Enhancing Security of Advanced Metering Infrastructure by Introducing Threshold Attendance Protocol

Artur Makutunowicz\textsuperscript{1} and Jerzy Konorski\textsuperscript{2}

\textsuperscript{1} VeriFone Sp. z o.o., Warsaw, Poland
\textsuperscript{2} Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

Abstract—The industry pushes towards Smart grid systems in order to resolve current limitations of the unidirectional legacy power grid infrastructure. By introducing Advanced Metering Infrastructure (AMI) as an integral part of the Smart grid solution, the utility company obtains an invaluable tool to optimize its network, lower the operational costs, and improve quality of service. Unfortunately, introducing two-way communication poses a security risk to the power grid infrastructure. In this paper the authors consider a Threshold Attendance Protocol (TAP) acting in a reverted security paradigm. Its main idea is to keep the network load at a predictable level at all times. To achieve that, TAP in AMI environment is embedded and the solution using real-life simulation parameters is validated.

Keywords—Secret sharing, Security, Smart grid.

1. Introduction

Legacy electric power grid delivers electricity to the end user in an unidirectional way. It has worked effectively for many decades, but currently demand and differentiation increases so that a more interactive method of delivering electricity is required. A Smart grid offers a solution by introducing Advanced Metering Infrastructure (AMI) as its core part, which facilitates two-way communication between the utility companies and the customer. When the customer is allowed to interact with the power systems, it poses a potential risk to power grid stability. There are many security challenges related mostly to frauds. By introducing more capabilities, including the remote load disconnect command, a power grid might face a new type of Denial of Service attacks. The main goal of the adversary is to make the power grid unstable by disconnecting load from a large number of smart meters. As a result, the grid might enter a non-optimal state or a blackout might happen. The remote disconnect messages can be sent by the central station also referred to as Metering Data Management System (MDMS). In this paper the authors assume MDMS is vulnerable to attacks and can start sending unauthorized load disconnect messages. There might be numerous reasons of such behavior, including, but not limited to, external attack, untrained operator, or a software bug.

To prevent such attacks, a protocol named Threshold Attendance Protocol (TAP) is proposed. It solves the problem of unauthorized load change provided the network elements, with a possible exception of MDMS, follow the protocol. In this paper TAP in the AMI environment is set, its implementation is described, and simulation results with extensive comments is provided. The authors believe it is a step towards making TAP a more real-life solution.

The rest of this paper is structured as follows. In the following subsection related work focusing on AMI and its real-world implementations is discussed. Section 2 outlines TAP. In Section 3 the basics of AMI and smart grid are explained. Subsequently, in Section 4 implementation details are discussed and the communication model is described. Section 5 focuses on experiments with TAP involving real-world parameters. The paper is concluded in Section 6, providing a short summary and suggesting future areas of research.

1.1. Related Work

There are multiple areas of research related to the content presented in this paper. As a general idea, Threshold Attendance Protocol has been defined in [1], but it was not embedded in AMI environment and had no real-world parameters defined. Its sole purpose was to present the idea and reason behind the solution validity. Shamir secret sharing scheme [2], around which TAP is built, has been proven to work correctly and reliably in many different applications, ranging from securing broadcast transmissions [3] to digital cash protocols [4].

AMI and Smart grid are extensively studied areas of research. Several lessons learned from the actual implementation have been presented in [5]. Smart grid was first introduced in [6]. Advanced Metering Infrastructure with the focus on the communication architecture was described in [7]. Related communication technologies and capacity planning were discussed in [8] and [9].

Smart grid and AMI security challenges were described in numerous papers, e.g., [10]–[12]. However, their focus was on the physical security within a standard security paradigm, whereby the central station can be trusted, whereas the smart meters cannot. In this paper a reverted security paradigm is assumed, whereby the central station
(MDMS) may turn rogue whereas the nodes (Data Concentrators) can be considered secure. It seems to be a reasonable assumption, as the Data Concentrators (contrary to smart meters) are located in safe and monitored places, usually owned or managed by a utility company. Maintaining a global consensus-like condition in a distributed system with adversaries, of which threshold attendance is a special case, is generally treated in [13].

2. Threshold Attendance Protocol

The main goal of the Threshold Attendance Protocol, called TAP, is to keep the predefined number (also referred to as the threshold, \( T \)) of nodes in active state at all times. TAP allows a distributed system where a central entity is a potential adversary to nevertheless keep the global condition. TAP is a message-based protocol using Shamir’s secret sharing scheme to agree on the global condition between distributed nodes in a secure way.

There exist two types of devices in TAP architecture: node and DSCC (Data Sink/Command Center). Each node controls a generic device that can be in or out of service, referred to as Enabled and Disabled states, respectively. When a node decides to disable the associated device (this event might be triggered either internally or externally) it has to ask all the remaining nodes for a permission to make sure the global attendance condition will not be violated. If the network consists of \( n \) nodes with a threshold \( T \), then at least \( T \) nodes have to endorse the original request, and guarantee to stay Enabled while the requesting node can go to Disabled state.

DSCC acts primarily as a forwarding device, i.e., the nodes can communicate via the DSCC only, and direct communication is not possible. However, the DSCC can work in two modes: it can either be well-behaved, i.e., employ optimization algorithms in order to minimize the delays and total number of messages in the network, or it can become rogue and try to jeopardize the global condition by trying to disable more than \( n-T \) nodes. Therefore, one of the main assumptions is to maintain the global condition at all times, while during normal operation the traffic flow should be optimized. When the DSCC launches an attack, the efficiency might be diminished, but the threshold attendance condition will be still met.

All the intelligence is embedded into DSCC. The nodes are simple, resource-constrained devices and are not capable of running any complicated algorithms.

TAP is built around a well-known threshold cryptography scheme – Shamir secret sharing [2]. Its main idea is to partition a secret key into several chunks and distribute them to multiple entities. In order to recover the key, a predefined number of chunks is required. The key can be reconstructed using many techniques such as Lagrange interpolation or solving a system of linear equations.

The TAP message flow is fully asynchronous. A simplified generic scenario is provided below and also presented in Fig. 1 [1]:

- A node \( R \) sends a Disable request message when it decides to go out of service. After sending the message, the node remains in-service. A confirmation from the network is required to change the state to Disabled.
- The DSCC receives the Disable request message and forwards it as Disable Endorsement Request to at least \( T \) nodes excluding the requesting node \( R \).
- When a node in Enabled state receives the Disable endorsement request, it generates a session key based on the master secret key, divides it into chunks and sends one chunk encapsulated in Disable endorsement message.
- Once DSCC collects at least \( T \) valid Disable endorsement messages it recovers the session key and sends Disable authorization message to the requesting node \( R \).
- Node \( R \) validates the Disable authorization message and transitions to Disabled state.
- Node \( R \) becomes Enabled after an external event occurs.
- Node \( R \) sends Enable Notification to the DSCC.

More detailed TAP description is provided in [1].

3. Advanced Metering Infrastructure

This section is an introduction to the concept of smart grid, AMI being its most recognizable part. The authors define the AMI architecture, touch upon its benefits and emphasize security challenges associated with two-way communication between the utility company and the customer. The smart grid can be defined as a modern electric power grid for improved resiliency, including self-healing capa-
abilities [14], security, with renewable energy sources [6]. Smart grid can be also thought as an umbrella term encompassing dozens of technologies and protocols, with extensive use of Information and Communication Technologies. At the highest level it can be divided into three main parts: Substation Automation, Phasor Measurement Units, and Advanced Metering Infrastructure, the latter being the most recognized part of the smart grid solution. As a result of introducing AMI, the utility company can benefit in several ways from the most trivial automated meter reading to enhanced demand prediction and load-balancing capabilities.

The basic communication architecture and key devices are depicted in Fig. 2. AMI consists of the following parts: central data storage server – Metering data management system (MDMS), Data concentrators (DCs), smart meters (SM), and communication network [5]. The communication system typically uses a three-tier architecture. If improved scalability is required, more fine-grained hierarchy levels may be introduced. Bidirectional communication takes place between consumer and the utility company in order to achieve the goals defined by the smart grid.

![High-level Advanced Metering Infrastructure diagram](Image)

Fig. 2. High-level Advanced Metering Infrastructure diagram.

Smart meters are devices installed at the customer’s premises and communicate directly with the Data concentrators. There are multiple communication technologies (both wired and wireless), but Power Line Communication-based ones seem to be most promising. The main benefit of PLC communication is a possibility of reusing existing power lines, thus installation costs and overall solution complexity are decreased. PLC might not be suitable for applications that require high bandwidth as the maximum raw data rate is about 20 Kb/s [8]. Data concentrator to smart meters connectivity details are outside the scope of this paper.

The main function of the DC is to act as an intelligent proxy between the SM and the MDMS. All data between SM and MDMS are exchanged via DC. Typically DC uses PLC on the smart meters side and cellular (GPRS, UMTS, LTE) technology on the WAN side. The WAN segment of the communication system is not usually owned by the utility company and is provided as a service by the specialized provider (both wireless and wired).

The MDMS acts as the central entity of the network. It allows the operators to connect, disconnect, manage, and monitor loads, detect any tamper attempts on the equipment, investigate faults etc. It might also serve as an interface to other systems (e.g. billing).

Taking into account the huge number of nodes, heterogeneous environment, scalability and reliability requirements, a protocol suite of choice is TCP/IP – it is scalable, field-tested, and known for its interoperability with existing network technologies.

There might be multiple applications using AMI, but two basic types of communication messages can be defined: data (interval data read) and maintenance (such as disconnect/connect requests, connectivity tests, firmware upgrade). Interval data read messages consume the largest share of bandwidth as they are sent most frequently (every 15–60 minutes) and by every meter [9].

Another benefit of AMI is the capability of extensive power quality monitoring. It can be achieved by sampling current and voltage waveforms, then analyzing the results in real-time. Despite the power line monitoring, smart meters can provide invaluable stream of data required to optimize energy generation, scheduling, and demand planning [15].

Security challenges arising from the introduction of two-way communication are numerous. Smart meters are located at customer’s premises, which brings about a non-negligible risk of fraud (energy theft) and can be related to the theft of information, denial of service, and manipulation of the service [10]. However, it does not pose a significant risk on the infrastructure as a whole, as long as the number of fraudulent customers is very low. A much higher risk to the critical infrastructure is an attack vector involving the adversary taking over the MDMS and sending an unauthorized load reduction request to the DCs. The MDMS might be easier to attack as it uses off-the-shelf hardware and operating system. When a significant number of smart meters is disabled, a potential catastrophic event might occur (uneconomical operation being the least severe). Thus it is crucial to ascertain that at least a predefined number of nodes are enabled at any given time.

Power demand can be predicted with a sufficient accuracy and demand is adjusted accordingly. As long as the demand follows a predictable trend (e.g., obtained by a historical observations), the grid remains balanced [16].

Unexpected change in the system behavior caused by the demand increase or decrease may have serious ramifications. Especially lowering power demand by disabling a significant number of customers is not desirable and might cause severe problems with the infrastructure, ranging from shifting the power grid system from its economically optimal state to a system-wide instability.
Smart meters have the ability of accepting remote disconnect/reconnect commands. This is helpful when used in an authorized way. However, the problem occurs when the central management station turns rogue, as a result of attack, software malfunction, or mistakes by untrained personnel and issues a huge number of load disconnect messages, causing the network to become unbalanced. By introducing TAP it is impossible to disable more meters than a predefined threshold.

4. Simulation

In this section the AMI model used during simulations is described and details about the simulator implementation are provided.

The simulation application is implemented using OMNeT++ [17] and INET (http://inet.omnetpp.org) frameworks. There are two distinct parts of the simulation: Node and MDMS (also referenced as DSCC in the TAP real AMI). Each of these parts acts as a custom UDP application in INET framework sitting on the top of Standard Host module. OpenSSL library (www.openssl.org) was used as a base for an implementation of the Shamir Secret sharing scheme.

The main goal when creating the simulator application was to accurately simulate TAP in an AMI environment. Based on the outcome of the simulation experiments, TAP’s correctness, scalability, and efficiency can be estimated. The results of the set of basic experiments are provided in the next section.

4.1. Model

The AMI network model for TAP consists of two main building blocks: network requirements and AMI parameters. The first group defines some basic assumptions about the communication environment for TAP. Based on the second provides a number of parameters required to run the simulation. The assumptions and parameters in both groups should be as close as possible to real-world implementations.

Let’s consider AMI to be a two-tier network with MDMS being the central entity and Data concentrators acting as nodes. Thus, without loss of generality, the influence of individual smart meters is suppressed in order to make the solution more scalable. With this assumption in mind, it is worth emphasizing that when DC gets a disconnect message, all smart meters associated with this DC are disconnected. Therefore MDMS can act on an area/cluster level of abstraction (DC being considered a cluster head), hiding unnecessary local details from the global view.

The basic assumptions about the AMI communication network are [1]:

- Both node and communication links are reliable. If a message is sent by MDMS it always gets to the node. The node is always operational and processes packets according to the TAP.
- MDMS is a central communication point. Regardless of the communication technology it is not possible for the nodes to communicate directly. MDMS forwards all the messages between nodes. However MDMS might delay, drop or alter the message as a result of turning rogue.
- Nodes are secure. Data concentrators (referred as a nodes) are physically secure, located in a guarded premises (e.g. power substations) and implement various anti-tamper techniques [18]. Each nodes stores a unique ID and shared key using a secure storage.
- The Threshold attendance level is constant. In real-world scenarios, it might vary throughout the network lifetime. However, in this model it is constant and known to all the network nodes. It is possible to avoid this restriction by using a protocol external to TAP, possibly based on a public-key infrastructure, but it is outside the scope of this paper.

In the proposed model, simple point-to-point links are used with Point-to-Point Protocol (the physical link type is not strictly defined). It is lossless, although this assumption can easily be relaxed. All the communication is UDP-based (TAP messages are encapsulated in UDP messages).

TAP parameters are chosen to closely reassemble real-life setups. Based on [9] following values were defined:

- Interval data read frequency. Each node sends Interval data read message every 60 minutes. In AMI networks there are two main approaches to interval data reading: sending periodic data messages (usually every 15–60 minutes) and aggregated batch transfer (once a day), the former being the preferred one.
- Remote disconnect/reconnect frequency. In AMI every node offers a remote cut-off and start-up capability. Disconnect/reconnect messages are not sent frequently (at most once a day) and typically few nodes send them every day. However in simulation a high fraction (10, 50, and 100%) of nodes send disconnect/reconnect messages is assumed.
- Message size and bandwidth. Every Interval Data Read message is 25 bytes long, which corresponds to the 400 MB per year approximation in [5]. However, in proposed model several message types contributing to the overall bandwidth, such as firmware updates tamper-detect messages, etc. are omitted as they are not relevant to the discussion.

Although the model reassembles real AMI environments quite accurately, there are still some assumptions (notably the constant $T$) that prevent the model from using it “as is” in the production AMI networks. The authors leave removing this obstacle to future work.
5. Experiments

There are two issues to be solved: solution validation and its performance/scalability. First a brief introduction to testing methodology is presented, next the experiments’ results are provided and discussed.

The simulations were launched on a regular PC with Ubuntu Linux installed. A month of the network life-time was simulated. It was a trade-off between the simulation time and statistical credibility of the results. A set of default parameters is laid out in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>30, 100</td>
</tr>
<tr>
<td>Fraction of nodes sending Disable request</td>
<td>1.0, 0.5, 0.1</td>
</tr>
<tr>
<td>Threshold attendance</td>
<td>5, 10, 15</td>
</tr>
<tr>
<td>Average interval between Data messages</td>
<td>60 minutes</td>
</tr>
<tr>
<td>Average interval between Disable request messages</td>
<td>12 hours</td>
</tr>
<tr>
<td>Average time spend in the Disable state</td>
<td>8 hours</td>
</tr>
</tbody>
</table>

Each simulation run was repeated 3 to 5 times with a different random number generator seed and the output values were averaged to create the figures.

5.1. Validation

The model and implementation is required to be validated by finding that the number of active (enabled) nodes never falls below $T$. In the first experiment two set of parameters were used: $T = 5$ in experiment depicted in Fig. 3 and $T = 15$ in experiment shown in Fig. 4. The remaining parameters were configured to the default values.

Figure 3 describes an experiment when the “natural” behavior of the system is not affected. The number of active nodes is strictly related to other parameters (e.g., frequency of sending Disable Request messages) and remains above the threshold at all times. However, when the threshold is increased from 5 to 15 (Fig. 4), one of the scenarios ($n = 30, T = 15$) changes its characteristics. Some of the nodes need to wait in order to transition to Disabled state.

Figure 3. Number of active nodes in time. Threshold attendance 5. 100% of the nodes sending Disable request messages.

Figure 4. Number of active nodes in time. Threshold attendance 15. 100% of the nodes sending Disable request messages.

Figure 5. Number of active nodes in time. Threshold attendance 5. 10% of the nodes sending Disable request messages.

Figure 6. Number of active nodes in time. Threshold attendance 15. 10% of the nodes sending Disable request messages.
as the global condition cannot be violated. This is the price to pay for the increased security as nodes cannot transition to Disabled states uncontrollably. It is worth emphasizing that both scenarios should be treated as worst-case, since the ratio of nodes requesting transition to Disable state $p = 1.0$ is nowhere near realistic environments. When the parameter $p$ is lowered to to 10% ($p = 0.1$), the active node count stays way above the threshold (cf. Figs. 5 and 6 respectively). Multiple simulations with variable threshold parameters was performed and in no scenario was the global condition violated.

5.2. Scalability and Efficiency

Protocols designed for AMI networks must be scalable as the number of nodes tends to be large and adding new nodes should not influence the whole system behavior in a significant way.

One of the experiments was designed in order to check the scalability of the solution. The total size of both Interval data read and TAP messages sent by each node was calculated and plots the average value in Figs. 7–9. Each figure shows results generated for two sets of configuration parameters: regular and aggressive. Regular uses 12 hours delay between Disable request messages and each node spends 8 hours in the Disabled state, while aggressive uses 8 hours and 6 hours, respectively. More details about the traffic share between Data/TAP messages can be found in Table 2.

In each scenario, the total size of Interval data read messages is almost constant and does not depend on TAP parameters (these messages are sent using regular intervals). However, TAP messages, required to guarantee the global condition consume a significant amount of bandwidth in comparison to data messages. Based on these results, it is easy to conclude that the total size of TAP messages grows in a linear fashion and depends on fraction of nodes taking part in the message exchange, and Disable request messages sending frequency (note that each message triggers a large amount of Disable endorsement request messages). In a real-life scenario the traffic shares (Table 2) might be different if the Interval data read messages are sent more frequently, as the typical interval between such messages is 15 minutes [19].

Obviously, the scalability of the solution is not satisfactory, but several factors might be changed to improve it. Currently, MDMS uses a trivial algorithm of sending Disable endorsement request messages as these messages are
sent to all nodes (including those in the Disabled state). MDMS might be modified to send to active nodes only. In order to achieve it, state-tracking capabilities should be introduced in MDMS. TAP allows many such improvements if they depend on MDMS implementation only. In order to improve the scalability on the architectural level, some well-known techniques as clustering or hierarchization might be introduced. These approaches combined with an aggregation at the cluster head level might be quite effective in enhancing solution scalability. The authors are going to address this issue in future work.

6. Summary and Future Work

In this paper an introduction to Advanced Metering Infrastructure is presented, focusing on the communication and security challenges. Threshold Attendance Protocol was described as a way to solve security issues in a reverted security paradigm. The authors also described proposed implementation of TAP and discussed a number of experiments with real-life scenarios.

The main contribution of this paper is taking TAP one step towards real-life scenarios. TAP in the AMI environment was embedded and moved from a general message passing implementation to UDP-based communication system using full TCP/IP stack. The results of the simulation were analyzed in order to confirm the solution is valid and to reason about its scalability.

Several possible areas of further research can be identified. In particular, the scalability of the solution might be an issue, hence more efficient communication schemes and MDMS operation modes should be defined. Currently TAP uses UDP-based communication. However, it might be worth considering a different transport protocol (e.g. SCTP) and investigate its efficiency. In order to bring the model closer to real life, a way of changing $T$ during the network lifetime should be invented.

References


Artur Makutunowicz received his M.Sc. degree in Computer Networks from Gdansk University of Technology, Poland, in 2012. He is currently employed in VeriFone as a Senior Network Engineer. His main areas of interest are Smart Grids, Software Defined Networks, and Network Devices Architectures.

E-mail: artur@makutunowicz.net

VeriFone Sp. z o.o.
Domaniewska st 44
02-672 Warsaw, Poland
Enhancing Security of Advanced Metering Infrastructure by Introducing Threshold Attendance Protocol

Jerzy Konorski received his M.Sc. degree in Telecommunications from Gdańsk University of Technology, Poland, and his Ph.D. degree in computer science from the Polish Academy of Sciences, Warsaw, Poland. He is currently with the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, where he conducts research and teaching in computer networking, probability, optimization methods, operational research, performance evaluation, and distributed systems. He has authored or co-authored about 150 scientific papers and led several national and U.S. Government-funded projects, including “Teaching Program for telecommunications”, “Cooperation Security in Wireless Networks”, and “User Misbehavior in Distributed Computer Systems and Networks”, “Information Transfer in Wireless Networks”, and “Information and Cooperation in Self-Organizing Networks”. He was also a task leader in three other projects funded by the European Union and National Science Centre, Poland. Dr. Konorski was co-editor of IFIP PWC 2000 and WMNC 2009 proceedings, and has served on the TPC for over 40 international networking and distributed systems conferences. His current work focuses on applications of game theory in wireless networks and low-level network security architectures, with a focus on centralized and distributed reputation systems. E-mail: jekon@eti.pg.gda.pl Faculty of Electronics, Telecommunications and Informatics Gdańsk University of Technology Narutowicza st 11/12 80-952 Gdańsk, Poland