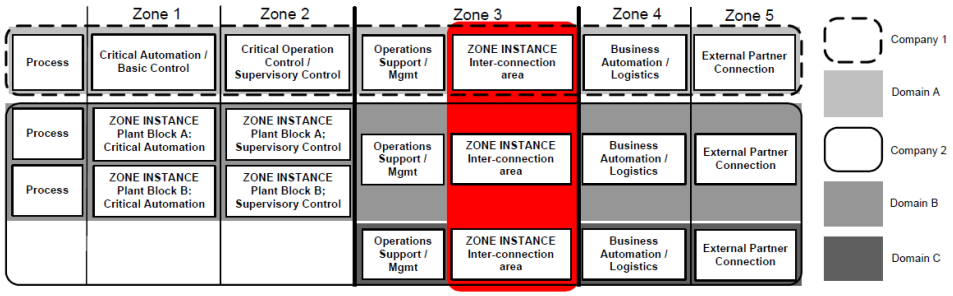

Figure 5.3: Zoning principles as an element of Cyber Security architecture [249]

Barbosa and Pras [30] define the model of an anomaly-based Intrusion Detection System which uses flows to model the network traffic. They make the assumption that network traffic in SCADA networks is *well behaved*, compared to traditional ICT systems. Their assumption is based on three important factors: the number of network devices, the number of communication protocols employed and the communication patterns. The number of devices on the network is somewhat fixed, without large variations during the lifetime of the SCADA network, so the identity of the actors communicating

[2] http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

in this network can be verified somewhat easily. The number of communication protocols employed in the network is also limited to the protocols used between the different devices present here. Apart from that this network should have a regular communication pattern, consisting mainly of the traffic between the Human Machine Interfaces and Remote Terminal Units and Programmable Logic Controllers. The services that are normally used in a traditional ICT network should be disabled, as to not interfere with the normal operation in the SCADA network.

When referring to the communication between Human Machine Interfaces and Programmable Logic Controllers, a protocol which is widely used in SCADA systems is MODBUS.[3] It was developed in 1979 and it is today one of the most used protocols in this domain. Cheung et al. [58] present three model-based techniques for an Intrusion Detection System for the SCADA system, together with a prototype for the MODBUS protocol. They again exploit the fact that the network topology is fixed and the network traffic patterns are regular. The motivation for choosing the model-based techniques (anomaly-based) over the signature-based ones is the limited degree of protocol knowledge and the limited number of known attacks that employ this protocol. The lack of this knowledge makes the option of employing anomaly-based techniques more appropriate for this domain. Based on these results, Valdes and Cheung [229] present a visualization tool that can better present MODBUS communication patterns to the human operator in order for him to better understand and identify anomalous traffic between SCADA devices.

Anomaly-based detection based on Bloom filters for the MODBUS protocol is also proposed by Parthasarathy and Kundur [180] with good advantages such as reduced memory requirements, zero false positive errors while at the same time also providing anonymity. Their solution is distributed among a number of devices in the network and they claim that it remains partially operational in the case that some of the field devices are compromised.

5.4.2 Intrusion detection systems for the AMI environment

Compared to SCADA networks, the advanced metering infrastructure is relatively young. Only some western European countries have fully deployed it completely,[4] and others will start the deployment process in the near future. For that reason, little is known about real-life attacks against the advanced metering infrastructure. Most of the threats and problems that can appear here have been presented in literature from a theoretical point of view. One important aspect is to protect customers' privacy, because it is possible to

[3] www.modbus.com

[4] http://smartenergyuniverse.com/short-takes/16234-advanced-metering-infrastructure-in-europe

infer sensitive information about their lifestyle [166]. It is discussed further in Chapter 3.

Carpenter et al. [49] present a number of vulnerabilities present in devices in the Advanced Metering Infrastructure together with an attack methodology starting from reconnaissance, initial and advanced analysis and also exploitation of these vulnerabilities. The exploitation scenarios are a good source of information for IDS development for devices in AMI.

Grochocki et al. [105] discuss attacks as well as how they can be detected best. They give an overview of the three main case studies presented in the literature related to attacks against the advanced metering infrastructure: Distributed Denial of Service (*DDoS*), attacks against the Data Collection Unit (*DCU*), attacks against customer's privacy and sending malicious remote disconnect commands ("remote shut-off") via the Data Collection Unit. The Data Collection Unit is an important component of AMI, and its purpose is to gather data from the Smart Meters and act as a relay between them and the Control Center. DDoS attacks starting from compromised smart meters and targeting the DCU can cause severe problems to the communication process in AMI. Figure 5.4 from [105] presents the attack tree developed by the authors for this specific attack case.

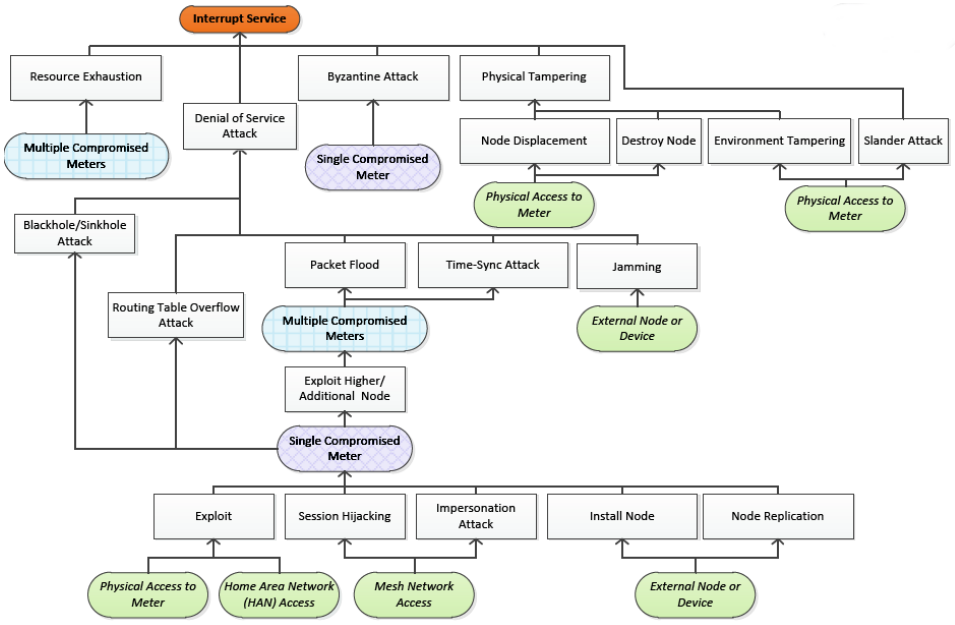

Figure 5.4: Denial of Service Attack Tree [105]

Sending malicious commands via the Data Collection Unit is another important attack scenario that can be performed in the Advanced Metering Infrastructure and detecting this kind of attacks should be one of the base requirements when developing an Intrusion Detection System.

Together with setting the scenario for the test cases, Grochocki et al. also present a table containing individual attack techniques and most important, a list of information required to detect them. This list can be used as an input for developing Intrusion Detection Systems for AMI. They also discuss the different places where Intrusion Detection System can be deployed in the Advanced Metering Infrastructure, together with advantages and disadvantages. As it is presented in [251], the nodes where the Intrusion Detection Systems are installed are spread all over the network.

The simplest and cheapest solution identified is to deploy a central IDS whose purpose is to monitor all traffic that flows to and from the advanced metering infrastructure. The disadvantage of this system is that it can not detect events that take place only inside the AMI network such as illegal firmware installs and Denial of Service attacks against the Data Collection Units.

Another solution is to embed the IDS directly in the smart meters. This solution has the advantage of offering a high coverage of the AMI network and increased accuracy based on this. The disadvantage is that an IDS may require additional computational resources which most of the current generation smart meters do not have yet.

Using dedicated IDS nodes is a better alternative to the one presented before because it employs devices that have the necessary computational resources to perform the role required by an IDS. The last solution presented by Zhang et al. [251] is that of designing a hybrid sensing infrastructure that would use both a central IDS and embedded IDS nodes or dedicated ones. The major advantage of this solution is the complete coverage of the advanced metering infrastructure by employing a diversity of IDS nodes, and this will also lead to better detection capability. An important practical aspect that needs to be solved is choosing the correct place where to deploy these dedicated sensors, in order to achieve the best coverage.

Raciti and Nadjm-Tehrani [187] also express the need of an embedded IDS solution for the advanced metering infrastructure. They start from the premises that correct storage and communication of smart metering data is a very important aspect from the security and privacy point of view. They present a trusted sensor meter platform that is responsible for storing and reporting the data but also stress the fact that an additional embedded system whose purpose is to detect anomalies is strongly needed. The authors developed a prototype for this system which was tested on four types of created attacks against the internal registers of the smart meter: data manipulation attack, recalibration attack (changing the values of some registers), reset attack (deleting the records regarding consumed energy) and sleep mode attack (the meter is put into sleep mode and the energy consumed is not registered). Their results show that after tuning the parameters for the clustering-based anomaly detection algorithm, the prototype achieves a high detection rate with a low value for false positives.

In a recent research paper, Mitchell and Chen [162] present a behavior-rule based intrusion detection system called BRIDS. Their system covers smart meters, data collection units (called *data aggregation points*) and also devices from the SCADA network (celled the *head-end*). They stress the fact that due to the reduced number of known attack signatures, a behavior-rule (anomaly) based Intrusion Detection System is preferred over a signature-based one. The authors present tables containing the behavior rules for each of the devices monitored, each behavior rule is then translated into different states, some of which are good or normal states while others can be attack states. The component states are then identified together with their correct ranges. The performance of their system is compared with the one of other anomaly-based Intrusion Detection Systems and the results show an improved performance for the BRIDS system.

5.5 Challenges facing intrusion detection research

During the previous sections, a number of obstacles were highlighted that must be overcome by an Intrusion Detection System for the smart grid. These obstacles are caused mainly by the differences between the Information and Communications Technology domain where the IDS was developed and used first and the smart grid domain which uses technology and knowledge both from the Electrical Engineering domain and the ICT domain.

Current research advocates the need for behavior-based intrusion detection systems for the smart grid due to the limited number of known attack signatures that a signature-based intrusion detection system could use; a consequence given the relatively young age of this domain.

Another impediment for the IDS is that the smart grid employs a large number of communication protocols in all of its sections, most of them proprietary protocols, or open standards but with vendor specific implementations. The correct behavior of all the devices present in the smart grid and of the communication between them must be correctly captured by the Intrusion Detection System, and this might be hard to accomplish by only one system. Multiple Intrusion Detection Systems, each with its own scope, may need to be deployed across the smart grid in order to provide full coverage.

The size of the smart grid network is another important aspect when developing an Intrusion Detection System. As mentioned in Section 5.3, the size of the SCADA network is relatively fixed and so is its network topology, but things are different for the advanced metering infrastructure. The AMI network is growing in order to accommodate new customer subscriptions and the network topology is not fixed, usually AMI communication is made through a meshed radio network where some smart meters also play the role of communication relays [228] as discussed in Chapter 4. The grid operator has mainly two options when choosing the correct solution for the advanced

metering infrastructure network. She can either build its own communication network, which depending on the size can be expensive, or lease an already existing one, such as a GPRS network from a mobile operator. Deploying an IDS in such a leased network may be difficult as the grid operator would need support from the network owner. Even if the network is owned by the grid operator, the use of encryption is necessary in order to achieve confidentiality and to protect customer's privacy. In this case, the IDS should have the knowledge of the encryption keys used in this communication, and the decryption process may require extra computational overhead on the node running the IDS. This can be extremely difficult to implement on embedded IDS systems, such as the ones mentioned in Section 5.4.2 and may require deployment of additional devices for dedicated IDS systems.

5.6 Summary

In this chapter we presented important research regarding intrusion detection systems tailored to the smart grid, differentiating between the two main parts of the smart grid: the SCADA systems and the advanced metering infrastructure. The IDS research related to the smart grid is still relatively limited. Its development commenced in the need to secure SCADA systems, and it is now expanding to also include IDS tailored to the AMI. Results so far show that the balance is towards using behavior-based IDS for both the SCADA systems and the AMI networks. One of the main reasons for this choice is the lack of known attack signatures for the domain. The multitude of devices and communication protocols used in the smart grid and the continuous expansion of the advanced metering infrastructure network are some of the main challenges that drive the research direction of intrusion detection systems for this domain.

6 Scalable data processing in smart grids

6.1 Introduction

This chapter focuses on data processing requirements and data intensive computing in the context of smart grids. In its current implementation, data processing solutions in smart grids are centralized and will not scale accordingly to the increasing amount of sources and data. To this end, distributed and parallel data processing are needed for better scalability while complying with given time requirements, be it for grid stability or for security monitoring purposes.

The twofold goal of data processing is to extract useful information while relying on a fast, on-line protocol that enables for such information to be leveraged (e.g., to spot and mitigate threats as soon as possible). As will be discussed, the data streaming processing paradigm (designed for fast analysis over streaming sequences of data) is appropriate in the context of smart grids because of its high capacity and low latency processing guarantees.

Data processing in the context of smart grid security applications has two distinct dimensions. On one hand, an online and scalable processing of data produced by devices deployed in the grid could provide reliable state and stability monitoring. On the other hand, such processing could also be leveraged by applications monitoring the traffic in order to detect and mitigate external threats. Apart from these security applications we also discuss two applications of particular importance: Special Protection Schemes (SPS) and Phasor Measurement Unit (PMU) analysis.

We first introduce the background of efficient computation, including parallel processing and the data streaming model, and discuss its suitability in the context of smart grids in Sections 6.2 and 6.3. The material will be somewhat known for readers coming from distributed or parallel computing, but serves as a quick introduction for readers with background in the security field.

Subsequently, we focus in Section 6.4 on how it can be used by applications monitoring and controlling the grid state and stability. We discuss data processing in the context of defense frameworks in Section 6.5. Further, in Section 6.6, we present existing machine learning techniques applied to smart grid monitoring. We conclude discussing how cloud infrastructures can be leveraged in the context of smart grids applications in Section 6.7.

6.2 Data streaming overview

Applications in the context of smart grids share commonalities with the scenarios that gave birth to the data streaming research field. This section introduces the basics of this processing paradigm and discusses its applicability to smart grids.

6.2.1 System model

Approximately from the year 2000, applications demanding continuous processing of streams of data in a real-time fashion started pushing the limits of traditional "store-then-process" approaches. These are solutions that rely exclusively on databases (DBs) to store and process information and whose processing costs, due to expensive write and read disk operations, become prohibitive [218]. Examples of such application areas include network traffic analysis, fraud detection applications in the context of mobile phone networks and analysis of financial markets tickets, among others. In these areas, huge amounts of data (in the order of hundred of thousand messages per second) are continuously injected into the system and should be processed in a real-time fashion in order to extract valuable information (e.g., to spot a network intrusion and block it). Further complicating the analysis, in some scenarios, such as sensor networks, data is assumed to have variable incoming rates and distributions due to the distributed and heterogeneous nature of the data sources.

In data streaming, incoming data is processed by means of *continuous queries* (or simply queries), which differ from their DB counterpart since they are not issued punctually but rather stand continuously and process information on the fly, updating their computation and generating results accordingly. A stream is defined as an unbounded sequence of tuples sharing the same schema, composed by attributes $(A_1, A_2, \ldots, A_n)$. Queries are usually defined as Directed Acyclic Graphs (DAGs) where vertices represent operators and edges specify how tuples flow between them [20]. Data streaming operators are divided into *stateless* and *stateful*. Stateless operators do not keep any state and process each tuple individually. Stateless operators include *Filter*, used to route or discard tuples, *Union*, used to merge multiple input streams into a single output stream, and *Map*, used to transform

tuples' schema. On the other hand, stateful operators perform operations on sequences of tuples. Stateful operators include the *Aggregate* operator, used to apply aggregation functions (such as *avg*,*max*,*min* or *count*) and the *Join* operator, used to match tuples from different streams. Due to the unbounded nature of streams, such operators perform their computations over subsets of subsequent tuples referred to as *windows*. Windows can be *time-based* (e.g., to compute an average value over the tuples received in the last 10 minutes) or *tuple-based* (e.g., to compute an average value given the last 20 received tuples). Windows are usually defined by parameters *size* and *advance*, which specify the extent of the window and the amount of information discarded each time the window slides forward. As an example, an average computed over a time-based window with size and advance of 60 and 10 seconds, respectively, will produce a result computed over the last minute every 10 seconds.

Figure 6.1 presents a sample query that could be used in the context of smart grids to spot customers whose daily average consumption double after two consecutive days. In the example, the input data is provided by smart meters and the tuples schema is composed by attributes `Meter ID`, `Timestamp` and `Reading`. The query is composed by three operators. The stateful aggregate operator A_1 is used to compute the average consumption on a per-meter, per-day basis. It relies on a time-based window whose size and advance are 1 day (windows whose advance is equal to the size parameter are usually referred to as *tumbling* windows). The results produced by A_1, composed by attributes `Meter ID`, `Day` and `Average Consumption` are then processed by the aggregate operator A_2, which computes the consumption increase for each pair of consecutive days on a per-meter basis. In this case, A_2 could rely on a tuple-based window of size 2 and advance 1 (recall that 1 tuple is produced for each day). Windows whose advance is lower than the size parameter are usually referred to as *sliding* windows. Finally, the Filter operator F_1 is used to forward tuples produced by A_2 (composed by attributed `Meter ID`, `Day 1`, `Day 2` and `Increase`) whose increase exceeds 100% (that is, tuples referring to customers whose consumption has doubled).

6.2.2 Stream processing engines evolution

Pioneer Stream Processing Engines (SPEs) [20, 41] were designed as centralized solutions, where all the operators of a given query are deployed and executed in the same node. Subsequently, centralized SPEs have evolved to distributed SPEs [19], allowing for operators belonging to the same query to be run at different nodes (*inter-operator parallelism*), and parallel-distributed SPEs [14, 16] , where single operators can be run in parallel by multiple nodes (*intra-operator* parallelism). Nowadays, elastic SPEs [106] (which are able to adjust resources according to the system load) are being studied

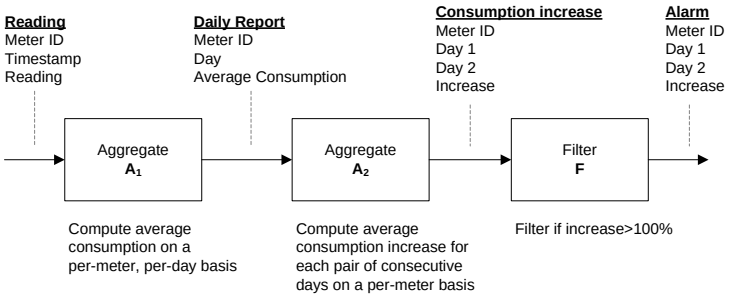

Figure 6.1: Sample query to spot smart meters whose daily average consumption doubles in two consecutive days

to take advantage of the cloud infrastructure. The idea is to reduce costs increasing or decreasing the number of processing nodes according to the system load. Figure 6.2 presents how the sample query in Figure 6.1 could be deployed and run by a centralized, a distributed, a parallel-distributed and an elastic SPE.

Smart grids share several aspects with the scenarios that motivated data streaming. The metering devices deployed in the infrastructure generate continuous streams of data. The rate at which data is generated can vary over time (e.g., due to different reading periods, depending on the time of the day or the facility where a meter is installed). Moreover, information generated by the devices deployed in the network should be analyzed in a real-time fashion by threat monitoring applications, but also for other use such as for real time pricing or network stability analysis. As we discuss in the following, smart grids applications can benefit from the improvements of each evolution step.

One of the pioneer data streaming solutions focused on sensor networks and investigated how to improve sensors databases by means of distributed information processing [41]. In this context, a sensor network is employed to measure physical signals such as heat, light or pressure and report it to a central server in charge of processing them. Early pre-processing of data at the sensors reduces the amount of information sent by each sensor to the central server in charge of monitoring the entire network, allowing thus for a larger scalability.

The above consideration also applies to smart grids, where smart meters would constitute the sensors and the centralized analysis is usually run at the utility server. As an example, a utility interested in the average consumption of the users over periods of one day could instruct the smart meters to process locally hourly produced readings and forward a single aggregated value in order to reduce bandwidth consumption.

The overall amount of data exchanged between each sensor and the central monitoring component (and therefore the bandwidth consumption) can

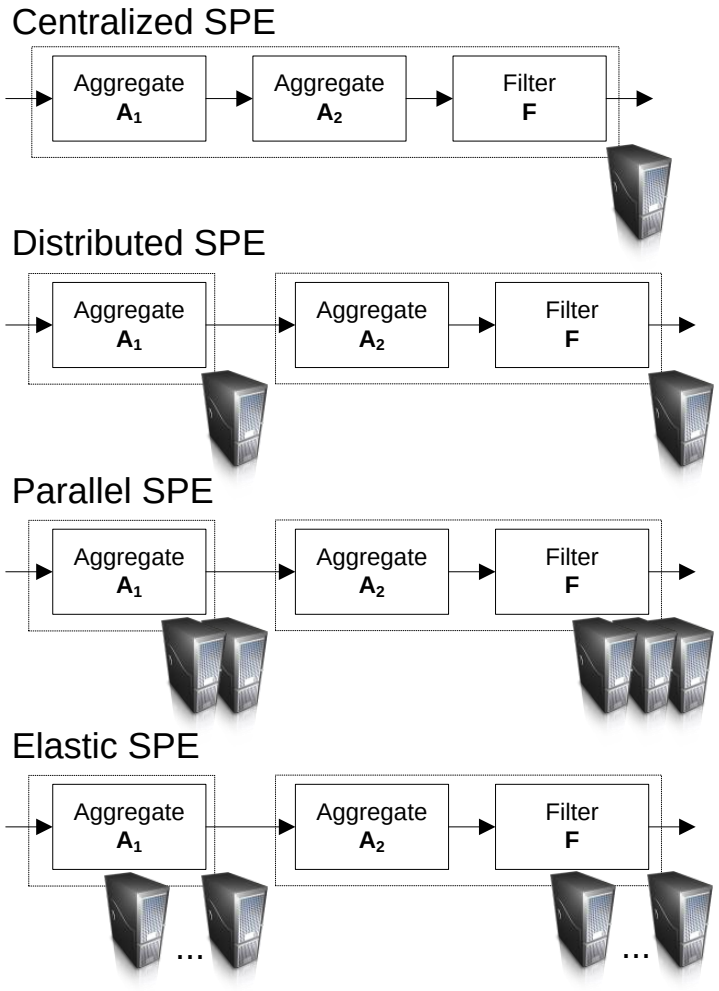

Figure 6.2: Evolution of SPEs

be further reduced by relying on *in-network aggregation* [154, 112, 171]. In-network aggregation focuses on processing tasks performed by hierarchical topologies where information flowing from multiple entities up to a central sink node can be aggregated while being forwarded. Figure 6.3 presents an example of a hierarchical topology on top of which in-network aggregation can be run. In the example, data is generated by sensor nodes S (smart meters) and forwarded to a centralized *Sink* node by *Cluster Head* (CH) nodes (data concentrator units in the smart grid); data aggregation could be performed by CHs to reduce the overall bandwidth consumption and reduce processing latency times. As an example, when looking for the highest reading over a given time, each CH could forward its local maximum value to the sink node. The presented topology resembles the one existing in smart

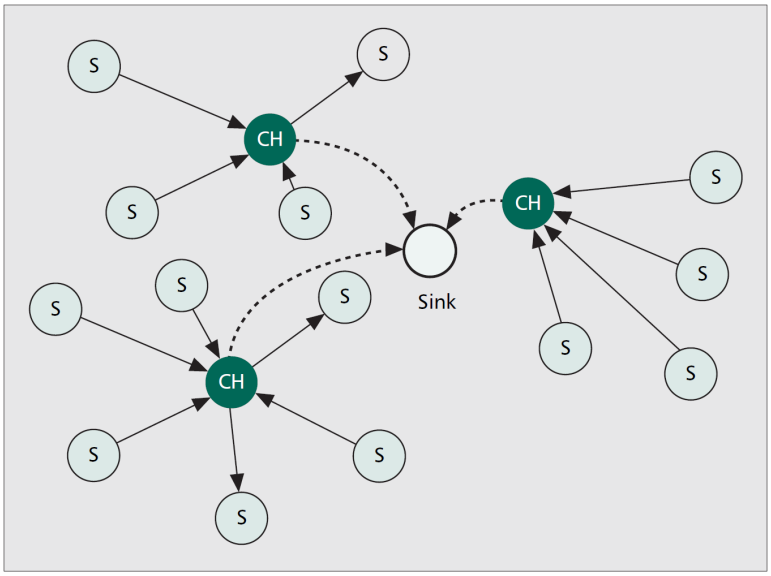

Figure 6.3: Example of in-network aggregation ([87]), where *CH* represent the cluster heads.

grids, where multiple sensor nodes (e.g., the smart meters) generate data that is forwarded to a central server by concentrator units that could perform data aggregation on the fly (see for example Figure 4.1). We refer the reader to [87] for a comprehensive survey about in-network aggregation in the context of sensor networks.

The need for high capacity processing (i.e., parallel-distributed data streaming processing) and the advantages of cloud infrastructures (such as its scalability and the economical benefits for companies shifting from traditional to smart grids) are also taken into account in the context of smart grids, as discussed in [204, 205]. We analyze such aspects in depth in Section 6.4.

6.3 Parallel data processing overview

The massive set of data that is generated by different aspects of the smart grid, and the need/possibility for more complex analysis, requires large amounts of computing power for real-time processing. In this section we give a short overview on parallel data processing and its role as a fundamental component to handle the challenges of big data and complex analysis in the smart grid.

6.3.1 Levels of parallelism

It has been, and still is, difficult to increase the performance of a single processor core. This has lead to multi-core processors becoming the new standard. Many-core co-processors, such as graphics processors or the Intel Xeon Phi, are also becoming more prevalent. The single, power expensive core, has been replaced with multiple low power cores, which brings more processing power at a lower energy cost. But this requires that we can take advantage of the potential parallelism in our algorithms, as a sequential program will be hurt by the lower single core performance. This parallelization can be provided at multiple levels and the best performance is often achieved when these levels can be combined and be made to scale to the capabilities of the available hardware [91, 79].

At a *high level*, some problems are of a so-called embarrassingly parallel type. That is, they can be decomposed into parts that can be worked on completely independently, by different processing units, without needing any kind of communication or synchronization between them. An example of this could be summing up the hourly energy consumption over a year for a set of costumers. Each processing unit could in such a scenario work independently on different customers' data with little need for communication.

Most problems are however not embarrassingly parallel, or there are time constraints, or the data is unbalanced. These types of problems need more fine-grained parallelism. At the *middle level* one need to decompose the problem into multiple small tasks. These tasks then need to be distributed to the different processing units to achieve load balance. Finally, there needs to be support for the communication between different tasks, which requires synchronization. With the introduction of graphics processors as general purpose computing units, and the Intel Xeon Phi, there is now also a large focus on *low-level* parallelism in the form of SIMD instructions and memory access coalescing.

6.3.2 Parallelization

The parallelization of an algorithm could be divided into three parts. We first need to decompose the problem into multiple tasks that can be executed concurrently, or decompose the input data into independent blocks that can be worked on concurrently. These tasks then need to be distributed to all available processing units in a manner such that we achieve an even load balance. The final part is to deal efficiently with the necessary communication between the different tasks. This communication is frequently performed via *concurrent data structures*, which need to provide efficient synchronization to not become a bottleneck.

6.3.3 Lock-free synchronization

The standard way of implementing concurrent data structures is often by the use of basic synchronization constructs such as locks and semaphores. Such blocking data structures that rely on mutual exclusion are often easier to design than their non-blocking counterpart, but a lot of time is spent in the actual synchronization, due to busy waiting and convoying. Busy waiting occurs when multiple processes repeatedly checks if, for example, a lock has been released or not, wasting bandwidth in the process. This lock contention can be very expensive. The convoying problem occurs when a process is preempted and is unable to release the lock quick enough. This causes other processes to have to wait longer than necessary, potentially slowing the whole program down.

Lock-free synchronization, on the other hand, allows access from several processes at the same time without using mutual exclusion. So since a process can't block another process they avoid convoys and lock contention. Such objects also offer higher fault-tolerance since one process can always continue, whereas in a blocking scenario, if the process holding the lock would crash, the data structure would be locked permanently for all other processes. A lock-free solution also eliminates the risk of deadlocks and have been shown to work well in practice [220, 225, 226]. They have been included in Intel's Threading Building Blocks Framework [123], the NOBLE library [220], the Java concurrency package [144], the parallel extensions to the Microsoft .NET Framework [159] and the PEPPHER synchronization library [33]. A recent study have also shown that lock-free data structures are more power efficient [120].

6.3.4 Emerging many-core platforms

Graphics processors have become a new target in the search for extra computational power. In normal use, graphics processors are often faced with embarrassingly parallel problems, such as performing matrix multiplications to transform 3D vertices into 2D space. This has caused them to target parallelism much earlier than generic processors, which have mostly dealt with sequential programs. Intel has recently released the Xeon Phi, which provides a similar many-core processor platform as the one provided by graphics processors.

An important feature of many-core processors is the possibility to concurrently execute the same instruction on multiple data, known as SIMD computing. Most parallel algorithms and concurrent data structures are, as their name implies, designed to support multiple concurrent operations, but when used on a many-core processor they also need to support concurrent instructions within an operation. This is not straightforward, as most have been designed for scalar processors. Considering that SIMD instructions

play an instrumental role in the parallel performance offered by many-core processors, it is necessary that data structures and algorithms are redesigned before being deployed on many-core processors.

The combination of highly parallel many-core processors, for easy to parallelize algorithms, and conventional multi-core processors with large caches, for mostly sequential code, will become the future standard. Together with data streaming it provides a very powerful platform for real-time processing of large amounts of data.

6.4 Data processing in smart grids state monitoring

In this section, we discuss how smart grids applications can leverage the latest advances in streaming processing engines (SPEs) and parallel multi-core computation. State-of-the-art SPEs (such as Storm [14], Yahoo S4 [16] or StreamBase [15]) enable for scalable processing of hundred of thousands messages per second and can cope with low latency processing requirements of applications such as Special Protection Schemes (SPS) [137, 100] or synchrophasors data analysis [115, 72, 68, 175], which are discussed in the following sections.

New efficient algorithms for many-core processors have shown great potential in speeding up analysis and data mining. As an example, two commonly used machine learning algorithms, deep belief networks and sparse coding, have been adapted to take advantage of the parallelism on graphics processors and show a 15-fold speedup [188]. Support Vector Machines [99], have been implemented on graphics processors and show a 100-fold performance improvement compared to standard CPU implementations. Sorting, an important component of numerous algorithms and programs, for example within machine learning, analysis, finance and database applications, has been implemented for graphics processors and vastly outperform sequential CPU versions [53]. It can be used to aggregate data that needs to be processed in a certain order, so called order-sensitive operations.

In its current implementation, data processing solutions in smart grids are centralized and will not scale accordingly to the increasing amount of sources and data. To this end, distributed and parallel data processing will enable for higher scalability while complying with given time requirements. As presented in [137], power grids are currently deployed with a centralized information infrastructure, usually referred to as Energy Management System (EMS). Such a centralized component is in charge of collecting the data from the grid and analyze it in order to monitor its stability. All the substations deployed in the grid are directly connected to the EMS by Remote Terminal Units (RTU), interfaces through which information and reports can be collected from the different components running at each substation. The communication between the EMS and the RTUs usually happens in a round

robin fashion, where the former connects to the latter and processes the data once all the RTUs have been contacted.

This processing paradigm does not scale accordingly to 1) the number of substations from which data is collected, 2) the number of monitoring devices deployed at each substation and 3) the amount of data generated by each of the monitoring devices. To this end, distributed and parallel data processing (provided by modern SPEs) could enable for higher scalability. Data could be preprocessed by each substation independently (distributed processing) reducing the amount of data retrieved by the EMS, while data from different substations could be processed at the same time (parallel processing) if it does not need to be correlated. Parallel data processing could also be improved further by using lock-free data structures for communication and data pipelines in the stream processing engines [51].

Three applications are of particular importance and show the benefits the parallel and data streaming processing paradigm could enable: Special Protection Schemes (SPS), Phasor Measurement Unit (PMU) analysis and improved optimization control.

6.4.1 Special protection schemes

SPS are hard-wired and localized monitoring protocols used by substations to provide fast reaction to predefined critical events (e.g., to react to a disturbance or to restore the power grid after a blackout). As defined by the IEEE 1646 standard, SPS schemes usually impose processing latency thresholds in the range of 8-12 ms [137] (this motivates why such solutions are hard wired and local to substations). As mentioned in [100], the local analysis performed by SPS could also help the higher-level analysis performed by the EMS. In order to enable this, an online and fast processing of such data must be guaranteed to comply with the given latency bounds. In this context, a parallel-distributed SPE could be employed to process in parallel (and therefore minimize the overall latency) the data collected from distinct substations.

6.4.2 Phasor measurement units

A second application that could benefit from scalable data streaming and parallel processing is PMU data analysis, as discussed in [100]. PMU devices allow for real time monitoring of phasor quantities such as voltages, currents and frequency. These values are usually used to study the stability of the network. PMUs have been used for decades, since their log has been shown to be useful when running postmortem analysis after incidents such as blackouts and to discover the originating cause. Nowadays, the increased communication capacity proper of smart grids could be leveraged to process such data in an online fashion. This analysis could lead

to early evidence of systems becoming unstable (e.g., continuously growing phase angles could indicate an increasing probability of failure with 20 or 30 minutes advance). As discussed in [100], PMUs produce 50 to 60 samples per second, where each measurement is timestamped (using GPS synchronization). In the last five years, thousands of new PMUs have been deployed. Such an increasing number of devices (together with the high rate at which measurements are generated) might lead to a scalability problem. Processing of PMU measurements should be performed in time frames of seconds or ten of seconds and usually involves expensive computations such as Fourier transformations and large matrix operations. Similarly to SPS, PMU measurements could be processed in parallel in order to achieve a greater scalability. Fourier transformations and matrix operations are also especially suitable to be parallelized. Correctly applied for machine learning they could provide an operator with increased situational awareness of events in the smart grid.

Some related work exists where PMU data analysis is studied by means of machine learning techniques using data streaming and multi-core algorithms. We discuss such techniques in Section 6.6.

6.4.3 Optimization control

Renewable energy from sources such as wind, solar and water play an important role in meeting the world's growing power demands. It is however hard to predict the actual power production at a given time for wind and solar plants. This requires expensive backup systems that can take over at times when the production is low. New solutions are becoming possible with the introduction of the smart grid, such as, for example, helping the consumer to adapt the power consumption to times when there is much renewable energy.

To achieve a more intelligent distribution of energy, it is required that the power flows in the grid becomes more dynamic and that the uncertainty of the renewable energy sources is taken into account. Currently power flows are calculated and updated in a static manner. These are computationally heavy operations, so to take the step to real-time dynamic calculation requires that new algorithms and methods be developed, to utilize efficient parallel computation on multi-core processors. The better the parallel design, the more complex the analysis can be while still being performed in real-time, resulting in better usage of the energy production capabilities [148]. The design work can be greatly simplified by taking advantage of existing system components that can guarantee efficiency at the run-time system level [98], data structures [220, 52] and basic algorithms such as sorting [53].

6.5 Data processing in smart grids defense frameworks

In this section, we discuss threat detection and mitigation frameworks in the context of smart grids. In such frameworks, effective and online processing of data is crucial to react to threats immediately after they have been detected (that is, to enable online mitigation of threats). We first present the characteristics of the smart grid infrastructure that could motivate an attack, we then continue by focusing on specific threats that involve the communication between smart grid devices and discuss the requirements of a defense frameworks. We conclude presenting existing related work about threats defense in smart grids.

6.5.1 Data stream processing for intrusion detection systems

As presented in Chapter 5, intrusion detection systems (IDSs) are used to spot (and possibly mitigate) threats while monitoring a system (e.g., in network traffic analysis to detect DoS or DDoS attacks). In order to describe a general defense framework that could identify previously unseen attack signatures, we focus on anomaly-based IDSs in this section. As stated previously, anomaly-based IDS are usually defined by two main components. A modeling component analyzes the data in order to build and maintain the profiles of normal and expected traffic while a monitoring component compares such expected patterns with the ones currently observed in the data in order to spot possible threats.

Existing IDSs will face new challenges when applied to smart grids. More specifically, as discussed in Chapter 5 and [141], the characteristics of smart grids that must be addressed by an IDS will include: 1) support for legacy protocols (without affecting their real time performance), 2) scalability, 3) support for legacy hardware proper of smart grids, which usually has reduced processing power, 4) compliance to existing standards, 5) adaptivity to the evolving topological model, 6) unaffected deterministic processing of data by SCADA networks and 7) reliable discovering of threats. An IDS that relies on *parallel and distributed data stream processing* would be an appropriate solution to cope with requirements 1) and 2). As discussed in [35], a requirement for an IDS running in the smart grid is to embrace the heterogeneous communication technologies forming the network. Such network is presented in Figure 6.4.

The *Wide Area Network* (WAN) forms the communication link between the local utility network and data concentrators (components in charge of gathering and forwarding smart meter data) and usually relies on long-range high-bandwidth communication technologies. *Neighborhood Area Networks* (NANs) connect concentrators and smart meters, which in turn inter-

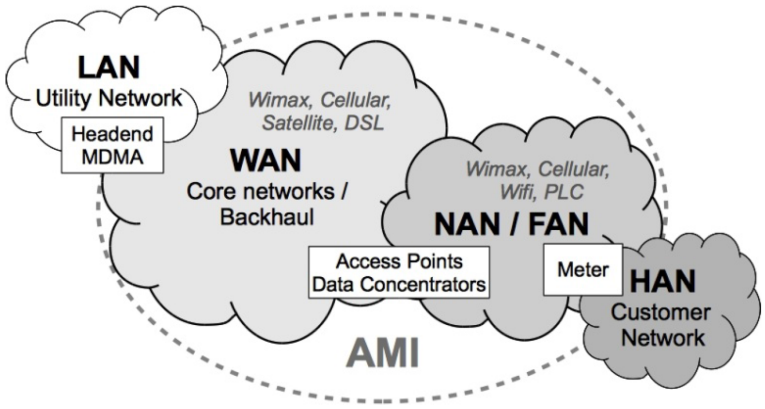

Figure 6.4: Overview of smart grid network from [35]

face with *Home Area Networks* (HANs). *Field Area Networks* (FANs) might be used by the utility workforce to locally connect to equipment. As discussed in [35], the first requirement for a detection framework deployed in such a network is to be distributed. At each different network, specific messages can be shared by devices that are not forwarded to other networks. To this extent, centralized solutions could not access the overall information and would therefore be blind to threats internal to other networks. At the same time, the processing of the data from the different networks should be parallel as far as the data do not need to be correlated (e.g., validation of energy consumption reading of distinct meters could be carried out in parallel at the same time). State-of-the-art SPEs providing parallel and distributed processing would comply with these requirements.

To the best of our knowledge, IDSs have been discussed as possible defense frameworks in the context of smart grids but not many concrete solutions have yet been proposed or deployed. One of the existing solutions that discussed the use of IDSs in this context is presented in [83]. In their work, the authors discuss the need for such a framework and present how already existing solutions seem promising in this new context. In their work, the authors rely on the data streaming processing paradigm in order to process data. In order to provide online processing of data, the authors focus on evolving stream mining techniques.

In their protocol, the authors also discuss the need for different modular IDSs to be placed at different positions inside the network. Different IDS are deployed for individual smart meters, data concentrators and the head-end central system. The rationale for these different deployments is the different nature of the data generated or collected by each of these devices. An overview of the characteristics of the data observed at each device is presented in Figure 6.5

Table 4. Performance comparison for classifiers

Classifier vs. Performance measures	OzaBag Adwin	Leveraging Bag	LimAtt Classifier	Single Classifier Drift
Training Data Set				
Accuracy(%)	98.61	99.41	99.49	99.16
Kappa Statistic(%)	97.57	98.97	99.12	98.54
Test Data Set				
Accuracy(%)	96.05	95.65	96.58	93.97
Kappa Statistic(%)	94.21	93.60	95.01	91.1
FPR(%)	2.15	1.6	2.48	2.49
FNR(%)	5.55	6.85	3.39	8.23

Figure 6.5: Data observed by IDS deployed at meters, concentrators and the central system [83]

As mentioned, the authors do not define a new protocol to mine the data in order to maintain profiles and spot threats. In the evaluation, the evolving data stream mining classifiers of the MOA (Massive Online Analysis) framework [12] are evaluated using a modified version of the KDD Cup 1999 data set [222]. From the original set of 16 classifiers, a subset of four classifiers is selected. As shown in Figure 6.6, these four classifiers provide similar results in terms of the precision with which normal and threat data is recognized while processing it.

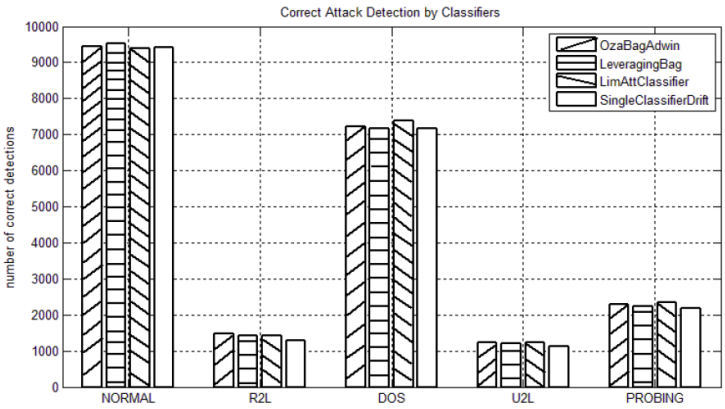

Figure 6.6: MOA classifiers precision comparison [83]

6.6 Machine learning techniques in smart grids data processing

In this section, we discuss existing work related to machine learning techniques applied to data streaming in smart grids. We focus on two scenarios: load consumption forecast and synchrophasor data analysis. In both scenarios, the existing work discusses the need for scalable and online processing of data. As discussed in Section 2.1, load forecast applications might be actively involved in a DDoS attack. In such attacks, false consumption data could be generated by each sensor in order to produce a wrong estimate of the needed energy, which in turn might lead to outages during demand peaks. On the other hand, a reliable forecast mechanism could be employed in a defense framework in charge of detecting or mitigating possible attacks. That is, reliable predictions could be used to spot suspicious deviations from expected energy consumptions.

6.6.1 Energy consumption load forecast

Electricity load forecast has been researched extensively in the past, a survey of the existing techniques is presented in [23]. As discussed in [73], load forecast can be categorized in different research areas depending on the time scale of the predictions. Traditionally, research has focused on short-term demand forecast (predicting the load consumption within time frames of hours to weeks). Medium-term (predicting the load consumption within time frames of weeks to months) and long-term (predicting the load consumption within time frames of months to years) forecasts have also been studied, although less solutions have been proposed.

During the last years, different classification and learning methods have been studied, with an increasing focus on the need for parallel and on-line processing. Results show that machine learning algorithms that can utilize many-core processors [188, 99] and lock-free synchronization [148] can achieve more than a ten-fold speedup compared to sequential approaches. We discuss in this section two existing techniques for load forecast that address this need for parallel and streaming processing.

Forecast techniques could be employed in the context of anomaly based IDSs. Given a reliable load consumption forecasting protocol, the live consumption of each customer could be compared with the expected in order to spot suspicious deviations. Such deviations might be indicators of threats depending on their scale. As an example, a small number of customers showing a suspicious change in their consumption pattern might be a sign of smart meters tampering in order to reduce bill costs. On the other hand, a considerable number of customers showing a suspicious load consump-

tion pattern might be a sign of a DDoS attack targeting the consumption estimation.

The first technique we take into account, discussed in [138], defines a protocol where load consumption is predicted relying on a regression technique. The idea is to model load consumption as a random variable and to study its conditional probability given some consumption observed in the past (e.g., to estimate the load consumption of one day given the load consumption of the previous one). The presented technique does not rely on the data streaming processing paradigm; nevertheless, as pointed out by the authors, parallel processing is needed in order to provide results in a timely fashion. In order to forecast the load consumption, the presented protocol relies on local kernel regression, based on the Nadaraya-Watson estimator. This kernel regression technique studies the conditional expectation $P(Y|X)$ of two random variables X and Y, modeling Y by means of a function $m(X)$. This function is defined by a *bandwidth* parameter that specifies the range of X values taken into account in order to model Y. Choosing the right value for the *bandwidth* parameter is challenging since small bandwidth values usually lead to over-fitting (that is, the forecast model is reliable only when observing values close to the ones used to train it) while big values might generalize too much. In their work, the authors proposed an evolutionary based approach to find the best modeling function. In evolutionary based approaches, populations of candidate solutions are generated randomly and, by means of reproduction, mutation, recombination and selection operations (as in biological evolution), populations evolve over time. At each evolution step, the best individuals are chosen (based on a given fitness function) and used to generate the following population. The population evolution stops after a given number of when a given quality for the individuals is met (we refer the reader to [110] for more details about the evolutionary approach). In their work, the authors discuss possible parallelization strategies in order to make the proposed technique applicable to large-scale smart grids. Two possible parallelization are defined. On the one hand, the kernel regression computations could be parallelized (that is, by assigning kernel computations functions to multiple machines). On the other hand, the overall forecast process could be parallelized distributing the computations performed by the evolutionary technique (generation and recombination of populations). In the evaluation, focusing on mid-term load forecasts for a time frame of one week, the proposed technique achieves lower error rates than its fixed counterpart where the bandwidth is chosen statically. The protocol has been evaluated with real power consumption data over three succeeding Sundays in January and February 2010. As discussed in the evaluation, the average absolute deviation from the training set (referred to as e_{TR}) and the test set (referred to as e_{TE}) decrease when employing the evolutionary technique with respect to a bandwidth parameter chosen statically (approximately from 80% to 60%). In the evaluation,

the authors also compare the proposed technique with other learning techniques such as Least Median Square Regression (LMS) or back-propagation neural networks, obtaining results comparable for the former and substantially better than the latter. Figure 6.7 presents the evaluation results, where rows A, B and C refer to the different days in the data set and columns present the kernel regression with fixed bandwidth, the evolutionary kernel regression, the LMS and the back-propagation neural networks results, respectively.

	$K=1$, $h=5.0$		$K=5$, LOO-CV				LMS		NN	
	e_{TR}	e_{TE}	best e_{TR}	med e_{TR}	best e_{TE}	med e_{TE}	e_{TR}	e_{TE}	e_{TR}	e_{TE}
A	93.89	86.88	66.91	67.39	63.88	64.44	74.08	64.78	134.67	124.51
B	91.06	118.66	58.06	60.79	104.33	104.72	57.79	105.58	104.64	154.01
C	97.44	86.00	72.43	79.58	56.04	64.00	71.95	58.79	142.95	159.31

Figure 6.7: Evaluation of kernel regression with evolutionary approach [138]

In order to address the time requirements of the processing protocol and enable for online forecast, the authors in [191, 94] propose a different hybrid technique composed by both an online and an offline component. The online phase is in charge of processing streams of data and reduce the amount of information forwarded to the offline components, in charge of predicting the load consumption. The technique is able to forecast consumption at different time scales of hours, days and weeks. Clustering is used to group similar streams into clusters (where similarity is based on the consumption pattern observed over a time window). Once aggregated, data is processed by a neural network (ANN) in order to forecast the load consumption for the given time frames [113]. The general idea is to find customers that behave in a similar way and for which a general prediction is close enough to the actual one.

One of the challenges in this approach is how to find the right number of clusters into which the data is partitioned. If given a priori, such a number could be too high, leading to an unnecessary number of clusters or too small, leading to an inappropriate number of sub-clusters that does not represent the real groups appearing in the data. Clustering protocols that do not require an a-priori number of clusters, referred to as variable clustering protocols, have been studied to overcome this limitation. In their work, the authors propose ODAC [192], a top-down hierarchical variable clustering protocol.

Clusters are maintained (and expanded/aggregated) by an hierarchical structure, as shown in Figure 6.8. All the streams are initially grouped together. The distance between a pair a, b of streams belonging to the same cluster is defined as their *dissimilarity*, expressed by the *Rooted Normalized*

One-Minus-Correlation

$$rnomc(a,b) = \sqrt{\frac{1 - corr(a,b)}{2}}$$

where $corr(a,b)$ is Pearson's correlation coefficient. This initial group is split into subgroups depending on its diameter, defined as the maximum distance between any pair of streams. Two main actions are defined to increase and decrease the overall number of clusters. The *expansion* operation is applied when the diameter of the cluster goes below a given threshold. On the other hand, the *aggregation* operation is used to merge two subclusters into a single one. Subclusters are merged when the diameter of a child subcluster exceeds the one of the parent subcluster, indicating changes in the data have occurred and the old split decision is no longer valid.

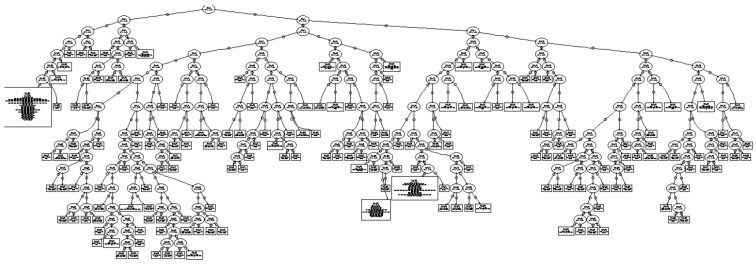

Figure 6.8: ODAC Hierarchical Clustering [191]

For each cluster, an ANN is trained in order to predict the load consumption. As discussed by the authors, at each time t, the ANN predicts the load at time $t + k$ while it updates its model back-propagating the predictions made for the moment $t - k$ (that is, past predictions are used to tune the model on the fly). For the evaluation, the authors rely on a network of 2500 sensors and a data set covering a period of three years, where consumption readings happen on a hourly basis. From the evaluation, the overall prediction precision is good, having a median prediction error of approximately 4% for hourly predictions which degrades to 6% for weekly and yearly predictions.

6.6.2 Synchrophasor data analysis

Similarly to load forecasting applications, machine learning algorithms have been used to study the data produced by PMUs.

In [68] the authors discuss an online dimension reduction of PMU data to enable real time analysis of the grid status. In the paper, the Principal Component Analysis (PCA) technique is used to reduce the synchrophasor data. The PCA technique is used to reduce a set of a given (usually high)

number of variables to a smaller set containing the variables showing the higher variance. The protocol is defined as unsupervised and single pass to allow for online streaming processing of data. The authors suggest that, once reduced to its principal components, synchrophasor data can be fed in a real time fashion to machine learning algorithms in charge of maintaining models for the grid status. A candidate machine learning could be classification type Decision Trees (DT), as discussed in [175]. DT are used in data mining as a model to predict the value of a given target variable based on previous observations of a set of input variables (used to train the model). The information is maintained in a tree structure where each interior node defines a test on one of the input variables while leaf nodes specify the outcoming target value. In the paper, the authors rely on DTs to study the stability of the grid. DTs are trained using angle phase attributes from the different PMUs, since a grid which is becoming unstable is characterized by widening voltage angles. In order to train the DTs, data collected over the past (during which outages happen) is used. Data is labeled offline and each label represents the margin to voltage collapse given the current phase. Moreover, each node is shaded proportionally with voltage secure (light shade) and insecure (dark shade) entries. Figure 6.9 presents a sample DT trained with the given data. The number of training entries falling into each node is specified on top of each node.

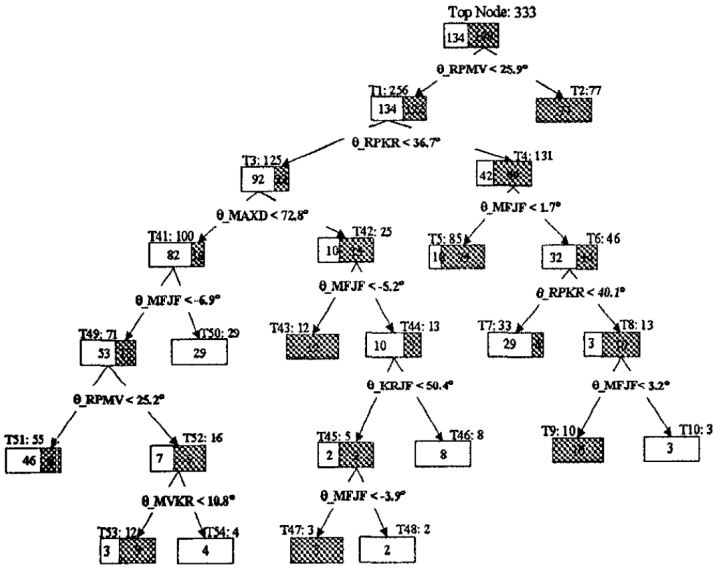

Figure 6.9: Sample Decision Tree [175]

An interesting connection point between research in the context of PMU data analysis based on DTs and data streaming is the work proposed in [78, 118]. In [78], the authors discuss how to build DTs in an online man-

ner while processing streaming data. In its basic learning form, each entry of the training data must be analyzed several times in order to build the DT that best predicts new incoming measurements. Nevertheless, such multi-pass analysis does not fit with the data streaming processing model, where single-pass analysis is enforced to minimize processing latency. In the paper, the authors discuss a single-pass protocol to learn DTs while processing data streams. Thanks to the Hoeffding bound theorem, the authors demonstrate how the learned DT can be as good as needed (with respect to the DT learned by a multi-pass protocol). The initial results assume a stationary data distribution, which might not always be the case in data streaming applications. This assumption is relaxed in their following work [118], where the authors discuss how to build a DT when data evolve over time.

6.7 Leveraging cloud infrastructures in smart grids

Cloud infrastructures provide several advantages for smart grid applications [206, 252, 205, 194, 204]. One of the advantages is the economical benefit of cloud models such as Infrastructure as a Service (IaaS) or Platform as a Service (PaaS) models. With the latter, companies shifting from traditional grids to smart grids can reduce investments, avoiding costs such as installation or hardware and software maintenance. Moreover, the elastic nature of cloud infrastructures allows for initial setups that can then be increased accordingly to the processing requirements. To this end, elastic infrastructures have shown to scale well and are an appropriate solution to process multi-million sources, as in the context of smart grids.

Research to leverage cloud infrastructures in the context of stream processing has been addressed in the last years and solutions that adapt to such infrastructures exist.

System provisioning is a valid example of how SPEs together with cloud infrastructures can reduce processing costs. When deploying an infrastructure in charge of processing data streams, system over-provisioning or under-provisioning is usually adopted depending on the data rate fluctuations and on quality of service standards defined by the utility. When over-provisioned, a system relies on the resources it needs to process the data at this maximum rate; nevertheless, the system is not fully utilized and costs are not optimized each time data does not come at its maximum rate. On the other hand, under provisioned systems are usually tuned to process data coming at its average rate, while they may violate the given time boundaries during load peaks. Elastic SPEs can be used to avoid both under and over provisioning, adjusting the used resources depending on the current system load. Existing elastic stream processing engines provide elasticity in terms of queries (where new nodes can be provisioned or decommissioned when new queries are deployed in the systems [152]), or in term of processing

capacity (where nodes or threads can be provisioned or decommissioned to running queries in order for them to cope with the given processing time boundaries [198, 106])

In [205], the authors discuss the use of cloud infrastructure to enable online portals to share real time energy usage and power pricing information between producer utilities and consumers. Such data could also be aggregated and shared with third-party applications (examples of these third-party applications include monitoring applications such as Google PowerMeter [9] or Microsoft Hohm [11]). At the same time, other applications could provide information used by utility forecast applications, such as meteorological parameters like cloud coverage or solar radiation intensity maps (we refer the reader to [252] for a detailed description of such applications). In this sense, the elastic features of the cloud infrastructure could help in adjusting the setup size accordingly to the number of entities accessing the information. As stated by the authors, an important concern in the context is privacy preservation. An interesting taxonomy of security and privacy issues is presented in [206] (as discussed in Chapter 2 and 3, stealing of personal information is one of the threat motivations). In their work, the authors take into account the main concerns of the three entities relying on the cloud infrastructure: the utilities producing the energy, the consumers and third-party applications, as summarized in Figure 6.10.

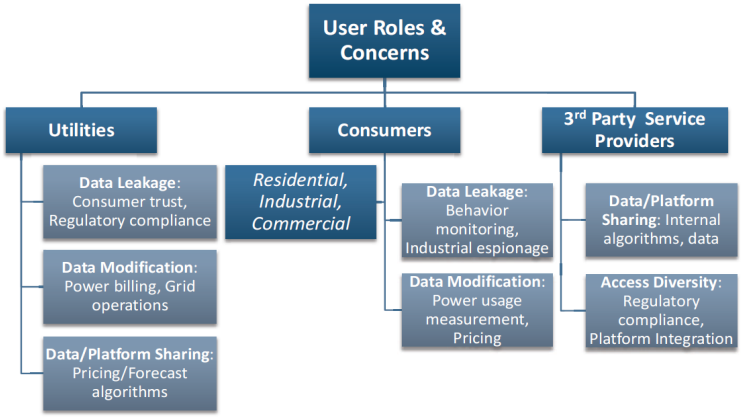

Figure 6.10: Security and privacy concerns in the context of cloud applications [206]

Figure 6.11 presents how the information could be accessed in a cloud-centric smart grid. The framework defines two separate APIs used to push to and get data from the cloud environment. The PUT interface could be used by the different sensors deployed in the network (referred to as S_i in the figure) while the GET interface is used by the different entities such

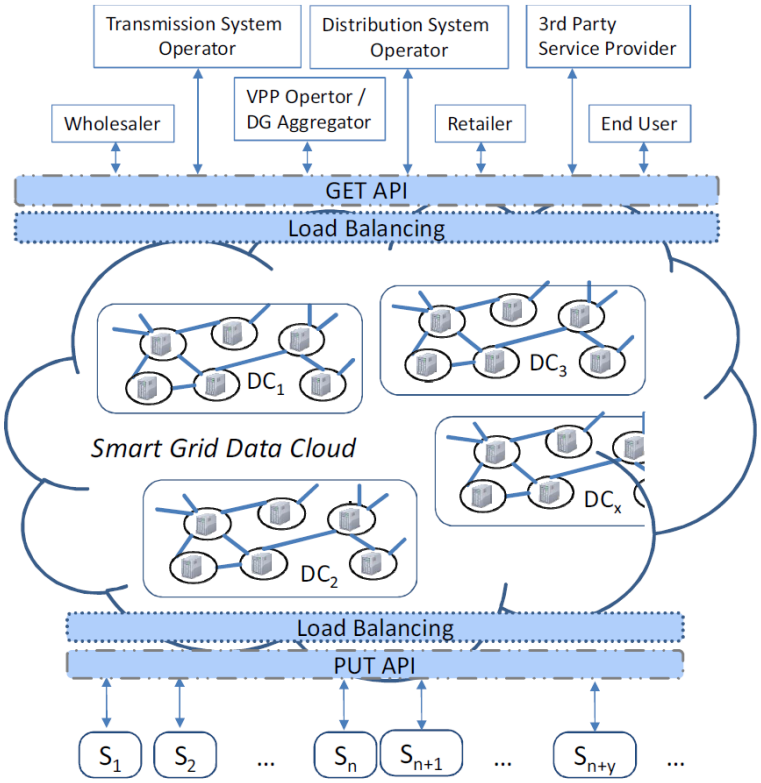

Figure 6.11: Leverage cloud infrastructure in the context of smart grids data management [194]

as energy wholesalers or retailers, end user accessing consumption data or transmission system operators.

With respect to data streaming applications, such a framework should define a transparent integration with a streaming data model, as discussed in [252]. Existing cloud infrastructures define interfaces where data such as files or structured collections can be shared but no transparent integration exists for streams of data. In their work, the authors discuss a work-flow framework that includes data streams as a possible input (unbounded series of binary data), taking into account techniques to ensure high performance and to enhance reliability.

It should be noted that adaptivity in the context of smart grids could not only refer to the need for the processing to adapt to the varying rates at which data is generated, but also for the rates at which events are generated to adapt depending on the system state. This concept is introduced by the authors of [204]. In their work, the authors discuss a load forecast protocol where smart meters can adjust the rate at which events are generated depending on the overall energy consumption. As discussed by the

authors, given a threshold for the sustainable energy consumption, forecast methods would require data at a finer granularity as consumption gets close to its limit, while a coarse-grained analysis could be used when energy consumption is considerably lower than the given threshold. To this end, the granularity at which measurements are produced by sensors should vary over time depending on the overall consumption.

The rate control logic defined by the authors (used to specify the rate at which events are created by smart meters) is defined as

$$\text{StreamRate} \propto \frac{1}{\text{AvailableCapacity} - \text{CurrentUsage}}$$

That is, the closer the user's usage to the overall available capacity, the higher the rate at which reports are generated by smart meters. In the protocol evaluation, three different queries are run to monitor the overall energy consumption. The first query monitors whether the power consumption of a user exceeds a given threshold. A second query is defined to check whether the power consumption of a user increases more than 25% from the previous measurement. Finally, a third query checks if the overall consumption over a window of 15 minutes exceeds 90% of the available bandwidth. Thanks to the rate control logic, the authors show how these queries meet the defined accuracy requirements while consuming 50% lesser bandwidth. In their implementation, the authors rely on the Eucalyptus private cloud infrastructure [8] and the IBM InfoSphere SPE [10].

6.8 Summary

In this chapter, we discussed the need for scalable processing of data in the context of smart grids. As presented, the fast online processing provided by the data streaming protocol is an appropriate candidate to process the data produced by devices deployed in smart grids and could increase security by means of state monitoring and threat defense. Moreover, several machine learning techniques (usually adopted to model complex systems) that have been studied in the context of data streaming could be applied in this new context. Efficient multi-core computation is a necessity in this context with the large volumes of data. The chapter points out example uses and application areas, including monitoring and optimization.

7

Trusted computing in smart meter environments: the TOISE project

TOISE[1] is an European funded project whose objective is to define, develop and validate trusted hardware and firmware mechanisms, which must be applicable both to lightweight embedded microelectronic devices and as security anchors within the related embedded computing platforms. This is a necessary enabling block for the secure and tamper-resistant solutions needed by applications such as smart grids, smart low-energy home appliances, environmental or infrastructural sensor networks, and more generally for managing trusted components and granting security over wired and wireless networks.

In the smart grid area, TOISE focuses on two applicative scenarios: smart metering and home multimedia gateways. The smart metering scenario, in particular, focuses on the basics: automatic measurement of electric power consumption, notification to the power provider, automated billing, automated load balancing in the power grid (e.g. disconnection and reconnection of the consumer loads as needed, management of emergencies and faults). While the basic scenario is focused on power metering, non–electric resources, such as gas, water, heating, are also taken into account.

Multimedia streaming is a specialization of the smart metering scenario. It consists of locally storing and distributing copyrighted multimedia data by means of the power grid, and possibly of downloading such data by means of the power grid as an alternative to using a traditional data network. Billing is obviously also an issue for multimedia streaming.

The underlying architecture is pretty simple. A power line connection distributes electric power and data together, using the existing electric power grid. It is characterized by having a higher reliability than wireless communication technologies. The smart meter is connected, again via a power

[1] http://www.toise.eu/

line, to a power provider or distributor. A wireless connection, e.g., Zigbee, may be used for the local communications among nearby meters (e.g., those displaced in a building) that measure non-electric resources such as gas or water, for which the electric power meter works as a gateway to the power line connection.

The security requirements in this scenario [31] are related to how data are used by the various actors on a communication line, and how they are maintained in a database respecting the CIA paradigm (as discussed in Section 5.2). Security threats are mainly located in the peripheral circuits and the buildings, as the power provider and the multimedia provider are trusted elements in the scenario. A few threats may also exist at the provider level, namely in the interoperability of multi-provider platforms, in the authentication, etc. The provider must handle secret keys to be shared with all the other actors. The provider can also act as a Certification Authority (CA).

Most threats are on the consumer side. In fact the consumer is subject to attacks on smart meters. Typically such attacks are of the following types: physical tampering, side channel analysis (e.g., power attacks, fault injection-based attacks, etc), network attacks against the flow of data towards the upper levels, user masquerading, hijacking, etc. Also confidentiality and privacy attacks occur here. Authentication must be provided for the meters that have commands such as enable/disable the meter itself, in order to prevent the unauthorized detachment of a single consumer or even of a whole neighborhood. The data collected from consumers should be aggregated for statistical purpose only, with no access to individual records in order to prevent an unauthorized consumer profiling.

All the attacks of the smart metering scenario, may occur in the multimedia streaming scenario as well. Moreover, there are attacks on the confidentiality and integrity of large volumes of application data, in the case such data are copyrighted. A typical multimedia home gateway will also be able to run applications, and will have a more refined operating system than that of a smart meter. Thus the multimedia gateway must also have a secure boot procedure.

For these two applications, TOISE is developing and testing through demonstrators, two embedded systems that will perform the smart metering and multimedia home gateway services, equipped with the necessary security functions to face the threats described above. These functions are mainly based on cryptography, both at the hardware and software level, and on the countermeasures for facing security attacks, of crypto-analytical and side-channel type. More precisely, two System-On-Chips (SOCs) are envisioned for the smart grids: a metering SOC with a small processor and a few embedded security IPs (both in HW and in SW) based on cryptography, to protect metering information and communication; and a home gateway SOC, based on the HomePlug AV powerline communication standard (versions 1 and 2), with a more powerful processor with embedded cryp-

tographic IPs and a secure boot procedure, to run a secure Linux operating system and the related applications. An overview of the TOISE approach can be seen in Figure 7.1.

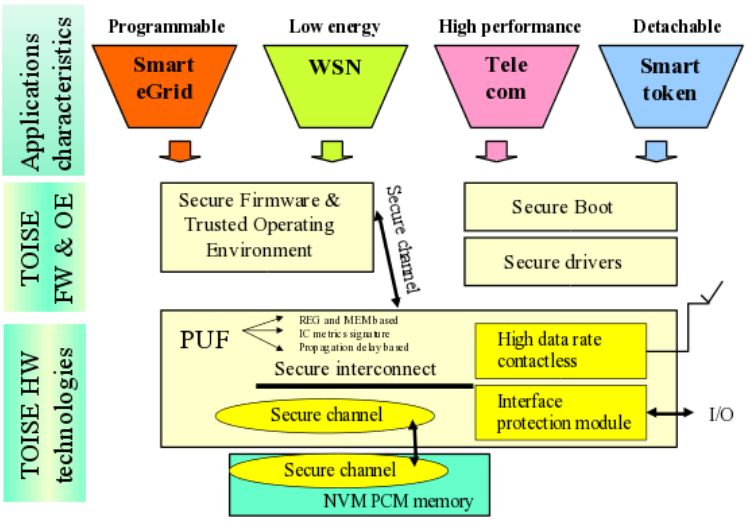

Figure 7.1: An overview of the TOISE project.

TOISE brings together a few European manufacturing-based semiconductor companies, such as STMicroelectronics and Micron, as well as several system actors, such as Eads, Gemalto, Hellenic Aerospace, Proton and Thales, in order to develop safe and secure solutions for smart grids in general. Furthermore, in TOISE a few Small and Medium Enterprises (SMEs) develop enabling blocks: Secure IC and Magillem Design. SMEs also contribute to apply the technology to the related targeted applications: AZCom and TST. CEA LETI, as a security evaluation center, performs some security tests. A few universities contribute various research aspects related to cryptographic methodologies and technologies: University Milano-Bicocca, University of Cantabria, ParisTech and Politecnico di Milano. Finally, seven research labs from the participating countries develop the further enabling research.

8

Multisensor security system for future smart homes

This chapter gives a short overview of some achievements related to multisensor security systems for future smart homes. Smart homes provide added value to the costumer domain of the smart grid by giving more control over the power consumption to the customer, but it raises security concerns. The work presented is a result of *SysSec* consortium partnering with a national research project "A Feasibility Study on Cyber Threats Identification and their Relationship with Users' Behavioral Dynamics in Future Smart Homes", DFNI-T01/4[1], a joint effort with the industry. Generally, DFNI-T01/4 aims at studying and methodologically develop a framework for the possibilities of identifying cyber threats in future smart homes. The target is an experimentally selected scenario context, accentuating on both technologies and human factor specificities.

8.1 Introduction

The availability of low power consumption electronic components nowadays allows easy development of smart multisensory devices for environment parameters detection and real-time data processing in relation to instant human health and security. In addition to the direct threats due to accidents, emergency medical problem or intrusion into a space, such system provides information about the way in which each measured environment physical parameter affects inhabitants' health, self-assessment and performance. One of the environmental factors affecting the health status of the people is the change in atmospheric pressure, which is often associated with insomnia, fatigue and dizziness [122]; Concentration of particles is related to increased risk for lung disease, susceptibility to pulmonary infections and reduced lung function [132, 170]; Low relative humidity is a cause of dry

[1] http://www.smarthomesbg.com/

or irritated mucous membranes of the eyes and respiratory system [241]; Temperature is another important factor in the room environment, which in combination with humidity affects the performance and overall health [167] and this has to be taken into account among other parameters, e.g. the elevated CO2 concentrations that lead to poor performance and fatigue [190]. On the other hand, CO is connected permanently with the hemoglobin in the blood, thus preventing the transfer of oxygen to the tissues, leading to a disturbance of the functions of the nervous and cardiovascular systems [24], i.e., the poisoning of the body.

In modern housing and workplaces the low frequency vibrations distort the sense of comfort [173], and noise in the environment producing a high risk factor for high blood pressure, having temporary or permanent effects on the cardiovascular system [56] that is also leading to sleep disturbances [169]. The level of illumination in residential and working areas also affects self-assessment, performance and psychological status of people [230].

With the advent of mobile technology in everyday life, especially in the last two decades, the level of electromagnetic load in big cities has dramatically increased. The effect on the human body is still under investigation [96], but many authors including the EC [4] suggest the existence of a significant effect of electromagnetic fields (1Hz to 300 GHz) on human health.

8.2 A security system for the future smart home

Generally, a security system for future smart homes has to meet the requirements of the three areas: reporting disaster, medical emergencies, notification of burglary, and tracking/ analysis of parameters indirectly related to health, self-assessment and performance. Additionally, the system has to allow upgrade with other sensor modules without lowering its quality and effectiveness. In response to these requirements, we propose a system of radio connected multisensory modules providing real-time information for environment changes.

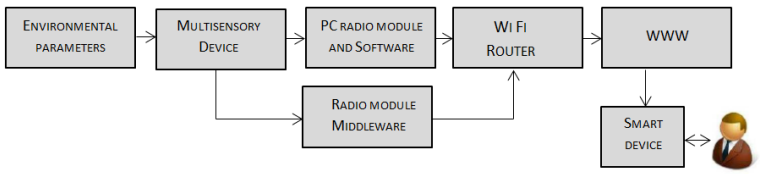

Figure 8.1: General scheme of the functioning of the security system in modern smart homes

The present security system encompasses three types of multi-sensor modules, which monitor the changes in various parameters of smart homes, as well as, paying attention to the software for processing, analysis, transmission and visualization of the data for notification of disasters, accidents and emergencies. The system also allows observation and analysis of particular parameters indirectly related to human health, self-assessment and work efficiency. We assume that the system should comprise three types of multi-sensors. The first multi-sensor type is a device combining a maximal number of sensory units in a single module (see Figure 8.2).

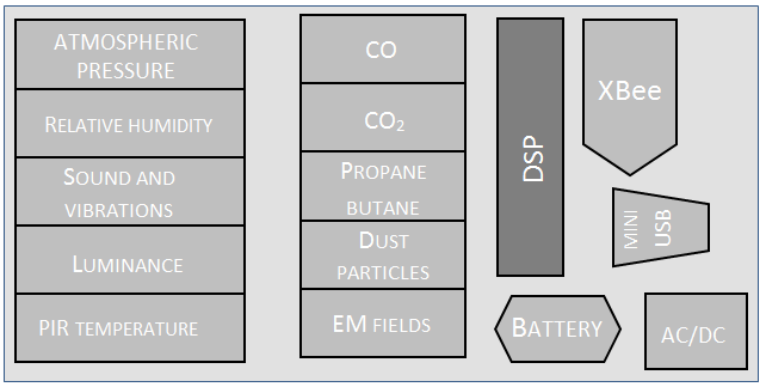

Figure 8.2: Multi-sensor system with three possible sensor elements included

The device processes the input information from different sensors in real time and transmits it to a PC via a wireless XBee Mesh network. The multi-sensor system of the first type has its power supply from the electric grid, whereas there is an option to switch to autonomous energy supply mode in cases of electric failures. The second multi-sensor system type is designed for different living spaces and integrates a smaller numbers of sensor units. In Figure 8.3, the multisensors of the second type are shown, designated for the kitchen (Figure 8.3-A), the bedroom (Figure 8.3-B) and the living room (Figure 8.3-C).

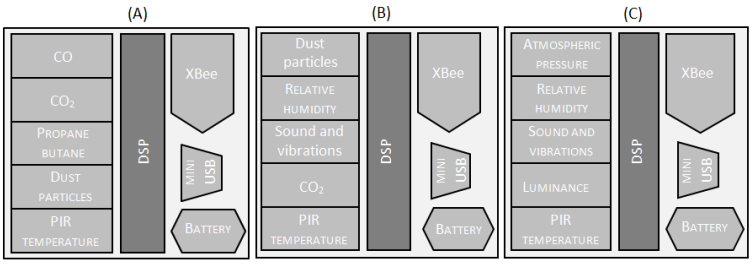

Figure 8.3: Multi-sensors with different combinations of specialized sensors

The second type of multi-sensors is designed to be supplied from the electric grid, but the embedded battery allows automatic switching to autonomous power supply mode. The third type of multi-sensors (Figure 8.4) are devices built upon only one sensor, managed by a processor and transmitting a signal in real time via an XBee network. This type of sensor has the lowest energy consumption and provides the most versatile usage and distribution possibilities, according to the particular needs of the user.

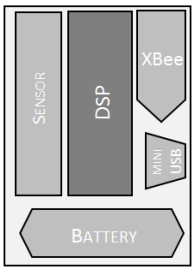

Figure 8.4: Sensor device with a single sensor element using MESH-type network

Multi-sensors of the first, second and third type can easily be combined into the system without any particular influence on the whole system's functionality. The radio connection between separate multi-sensors is realized via embedded XBee modules. The architecture of the XBee Mesh network allows adding/removing sensor units without affecting the work of the rest of the units. Sensors for the electricity consumption of any of the plugs in the space are included. That sensor permits to remotely check the energy consumption as it transmits the data via an XBee radio on the same MESH network. The data obtained from that sensor can be used to look for electric appliances that mistakenly have been left with power even though the inhabitants are not at home. Thus, the sensors for electricity consumption prevent fires and can be used to improve the domestic energy efficiency.

XBee networks: The wireless communication is built with XBee radio modules [5] based on the IEEE 802 protocol (Figure 8.5). They form a so-called *wireless MESH network*, i.e. a network where each radio module not only broadcasts its own data, but also retranslates the data of other radio modules. The MESH network organization allows for most transmissions to be successfully distributed, even if one of the network nodes goes down [36].

The ability to send information to a web-based user portal and to receive and execute user commands provide user access to the system from every point in the world with Internet communication.

Data transportation: The sensory data obtained by the environmental sensors have to be transmitted to the application in a way that allows versatility in sensor installation and ease of access by the users. To this end a

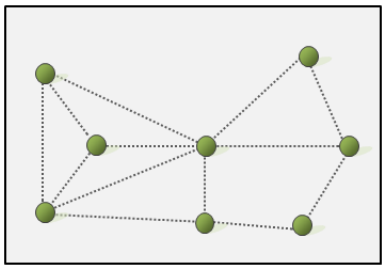

Figure 8.5: MESH Network

two level communication infrastructure is used: one level that groups the environmental sensors into a wireless MESH network and another for distributing the data on the global network (Figure 8.6).

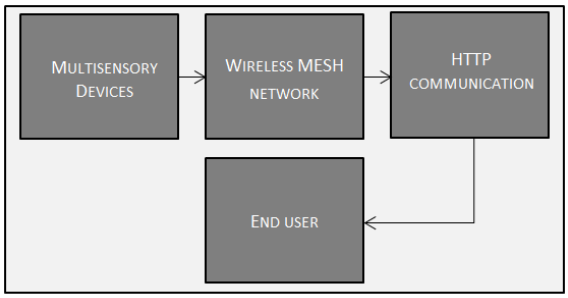

Figure 8.6: System Communications Architecture

Visualization: The multisensory data are stored in a database and is accessible from smart devices with Internet connectivity. The visualization is intuitive and easy for understanding. A web based application (Figure 8.7) has been developed with the use of the jqPlot [6] a plotting plugin for the jQuery [7] Javascript framework.

The data update is organized on regular time intervals. For a test system at intervals of 10 seconds an AJAX request is sent to the database. The minimal time interval is 1 sec and the maximum is limited to six hours. The latest data are visualized along with eight hours history for comparison. For each of the monitored environmental parameters a data range can be defined and if the current value is outside this range the application marks these values in red. The application provides ability for data review based on different date and time filters, including past periods, hours statistics, days and even months.

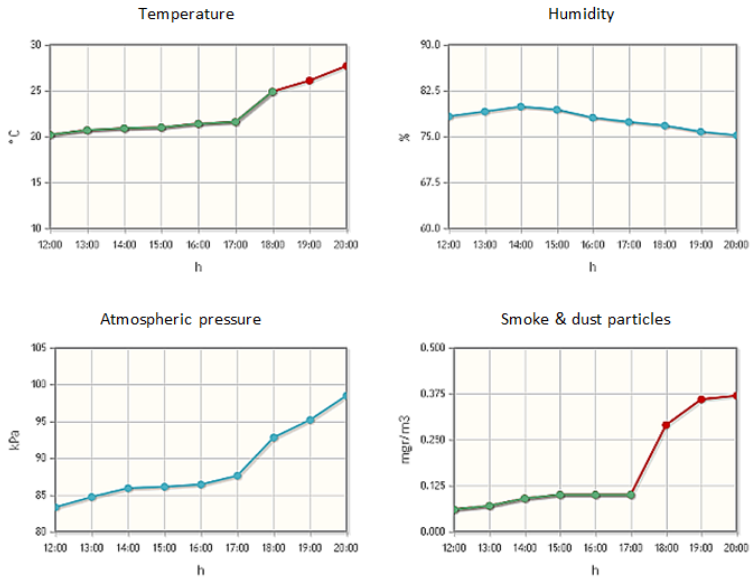

Figure 8.7: Multisensory devices web data visualization

8.3 Conclusions

The system for monitoring the parameters of the future smart home has the following advantages.

- **Flexibility:** The system permits the usage of a combination of sensor-, multi-sensor and coordinating devices without hindering the normal data transmission. The former grants the users with freedom to position the sensors in the space, according to their own preferences, which can also differ from the formerly described applications.

  There are overall three types of sensor modules and two types of coordinating devices that allow a convenient and flexible installation of a sensor network in closed spaces according to the needs and preferences of the particular user's smart home design.

- **Functionality:** The users of the system have the possibility to set the working mode of each separate sensor, and the system as a whole, online or directly via a smart device with web access. The changes to the working modes of the system and for each of the sensors are possible due to the specialized developed software using a web-based platform.

- **Continuity:** The sensors automatically transmit their signals to the coordinating unit in given time intervals. The intervals can be defined individually for each separate sensor among the multi-sensor modules.

The minimum interval for transmitting new measurements is 1s, and the maximum is restricted to 6h.

- **Energy Efficiency:** The sensors can work in a low consumption mode from which they are "awakened" by the user on request to measure and send data. After the request is fulfilled, the sensors switch back to low consumption sleeping mode.

9 Improving the analysis of embedded devices

9.1 Introduction

Embedded systems represent the majority of the devices and components that compose the smart grid. An embedded system consists of a number of interdependent hardware and software components, often designed to interact with a specific environment (e.g., a car, a pacemaker, a television, or an industrial control system). Those components are often based on basic blocks, such as CPUs and bus controllers, which are integrated into a complete custom system. When produced in large quantities, such customization results in a considerable cost reduction. For large quantities, custom built chip (ASIC) are preferred as they allow improved customization, better integration, and a reduction of the total number of parts. Such chips, also called Systems on a Chip (SoC), are often built from a standard CPU core to which both standard and custom hardware blocks are added. Standard blocks, commonly called *IP Cores*, are often in the form of a single component that can be integrated into a more complex design (e.g., memory controllers or standard peripherals). Custom hardware blocks are instead often developed for a specific purpose, device, and manufacturer. For example, a mobile phone modem may contain a custom voice processing DSP, an accelerator for the GSM proprietary hardware cryptography (A5/1) and an off-the-shelf USB controller.

Over the years, such SoCs have significantly grown in complexity. Nowadays, they often include Multiple Processors (MPSoC) and complex, custom, hardware devices. As a consequence, virtually every embedded system relies on a different, application specific, system. As a witness of this phenomenon, the website of ARM Ltd., which provides one of the most common CPU core used in embedded systems, lists about 200 silicon partners.

Unfortunately, the increasing pervasiveness and connectivity of embedded devices significantly increased their exposure to attacks and misuses.

Such systems are often designed without security in mind and visible features, low time to market, and reduction of costs are the common driving forces of their engineering teams. As a consequence, we recently observed an increasing number of reports of embedded systems compromises, with often devastating consequences [48, 57, 174, 168, 40, 29, 224, 182, 76, 66].

To make things worse, such systems often play an important role in security-relevant scenarios, as part of safety critical devices, integrated in home networks, or by handling personal user information. Therefore, it is extremely important to develop the required tools and techniques to analyze embedded systems from a security point of view.

In the traditional IT world, dynamic analysis systems play a crucial role in many security activities - ranging from malware analysis and reverse engineering, to vulnerability discovery and incident handling. Unfortunately there is no equivalent in the embedded system world. If an attacker compromises the firmware of a device (e.g., a smart meter or a PLC in a Stuxnet-like attack scenario [174]) even vendors often do not have the tools to dynamically analyze the behavior of the malicious code.

Dynamic analysis allows the analyst to overcome many limitation of static analysis (e.g., packed or obfuscated code) and to perform a wide range of more sophisticated examinations - including taint propagation [130, 233], symbolic and concolic execution [199, 61, 45, 69], unpacking [131], malware sandboxing [1, 3], and whitebox fuzzing [102, 103].

Unfortunately, all these techniques and their benefits are still not available in the world of embedded systems. The reason is that in the majority of the cases they require an emulator to execute the code and possibly monitor or alter its execution. However, the large number of custom and proprietary hardware components make the task of building an accurate emulator a daunting process. If we then consider that additional modules and hardware plugins should be developed for each embedded system on the market, we can easily understand the infeasibility of this approach.

To fill this gap we are currently working on a technique that will overcome the limitation of a pure firmware emulation. The idea is to connect together the physical device with an external emulator. By injecting a special software proxy in the embedded device, we can execute the firmware instructions inside the emulator while channeling the I/O operations to the physical hardware.

9.2 Dynamic Firmware Analysis

While the security analysis of an embedded device's firmware is still a new and emerging field, several techniques have been proposed in the past to support the debugging and troubleshooting of embedded systems.

Hardware debugging features (mostly built around In-circuit Emulators and JTAG-based hardware debuggers) are nowadays included in many embedded devices to simplify the debugging procedure. However, the task remains extremely challenging and it often requires dedicated hardware and a profound knowledge of the system under test. Several common debugging interfaces have been proposed, including the Background Debug Mode (BDM) and the ARM CoreSight debug and trace technology. Architecture-independent standards for debugging embedded devices also exist, such as the IEEE-ISTO 5001-2003 NEXUS [2]. Most of these technologies allow the operator to access, copy, and manipulate the state of the memory, to insert breakpoints, single step through the code, and collect instructions or data traces.

When available, hardware debugging interfaces can be used to perform certain types of dynamic analysis. However, they are often limited in their functionality and do not allow the operator to perform complex operations, such as taint propagation or symbolic execution. In fact, these advanced dynamic analysis techniques require an instruction set emulator to emulate the firmware of the embedded target. Unfortunately, the proper emulation of the firmware requires the analysis environment to correctly emulate also all the peripherals devices. Without such support, the emulated firmware would often hang, crash, or in the best case scenario, show a different behavior than on the real hardware. Such deviations can be due, for example, to incorrect memory mappings, active polling on a value that should be changed by the hardware, or the lack of the proper hardware-generated interrupts or DMA operations.

To overcome these problems, researchers and engineers have devised three solutions, each with its own limitations and drawbacks:

- *Complete Hardware Emulation*
  Chipounov [60] and Kuznetsov et al. [142] proposed the implementation of an emulated PCI bus and network card that returns symbolic values. This approach has the main drawback that it requires to emulate the device properly. While this is not much of a problem for well understood devices, like a PCI network card supported by most PC emulation software, it can be a real challenge in embedded systems and can be just impossible when the hardware is not documented. Unfortunately, lack of documentation is the rule in the embedded world, especially in complex proprietary SoCs.

  In some cases, accurate system emulators are developed as part of the product development to allow the firmware development team to develop software while the final hardware is still not available. However, those emulators are usually unavailable outside the development team and they are often not designed for code instrumentation, mak-

ing them unable to perform basic security analysis (tainting, symbolic execution...).

- *Hardware Over-approximation*
  A simpler approach consists in using a generic, approximated, model of the hardware. For example, by assuming that interrupts can happen at any time or that reading an IO port can return any value. This approach is easy to implement because it does not require a deep knowledge of the real hardware, but it can clearly lead to false positives, (e.g., values that will never be returned by the real system) or misbehavior of the emulated code (when a particular value is required). This approach is commonly used when analyzing small systems and programs that are typically limited to a few hundreds of lines of code, as showed by Schlich [197] and Davidson et al. [69]. However, on larger programs and on complex peripherals this approach will invariably lead to a state explosion that will prevent any useful analysis.

- *Firmware Adaptation*
  Another approach consists in adapting the firmware (or in extracting limited parts of its code) in order to emulate it in a generic emulator. While this is possible in some specific cases, for example in Linux-based embedded devices, this technique does not allow for an holistic analysis and may still be limited by the presence of custom peripherals. Moreover, this approach is not possible for monolithic firmwares that cannot be easily split in independent parts - unfortunately a very common case in low-end embedded systems.[67]

In the next section we present our work-in-progress technique based on an hybrid combination of the actual hardware with a generic CPU emulator. Our approach enables to perform advanced dynamic analysis of embedded systems, even when very little information is available about the firmware and when the hardware of the device does not provide even the basic debugging support.

9.3 Embedded System Emulation

Our approach aims at enabling independent dynamic analysis of embedded software. It can assist in several activities, including reverse engineering, bug finding through fuzzing or symbolic execution, vulnerability assessment, and root cause analysis of known vulnerabilities.

Our framework is built around the concept of executing the firmware into an emulator while forwarding the IO access to the real device.

The emulator is instrumented to forward accesses of memory regions to our engine, which in turn forwards them to the target device. The internal

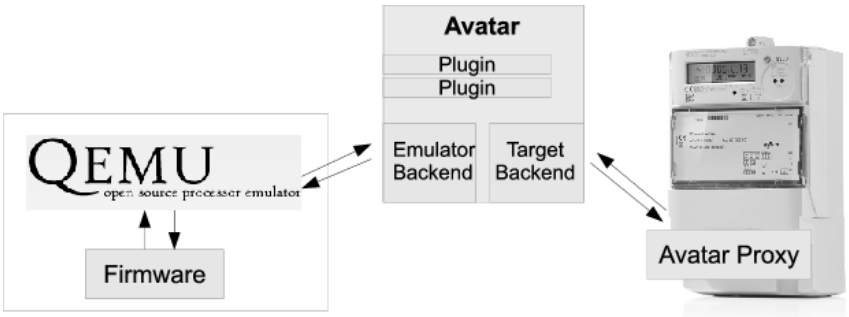

Figure 9.1: Architecture Overview

architecture of the system is completely event-based, allowing plugins to tap into the data stream and even modify data as it is passed between the emulator and the target.

We have developed an emulator plugin for Qemu/S2E. This plugin talks to a GDB instance over the GDB/MI interface to control the firmware's execution in Qemu: It can get or set registers, set breakpoints and read or write memory. Additionally it is connected to a plugin in S2E which forwards memory accesses of selected memory regions to our system. More advanced functionality of S2E will be made available through the QMP interface, a JSON-based request-response protocol which will allow to execute Lua functions to control symbolic execution. Currently we only support Qemu/S2E as emulator, but the interface for emulators is generic and allows other emulators to be added easily.

On the target side, we developed three backends:

- A plugin that uses the GDB serial protocol to communicate with GDB servers (e.g. a debugger stub on the device or a JTAG GDB server)
- A plugin to support low-level access to OpenOCD's JTAG debugging via a telnet-like protocol
- A plugin that uses a custom binary protocol which is more efficient than the GDB protocol to interact with our debugger stub on the target

Additionally, we also developed a plugin to optimize the access to memory. When emulator executes the firmware while forwarding all memory accesses, this plugin observes the memory addresses accessed by the requests as well as the current program counter and stack pointer addresses. If a value is read from or written to the target, it is cached by the plugin. When the next read for the same address happens, the plugin verifies whether the read value corresponds to the expected one from the cache. A difference between the cached value and the read value is interpreted as memory-mapped IO access.

For each of the categories read access, write access, stack access, execute access and IO access the plugin keeps a per-page count (the page size is configurable) and derives the type of the memory range (code, read-only data, read-write data, stack, IO memory) at the end of the firmware execution.

9.3.1 Context Switching

While it is possible to perform the complete analysis on the emulator, it is sometimes interesting to let the program run completely on the target device. This allows to proceed without any delay or interaction with the emulator until the analysis point is reached. For example, during the loading phase of the embedded system there may be an intensive IO access phase. This approach is also useful when the device under analysis needs to perform some network interaction that has to be completed without delays.

Starting the Analysis at a selected point

In this case the firmware starts on the physical device and runs natively until a certain pre-defined event occurs (e.g., a breakpoint is reached or an exception is raised). At this point, the execution on the physical device is frozen and the state is transferred to the emulator, where the execution is resumed.

Returning execution to the Hardware

After the required analysis is performed on the emulator, the execution of the firmware can be transferred back to continue on the real device. In this case the memory state of the virtual environment is copied back to the physical device. Depending on the analyst's needs, it is possible to switch again to the emulator at a later stage.

9.4 Full-Separation Mode

At the current stage, the system is completely implemented and it has been tested on a number of different physical devices. The prototype now works in what we call "full-separation mode", in which the code is executed in the emulator and the (memory) state is kept in the physical device. In other words, for each instruction that is executed in the emulator, the accessed memory addresses are fetched from and written to the real memory of the embedded system. At the same time, interrupts are intercepted by our proxy in the physical system and forwarded back to the emulator.

This approach allows, in theory, to emulate any firmware. In practice, there are two main problems. First, the throughput is often so low (few

instructions per second) that the analysis time becomes intractable. Second, many physical devices have time-critical sections that need to be executed in a short amount of time or the execution would fail. For instance, certain interrupts are generated at a very high frequency and the corresponding handler needs to be executed before the next interrupt is generated.

For this reason, we are now implementing two new functionalities in our analysis framework. On one side, we plan to transfer part of the state of the application from the physical device to the emulator, in order to reduce the memory I/O operations that need to be forwarded to the physical device. On the other side, we are implementing a new technique to isolate certain functions in the code and move them back to be executed natively on the hardware. Even though the implementation of these optimization techniques is not completed, preliminary experiments seem to confirm that they would successfully speed up the system to a point in which complex dynamic analysis of embedded devices becomes possible.

9.5 Conclusions

Smart grids greatly rely on embedded systems to perform a variety of computation and automation tasks. Unfortunately, while for normal computers there are many static and dynamic analysis tools available, on the embedded world these tools are still missing. It is often the case that even the vendors do not have the right emulators required to properly study compromised devices, or malicious firmwares installed by an attacker.

This lack of tools greatly reduces the ability of researchers to investigate the security of embedded devices. For this reason, we are now working on a novel technique to fill this gap by making advanced dynamic analysis possible on unknown firmwares. Our current prototype shows promising results and we hope to be able to overcome the current performance limitations in the near future.

10 Remote control of smart meters: friend or foe?

10.1 Introduction

Even though the transition to the smart grid offers new functionalities, it also brings security concerns in how the technology can be misused by a malicious adversary. We have detailed some of these problems in earlier chapters and highlighted the need for an interdisciplinary approach to analyze possible solutions. In this chapter we describe and analyze just such a problem. We present a scenario where the adversary targets the smart meters installed in the electrical distribution network. Our main focus is not on the attack on the smart meters themselves, but rather on the impact (and resulting damage) that the adversary can cause to the electric grid *if* he would manage to control a number of smart meters in a single neighborhood or even several cities within a country.

As explained in Section 1.2, the EU mandates that all the metering devices present in the traditional energy distribution network should be replaced with smart meters by 2020, in an attempt to better control and monitor the energy consumption. These meters have a range of functions, allowing remote reading of consumption but also the possibility to remotely turn power on or off. The distribution company can, for example, cut the power from customers that have defaulted on their payments by simply issuing a remote command to the smart meter.

The scenario is localized to a neighborhood modeled as a *power island* (see Section 10.2). Here, the attacker's purpose is to create havoc and damage the electrical appliances in the neighborhood by changing the voltage in the network through his control over the smart meters, or more specifically over their ON/OFF switch.

The rest of this chapter is organized as follows. In Section 10.2 we give a short background to important properties of electrical networks. In

Section 10.3, we outline the scenario and the possible consequences and describe the simulation results in Section 10.4 before we conclude the chapter.

10.2 Background

10.2.1 Power quality

One of the important factors in the design of a distribution grid is power quality, i.e. limits for power supply frequency and voltage magnitude, so that electric and electronic equipments can function without damage when connected to the outlet. It is important to account for the transmission line's loss to ensure power quality standards for each *feeder*, where a feeder is a portion of the grid that provides power transportation capabilities to service areas. In real-world conditions with variable loads and unpredictable power generation levels, power quality issues need to be handled in order to respect the specification of appliances. According to the European "Voltage Characteristics in Public Distribution Systems" including EN 50160 and EN 61000 standards [156], there are specific voltage requirements and frequency regulations for different situations. There are requirements for an acceptable variation of voltage magnitude (from $220 - 240$V nominal value, depending on country) and power frequency (50Hz nominal value). Variations allowed for the frequency must be on average $\pm 1\%$ ($49.5 - 50.5$Hz) in 95% of a week time and $-6\%/+4\%$ ($47 - 52$Hz) at all times. Voltage magnitude variation should be within $\pm 10\%$ of the nominal voltage in 95% of a week time and all average values should not go outside $-15\%/+10\%$ of the nominal voltage. In the case that an energy provider does not respect power quality standards, grid problems may appear and cascade to unforeseen consequences. Therefore, power quality specification are enforced by financial penalties on the energy providers. Problems related to the quality of the voltage may manifest themselves as short interruptions, flickers, voltage dips, supply voltage variations and harmonic disturbances. For more information, we refer the reader to [90], which explains in detail how these phenomena manifest.

In the attack scenario that will be described in detail in Section 10.3, the voltage magnitude is pushed outside of the standard value limits. By inducing abrupt variations in the loads at precise points in time, for example when the grid becomes underloaded (too much available power), the voltage magnitude can be pushed outside the range of safe values. Formally, this is a consequence of the power flow equations relating individual nodes or bus power properties, used for the simulation in Section 10.4. These equations are briefly outlined below.

10.2.2 Power flow equations

As previously mentioned, the electrical grid is composed of *nodes* or *buses*, where each node is connected by a *transmission line*. In order to model the electrical infrastructure and electricity flow, each transmission line is characterized by physical properties (resistance, capacity and inductance). The line's specific properties can be measured, and the loss characterizing the line, also known as impedance can be calculated. A square matrix, whose dimension depends on the number of nodes in the network, can be built to represent the impedance between nodes (the Z matrix or impedance matrix). However, the Y matrix (or admittance matrix), the inverse of the impedance matrix, is used in practice. With the following power flow equations, voltage magnitude and angle at each node, as well as real and reactive power flowing in and out each node, can be computed as:

$$\begin{aligned} P_i &= \sum_{i=1}^{N} Y_{ij} V_i V_j \cos\left(\theta_{ij} + \delta_j - \delta_i\right), \\ Q_i &= -\sum_{i=1}^{N} Y_{ij} V_i V_j \sin\left(\theta_{ij} + \delta_j - \delta_i\right), \end{aligned}$$

where P_i and Q_i are the real and the reactive power at node i, respectively; Y_{ij} and θ_{ij} are the magnitude and angle of the admittance between node i and node j; V_i and V_j are the voltage magnitudes at node i and node j; δ_i and δ_j are the phase angles at node i and node j.

Based on these values, voltage regulations can be enforced for each feeder in the grid [163]. Grid control and regulation in a centralized system can be solved by central operators in the SCADA (Supervisory Control And Data Acquisition) system. However, in the context of distributed generation where customers can produce their own energy, and thus become providers for their neighbors, serious problems may arise; this is the actual topic explored in our scenario. For instance, since power injection into a grid is in some regions freely allowed (with the required standard specifications to be met), a node's voltage magnitude may vary even more due to abrupt changes in consumption.

10.2.3 Power islands

One of the main changes involved in the transition to the smart grid is the distributed generation of energy. In [21], distributed generation is defined as: "[...] an electric power source connected directly to the distribution network or on the customer side of the meter". Many customers will opt for installing a renewable energy production facility on their domain, for example a wind turbine or photovoltaic panels, in order to obtain some independence from their main energy provider; some may even sell the surplus energy produced.

The traditional distribution network can be modeled as a tree-like structure, but with the changes of local producers of energy the best model is more flat, like interconnected power islands. Our scenario takes place in a power island (Figure 10.1), which has become self sufficient and surplus energy is injected back into the grid.

10.3 Voltage variation in a neighborhood

With the overview given above, let us now consider the actual attack scenario. We focus on the main steps of the scenario and the impact of the final attack, and not particularly on the details of the actual attack against the smart meters themselves. Such attacks have been documented elsewhere and they are partly described in Chapter 2.

The scenario focuses on a small neighborhood with only one power transformer. Some houses in the neighborhood produce their own renewable energy, and the excess energy is injected back into the network. Every customer has a smart meter installed, that in turn communicates with the data concentrator attached to the neighborhood power transformer. The power consumption is reported to the data concentrator once every 15 minutes, and the smart meters can receive commands to turn on or off the electricity for the customer at any moment.[1] All communication is encrypted with a symmetric encryption key but no other security mechanism is running on the smart meters. The neighborhood is shown in Figure 10.1.

A common misconception about security is that *if* encryption is used, the network and the devices are safe from attacks. However, the devices may still be vulnerable to an exploit (buffer overflow), the protocols may not be well implemented, an oversight may lead to no change of the default settings, or there might even be an inside leak from the electricity company in question. As a parallel, Stuxnet used valid signatures [85] in its infection.

Thus, the attacker is presumed to have a good knowledge of the smart meter deployment as well as its shared encryption key. The attacker can then take control over a number of smart meters in the neighborhood. Arbitrarily turning on or off electricity for customers would at least inconvenience customers, and cause a financial loss to the electricity company. However, the purpose of this scenario, further explored in the next section with a simulation, is to determine if an attacker can make the voltage of the network go outside the tolerance limits of +10% and -15% around the optimal value, being 230V for Europe.

[1] The smart meter has its own power supply, so it does not depend on whether the electricity in the house is on or off.

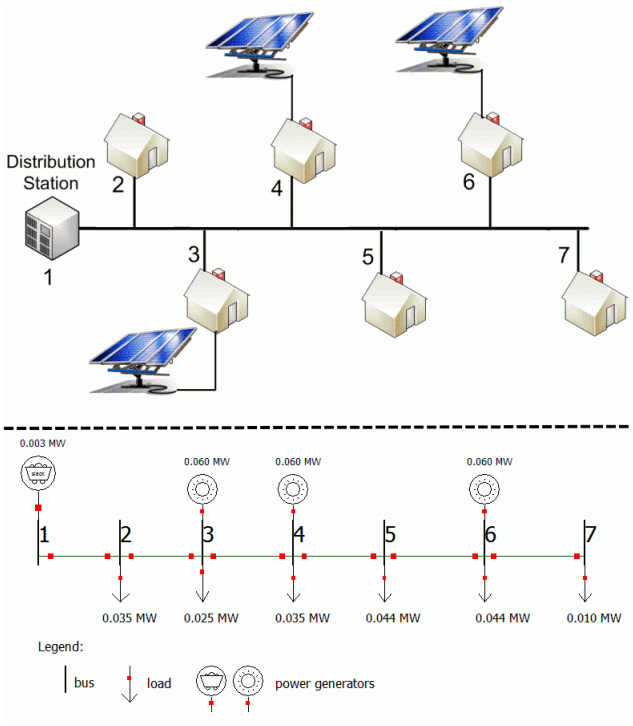

Figure 10.1: Neighborhood overview (top) and electrical network model overview (bottom)

10.4 Simulating the effects

To demonstrate the viability of the voltage variation attack scenario, we use the PowerWorld Simulator software suite [64] to model a realistic grid configuration (e.g. modeling transmission lines, generators, loads and solving the resulting power flow equations). The upper of Figure 10.1 presents an intuitive overview of the neighborhood while the lower subfigure shows the overview of the resulting electrical network model used in the simulation.

10.4.1 Simulation setup

The neighborhood is a typical country-side distribution grid, with several buildings and their facilities served by one power substation (marked as node one in the figure). The six buildings (numbered two to seven) have a relatively higher than normal individual instantaneous energy consumption from 10 to 50kW. There are three renewable energy production facilities in the neighborhood, connected at nodes three, four, and six respectively. These facilities can produce more energy than required for the local neighborhood, so sometimes the surplus is injected into the electrical network. The bars numbered from one to seven in the lower subfigure are called

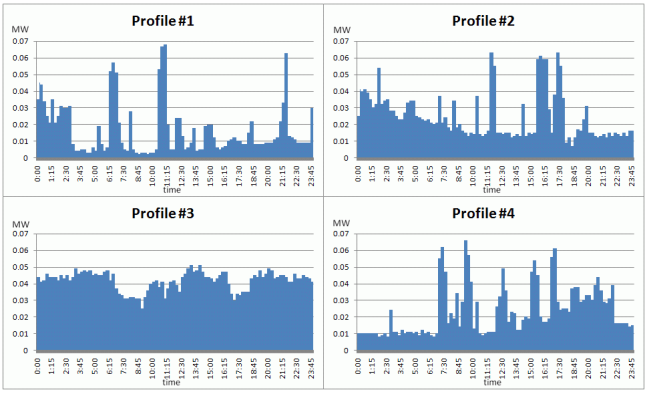

Figure 10.2: The consumption profiles for four different customers

buses. Buses are points in the electrical system where certain electrical attributes such as voltage, power and current can be evaluated ("p.u." signifies the voltage per unit value of each bus). Every building that consumes energy is modeled as a load, every renewable energy facility is modeled as a generator and the electrical lines connecting the nodes are modeled as transmission lines with proper loss. We utilize four real-time daily consumption profiles for the customers according to [136]. In Figure 10.2, the consumption profiles are based on 24 hour consumption patterns with 15-minute interval measurements. These profiles are characterized by peaks during the rush hours (in the morning, at lunch and in the evening), depending on each household's appliances in use. We let customers #2 and #4 use the consumption profile one, customer #3 use profile two, customer #5 and #6 use profile three and finally customer #7 use profile four (chosen arbitrarily).

The simulation runs during 24 simulation seconds, equivalent to 24h with the load variation described above. Changing either the load profiles or the grid configuration will change the end result.

10.4.2 Simulation results under normal conditions

In Figure 10.3 we present the voltage magnitude variation of Bus #7, i.e. the load for customer seven. The grid's behavior (loss, power injections, loads) is responsible for the voltage maximum and minimum points seen in the figure. The voltage peaks are the result of large amounts of power in the grid and few consumers, while the voltage minimum points are a result of many consumers and little available power. In normal running conditions, the voltage profile of Bus #7 respects regulations and the voltage magnitude never goes beyond $+8\%/-6\%$ of the nominal value.

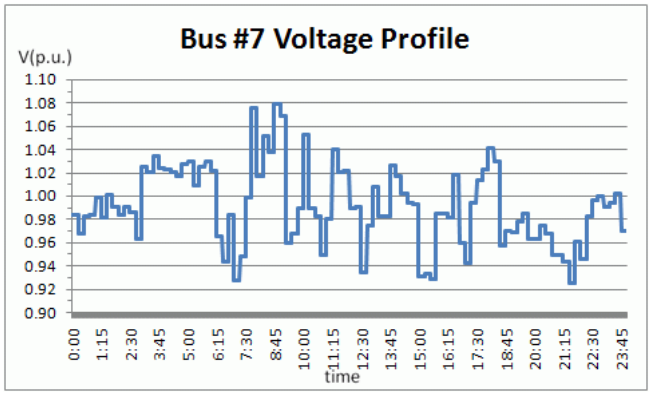

Figure 10.3: Voltage level at Bus #7 during the simulation

10.4.3 Simulation results while executing the attack

The goal of the adversary is to vary the voltage magnitude of Bus #7 outside the safety zone of $\pm 10\%$. This is achieved by manipulating the smart meters in the neighborhood to shut down power in only some parts. In a first attempt, the attacker gains control of the meter controlling the load on Bus #5. The attack is then launched at an appropriate point in time, for example at the voltage peak observed between time $7 - 10$ when the grid is vulnerable; there is then a high demand for energy (people are preparing for going to work) and the generators must compensate and push more power into the grid. If the attack is timed correctly, Load #5 will be interrupted during the established period, leading to more power being routed to the other buses. The result is shown in Figure 10.4, where the highlighted area represents the time of the attack. The voltage magnitude barely goes up 2% of the established barrier at 1.1 volts per unit and for very short period of times. This may not be enough to cause damage to the target and the attack cannot be deemed as a success.

In a second attempt, the attacker gains control of an additional smart meter (the one controlling Bus #6) and the result is showed in Figure 10.5. This time, the attacker manages to drive the voltage magnitude to peaks of $+17\%$ with a constant average value above the normal $+10\%$ between time 7 and 10. This increase in voltage should be enough to cause major damage, both physical and economical, to the customer in the absence of voltage regulators. As an observation, the number of smart meters that need to be compromised is not in direct relation to the total number of smart meters in the neighborhood, but with the power loads controlled by one of them. For example, a smart meter that controls a high-load household is more attractive for takeover since the strain it can reflect into the electrical grid is higher. The effect can be replicated any time by the attacker, by simply turning off the energy consumption in some buildings in a neighborhood, in

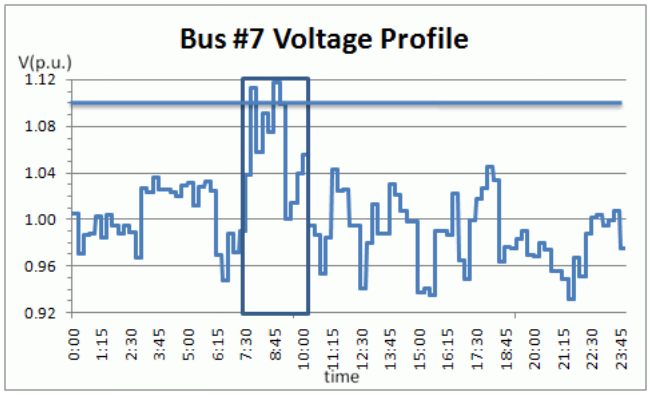

Figure 10.4: Bus #7 voltage after launching the attack on Bus #5 with the attack period emphasized with the square box. The safe voltage limit is at 1.1 V(p.u.) and the voltage is normalized.

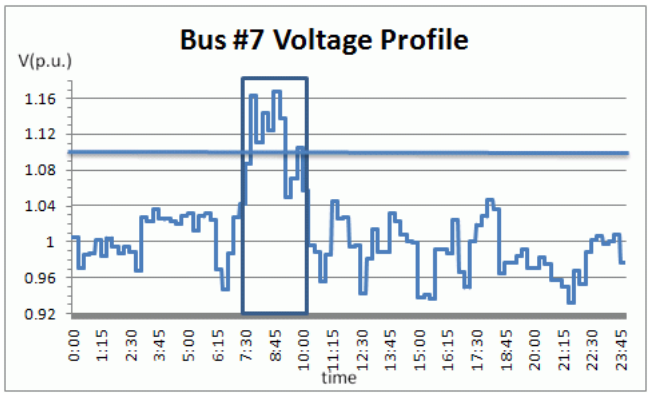

Figure 10.5: Bus #7 voltage after launching the attack on Bus #5 and #6 with the attack period emphasized with the square box. The safe voltage limit is at 1.1 V(p.u.) and the voltage is normalized.

turn damaging electric appliances in other buildings still connected to the network.

In the future, we would also like to explore possible economic losses caused by the attacker because even if no permanent damage is achieved, both customers and the electrical company may lose money when the attack is performed. Three ways to mitigate the attack is to either harden the smart meters, install voltage regulators at the customer's site or install adaptable renewable generation facilities.

10.5 Conclusions

In this chapter we presented a scenario where a skilled adversary may affect fundamental properties of the electrical grid by controlling a number of smart meters. By complementing and building on related research, we show how even attacks on the distribution network may affect grid stability. We concentrated our work by simulating a small *power island* to show feasibility, but we have also explored variants on a larger scale where the attacker tries to shift the *frequency* of the network instead of the voltage. Mitigation and defense techniques against these attacks were only mentioned briefly and is ongoing work.

With the current push for massive installment of smart meters as well as a continuous development of their capabilities, it is only a question of time before the infrastructure is attacked. Thus, one goal of this work is to look at the problems from an interdisciplinary point of view, considering both issues related to computer security and the electrical power domain. The smart grid straddles both these two domains and expertise on both areas are necessary to develop successful mitigation strategies.

Conclusion

In Europe and elsewhere, the electrical grid is being transitioned into the "smart grid" in order to increase flexibility and accommodate large scale energy production from renewable sources. This transition involves, among other steps, the installation of new, advanced equipment with, for example, the replacement of traditional domestic electrical meters with smart meters, and remote communication with devices, for example allowing remote access to an unsupervised energy production site.

The objective of this deliverable is to give a broad survey of the ongoing research related to the smart grid, especially related to the *SysSec* themes. As the smart grid will be a highly complex system of interacting systems, it is essential to have an outlook and discussion of its major building blocks and technologies and how they may influence the grid. Each of these technologies comes with its own benefits and concerns, increasing both the complexity and the functionality in the future electricity grids.

Naturally, this deliverable includes sections on networking technologies and issues that they imply as part of the infrastructure. As very large datasets, also including customer data, are produced to control the smart grid, the report naturally also includes sections for privacy concerns, as well as for scalable data processing in the smart grid with a focus on security processing; besides, intrusion detection has a natural part in these contexts.

At relevant places in these surveys we highlight research directions enabled and followed by *SysSec* partners, namely in intrusion detection and processing, referring to monitoring and control, as well as extracting information out of the data, in order to signal alarms and related events. We emphasize the role of streamed event processing and multicore/parallel computations in this context, especially due to the very large data volumes and the need for efficiency and timely action.

Finally, we also provide in some more detail surveys of a set of related research areas from partners of the *SysSec* consortium: the TOISE project, the future smart home, the need to be able to analyze the firmware of embedded devices, such as smart meters, and a simulated scenario where we look at the power quality of the network if a sufficient number of smart meters are compromised.

Bibliography

[1] Anubis: Analyzing unknown binaries. http://anubis.iseclab.org/.

[2] Ieee-isto 5001 - 2003 the nexus 5001 forum standard for a global embedded processor debug interface. IEEE - Industry Standards and Technology Organization, December 2003.

[3] Cwsandbox, 2008. http://www.cwsandbox.org.

[4] http://ec.europa.eu/health/ph_risk/committees/04_scenihr/docs/scenihr_o_006.pdf, accessed in 2013.

[5] http://www.zigbee.org/, accessed in 2013.

[6] http://www.jqplot.com/, accessed in 2013.

[7] http://www.jquery.com/, accessed in 2013.

[8] Eucalyptus. http://www.eucalyptus.com/eucalyptus-cloud/iaas, accessed in 2013.

[9] Google PowerMeter. http://www.google.com/powermeter/about/, accessed in 2013.

[10] IBM InfoSphere. http://www-01.ibm.com/software/data/infosphere/, accessed in 2013.

[11] Microsoft Hohm. http://www.microsoft.com/environment/, accessed in 2013.

[12] MOA Framework. http://researchcommons.waikato.ac.nz/handle/10289/4934, accessed in 2013.

[13] SONET / SDH Technical Summary, http://www.techfest.com/networking/wan/sonet.htm, accessed in 2013.

[14] Storm Project. http://storm-project.net/, accessed in 2013.

[15] StreamBase. http://www.streambase.com/, accessed in 2013.

[16] Yahoo S4. http://incubator.apache.org/s4/, accessed in 2013.

[17] ZigBee Alliance - ZigBee certified products, http://www.zigbee.org/Products/ByFunction/AllFunctions.aspx, accessed in 2013.

[18] ZigBee Alliance - ZigBee specification, http://www.zigbee.org/Specifications.aspx, accessed in 2013.

[19] D. J. Abadi, Y. Ahmad, M. Balazinska, U. Cetintemel, M. Cherniack, J.-H. Hwang, W. Lindner, A. S. Maskey, A. Rasin, E. Ryvkina, et al. The design of the borealis stream processing engine. CIDR, 2005.

[20] D. J. Abadi, D. Carney, U. Çetintemel, M. Cherniack, C. Convey, S. Lee, M. Stonebraker, N. Tatbul, and S. Zdonik. Aurora: a new model and architecture for data stream management. *The VLDB Journal – The International Journal on Very Large Data Bases*, 12(2):120–139, 2003.

[21] T. Ackermann, G. Andersson, and L. Söder. Distributed generation: a definition. *Electric Power Systems Research*, 57(3):195 – 204, 2001.

[22] I. Akyildiz and X. Wang. A survey on wireless mesh networks. *Communications Magazine, IEEE*, 43(9):S23–S30, 2005.

[23] H. K. Alfares and M. Nazeeruddin. Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science*, 33(1):23–34, 2002.

[24] Y. Amitai, Z. Zlotogorski, V. Golan-Katzav, A. Wexler, and D. Gross. Neuropsychological impairment from acute low-level exposure to carbon monoxide. *Archives of neurology*, 55(6):845, 1998.

[25] J. Anatory, N. Theethayi, and R. Thottappillil. Channel characterization for indoor power-line networks. *Power Delivery, IEEE Transactions on*, 24(4):1883–1888, 2009.

[26] D. G. Andersen. Mayday: distributed filtering for internet services. In *USITS'03: Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems*, pages 3–3, Berkeley, CA, USA, 2003. USENIX Association.

[27] J. P. Anderson. Computer security threat monitoring and surveillance. Technical report, James P Anderson Co., Box 42, Fort Washington, PA 19034, USA, Apr. 15, 1980.

[28] R. Anderson and S. Fuloria. Who controls the off switch? In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 96–101. IEEE, 2010.

[29] Anonymised for peer review. Implementation and implications of a stealth hard-drive backdoor.

[30] R. R. R. Barbosa and A. Pras. Intrusion detection in SCADA networks. In *Proceedings of the Mechanisms for autonomous management of networks and services, and 4th international conference on Autonomous infrastructure, management and security*, AIMS'10, pages 163–166, Berlin, Heidelberg, 2010. Springer-Verlag.

[31] A. Barenghi, G. Bertoni, L. Breveglieri, M. Fugini, and G. Pelosi. Smart metering in power grids: Application scenarios and security. In *Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES*, pages 1–8, 2011.

[32] J. Ben-Othman and Y. Saavedra Benitez. On securing HWMP using IBC. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5, 2011.

[33] S. Benkner, S. Pllana, J. Traff, P. Tsigas, U. Dolinsky, C. Augonnet, B. Bachmayer, C. Kessler, D. Moloney, and V. Osipov. PEPPHER: Efficient and Productive Usage of Hybrid Computing Systems. *IEEE Micro*, 2011.

[34] C. Bennett and S. Wicker. Decreased time delay and security enhancement recommendations for AMI smart meter networks. In *Innovative Smart Grid Technologies (ISGT), 2010*, pages 1–6, 2010.

[35] R. Berthier, W. H. Sanders, and H. Khurana. Intrusion detection for advanced metering infrastructures: Requirements and architectural directions. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 350–355. IEEE, 2010.

[36] B. Bilgin and V. Gungor. Performance evaluations of ZigBee in different smart grid environments. *Computer Networks*, 56(8):2196–2205, 2012.

[37] S. Biswas and R. Morris. ExOR: Opportunistic Multi-Hop Routing for Wireless Networks. In *SigComm: Proc. of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2005.

[38] C. A. Boano, T. Voigt, C. Noda, K. Römer, and M. A. Zúñiga. JamLab: Augmenting sensornet testbeds with realistic and controlled interference generation. In *Proceedings of the 10^{th} IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2011.

[39] J.-M. Bohli, C. Sorge, and O. Ugus. A privacy model for smart metering. In *Communications Workshops (ICC), 2010 IEEE International Conference on*, pages 1 –5, May 2010.

[40] H. Bojinov, E. Bursztein, and D. Boneh. Embedded management interfaces: Emerging massive insecurity. In *Blackhat 2009 Technical Briefing / whitepaper*, 2009.

[41] P. Bonnet, J. Gehrke, and P. Seshadri. Towards sensor database systems. *Lecture Notes in Computer Science*, pages 3–14, 2001.

[42] F. Borges, L. A. Martucci, and M. Muhlhauser. Analysis of privacy-enhancing protocols based on anonymity networks. pages 378–383, 2012.

[43] A. Breidthardt. German government wants nuclear exit by 2022 at latest. http://uk.reuters.com/article/2011/05/30/us-germany-nuclear-idUKTRE74Q2P120110530, May 2011.

[44] E. Buchmann, K. Böhm, T. Burghardt, and S. Kessler. Re-identification of smart meter data. *Personal and Ubiquitous Computing*, 17(4):653–662, 2013.

[45] C. Cadar, D. Dunbar, and D. Engler. KLEE unassisted and automatic generation of high-coverage tests for complex systems programs. In *OSDI*, 2008.

[46] A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd conference on Hot topics in security*, HOTSEC'08, pages 6:1–6:6, Berkeley, CA, USA, 2008. USENIX Association.

[47] A. A. Cárdenas, S. Amin, and G. Schwartz. Privacy-aware sampling for residential demand response programs. 2012.

[48] Carna Botnet. Internet census 2012, port scanning /0 using insecure embedded devices, 2012. http://internetcensus2012.bitbucket.org/paper.html.

[49] M. Carpenter, T. Goodspeed, B. Singletary, E. Skoudis, and J. Wright. Advanced Metering Infrastructure Attack Methodology. http://inguardians.com/pubs/AMI_Attack_Methodology.pdf, 2009.

[50] M. Çakiroğlu and A. T. Özcerit. Jamming detection mechanisms for wireless sensor networks. In *Proceedings of the 3rd international conference on Scalable information systems*, InfoScale '08, pages 4:1–4:8, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[51] D. Cederman, B. Chatterjee, and P. Tsigas. Understanding the performance of concurrent data structures on graphics processors. In *Euro-Par*, pages 883–894, 2012.

[52] D. Cederman, A. Gidenstam, P. H. Ha, H. Sundell, M. Papatriantafilou, and P. Tsigas. Lock-free concurrent data structures. *CoRR*, abs/1302.2757, 2013.

[53] D. Cederman and P. Tsigas. GPU-Quicksort: A practical Quicksort algorithm for graphics processors. *Journal of Experimental Algorithmics (JEA)*, 14:1.4–1.24, 2009.

[54] S. Chachulski, M. Jennings, S. Katti, and D. Katabi. Trading Structure for Randomness in Wireless Opportunistic Routing. In *SigComm: Proc. of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2007.

[55] Y. Challal, H. Bettahar, and A. Bouabdallah. A taxonomy of multicast data origin authentication: Issues and solutions. *Communications Surveys Tutorials, IEEE*, 6(3):34–57, 2004.

[56] T.-Y. Chang, T.-C. Su, S.-Y. Lin, R.-M. Jain, and C.-C. Chan. Effects of occupational noise exposure on 24-hour ambulatory vascular properties in male workers. *Environmental health perspectives*, 115(11):1660, 2007.

[57] S. Checkoway, D. McCoy, D. Anderson, B. Kantor, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno. Comprehensive Experimental Analysis of Automototive Attack Surfaces. In *Proceedings of the USENIX Security Symposium*, San Francisco, CA, August 2011.

[58] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes. Using model-based intrusion detection for SCADA networks. In *Proceedings of the SCADA Security Scientific Symposium*, pages 1–12, 2007.

[59] E. Chien, L. OMurchu, and N. Falliere. W32.Duqu: the precursor to the next stuxnet. In *USENIX conference on Large-Scale Exploits and Emergent Threats*. USENIX Association, Apr. 2012.

[60] V. Chipounov and G. Candea. Reverse Engineering of Binary Device Drivers with RevNIC. In *Proceedings of the 5th ACM SIGOPS/EuroSys European Conference on Computer Systems (EuroSys), Paris France, April 2010*, Paris, France, 2010.

[61] V. Chipounov, V. Kuznetsov, and G. Candea. The s2e platform: Design, implementation, and applications. *ACM Trans. Comput. Syst.*, 30(1):2:1–2:49, Feb. 2012.

[62] CISCO. Substation automation for the smart grid, http://www.cisco.com/en/US/prod/collateral/routers/ps10967/ps10977/white_paper_c11_603566.pdf, 2010.

[63] CNN. Researchers find smart meters could reveal favorite TV shows. http://news.cnet.com/8301-27080_3-57364883-245/researchers-find-smart-meters-could-reveal-favorite-tv-shows/, 2012.

[64] P. Corporation, June 2011.

[65] M. Costache, V. Tudor, M. Almgren, M. Papatriantafilou, and C. Saunders. Remote control of smart meters: Friend or foe? In *Computer Network Defense (EC2ND), 2011 Seventh European Conference on*, pages 49–56, 2011.

[66] A. Cui, M. Costello, and S. J. Stolfo. When firmware modifications attack: A case study of embedded exploitation. In *NDSS*, 2013.

[67] A. Cui and S. J. Stolfo. Defending legacy embedded systems with software symbiotes. In *The 14th International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2011.

[68] N. Dahal, R. L. King, and V. Madani. Online dimension reduction of synchrophasor data. In *Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES*, pages 1–7. IEEE, 2012.

[69] D. Davidson, B. Moench, S. Jha, and T. Ristenpart. FIE on firmware: Finding vulnerabilities in embedded systems using symbolic execution. In *Proceedings of the USENIX Security Symposium*, Washington, DC, August 2013.

[70] M. Davis. Smartgrid device security: Adventures in a new medium. http://www.blackhat.com/presentations/bh-usa-09/MDAVIS/BHUSA09-Davis-AMI-SLIDES.pdf, 2009.

[71] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris. A High-Throughput Path Metric for Multi-Hop Wireless Routing. In *MobiCom: Proc. of the ACM Int. Conference on Mobile Computing and Networking*, 2003.

[72] J. De La Ree, V. Centeno, J. S. Thorp, and A. Phadke. Synchronized phasor measurement applications in power systems. *Smart Grid, IEEE Transactions on*, 1(1):20–27, 2010.

[73] D. De Silva, X. Yu, D. Alahakoon, and G. Holmes. Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters. In *Electrical Machines and Systems (ICEMS), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[74] D. Dean, M. Franklin, and A. Stubblefield. An algebraic approach to IP traceback. *ACM Trans. Inf. Syst. Secur.*, 5(2):119–137, 2002.

[75] H. Debar, M. Dacier, and A. Wespi. Towards a taxonomy of intrusion-detection systems. *Computer Networks*, 31(8):805–822, Apr. 1999.

[76] G. Delugré. Closer to metal: Reverse engineering the broadcom netextreme's firmware. HACK.LU 2010.

[77] E. Directive. 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the EC*, 23:6, 1995.

[78] P. Domingos and G. Hulten. Mining high-speed data streams. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 71–80. ACM, 2000.

[79] P. Dubey. A platform 2015 workload model: Recognition, mining and synthesis moves computers to the era of tera. Technical report, Intel Corporation, 2005.

[80] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 238 –243, Oct. 2010.

[81] E. Ekici, I. Akyildiz, and M. Bender. A multicast routing algorithm for LEO satellite IP networks. *Networking, IEEE/ACM Transactions on*, 10(2):183–192, 2002.

[82] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec. The many faces of publish/subscribe. *ACM Comput. Surv.*, 35(2):114–131, June 2003.

[83] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez. Securing advanced metering infrastructure using intrusion detection system with data stream mining. In *Intelligence and Security Informatics*, pages 96–111. Springer, 2012.

[84] N. Falliere, L. O. Murchu, and E. Chien. W32. stuxnet dossier. Technical report, Symantec Corporation, 2011.

[85] N. Falliere, L. O. Murchu, and E. Chien. W32.Stuxnet Dossier, 2011.

[86] X. Fang, S. Misra, G. Xue, and D. Yang. Smart grid- the new and improved power grid: A survey. *Communications Surveys Tutorials, IEEE*, 14(4):944–980, 2012.

[87] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi. In-network aggregation techniques for wireless sensor networks: a survey. *Wireless Communications, IEEE*, 14(2):70–87, 2007.

[88] H. C. Ferreira, L. Lampe, J. Newbury, and T. G. Swart. *Power line communications: Theory and applications for narrowband and broadband communications over power lines*. John Wiley and Sons, 2011.

[89] R. Fonseca, O. Gnawali, K. Jamieson, and P. Levis. Four Bit Wireless Link Estimation. In *HotNets: Proc. of the Workshop on Hot Topics in Networks*, 2007.

[90] B. Franken, V. Ajodhia, K. Petrov, K. Keller, and C. Müller. Regulation of Voltage Quality. In *9th International Conference "Electric Power, Quality and Utilisation", Barcelona*, 2007.

[91] A. Freitas. A survey of parallel data mining. In *Proc 2nd Int Conf on the Practical Applications of Knowledge Discovery and Data Mining*, pages 287–300. The Practical Application Company, 1998.

[92] Z. Fu and M. Papatriantafilou. Off the wall: Lightweight distributed filtering to mitigate distributed denial of service attacks. In *Reliable Distributed Systems (SRDS), 2012 IEEE 31st Symposium on*, pages 207–212, 2012.

[93] S. Galli, A. Scaglione, and Z. Wang. Power line communications and the smart grid. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 303–308, 2010.

[94] J. Gama and P. P. Rodrigues. Stream-based electricity load forecast. In *Knowledge Discovery in Databases: PKDD 2007*, pages 446–453. Springer, 2007.

[95] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. P. Chen. A survey of communication/networking in smart grids. *Future Generation Computer Systems*, 28(2):391 – 404, 2012.

[96] S. J. Genuis. Fielding a current idea: exploring the public health impact of electromagnetic radiation. *Public Health*, 122(2):113–124, 2008.

[97] A. Ghassemi, S. Bavarian, and L. Lampe. Cognitive radio for smart grid communications. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 297–302, 2010.

[98] A. Gidenstam, M. Papatriantafilou, and P. Tsigas. NBmalloc: Allocating Memory in a Lock-Free Manner. *Algorithmica*, 2009.

[99] J. Giménez and L. Marquez. Svmtool: A general pos tagger generator based on support vector machines. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*. Citeseer, 2004.

[100] J. Giri, D. Sun, and R. Avila-Rosales. Wanted: A more intelligent grid. *Power and Energy Magazine, IEEE*, 7(2):34–40, 2009.

[101] O. Gnawali, R. Fonseca, K. Jamieson, D. Moss, and P. Levis. Collection Tree Protocol. In *SenSys: Proc. of the ACM Int. Conference on Embedded Networked Sensor Systems*, 2009.

[102] P. Godefroid, M. Y. Levin, and D. Molnar. Automated whitebox fuzz testing. In *Network Distributed Security Symposium (NDSS)*. Internet Society, 2008.

[103] P. Godefroid, M. Y. Levin, and D. Molnar. SAGE: whitebox fuzzing for security testing. *Communications of The ACM*, pages 40–44, 2012.

[104] A. Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005.

[105] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, W. H. Sanders, A. A. Cárdenas, and J. G. Jetcheva. AMI threats, intrusion detection requirements and deployment recommendations. In *Proceedings of the 3rd IEEE International Conference on Smart Grid Communications (SmartGridComm), Tainan City, Taiwan*, 2012.

[106] V. Gulisano, R. Jimenez-Peris, M. Patino-Martinez, C. Soriente, and P. Valduriez. Streamcloud: An elastic and scalable data streaming system. 2012.

[107] V. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. Hancke. Smart grid technologies: Communication technologies and standards. *Industrial Informatics, IEEE Transactions on*, 7(4):529–539, 2011.

[108] V. C. Gungor and F. C. Lambert. A survey on communication networks for electric system automation. *Comput. Netw.*, 50(7):877–897, May 2006.

[109] J. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter, and B. Heile. IEEE 802.15.4: a developing standard for low-power low-cost wireless personal area networks. *Network, IEEE*, 15(5):12–19, 2001.

[110] N. Hansen. The CMA evolution strategy: A tutorial. *Vu le*, 29, 2005.

[111] S. Haykin. Cognitive radio: brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2):201–220, 2005.

[112] J. M. Hellerstein, W. Hong, S. Madden, and K. Stanek. Beyond average: Toward sophisticated sensing with queries. In *Information Processing in Sensor Networks*, pages 63–79. Springer, 2003.

[113] H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural networks for short-term load forecasting: A review and evaluation. *Power Systems, IEEE Transactions on*, 16(1):44–55, 2001.

[114] C. Hochgraf, R. Tripathi, and S. Herzberg. Smart grid charger for electric vehicles using existing cellular networks and SMS text messages. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 167–172, 2010.

[115] D. Hu and V. Venkatasubramanian. New wide-area algorithms for detection and mitigation of angle instability using synchrophasors. In *Power Engineering Society General Meeting, 2007. IEEE*, pages 1–8. IEEE, 2007.

[116] Y. Hu and V.-K. Li. Satellite-based internet: a tutorial. *Communications Magazine, IEEE*, 39(3):154–162, 2001.

[117] Y.-C. Hu, A. Perrig, and M. Sirbu. SPV: secure path vector routing for securing BGP. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 179–192, New York, NY, USA, 2004. ACM.

[118] G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 97–106. ACM, 2001.

[119] K. Hung, W. K. Lee, V.-K. Li, K. Lui, P. W. T. Pong, K. K. Y. Wong, G. H. Yang, and J. Zhong. On wireless sensors communication for overhead transmission line monitoring in power delivery systems. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 309–314, 2010.

[120] N. Hunt, P. Sandhu, and L. Ceze. Characterizing the performance and energy efficiency of lock-free data structures. In *Interaction between Compilers and Computer Architectures (INTERACT), 2011 15th Workshop on*, pages 63 –70, feb. 2011.

[121] IEEE Standards Association. IEEE 2030 draft guide for smart grid interoperability of energy technology and information technology operation with the electric power system (EPS), and end-use applications and loads http://grouper.ieee.org/groups/scc21/2030/2030_index.html, 2011.

[122] C. Imray, A. Wright, A. Subudhi, and R. Roach. Acute mountain sickness: pathophysiology, prevention, and treatment. *Progress in cardiovascular diseases*, 52(6):467–484, 2010.

[123] Intel. Threading Building Blocks, 2009.

[124] International Electrotechnical Commission. IEC 61850: Communication networks and systems in substations, http://www.iec.ch/smartgrid/standards/, accessed in 2013.

[125] M. S. Islam, Y. J. Yoon, M. A. Hamid, and C. S. Hong. A secure hybrid wireless mesh protocol for 802.11s mesh network. In *Proceedings of the International Conference on Computational Science and Its Applications, Part I*, ICCSA '08, pages 972–985, Berlin, Heidelberg, 2008. Springer-Verlag.

[126] M. Jawurek, M. Johns, and K. Rieck. Smart metering de-pseudonymization. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 227–236. ACM, 2011.

[127] S. Jie, Z. Peng, and Q. Bing. Applications of McWiLL broadband multimedia trunk communication technology in Smart Grid. In *Optics Photonics and Energy Engineering (OPEE), 2010 International Conference on*, volume 1, pages 229–232, 2010.

[128] G. Kalogridis, R. Cepeda, S. Denic, T. Lewis, and C. Efthymiou. ElecPrivacy: Evaluating the Privacy Protection of Electricity Management Algorithms. *Smart Grid, IEEE Transactions on*, 2(4):750 –758, Dec. 2011.

[129] G. Kalogridis, C. Efthymiou, S. Denic, T. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 232 –237, Oct. 2010.

[130] M. G. Kang, S. McCamant, P. Poosankam, and D. Song. DTA++: Dynamic taint analysis with targeted control-flow propagation. In *Proceedings of the 18th Annual Network and Distributed System Security Symposium*, San Diego, CA, Feb. 2011.

[131] M. G. Kang, P. Poosankam, and H. Yin. Renovo: a hidden code extractor for packed executables. In *Proceedings of the 2007 ACM workshop on Recurring malcode*, WORM '07, pages 46–53, New York, NY, USA, 2007. ACM.

[132] C. J. Karr, C. B. Rudra, K. A. Miller, T. R. Gould, T. Larson, S. Sathyanarayana, and J. Q. Koenig. Infant exposure to fine particulate matter and traffic and risk of hospitalization for rsv bronchiolitis in a region with lower ambient air pollution. *Environmental research*, 109(3):321–327, 2009.

[133] R. Kemmerer and G. Vigna. Intrusion detection: a brief history and overview. *Computer*, 35(4):27–30, 2002.

[134] S. Kent, C. Lynn, J. Mikkelson, and K. Seo. Secure border gateway protocol (S-BGP. *IEEE Journal on Selected Areas in Communications*, 18:103–116, 2000.

[135] A. D. Keromytis, V. Misra, and D. Rubenstein. SOS: secure overlay services. *SIGCOMM Comput. Commun. Rev.*, 32(4):61–72, 2002.

[136] W. H. Kersting. *Distribution System Modeling and Analysis*. CRC Press, 2002.

[137] Y.-J. Kim, M. Thottan, V. Kolesnikov, and W. Lee. A secure decentralized data-centric information infrastructure for smart grid. *Communications Magazine, IEEE*, 48(11):58–65, 2010.

[138] O. Kramer, B. Satzger, and J. Lässig. Power prediction in smart grids with evolutionary local kernel regression. In *Hybrid Artificial Intelligence Systems*, pages 262–269. Springer, 2010.

[139] KrebsonSecurity. FBI: Smart Meter Hacks Likely to Spread. http://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/, April 2012. [last downloaded October 2012].

[140] K. Kursawe, G. Danezis, and M. Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *Privacy Enhancing Technologies*, pages 175–191. Springer, 2011.

[141] N. Kush, E. Foo, E. Ahmed, I. Ahmed, and A. Clark. Gap analysis of intrusion detection in smart grids. In *Proceedings of the 2nd International Cyber Resilience Conference*, pages 38–46. Secau-Security Research Centre, 2011.

[142] V. Kuznetsov, V. Chipounov, and G. Candea. Testing closed-source binary device drivers with ddt. In *Proceedings of the 2010 USENIX conference on USENIX annual technical conference*, USENIXATC'10, pages 12–12, Berkeley, CA, USA, 2010. USENIX Association.

[143] O. Landsiedel, E. Ghadimi, S. Duquennoy, and M. Johansson. Low power, low delay: Opportunistic routing meets duty cycling. In *IPSN'12: Proceedings of the 11th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2012.

[144] D. Lea. The Java Concurrency Package (JSR-166), 2009.

[145] M. LeMay, G. Gross, C. A. Gunter, and S. Garg. Unified architecture for large-scale attested metering. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, HICSS '07, pages 115–115, Washington, DC, USA, 2007. IEEE Computer Society.

[146] F. Li, B. Luo, and P. Liu. Secure information aggregation for smart grids using homomorphic encryption. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 327–332, 2010.

[147] H. Li, R. Mao, L. Lai, and R. Qiu. Compressed meter reading for delay-sensitive and secure load report in smart grid. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 114–119, 2010.

[148] L. Liu, E. Li, Y. Zhang, and Z. Tang. Optimization of frequent itemset mining on multiple-core processor. In *Proceedings of the 33rd international conference on Very large data bases*, pages 1275–1285. VLDB Endowment, 2007.

[149] Y. Liu, P. Ning, H. Dai, and A. Liu. Randomized differential DSSS: Jamming-resistant wireless broadcast communication. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, 2010.

[150] Y. Liu, P. Ning, and M. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM conference on Computer and communications security, Chicago, Illinois*, 2009.

[151] P. Ljungberg. Long term evolution for control system applications in a smart grid context, HYCON2 workshop on energy 2012, http://hycon2-ad2-wks.sciencesconf.org/conference/hycon2-ad2-wks/pages/LTE_FOR_UTILITIES_Hycon2_September2012_A1.pdf, 2012.

[152] S. Loesing, M. Hentschel, T. Kraska, and D. Kossmann. Stormy: an elastic and highly available streaming service in the cloud. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12. ACM, 2012.

[153] Z. Lu, X. Lu, W. Wang, and C. Wang. Review and evaluation of security threats on the communication networks in the smart grid. In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, pages 1830–1835, 2010.

[154] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TAG: A tiny aggregation service for ad-hoc sensor networks. *ACM SIGOPS Operating Systems Review*, 36(SI):131–146, 2002.

[155] M. Manzo, T. Roosta, and S. Sastry. Time synchronization attacks in sensor networks. In *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks*, SASN '05, pages 107–116, New York, NY, USA, 2005. ACM.

[156] H. Markiewicz and A. Klajn. Standard EN 50160 - voltage characteristics in public distribution systems, 2004.

[157] F. Mármol, C. Sorge, O. Ugus, and G. Pérez. Do not snoop my habits: preserving privacy in the smart grid. *Communications Magazine, IEEE*, 50(5):166–172, May 2012.

[158] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and P. McDaniel. Multi-vendor penetration testing in the advanced metering infrastructure. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 107–116. ACM, 2010.

[159] Microsoft. Parallel Computing Developer Center, 2009.

[160] B. Miller and D. Rowe. A survey of SCADA and critical infrastructure incidents. In *Annual conference on Research In Information Technology*. ACM Request Permissions, Oct. 2012.

[161] J. Mirkovic and P. Reiher. A taxonomy of DDoS attack and DDoS defense mechanisms. *SIGCOMM Comput. Commun. Rev.*, 34(2):39–53, 2004.

[162] R. Mitchell and I.-R. Chen. Behavior-rule based intrusion detection systems for safety critical smart grid applications, 2013.

[163] N. Mithulananthan, M. M. A. Salama, C. A. Canizares, and J. Reeve. Distribution system voltage regulation and var compensation for different static load models. In *International Journal of Electrical Engineering, vo1.37, no. 4, pp.384-395*, 2000.

[164] A. B. I. M. Mitzenmacher. Network applications of bloom filters: A survey. In *Internet Mathematics*, pages 636–646, 2002.

[165] Y. Mo, T.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[166] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*, pages 61–66. ACM, 2010.

[167] M. Morabito, A. Crisci, M. Moriondo, F. Profili, P. Francesconi, G. Trombi, M. Bindi, G. F. Gensini, and S. Orlandini. Air temperature-related human health outcomes: Current impact and estimations of future risks in Central Italy. *Science of the Total Environment*, 441:28–40, 2012.

[168] C. Mulliner, N. Golde, and J.-P. Seifert. SMS of Death: From Analyzing to Attacking Mobile Phones on a Large Scale. In *Proceedings of the 20th USENIX Security Symposium*, San Francisco, CA, USA, August 2011.

[169] A. Muzet. Environmental noise, sleep and health. *Sleep Medicine Reviews*, 11(2):135–142, 2007.

[170] L. Nakayama Wong, H. Aung, M. Lamé, T. Wegesser, and D. Wilson. Fine particulate matter from urban ambient and wildfire sources from California's San Joaquin Valley initiate differential inflammatory, oxidative stress, and xenobiotic responses in human bronchial epithelial cells. *Toxicology in Vitro*, 25(8):1895–1905, 2011.

[171] S. Nath, P. B. Gibbons, S. Seshan, and Z. R. Anderson. Synopsis diffusion for robust aggregation in sensor networks. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 250–262. ACM, 2004.

[172] National Institute of Standards and Technology. NIST framework and roadmap for smart grid interoperability standards, release 2.0 http://www.nist.gov/smartgrid/upload/NIST_Framework_Release_2-0_corr.pdf, 2012.

[173] S. Nhleko, M. Williams, and A. Blakeborough. Vibration perception and comfort levels for an audience occupying a grandstand with perceivable motion. In *27th International Modal Analysis Conference (IMAC XXVII)*, 2009.

[174] L. O. Nicolas Falliere. W32.Stuxnet Dossier, 2011.

[175] R. F. Nuqui, A. G. Phadke, R. Schulz, and N. Bhatt. Fast on-line voltage security monitoring using synchronized phasor measurements and decision trees. In *Power Engineering Society Winter Meeting, 2001. IEEE*, volume 3, pages 1347–1352. IEEE, 2001.

[176] Observ'ER. Worldwide energy production from renewable energy sources, Stats and Figures Series, 2010. http://www.energies-renouvelables.org/observer/html/inventaire/pdf/12e-inventaire-Chap01-Eng.pdf.

[177] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Proceedings of the 17th international conference on Theory and application of cryptographic techniques*, EUROCRYPT'99, pages 223–238, Berlin, Heidelberg, 1999. Springer-Verlag.

[178] M. Paolini. Empowering the smart grid with WiWAX, http://www.wimaxforum.org/sites/wimaxforum.org/files/document_library/SenzaFili_SmartGrid.pdf, 2010.

[179] B. Parno, A. Perrig, and D. Andersen. SNAPP: Stateless Network-Authenticated Path Pinning. In *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, Tokyo, Japan, mar 2008. IEEE Computer Society.

[180] S. Parthasarathy and D. Kundur. Bloom filter based intrusion detection for smart grid SCADA. In *Electrical Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on*, pages 1–6, 2012.

[181] T. Peng, C. Leckie, and K. Ramamohanarao. Survey of network-based defense mechanisms countering the DoS and DDoS problems. *ACM Computing Survey*, 39(1):3, 2007.

[182] Y.-A. Perez and L. Duflot. Can you still trust your network card? CanSecWest 2010.

[183] A. Perrig. The BiBa one-time signature and broadcast authentication protocol. In *Proceedings of the 8th ACM conference on Computer and Communications Security*, CCS '01, pages 28–37, New York, NY, USA, 2001. ACM.

[184] A. Perrig, R. Canetti, D. Song, and J. Tygar. Efficient and secure source authentication for multicast. In *Network and Distributed System Security Symposium, NDSS*, volume 1, pages 35–46, 2001.

[185] H. L. Z. C. X. Qin, C. Li, and H. Tan. Secure routing in wired networks and wireless ad hoc networks, http://www.cse.msstate.edu/~ramkumar/N3-Pilate.pdf, 2004.

[186] R. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, and G. Zheng. Cognitive radio network for the smart grid: Experimental system architecture, control algorithms, security, and microgrid testbed. *Smart Grid, IEEE Transactions on*, 2(4):724–740, 2011.

[187] M. Raciti and S. Nadjm-Tehrani. Embedded Cyber-Physical Anomaly Detection in Smart Meters. 2012.

[188] R. Raina, A. Madhavan, and A. Y. Ng. Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 873–880, New York, NY, USA, 2009. ACM.

[189] A. Rial and G. Danezis. Privacy-preserving smart metering, http://research.microsoft.com/apps/pubs/?id=141726, 2010.

[190] S. A. Rice. Human health risk assessment of CO2: Survivors of acute high-level exposure and populations sensitive to prolonged low-level exposure. *environments*, 3(5):7–15, 2004.

[191] P. P. Rodrigues and J. Gama. A system for analysis and prediction of electricity-load streams. *Intelligent Data Analysis*, 13(3):477–496, 2009.

[192] P. P. Rodrigues, J. Gama, and J. P. Pedroso. Hierarchical clustering of time-series data streams. *Knowledge and Data Engineering, IEEE Transactions on*, 20(5):615–627, 2008.

[193] S. Rohjans, M. Uslar, R. Bleiker, J. Gonzalez, M. Specht, T. Suding, and T. Weidelt. Survey of smart grid standardization studies and recommendations. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 583–588, 2010.

[194] S. Rusitschka, K. Eger, and C. Gerdes. Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 483–488. IEEE, 2010.

[195] K. Samarakoon and J. Ekanayake. Demand side primary frequency response support through smart meter control. In *Proceedings of the 44th International Universities Power Engineering Conference (UPEC)*, pages 1–5, Sept. 2009.

[196] S. Savage, D. Wetherall, A. Karlin, and T. Anderson. Practical network support for IP traceback. *SIGCOMM Comput. Commun. Rev.*, 30(4):295–306, 2000.

[197] B. Schlich. Model checking of software for microcontrollers. *ACM Trans. Embed. Comput. Syst.*, 9(4):36:1–36:27, Apr. 2010.

[198] S. Schneider, H. Andrade, B. Gedik, A. Biem, and K. Wu. Elastic scaling of data parallel operators in stream processing. In *Proceedings of the 2009 IEEE International Symposium on Parallel&Distributed Processing*, IPDPS '09. IEEE Computer Society, 2009.

[199] E. J. Schwartz, T. Avgerinos, and D. Brumley. All you ever wanted to know about dynamic taint analysis and forward symbolic execution (but might have been afraid to ask). In *Proceedings of the 2010 IEEE Symposium on Security and Privacy*, SP '10, pages 317–331, Washington, DC, USA, 2010. IEEE Computer Society.

[200] SecureState. World's first smart meter hacking framework released. http://www.securestate.com/Insights/Pages/press-release-securestate-smart-frame-released.aspx/, 2012.

[201] D. Seo, H. Lee, and A. Perrig. Secure and efficient capability-based power management in the smart grid. In *Parallel and Distributed Processing with Applications Workshops (ISPAW), 2011 Ninth IEEE International Symposium on*, pages 119–126, 2011.

[202] S. Sharma and S. K. Jena. A survey on secure hierarchical routing protocols in wireless sensor networks. In *Proceedings of the 2011 International Conference on Communication, Computing & Security*, ICCCS '11, pages 146–151, New York, NY, USA, 2011. ACM.

[203] F. Siddiqui, S. Zeadally, C. Alcaraz, and S. Galvao. Smart grid privacy: Issues and solutions. In *Computer Communications and Networks (ICCCN), 2012 21st International Conference on*, pages 1 –5, Aug. 2012.

[204] Y. Simmhan, B. Cao, M. Giakkoupis, and V. K. Prasanna. Adaptive rate stream processing for smart grid applications on clouds. In *Proceedings of the 2nd international workshop on Scientific cloud computing*, pages 33–38. ACM, 2011.

[205] Y. Simmhan, M. Giakkoupis, B. Cao, and V. Prasanna. On using cloud platforms in a software architecture for smart energy grids. In *IEEE International Conference on Cloud Computing (CloudCom)*. Citeseer, 2010.

[206] Y. Simmhan, A. G. Kumbhare, B. Cao, and V. Prasanna. An analysis of security and privacy issues in smart grid software architectures on clouds. In *Cloud Computing (CLOUD), 2011 IEEE International Conference on*, pages 582–589. IEEE, 2011.

[207] F. Skopik, Z. Ma, T. Bleier, and H. Gruneis. A Survey on Threats and Vulnerabilities in Smart Metering Infrastructures. *International Journal of Smart Grid and Clean Energy*, 2012.

[208] SmartGrids – European Technology Platform, June 2011.

[209] D. X. Song and A. Perrig. Advanced and authenticated marking schemes for IP traceback. In *IEEE INFOCOM 2001.*, volume 2, pages 878–886 vol.2, 2001.

[210] M. Souryal, C. Gentile, D. Griffith, D. Cypher, and N. Golmie. A methodology to evaluate wireless technologies for the smart grid. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 356–361, 2010.

[211] A. Sreesha, S. Somal, and I.-T. Lu. Cognitive radio based wireless sensor network architecture for smart grid utility. In *Systems, Applications and Technology Conference (LISAT), 2011 IEEE Long Island*, pages 1–7, 2011.

[212] K. Srinivasan, M. Jain, J. I. Choi, T. Azim, E. S. Kim, P. Levis, and B. Krishnamachari. The κ Factor: Inferring Protocol Performance using Inter-Link Reception Correlation. In *MobiCom: Proc. of the ACM Int. Conference on Mobile Computing and Networking*, 2010.

[213] K. Srinivasan, M. A. Kazandjieva, S. Agarwal, and P. Levis. The β Factor: Measuring Wireless Link Burstiness. In *SenSys: Proc. of the ACM Int. Conference on Embedded Networked Sensor Systems*, 2008.

[214] State Grid Corporation of China. SGCC framework and roadmap for strong and smart grid standards http://esci-ksp.org/?publication=sgcc-framework-and-roadmap-for-strong-smart-grid-standards, 2010.

[215] A. Stavrou and A. D. Keromytis. Countering DoS attacks with stateless multipath overlays. In *Proceedings of ACM CCS*, pages 249–259, New York, NY, USA, 2005. ACM.

[216] A. Stefanov and C.-C. Liu. Cyber-power system security in a smart grid environment. *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES*, 2012.

[217] M. Stegelmann and D. Kesdogan. GridPriv: A Smart Metering Architecture Offering k-Anonymity. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 419 –426, June 2012.

[218] M. Stonebraker, U. Çetintemel, and S. Zdonik. The 8 requirements of real-time stream processing. *ACM SIGMOD Record*, 34(4):42–47, 2005.

[219] M. Strasser, C. Pöpper, S. Capkun, and M. Cagalj. Jamming-resistant key establishment using uncoordinated frequency hopping. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP '08, pages 64–78, Washington, DC, USA, 2008. IEEE Computer Society.

[220] H. Sundell and P. Tsigas. NOBLE: A Non-Blocking Inter-Process Communication Library. In *Proceedings of the 6th Workshop on Languages, Compilers and Run-time Systems for Scalable Computers*, Lecture Notes in Computer Science. Springer Verlag, 2002.

[221] C. Szilagyi and P. Koopman. Low cost multicast authentication via validity voting in time-triggered embedded control networks. In *Proceedings of the 5th Workshop on Embedded Systems Security*, WESS '10, pages 10:1–10:10, New York, NY, USA, 2010. ACM.

[222] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani. A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications 2009*, 2009.

[223] M. Tomosada and Y. Sinohara. Virtual energy demand data: Estimating energy load and protecting consumers' privacy. In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1 –8, Jan. 2011.

[224] A. Triulzi. A SSH server in your NIC. PacSec 2008.

[225] P. Tsigas and Y. Zhang. Evaluating the Performance of Non-Blocking Synchronization on Shared-Memory Multiprocessors. *ACM SIGMETRICS Performance Evaluation Review*, 29(1):320–321, 2001.

[226] P. Tsigas and Y. Zhang. Integrating Non-Blocking Synchronisation in Parallel Applications: Performance Advantages and Methodologies. In *WOSP '02: Proceedings of the 3rd international workshop on Software and performance*, pages 55–67, New York, NY, USA, 2002. ACM.

[227] M. Uslar, S. Rohjans, R. Bleiker, J. Gonzalez, M. Specht, T. Suding, and T. Weidelt. Survey of smart grid standardization studies and recommendations; part 2. In *Innovative Smart Grid Technologies Conference Europe (ISGT Europe), 2010 IEEE PES*, pages 1–6, 2010.

[228] B. Vaidya, D. Makrakis, and H. Mouftah. Secure multipath routing for AMI network in Smart Grid. In *Performance Computing and Communications Conference (IPCCC), 2012 IEEE 31st International*, pages 408–415, 2012.

[229] A. Valdes and S. Cheung. Intrusion monitoring in process control systems. In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pages 1–7, 2009.

[230] J. A. Veitch. Psychological processes influencing lighting quality. *Journal of the Illuminating Engineering Society*, 30(1):124–140, 2001.

[231] Q. Wang, H. Khurana, Y. Huang, and K. Nahrstedt. Time valid one-time signature for time-critical multicast data authentication. In *INFOCOM 2009, IEEE*, pages 1233–1241, 2009.

[232] S. Wang, L. Cui, J. Que, D.-H. Choi, X. Jiang, S. Cheng, and L. Xie. A randomized response model for privacy preserving smart metering. *Smart Grid, IEEE Transactions on*, 3(3):1317 –1324, Sept. 2012.

[233] T. Wang, T. Wei, G. Gu, and W. Zou. Taintscope: A checksum-aware directed fuzzing tool for automatic software vulnerability detection. In *IEEE Symposium on Security and Privacy*, pages 497–512, 2010.

[234] W. Wang and Z. Lu. Survey cyber security in the smart grid: Survey and challenges. *Comput. Netw.*, 57(5):1344–1371, Apr. 2013.

[235] W. Wang, Y. Xu, and M. Khanna. A survey on the communication architectures in smart grid. *Computer Networks*, 55(15):3604 – 3629, 2011.

[236] X. Wang, M. Fei, and X. Li. Performance of chirp spread spectrum in wireless communication systems. In *Communication Systems, 2008. ICCS 2008. 11th IEEE Singapore International Conference on*, pages 466–469, 2008.

[237] X. Wang and P. Yi. Security framework for wireless communications in smart distribution grid. *Smart Grid, IEEE Transactions on*, 2(4):809–818, 2011.

[238] S. D. Warren and L. D. Brandeis. The right to privacy. *Harvard law review*, 4(5):193–220, 1890.

[239] D. Wei, Y. Lu, M. Jafari, P. M. Skare, and K. Rohde. Protecting Smart Grid Automation Systems Against Cyberattacks. *Smart Grid, IEEE Transactions on*, 2(4), 2011.

[240] T. Winter (Ed.), P. Thubert (Ed.), and R. A. Team. RPL: IPv6 routing protocol for low power and lossy networks. Internet Draft draft-ietf-roll-rpl-19, work in progress.

[241] P. Wolkoff and S. K. Kjærgaard. The dichotomy of relative humidity on indoor air quality. *Environment International*, 33(6):850–857, 2007.

[242] A. Woo, T. Tong, and D. Culler. Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks. In *SenSys: Proc. of the ACM Int. Conference on Embedded Networked Sensor Systems*, 2003.

[243] B. Wu, J. Chen, J. Wu, and M. Cardei. A survey of attacks and countermeasures in mobile ad hoc networks. In *Wireless Network Security*, pages 103–135. Springer, 2007.

[244] A. Yaar, A. Perrig, and D. Song. SIFF: A Stateless Internet Flow Filter to Mitigate DDoS Flooding Attacks. *IEEE Security and Privacy Symposium*, page 130, 2004.

[245] Y. Yan, Y. Qian, and H. Sharif. A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid. In *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, pages 909 –914, March 2011.

[246] Y. Yan, Y. Qian, and H. Sharif. A secure data aggregation and dispatch scheme for home area networks in smart grid. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6, Dec. 2011.

[247] Y. Yan, Y. Qian, H. Sharif, and D. Tipper. A survey on cyber security for smart grid communications. *Communications Surveys Tutorials, IEEE*, 14(4):998–1010, 2012.

[248] X. Yang, D. Wetherall, and T. Anderson. A DoS-limiting network architecture. In *Proceedings of the ACM SIGCOMM*, Aug. 2005.

[249] J. Zerbst, M. Schaefer, and I. Rinta-Jouppi. Zone principles as cyber security architecture element for smart grids. In *Innovative Smart Grid Technologies Conference Europe (ISGT Europe), 2010 IEEE PES*, pages 1–8, 2010.

[250] Y. Zhang, L. Wang, W. Sun, R. Green, and M. Alam. Artificial immune system based intrusion detection in a distributed hierarchical network architecture of smart grid. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–8, 2011.

[251] Y. Zhang, L. Wang, W. Sun, R. Green, and M. Alam. Distributed intrusion detection system in a multi-layer network architecture of smart grids. *Smart Grid, IEEE Transactions on*, 2(4):796–808, 2011.

[252] D. Zinn, Q. Hart, T. McPhillips, B. Ludascher, Y. Simmhan, M. Giakkoupis, and V. K. Prasanna. Towards reliable, performant workflows for streaming-applications on cloud platforms. In *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*, pages 235–244. IEEE, 2011.