VAASAN AMMATTIKORKEAKOULU
UNIVERSITY OF APPLIED SCIENCES

Antti Lahti

X10+ – INDUSTRIAL LEVEL DATA COLLECTION AND CLOUD CONNECTIVITY SOLUTION

Technology and Communication
2019

VAASAN AMMATTIKORKEAKOULU
Tieto- ja viestintätekniikka

TIIVISTELMÄ

Tekijä	Antti Lahti
Opinnäytetyön nimi	X10+ − teollisuustason ratkaisu tiedonkeruuseen ja pilviyhteyksiin
Vuosi	2019
Kieli	englanti
Sivumäärä	48 + 1 liite
Ohjaaja	Antti Virtanen

Opinnäytetyö tehtiin ARNON Oy:lle. Työn aiheena oli suunnitella palomuuriratkaisu teollisuuden laitevalmistajille tarjottavaan tiedonkeruujärjestelmään. Tiedonkeruujärjestelmän tärkein tehtävä on siirtää kerätyt tiedot asiakkaan pilvipalveluun. Palomuurien avulla on tarkoitus ehkäistä luvaton pääsy tietoverkkoon, mutta kuitenkin sallia ulospäin suuntautuva tiedonsiirto.

Opinnäytetyössä kehitettiin asiakkaan vaatimusten perusteella mahdollisimman tietoturvallinen ratkaisu tiedonsiirtoon suljetussa verkkoympäristössä sijaitsevan tiedonkeruulaitteiston ja asiakkaan pilvipalvelun välillä. Asiakkaan moninaisten tarpeiden vuoksi projektin aikana päädyttiin käyttämään tiedostopohjaista tiedonsiirtoa reaaliaikaisen tiedonsiirron sijaan. Suurin osa siirrettävästä tiedosta oli lähtökohtaisesti kirjoitettu tiedoistoihin, joten tehtäväksi jäi enemmänkin etsiä varmatoimisin ja turvallisin tapa toteuttaa tiedostojen siirtäminen. Tiedonsiirto tapahtuu yhdensuuntaisesti ulos suljetusta verkosta SFTP-protokollaa hyödyntäen. Tunnistautuminen palvelimelle toteutettiin sertifikaatin avulla.

Projektin lopputuloksena syntyi toimiva, asiakkaan tietoturvaosaston hyväksymä järjestelmä, joka on tuotantokäytössä tätä opinnäytetyötä kirjoitettaessa.

Avainsanat palomuuri, tiedonkeruu, pilvipalvelu, tiedonsiirto

VAASAN AMMATTIKORKEAKOULU
UNIVERSITY OF APPLIED SCIENCES
Tieto- ja viestintätekniikka

ABSTRACT

Author	Antti Lahti
Title	X10+ – Industrial Level Data Collection and Cloud Connectivity Solution
Year	2019
Language	English
Pages	48 + 1 Appendix
Name of Supervisor	Antti Virtanen

This Bachelor's thesis was done for ARNON Oy. The objective of the thesis was to assist in the design and implementation of a firewall solution for an industrial data collection system, which is offered to industrial equipment manufacturers. The main objective of the data collection system is to provide the possibility to transfer the data collected by SCADA systems over the Internet while maintaining the required standards for cybersecurity and without compromising cybersecurity of the plant itself. This thesis describes the firewall configurations used to create a connection between a data collection system located in the control network of a power plant and a customer's cloud service.

This thesis describes the firewall configurations used to create a connection between a data collection system located in the control network of a power plant and a customer's cloud service. Given the complexity of the customer's needs, it was decided that file transfer was to be used instead of real-time data transfer. Most of the data that was to be transferred, was already written into files, so the main focus was to find the most reliable and safe way to transfer files. The communication is outbound only using the SFTP-protocol. Server authentication was done using a digital certificate.

The result of the thesis is a method of data transfer that has been tested and accepted by the customer's cybersecurity team for use in future projects.

Keywords Firewall, data collection, cloud service, data transfer

CONTENTS

LIST OF ABBREVIATIONS

API	Application Protocol Interface
ARP	Address Resolution Protocol
AWS	Amazon Web Services
CAN	Controller Area Network communication bus
CIA	Confidentiality, Availability, and Integrity triad model of best practices in information technology
CIDR	Classless Inter-Domain Routing
CPU	Central Processing Unit
DHCP	Dynamic Host Protocol
DMZ	Demilitarized Zone
FTP	File Transfer Protocol
HMI	Human-Machine Interface
HTTPS	Hypertext Transfer Protocol Secure
ICMP	Internet Control Message Protocol
ICS	Industrial Control System
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISO	International Organization for Standardization
JSON	JavaScript Object Notation

LAN	Local Area Network
MMI	Man-Machine Interface
MTU	Master Terminal Unit
NAT	Network Address Translation
NTP	Network Time Protocol
OEM	Original Equipment Manufacturer
OPC UA	OPC Unified Architecture is a machine to a machine communication protocol for industrial automation developed by the OPC Foundation.
PLC	Programmable Logic Controller
REST	Representational State Transfer
RFC	Request for Comments (RFC) is a type of publication from the technology community.
RTU	Remote Terminal Unit
S7	Siemens SIMATIC STEP 7 program suite
SCADA	Supervisory Control and Data Acquisition
SFTP	SSH File Transfer Protocol
SSH	Secure Shell
SSL	Secure Sockets Layer
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol

VLAN	Virtual Local Area Network
VPN	Virtual Private Network
WAN	Wide Area Network
X10+	Industrial data collection system developed by ARNON

LIST OF FIGURES AND TABLES

LIST OF APPENDICES

1 INTRODUCTION

1.1 Overview

This thesis was done for ARNON Oy, a provider of industrial electrification and automation as a service to their customers – machine, equipment, and system providers mainly in marine, mining, and renewable energy segments. ARNON Oy also offers a unique portfolio of competencies starting from sensor, instrument, and electrification level through automation systems and remote connections to cloud services /1/.

The objective of this thesis was to assist in the design and implementation of a versatile industrial data collection system while maintaining the required standards for cybersecurity and network segmentation. This system can be installed in existing production facilities without expensive stoppages to production or updates to the existing SCADA (Supervisory Control and Data Acquisition) systems.

The primary purpose of the X10+ system is to collect all the critical data from a given facility and forward the information to the equipment manufacturer or plant operator's corporate office for further analysis. This data gives a nearly real-time view of the site in question and allows for proactive maintenance and optimization of the equipment to minimize costly downtime due to avoidable equipment failures.

Industrial facilities have local SCADA systems for controlling and monitoring the processes, but these systems are often behind several firewalls, for security reason, as the equipment is critical to the operation of the facility and are rarely patched with the latest security updates. The main reason for this thesis is to study the possibility to transfer the data collected by SCADA systems over the Internet without compromising cybersecurity of the plant itself.

1.2 ARNON Oy

ARNON Oy is a privately-owned company founded in 1978. Today ARNON is the biggest automation and switchgear supplier in Northern Europe. ARNON has over 200 experts working in six different locations in three countries: Finland, Sweden,

and Poland /1/. ARNON headquarters are located in Tampere, Finland. In 2018, the company's turnover was 51M€.

2 BACKGROUND AND PURPOSE OF THE PROJECT

Original equipment manufacturers (OEMs) and site operators have an ever-rising need to collect and transfer data from SCADA systems within industrial facilities, power plants, and even ships. The need for the data originates from OEMs desire to offer their customers proactive services, such as condition-based maintenance, to allow the customers to run the equipment optimally. These improvements increase equipment availability while reducing operating expenses and downtime due to unexpected equipment failures. In the case of an owner/operator, the need is to optimize the operation of the plant using the latest possible data available.

As will be discussed later, SCADA systems are littered with cybersecurity issues, and malicious operators are looking for ways to manipulate these issues to gain access to the equipment for monetary gain or purely espionage. Transferring data off-site from facilities means that the systems need to be connected to the outside world in one way or another, which in turn, presents a new attack vector for cybercriminals.

The objective of this thesis was to create a solution for data collection and transfer without compromising cybersecurity. The transfer of data is done in real time or with a small delay depending on the quality and speed of the Internet access available to the site.

The X10+ data collectors used in this thesis were configured to a customer's specification, and therefore, the function of the collectors is not the main focus of this thesis. The customer's IT personnel completed the configuration of data servers receiving the data from the plant as well as the creation of the certificate-based authentication, and these parts of the project are outside the scope of this thesis.

3 THEORETICAL BACKGROUND

This chapter describes SCADA systems in general, cyber security challenges inherent to SCADA systems, as well as software and protocols used in the X10+ data collection system.

3.1 SCADA Systems in General

SCADA systems are used to monitor and control a plant or equipment in industries such as telecommunications, water and waste control, energy, oil and gas refining, and transportation. These systems encompass the transfer of data between a SCADA central host computer and some Remote Terminal Units (RTUs) and Programmable Logic Controllers (PLCs), and the central host and the operator terminals /2/.

SCADA systems consist of:

- One or more field data interface devices, usually RTUs, or PLCs, which interface to field sensing devices and local control switchboxes and valve actuators /2/.
- A communications system used to transfer data between field data interface devices and control units and the computers in the SCADA central host. The system can be radio, telephone, cable, satellite or any combination of these /2/.
- A central host computer server or servers (sometimes called a SCADA Center, master station, or Master Terminal Unit (MTU) /2/.
- A collection of standard and custom software (sometimes called Human Machine Interface (HMI) software or Man Machine Interface (MMI) software) systems used to provide the SCADA central host and operator terminal application, support the communications system, and monitor and control remotely located field data interface devices /2/.

3.2 SCADA System Cyber Security Challenges

Traditionally SCADA system design has focused on creating isolated networks that are reliable, fast, and stable. SCADA and control networks rely on a consistent connection between master and slave devices. Slave devices are required to answer the poll from master devices within a given period on time. Connection or communication failures often result in equipment reverting to safe mode, shutdown procedures, or equipment failures. This system of communication is by design to make the process safe for the plant operators and in the case of critical infrastructure, for the public as well. /4, 10/

There is a fundamental difference in the communication works in a traditional IT system and an ICS Industrial Control System (ICS). IT systems can be described as "best effort" in that tasks are completed when they happen to be completed. ICS systems need to be "deterministic" where everything has to happen immediately to avoid the destruction of equipment. The traditional CIA Triad Model (Confidentiality, Integrity, and Availability) of best practices used in IT systems is turned upside down in ICS systems. As extra emphasis is given to availability and message integrity, ICS systems reverse the order to AIC − Availability, Integrity, and Confidentiality. /12/

Equipment and communication protocols are often proprietary, and both enjoy a high degree of security through obscurity. Additionally, SCADA systems and the accompanied control devices are built to have a service life equal to the life cycle of the plant or operation the equipment control. Given this paradigm, it has been completely acceptable in the SCADA world to operate a large number of legacy devices, both in hardware and software terms, without updates to the working system. Typically, engineers do not enjoy fixing something that is not broken. /4, 10/

Due to the rapid growth of the Internet, there has been convergence on the TCP/IP protocol suite as the dominant network protocol for business and industry. Vendors of industrial automation products have taken note and have equipped their new products with networking capabilities brought forth by TCP/IP. A concoction of new possibilities in connectivity, fading borders between previously separate industrial and corporate networks, the use of Internet technologies accessing SCADA

networks and IT outsourcing create a massive minefield of threats and potential attack vectors on which anyone can test one's own hacking skills. /4, 10/

In 2009, security researches in Australia looked at SCADA networks of Australian critical infrastructure providers and found the following common issues. /10/

General Issues:

- Connection of SCADA to corporate networks
- Governance
- Policy
- Physical Security

IT Specific issues:

- Un-patched hardware and software
- Lack of network segregation and segmentation
- Lack of sound authentication mechanisms
- Lack of monitoring, logging, and auditing

To put the severity of the situation into perspective, the list below shows eight critical infrastructure sectors according to the Department of Home Affairs in Australia. /11/

- banking and finance
- government
- communications
- energy
- food and grocery
- health
- transport
- water

A successful cyber attack preventing the functioning of any one of the above services would most likely cause, in any first world country, severe unrest with resultant economic impacts and even loss of life directly or indirectly.

3.3 Software and Protocols

This section describes the essential software and communication protocols used in this thesis. A layout of a straightforward SCADA system is shown in **Figure 1.**

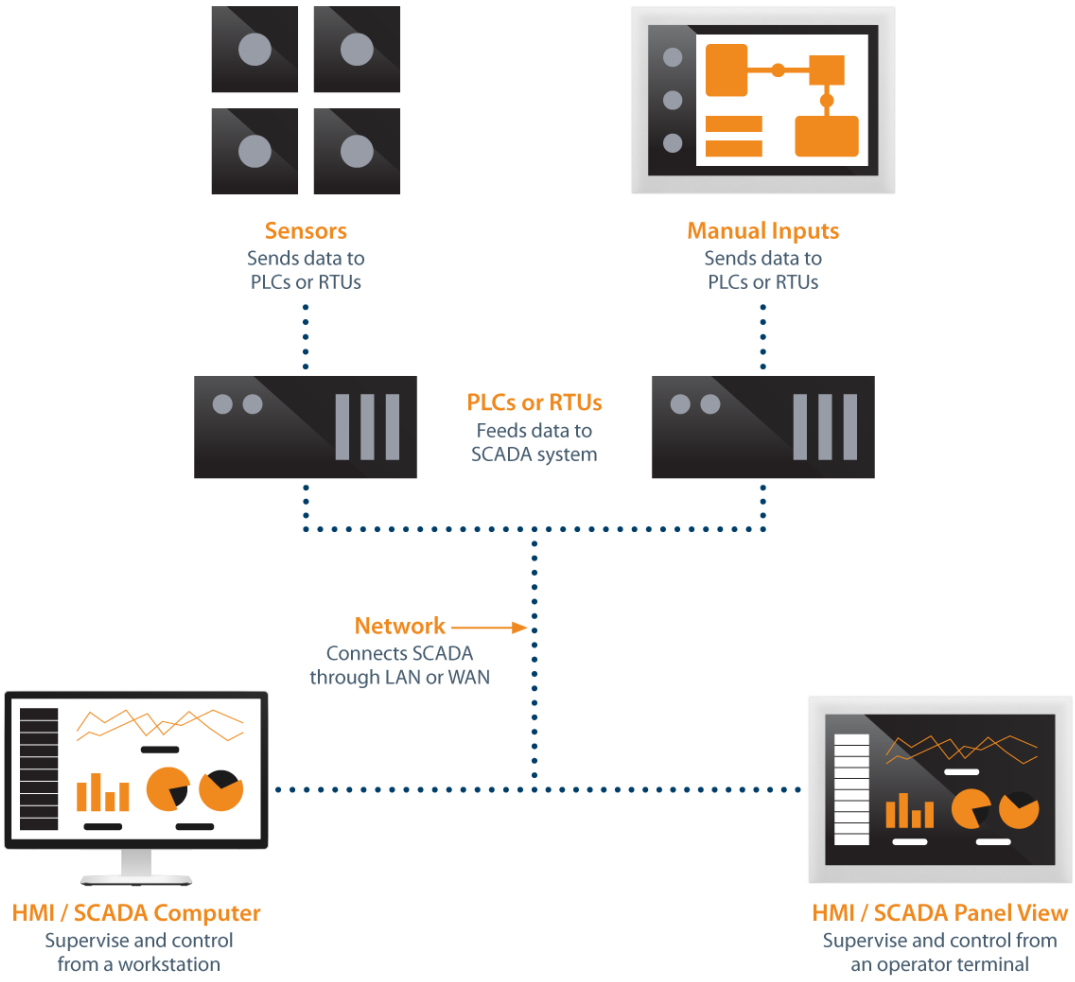

Figure 1. Simple SCADA system /3/.

3.3.1 Modbus TCP

Modbus is the de facto industrial communication standard developed in 1979. Modbus is an application-layer messaging protocol which provides client/server communication between devices connected on different types of buses or networks. Modbus is accessed at a reserved system port 502 on the TCP/IP (Transmission

Control Protocol / Internet Protocol) stack. The Modbus protocol uses a Client/Server communication technique where each Object Messaging request requires a corresponding Object Messaging response. /5/

3.3.2 Siemens S7

The Siemens S7 communication protocol allows communication Siemens PLCs (Programmable Logic Controllers) and other network-connected devices. Communication is possible through connection-oriented protocols: TCP native as per RFC 793, ISO on TCP as per RFC 1006 or connectionless protocol: UDP (User Datagram Protocol) as per RFC 768.

The S7 protocol over Industrial Ethernet should not be used for time-deterministic data flow. It is not possible to know when a remote CPU will respond to requests as the responses are asynchronous to the CPU cycle. /6-7/

3.3.3 SSH / Secure Shell

The SSH (Secure Shell) protocol is a standard method of secure login into remote computers and servers. Based on a client-server model, the connection is always established by the client, which initiates the connection by login into the server. Public key authentication and passwords are conventional methods of user authentication. A simplified setup of an SSH connection using public key authentication is shown in **Figure 2.** /9/

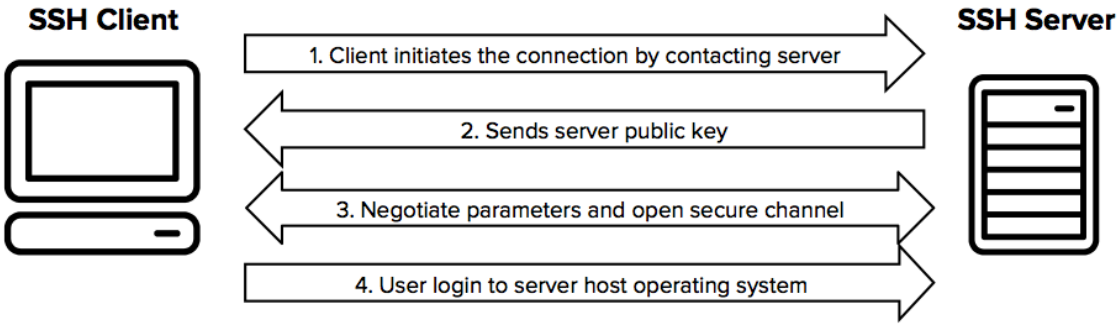

Figure 2. Simplified setup of an SSH connection /9/.

3.3.4 SFTP

SFTP (SSH File Transfer Protocol) is a client-server protocol designed for secure file transfer. As the integrity of the data is protected by using encryption and cryptographic hash functions, SFTP can be reliably used over the Internet and NAT (Network Address Translation). A significant benefit of SFTP is the ability to use public key authentication to log into the server, which allows for a fully automated file transfer process. /8/

3.3.5 HTTPS

HTTPS (Hypertext Transfer Protocol Secure) is a combination of the HTTP-protocol and the TLS/SSL-protocol (Transport Layer Security/Secure Sockets Layer). In HTTPS, information is secured by encrypting it before sending using TLS. The primary purposes of the encryption are to protect the integrity of data while it is in transit and to assure that communication with the intended source is protected against tampering and eavesdropping.

3.3.6 X10+ Data Collector Protocols

X10+ field and DMZ collectors support a multitude of communication protocols used in industrial automation, shown in **Figure 3.** Most protocols are supported directly, and some analog or serial protocols via protocol converters. The protocol converters typically convert the original signal to Modbus TCP.

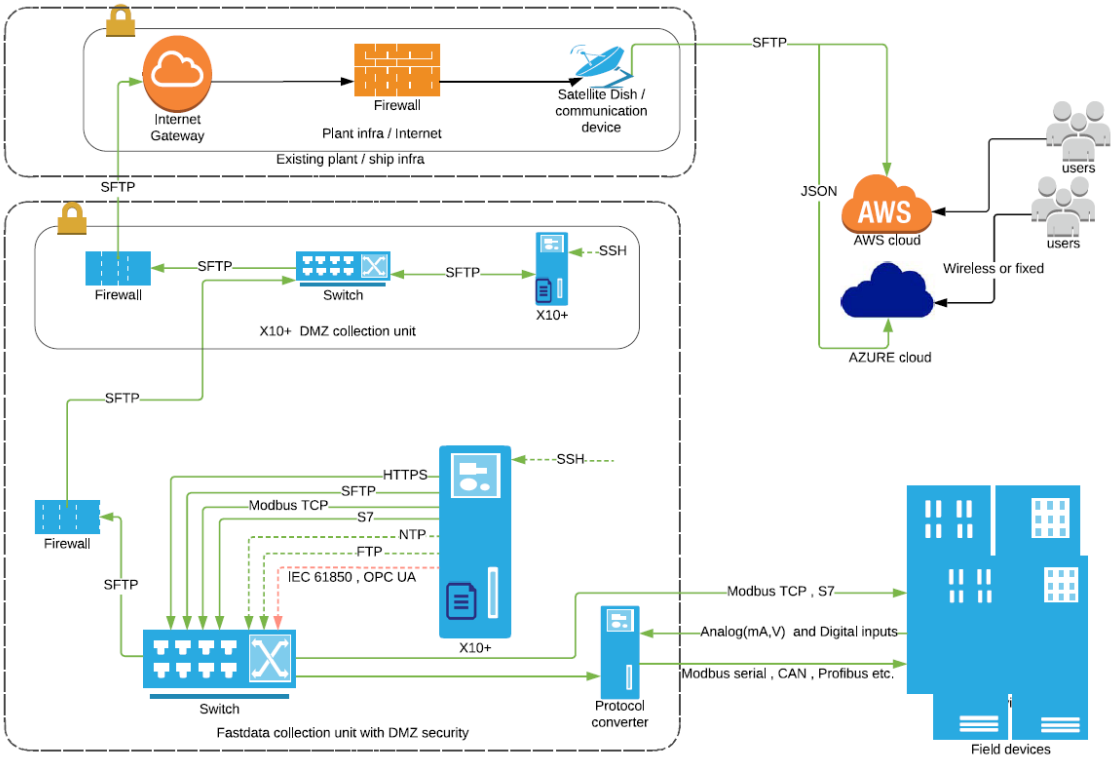

Figure 3. Overview of software protocols used in X10.

The X10+ DMZ collector supports current methods of data transfer using APIs (Application Protocol Interface). For example, data transfer to Azure is possible in JSON (JavaScript Object Notation) using a REST (Representational State Transfer) API. JSON is the most widely used data format for data interchange on the Internet.

The ability to use hostname-based routing has also been built into the system as it gives increased flexibility when routing data to storage services.

4 X10 DATA COLLECTION SYSTEM

4.1 Planning

ARNON's customer indicated that they had plans to create a new data collection system with a faster data transfer interval and cloud connectivity capability. Data was also to be transmitted in an entirely new format developed during the planning and production phase of the system. The customer's cybersecurity team and white hat hackers would oversee and test the proposal to make sure it was up to the standards set forth by the customer.

4.1.1 Requirements

It was decided that the X10+ field collectors would be placed in the control network beside the existing SCADA equipment to ensure easy compatibility with old and new production facilities. The field collectors read all available signals, create data files, collect log files from auxiliary equipment, and send the files to the DMZ collector (Demilitarized Zone). Some of the essential system functionalities, short descriptions of the functionalities, and the priority as discussed with the customer are listed in **Table 1.**

Table 1. X10+ data collection system requirements.

Functionality	Description	Priority
File format	File format specified by the customer to limit the use of bandwidth	1
Configurable data transfer interval	Data transferred in five-minute data blocks once the file is ready	1
Data collector configuration update	Possibility to update data collectors via the cloud	2
Software Patching	Possibility to push security updates to the data collectors via the cloud	3

User authentication	Automated login to the server using certificate-based user authentication	1
Price	Price to be kept reasonable, no commercial software which requires software licenses to be used	1
Data sampling	All analog and digital signals need to be read every second	1
Log file transfer	Possibility to collect log files from other equipment in the control network and send them to the cloud service for analyzes	3
Segmented network layout	Equipment in the control network to be protected by two layers of firewalls	1
No VPN connections	No VPN connections allowed	1

4.1.2 Hardware

Choosing hardware for this project was a balancing act between the customer requirements, relatively limited space in the production facility for mounting the equipment, and ARNON's product strategy. The selected hardware was expected to function in a multitude of different scenarios by having different configuration possibilities and meet the highest standards and approvals such as DNV GL.

4.1.3 Software

Earlier iterations of the X10+-data collector with all the custom-made software running on them have been based on Debian Linux. The software package has repeatedly been proved to be very stable, secure and capable of running on even very inexpensive single-board computers. Stable software packages for ARM and X86 computers are available.

4.1.4 Networks and Layout

The vast range and size of industrial Ethernet networks called for Ethernet switches and firewalls to have different product variants available for twisted-pair cables

(RJ45) and multimode fibers (SC). Twisted-pair cables are considered reliable within networks where the length of the cables is less than 100 meters. More extensive networks usually use fiber optic cables to prevent signal attenuation. Network speed usually is not an issue in most industrial networks as time-critical safety applications use a separate network. It is, however, becoming more common to use industrial safety applications over industrial Ethernet.

4.2 Implementation

4.2.1 Hardware

Limited by available space in the customer's plant, the decision was made to manufacture two 6U server racks and mount the equipment to the racks. Each network segment was built into a separate rack to limit the possibility of confusing the different network segments if cables are pulled out and reconnected in the future. Three X10+ field collectors were used for this project. The collectors are responsible for data collection of 9500+ signals every second from primary and auxiliary systems in the facility. The field collector rack is shown in **Figure 4.**

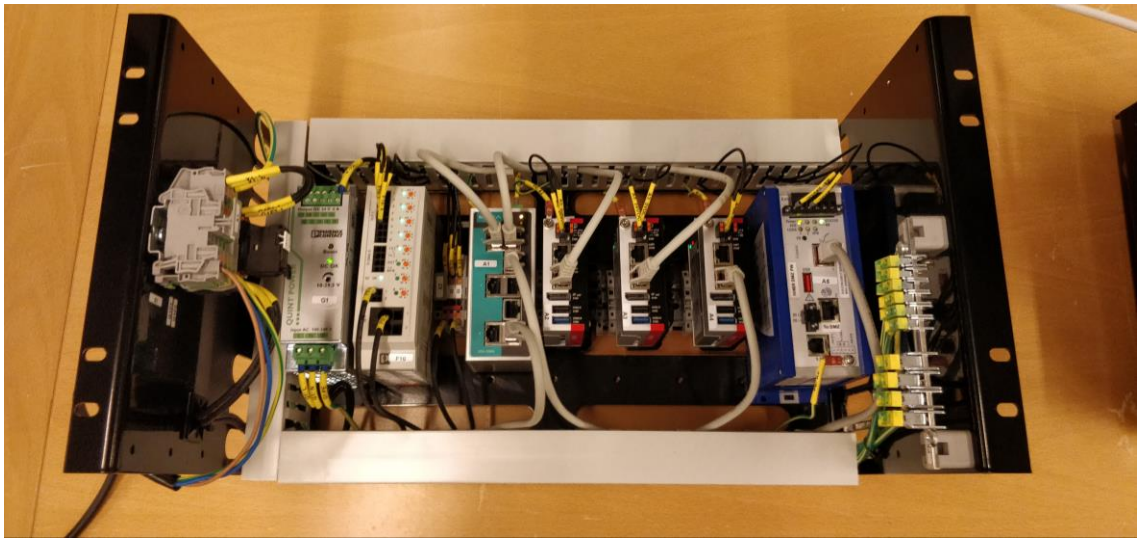

Figure 4. Field collectors mounted to the server rack.

The DMZ collector is used for relaying the data files, sent by the field collectors, to the cloud service. As only three files are sent every five minutes in addition to several log files once every 24 hours, one computer can easily handle the task. The DMZ collector is shown in **Figure 5.**

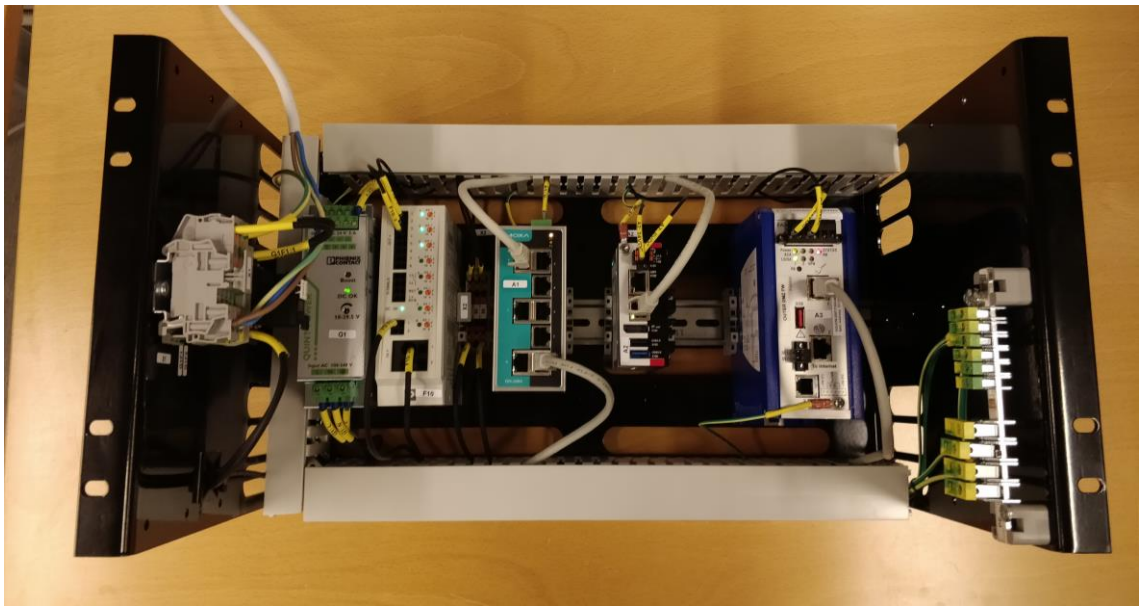

Figure 5. DMZ-collector mounted in the server rack.

4.2.2 Hirschmann Eagle ONE Firewall

The Hirschmann Eagle ONE is a DIN-rail mounted industrial security router with features, such as stateful packet inspection, VPN (Virtual Private Network) and 1:1 NAT. Other beneficial features are the redundant power supply, hardwired alarm capability, and remote activation of VPN via a digital input. Hirschmann Eagle ONE firewall is shown in **Figure 6.**

Figure 6. Hirschmann Eagle ONE Firewall.

4.2.3 Beckhoff C6150-0010 Industrial Computer

Previous experience and customer requirements made the selection of the computer one of the most critical tasks regarding the success of the project. Some of the

essential computer functionalities, short descriptions of the functionalities, and the priority as discussed with the customer are listed in **Table 2.**

Table 2. X10+ computer requirements.

Characteristic	Description	Priority
Real-time clock	PC must have a real-time clock with a changeable battery to prevent the clock from going out of sync during a power outage.	1
Storage	Enough storage space to backup 30 days of data in case of network connectivity issues.	1
Power supply	The PC power supply must be 24 VDC.	1
DIN-rail mounting	PC must have DIN-rail mounting capability for easy and flexible mounting.	1
Linux support	The computers use Debian Linux as their operating system, so this is an obvious requirement.	1
Powerful processor	The computer must be able to read and log the data in less half the time of the poll interval. If the poll interval is 1 s, the computer must complete all of the given tasks in 0.5 s or less.	1

After a careful review of the current Industrial PCs on the market, Beckhoff's ultra-compact C6150 Industrial PC series, shown in **Figure 7**, was the obvious choice for many reasons. Some of the most significant benefits were:

- Minimal DIN-rail footprint and overall size
- Linux support
- Battery-backed real-time clock
- 60GB SSD hard drive
- Intel® Atom™ CPU with four cores

Figure 7. Beckhoff C-06150 Industrial computer.

4.3 Network Configuration Using Firewalls

After several discussions with the customer, it was decided that minimal impact on the existing plant network was a high priority goal. As the plant in question was reasonably new, the field network was initially created with a sufficient amount of segmentation. Several firewalls and managed ring switches required updated rule sets to allow the X10+ field collectors access to the data.

The DMZ network segment was to be completely separate from the control network. Routing to the Internet would be done via the plant LAN-network and the plant operator's firewall. This network configuration required no changes to the plant operator's network equipment, and thus, the network's perimeter defense was left unchanged.

All IP addresses shown in **Table 3** below have been created explicitly for this thesis. Any resemblance to real life system setups is purely coincidental. The use of virtual IP addresses and NAT is explained later in this chapter. Please refer to **Figure 17** for a clarification on how the virtual addresses were used.

Table 3. IP addresses and network segments used in this thesis.

Network segment	IP Address	Device	Description
Control network	192.168.90.0/24	N/A	Control network segment
Control network	192.168.90.231/32	X10+ field collector 1	Network interface 1
Control network	192.168.90.232/32	X10+ field collector 2	Network interface 1
Control network	192.168.90.233/32	X10+ field collector 3	Network interface 1
Control network	192.168.90.229/32	Internal DMZ firewall	The outbound virtual IP address
Control network	192.168.90.230/32	Internal DMZ firewall	LAN port
DMZ	192.168.154.0/23	N/A	DMZ network segment
DMZ	192.168.155.188/32	Internal DMZ firewall	WAN port / Virtual IP address
DMZ	192.168.155.189/32	X10+ DMZ collector	Network interface 1
DMZ	192.168.155.190/32	External DMZ fire-wall	LAN port
DMZ	192.168.155.192/32	External DMZ fire-wall	The outbound virtual IP address
Customer network	0.0.0.0	N/A	
Customer network	0.0.0.100/32	External DMZ fire-wall	WAN port

Internet	34.208.4.84/32	Amazon Server	Amazon Web Services server cluster

Figure 24 in Appendix 1 shows a simplified layout drawing of the field, control, plant, and end customer network infrastructure of the project.

Figure 8 shows an overview of the firewall setup with all IP addresses. The schematic is read from left to right with the left being the control network, the middle is the DMZ, and right side of the schematic is the end-customers network with routing to AWS.

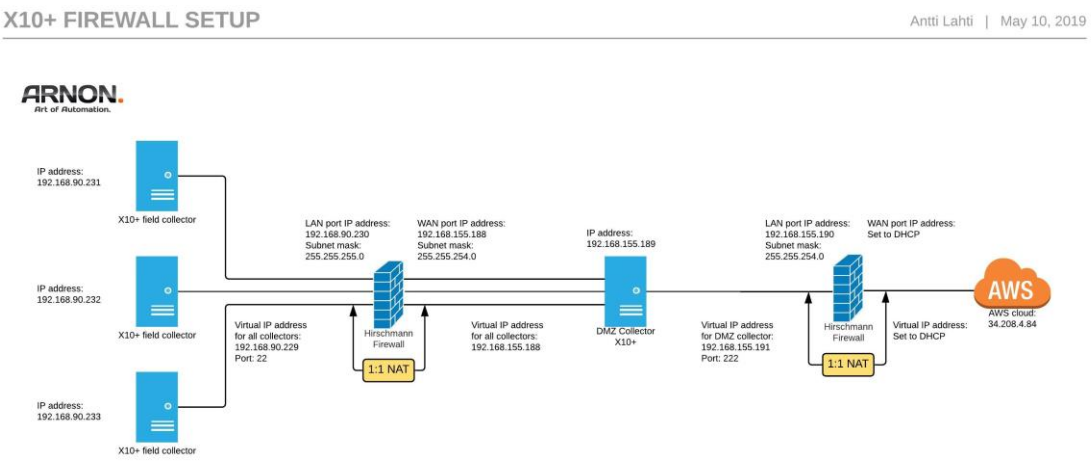

Figure 8. Firewall schematic.

4.3.1 Configuration Software

Hirschmann firewalls can be configured in several ways including command line, https-access via a web browser, using configuration files, and a purpose-built software called HiView which is an executable run locally on the computer. **Figure 9** shows several firewalls in HiView ready for configuration.

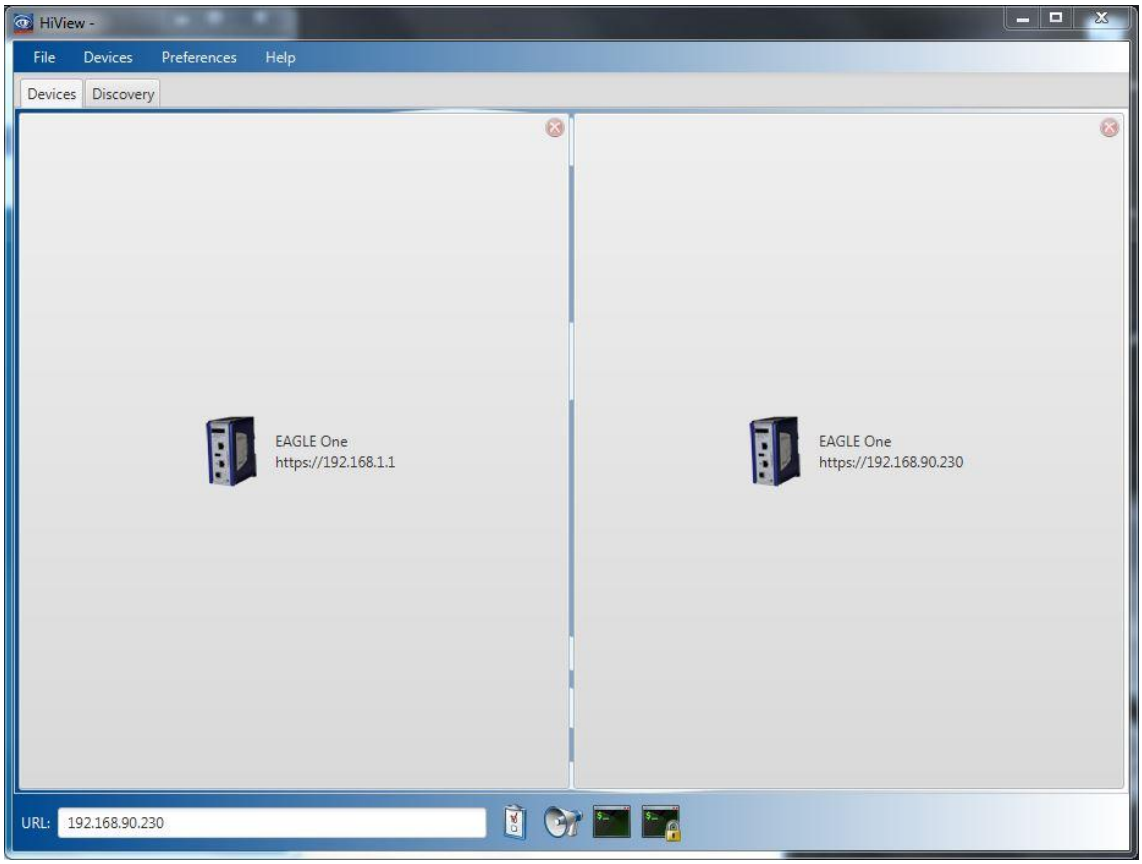

Figure 9. Firewall management software HiView by Hirschmann.

4.3.2 Internal DMZ Firewall Configuration

The function of the internal DMZ firewall is to allow communication from the control network into the DMZ and drop all communication from the DMZ to the control network. This chapter describes the steps taken to create the configuration.

The system interface window opens, once the user logs into the firewall (**Figure 10.**). This dialog is straightforward and only allows the user to change the name and location information of the firewall. This information should be input if the system includes more than one firewall as it makes replacing and reconfiguring the firewalls easier later on.

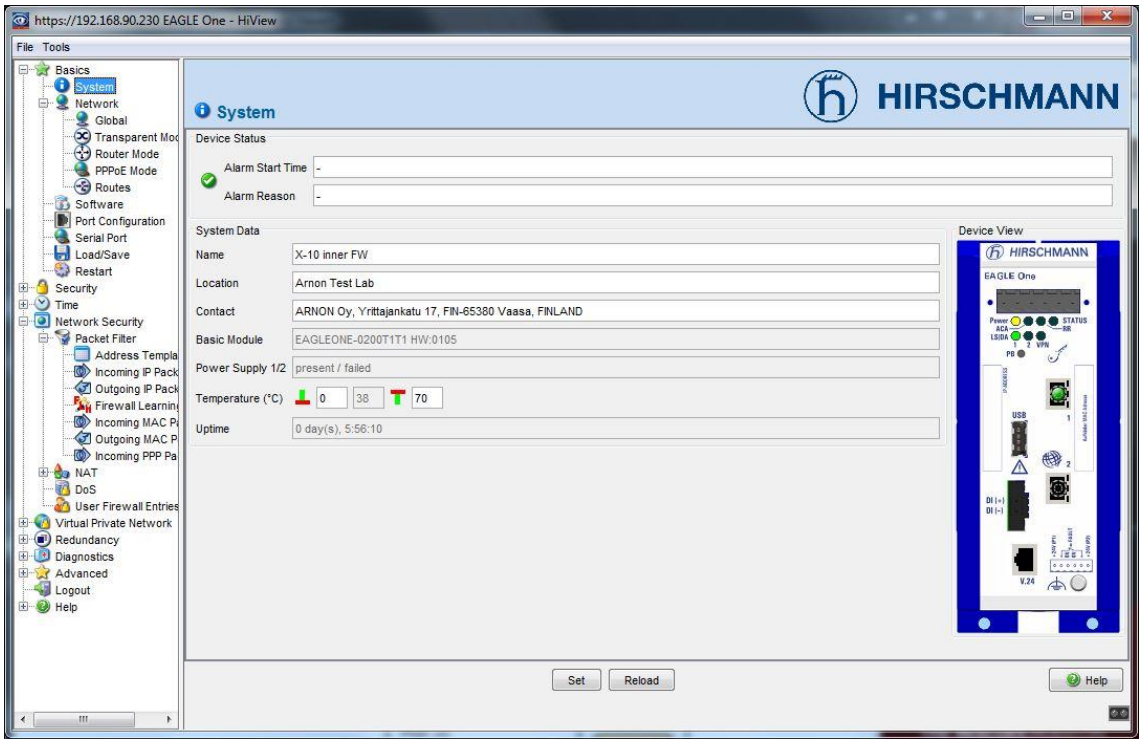

Figure 10. System interface window of the internal DMZ firewall.

The firewall was set to the router mode, and static IP addresses for the LAN and WAN ports were configured (**Figure 11.**). In the router mode, the device behaves like a router and transmits on layer 3 of the ISO/OSI layer model. The use of VLANs (Virtual Local Area Network) is also a possibility, but in that case, the firewall uses VLANs exclusively and disables all NAT functions.

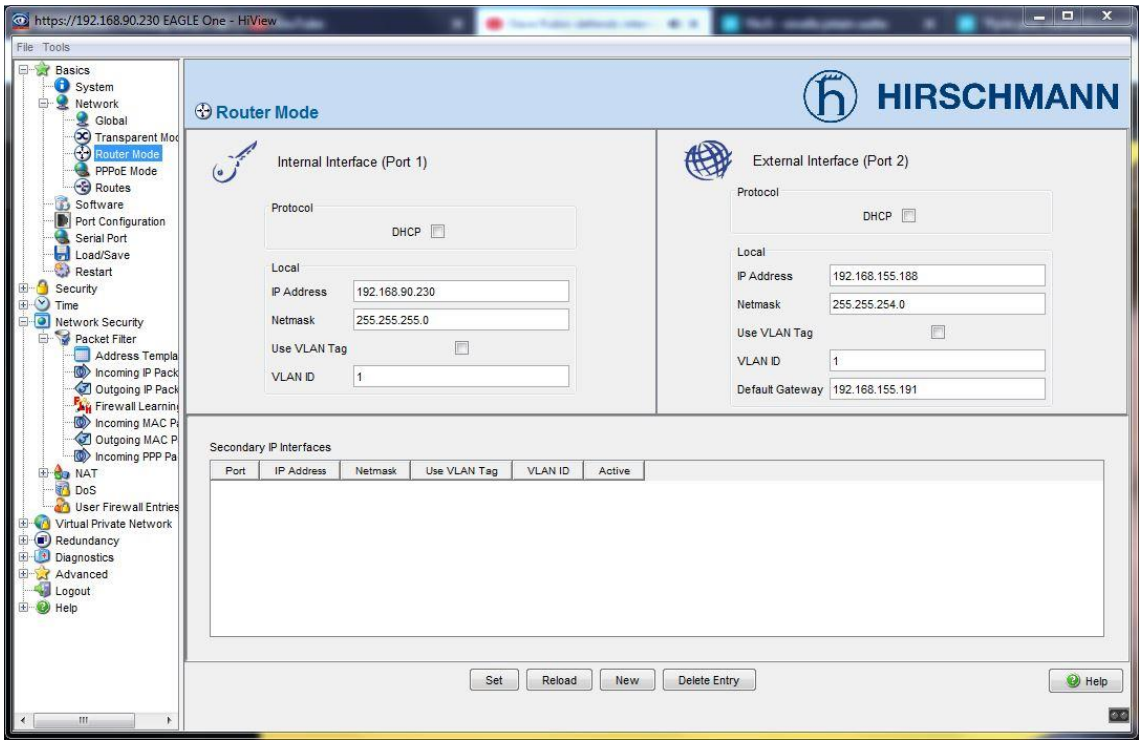

Figure 11. Firewall mode and IP address settings of the internal DMZ firewall.

Some industrial automation systems rely on legacy equipment due to many reasons, one of the biggest reasons being the cost of plant downtime caused by updating the equipment in addition to the cost of the new equipment itself. It is, therefore, possible to encounter devices with varying port speeds and operation modes. 10 Mbit/s half-duplex Ethernet ports can still be found, and communication errors will occur if auto-negotiation or careful manual configuration is not done. Below, in **Figure 12**, the firewall ports are turned on, and automatic configuration of the ports is activated. In this case, 100 Mbit/s full-duplex (FDX) is used.

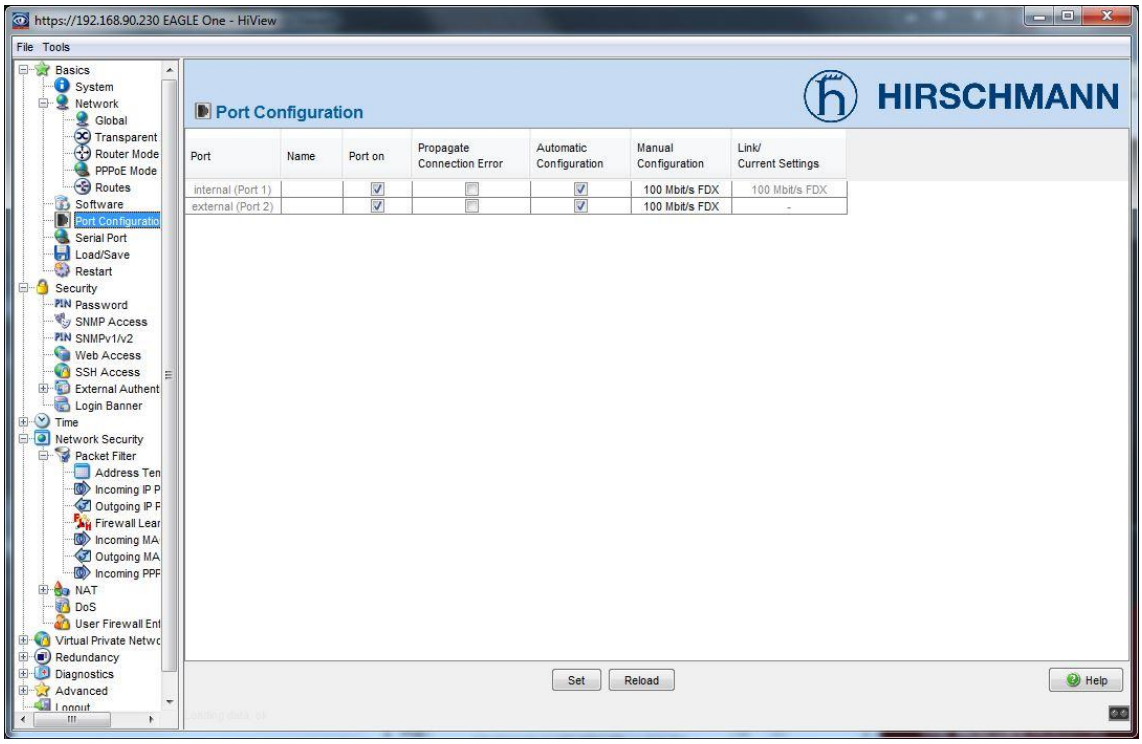

Figure 12. Firewall port configuration of the internal DMZ firewall.

The SSH server of the firewall was activated, and logging was enabled (**Figure 13**). The SSH server of the device allows authorized personnel to configure the device using the command line interface, to load firmware updates, configuration files or VPN certificates to the firewall using SFTP. Port configuration and SSH server settings for both the internal and the external firewall were identical.

The external port drops all inbound SSH connections due to security reasons.

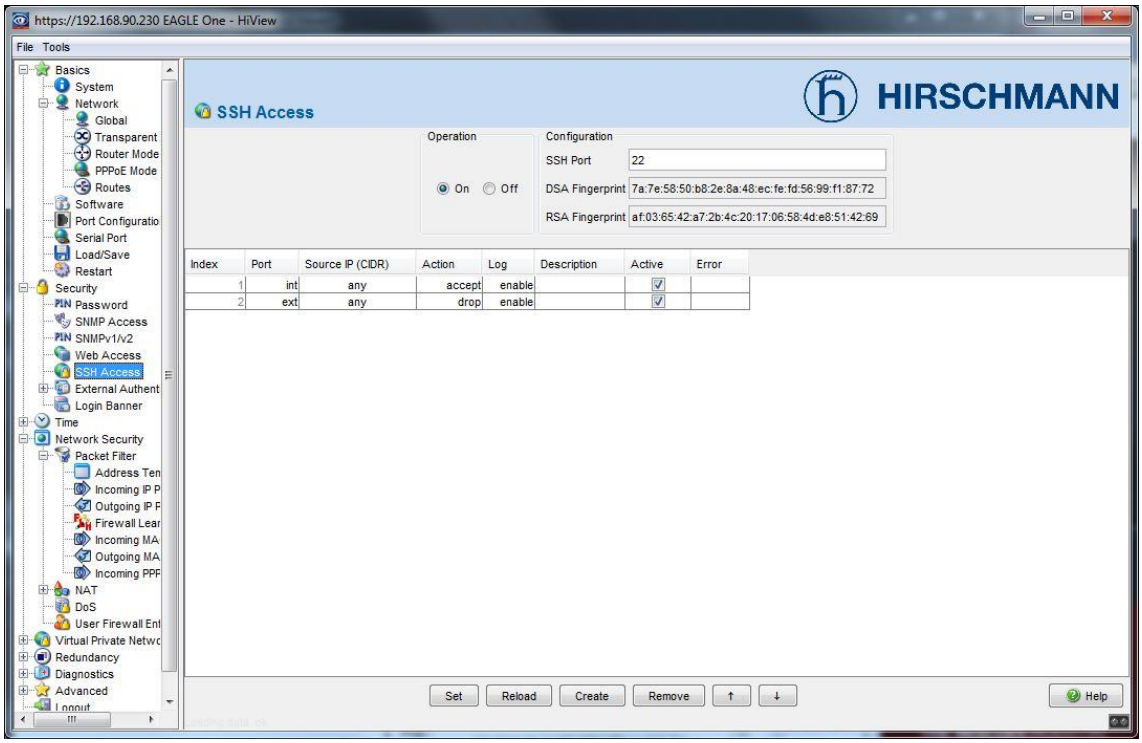

Figure 13. Firewall SSH access of the internal DMZ firewall.

Address templates were used to create and modify IP packet filter entries quickly and more efficiently (**Figure 14**). An address template consists of one or more address entries with the same name. The entries are made in Classless Inter-Domain Routing (CIDR) notation. For example, /24 after an IP address defines the address template to include 255 IP addresses, and /32 represents a single IP address.

The device automatically creates the suitable packet filter entries from a packet filter entry with variables. If one changes the address template for a variable, the device automatically modifies the packet filter entries created.

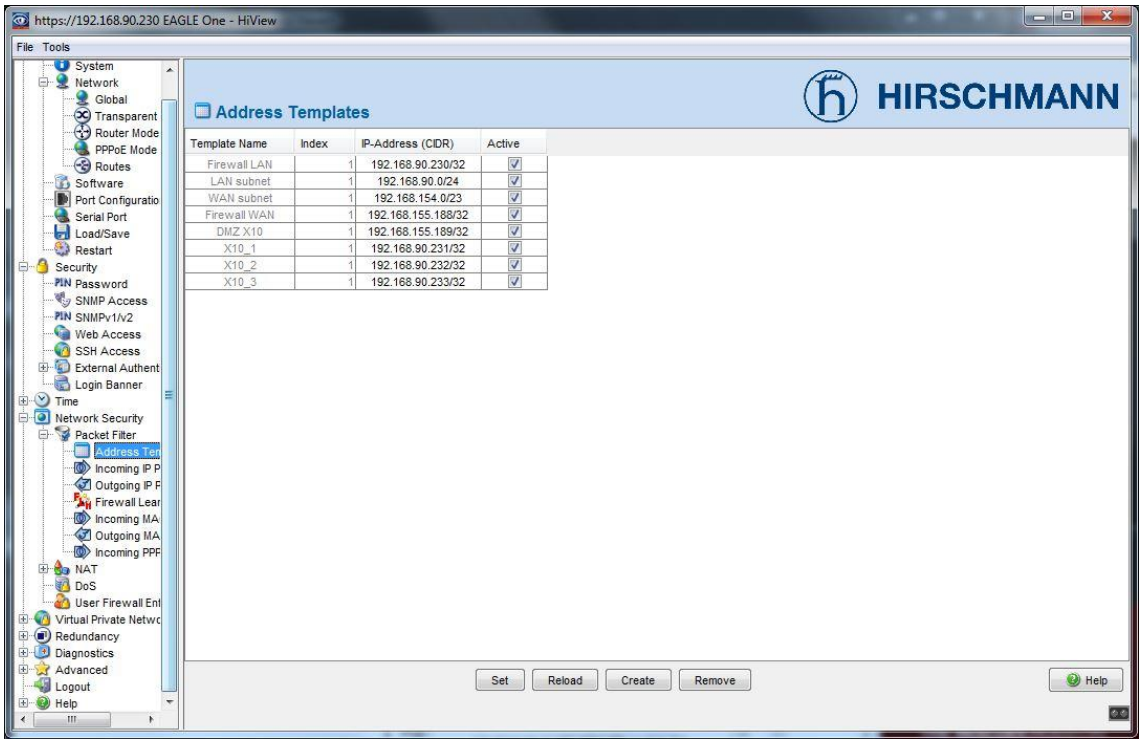

Figure 14. Address templates of the internal DMZ firewall.

One of the most critical tasks when configuring a firewall is to create rules for incoming IP packets. Firewalls usually check the incoming IP packets at the external and internal ports based on some or all of the following parameters:

- The logical port
- The source IP address
- The logical destination port
- The destination IP address
- The transmission protocol

Every packet that does not match any of the rules in the table is automatically dropped. Firewalls commonly contain an invisible rule for this, and in most cases, it cannot be deleted or changed.

Incoming IP packets are the traffic trying to access the WAN port of the firewall. In this case, the objective was only to push data outwards using SFTP, so no traffic inwards was allowed (**Figure 15**). Regardless of the abovementioned invisible rule, two visible rules were created to drop all incoming traffic.

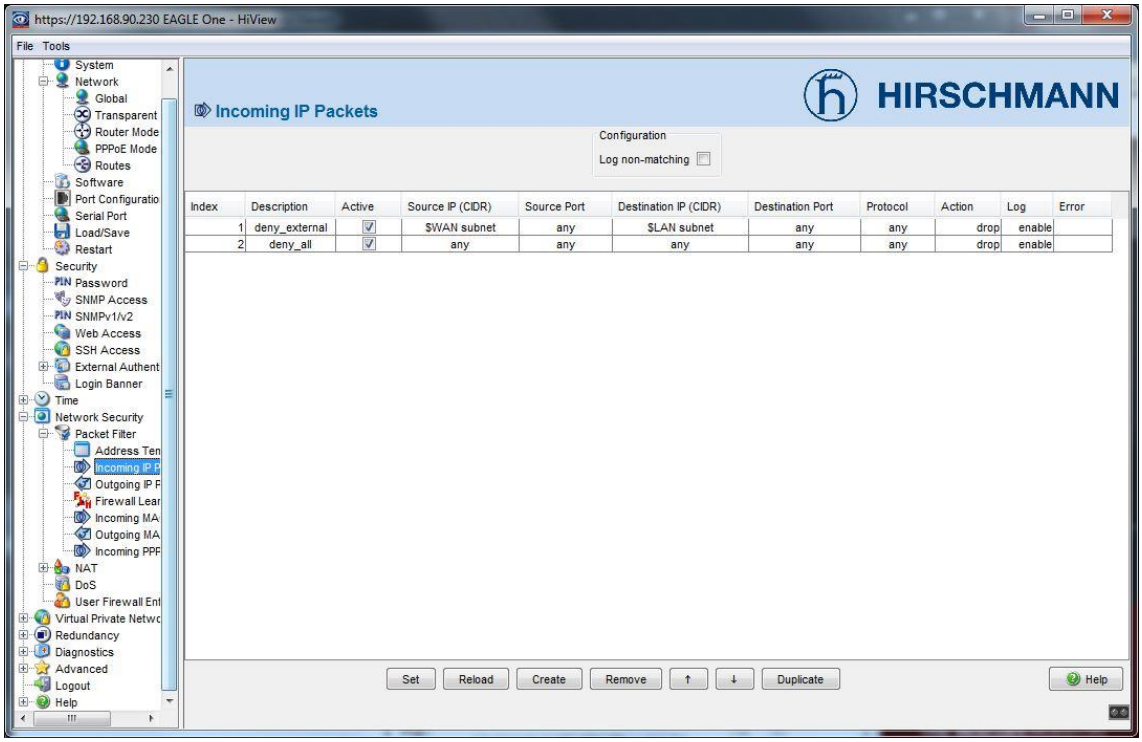

Figure 15. Incoming IP rules of the internal DMZ firewall.

Outgoing IP packets are configured to allow the transmission of data and ping for troubleshooting network connectivity issues, should such issues occur (**Figure 16**). All other traffic to any other port is again dropped.

It is important to note that the rules are followed from the top down. For example, if something is disabled in rule 2, it will not be allowed in any of the following rules. This is where many of the commonly seen configuration errors happen.

The devices in the control network are allowed to ping the firewall LAN port and X10+ DMZ collector. Even though the X10+ DMZ collector is in a separate network (192.168.155.189), it can be reached by pinging the virtual IP address (192.168.90.229) created in the firewall settings.

The actual data transfer is done through port 22 using the SFTP protocol.

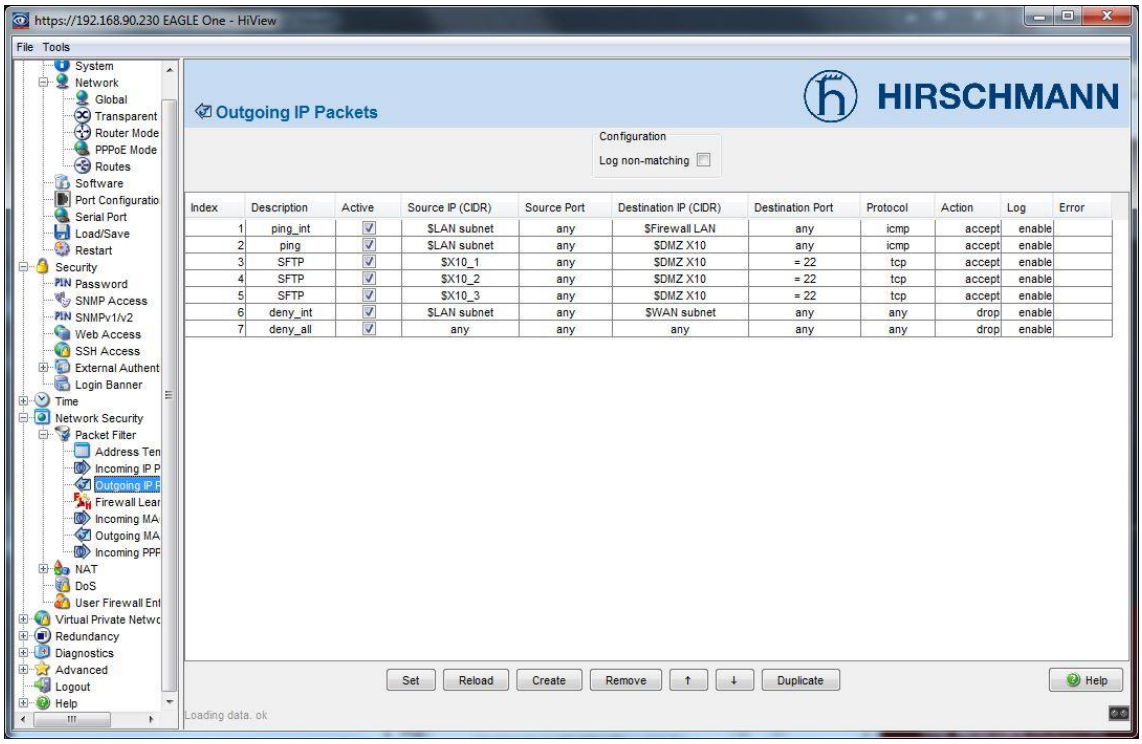

Figure 16. Outgoing IP rules of the internal DMZ firewall.

In addition to incoming and outgoing IP packet rules, NAT rules are required as the devices are located in different networks. Up to 128 entries can be created, and again one entry can encompass one or several addresses within a network depending on the netmask used in each rule.

With 1:1 NAT, the firewall operates as a router and allocates an additional IP address in the external network for a terminal device in the internal network. Additionally, as a proxy, the firewall answers the ARP (Address Resolution Protocol) queries for the additional IP address in the external network.

For outgoing data packets, the firewall replaces the internal source IP address of the terminal device with its external IP address. For incoming data packets, it replaces the external destination IP address with the internal IP address.

An “Invert Direction” selection is made for one of the rules to allocate an additional IP address in the internal network via proxy ARP. Thus, terminal devices in the internal network can communicate with external terminal devices without gateway entries.

If Double-NAT is selected and the source address is implemented in the packets, the device also replaces the destination address if there is a corresponding rule. Thus, terminal devices in both the internal and external networks can communicate with terminal devices in the other network without gateway entries.

In this case, the firewall reserves two addresses for NAT. 192.168.90.229 in the internal network and 192.168.155.189 in the external network (**Figure 17**). The field collectors were all assigned a NAT rule from the device address to the virtual address in the external network, i.e., the DMZ. The virtual addresses were linked together with one NAT rule which had the invert direction function activated to allow for proxy ARP.

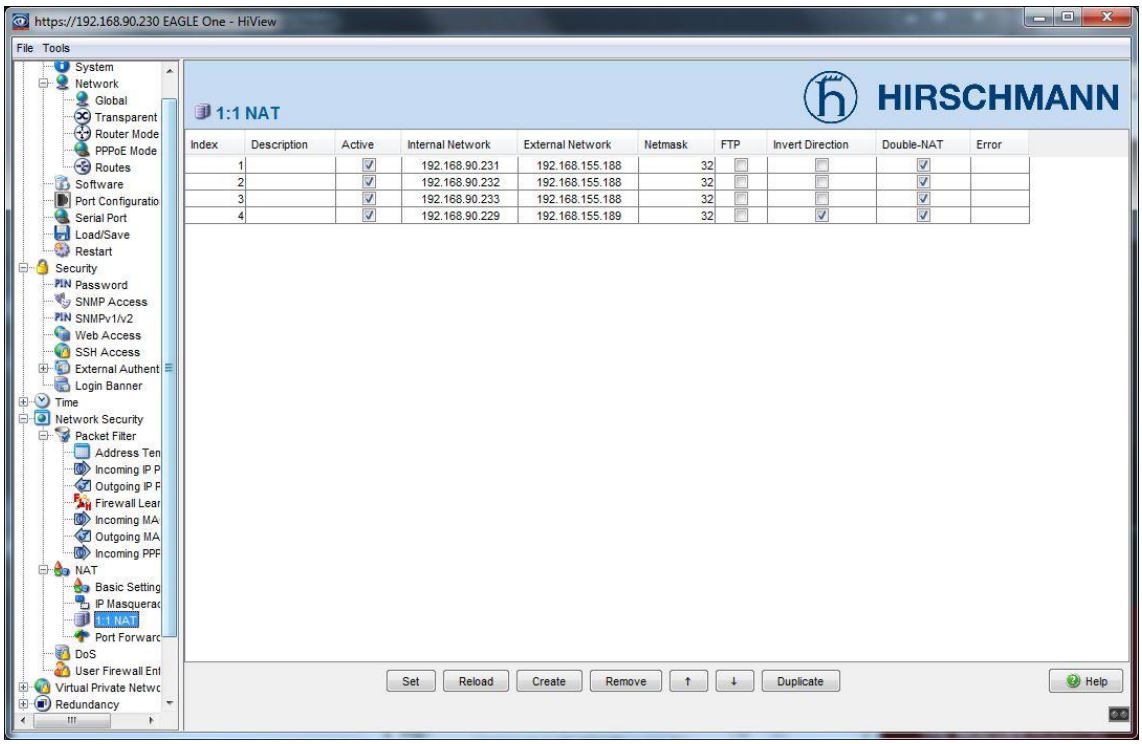

Figure 17. NAT ruleset of the internal DMZ firewall.

4.3.3 External DMZ Firewall Configuration

The function of the external DMZ firewall is to face the customer's local area network, which allows access to the Internet. There is no access from the customer's network into the DMZ, and only one port is open outwards to allow data transfer to AWS. This chapter describes the steps taken to create the configuration.

The system interface window of the external DMZ firewall is identical to the internal one. Only the name of the firewall has been changed (**Figure 18**).

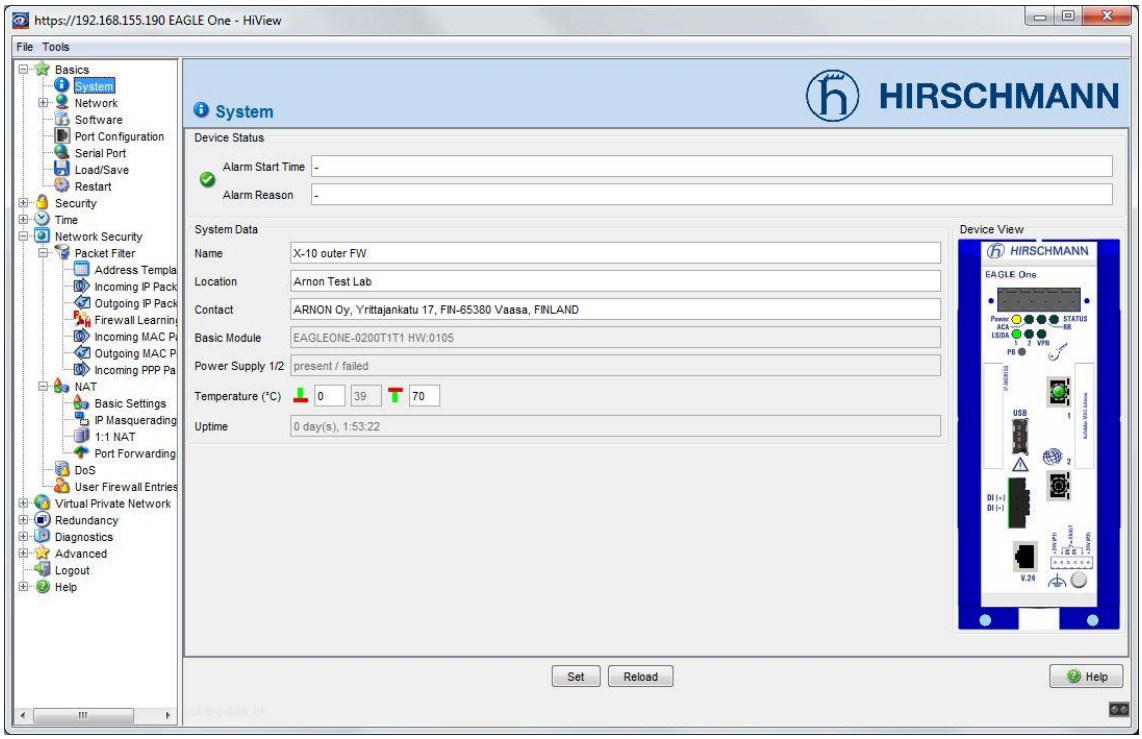

Figure 18. System interface window of the external DMZ firewall.

The external DMZ firewall was also set to the router mode, and the LAN port was set to a static address. The WAN port, however, was set to DHCP. The reason for using DHCP was that the IP range and subnet mask of the customer network were unknown. In such a case, it is easiest to leave the firewall in DHCP and make the last configuration at the site during the commissioning of the system. The configuration is shown in **Figure 19.**

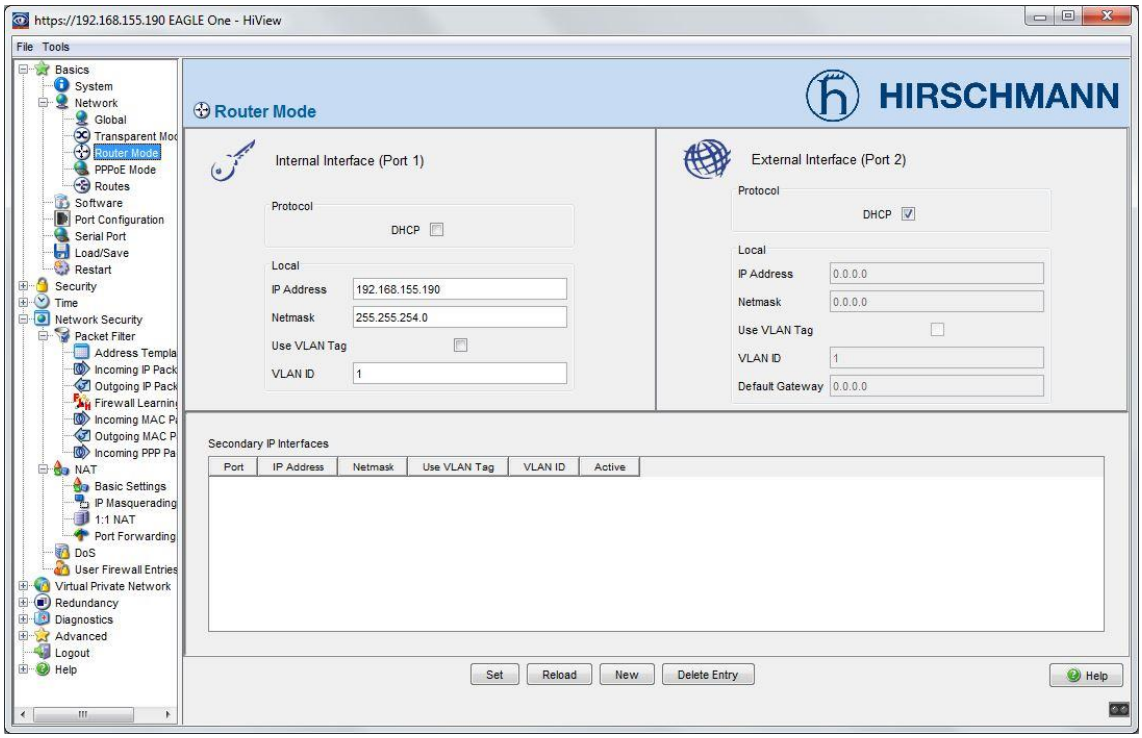

Figure 19. Firewall mode and IP address settings of the external DMZ firewall.

The address template of the external DMZ firewall is shown in **Figure 20.** WAN subnet and firewall WAN addresses were configured with placeholder addresses, as those would need to be changed during the commissioning when the actual customer network information was available. AWS address 34.208.4.84/32 shows the final destination of the data.

It this vital to point out that a more extensive scale use of the plant to cloud data collection systems must use a hostname-based routing (i.e., api.cloudservice.com) instead of static IP addresses as changes made to the cloud service's server cluster can result in the change of the server IP address resulting in the loss of connectivity of all IP address based equipment.

Hostname-based routing sets different requirements for the network and network equipment. A DNS-server must be accessible by the external DMZ firewall, and the firewall itself must be able to relay the DNS information to its own LAN-port. This discussion is one that needs to be had, but it is outside the scope of this thesis.

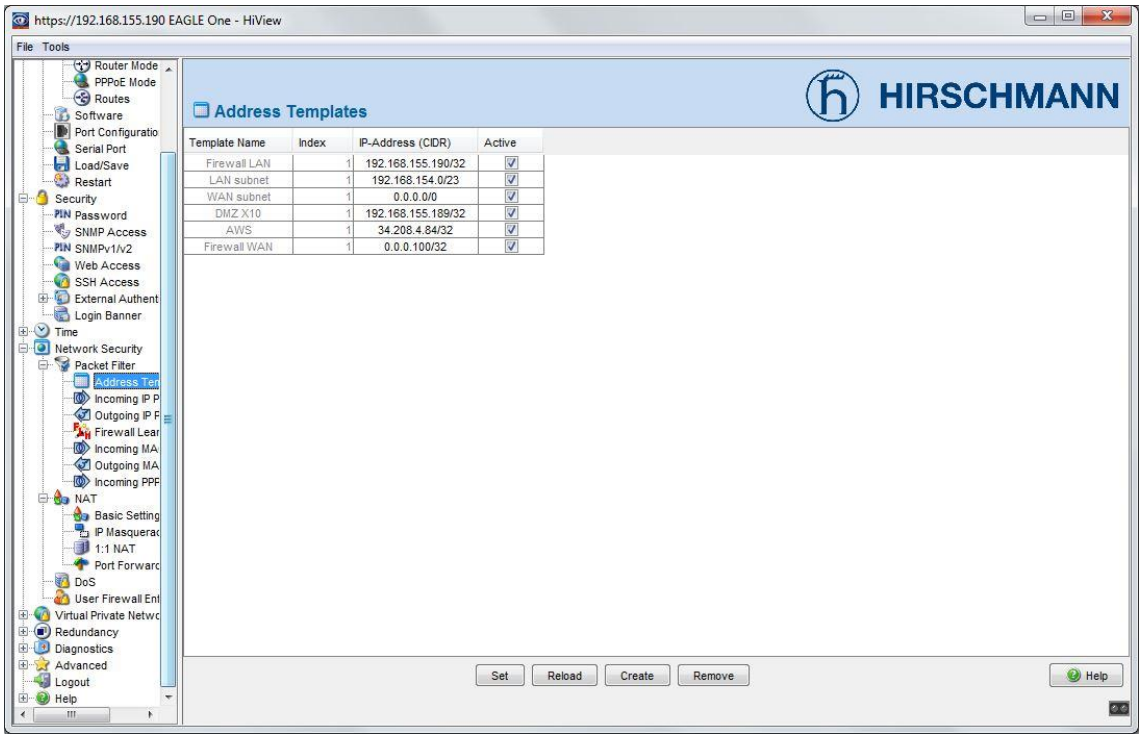

Figure 20. Address templates of the external DMZ firewall.

As with the internal firewall, all incoming IP packets are dropped by the external firewall. Again, visible rules were configured despite the invisible default rule configured into the firewall. The same configuration as with the internal firewall is shown in **Figure 21.**

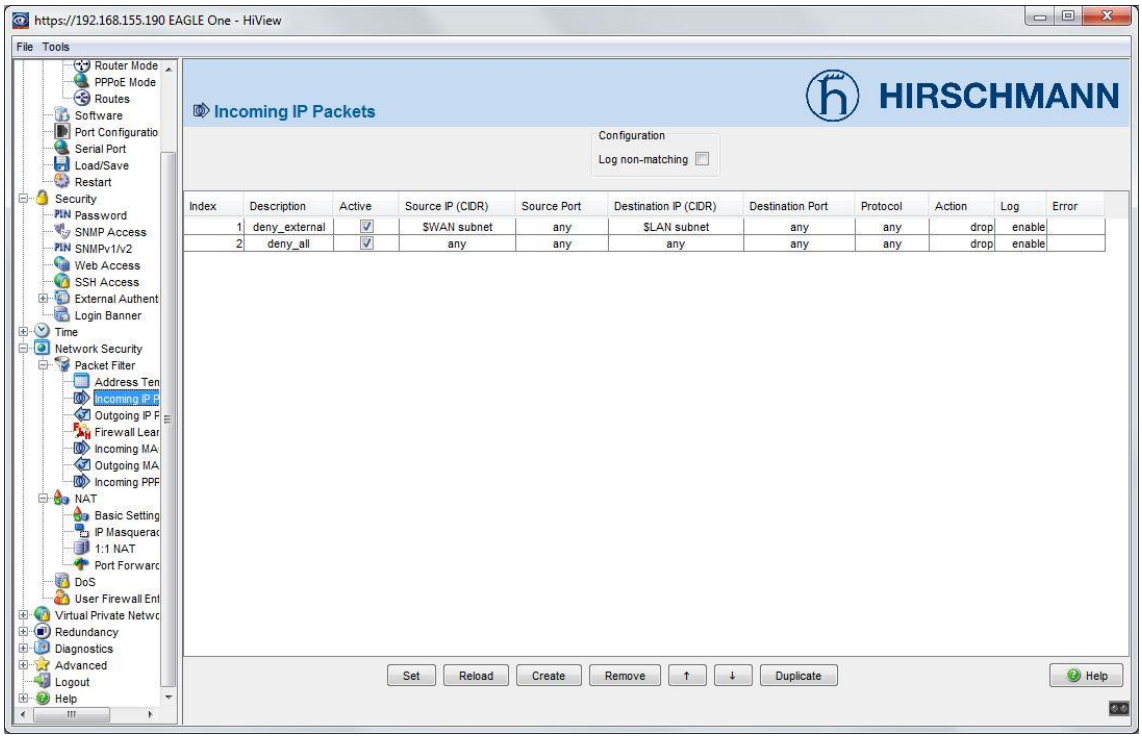

Figure 21. Incoming IP rules of the external DMZ firewall.

Outgoing IP packets were configured to allow the transmission of data and ping for troubleshooting network connectivity issues, should such issues occur (**Figure 22**). This external DMZ firewall is dedicated only for the X10+ device, so all other outgoing traffic from the DMZ network to any other ports is dropped.

Data transfer to the AWS cloud service is done through port 222 using the SFTP protocol. The default port for SSH related services is port 22, but it is often changed as port 22 receives a fair amount of malicious traffic from botnets and other scrupulous actors.

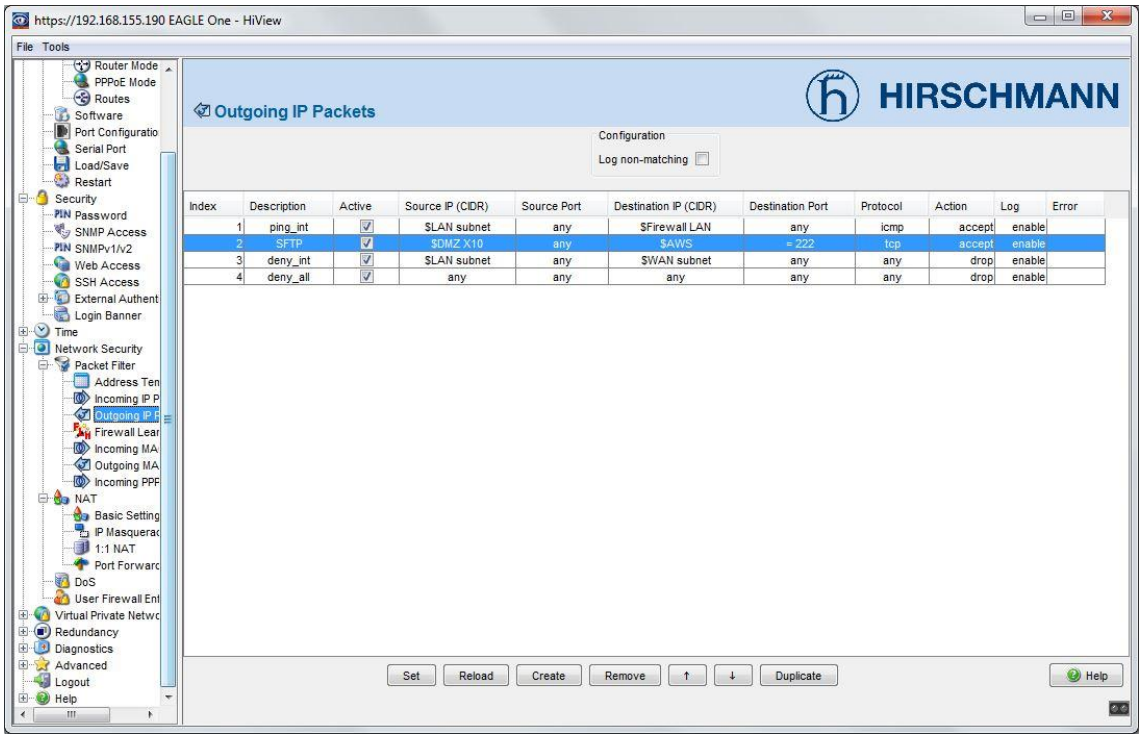

Figure 22. Outgoing IP rules of the external DMZ firewall.

The external DMZ firewall NAT rules reserve address 192.168.155.191 from the DMZ and one address from the customer's network. Again, the placeholder address of 0.0.0.0 is used to depict the virtual IP in the customer's network (**Figure 23**).

The X10+ collector in the DMZ was given one NAT rule, and the virtual IP address of the DMZ is linked directly to the IP address of the AWS cloud service.

These settings were modified during commissioning as this address configuration will not work simply by plugging in a network cable.

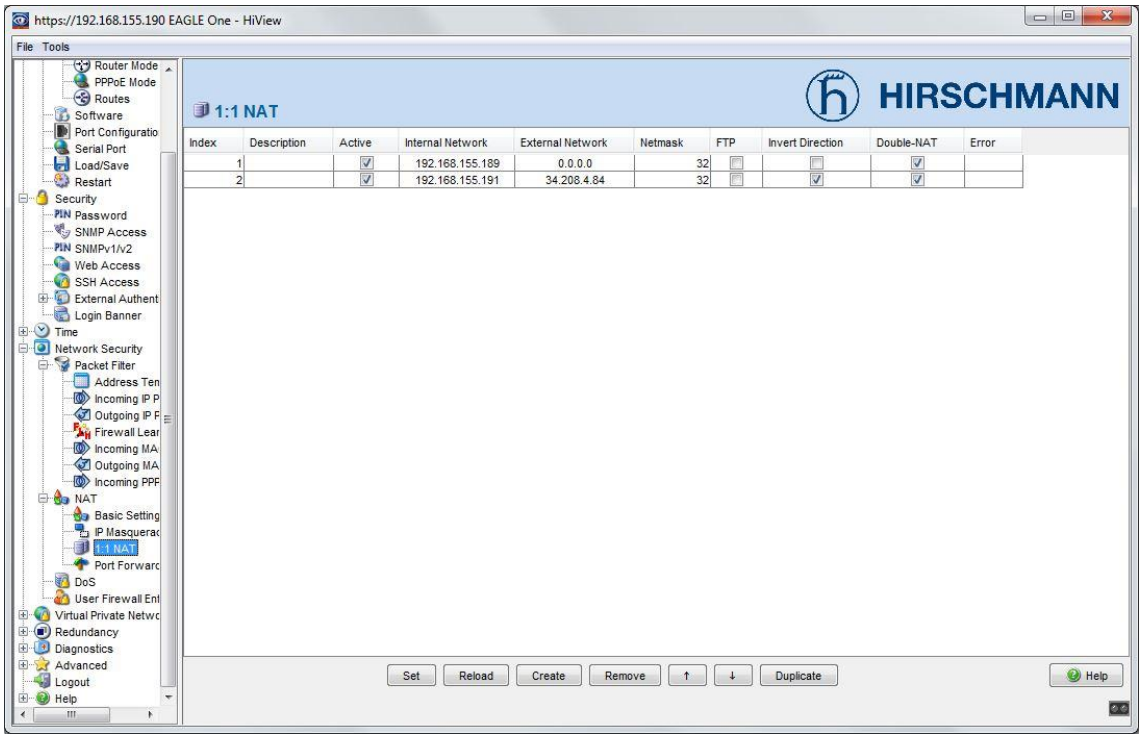

Figure 23. NAT ruleset of the external DMZ firewall.

5 REVIEW OF THE PROJECT AND THE BENEFITS OF THE SYSTEM

The key results of the project and the benefits of the system are reviewed in this chapter. The requirement specification, created as a result of numerous meetings, was used as a basic blueprint for the project.

5.1 Project Summary

The final result of the project is a working data collection system with cloud connectivity. Data is transferred in near real-time as initially requested by the customer. The X10+ data collection and transfer system were also accepted and certified by the customer's cybersecurity team, which allows the use of the system in future projects. The initially requested functionalities and the results at the time of this thesis being finalized are summarized in **Table 4**.

Table 4. Project summary.

Functionality	Priority	Result
File format	1	Customer specified file format successfully implemented.
Configurable data transfer interval	1	Data successfully transferred at five-minute intervals. Interval is configurable for each project depending on the quality of the Internet connection.
Data collector configuration update	2	Configuration via the cloud was not implemented. Most discussions are required to agree upon a cyber-secure method of the downstream file transfer.
Software Patching	3	Software security updates via the cloud were not implemented. Most discussions are required to agree upon a cyber-secure method of the downstream file transfer.
User authentication	1	Certificate-based user authentication was successfully implemented.
Price	1	No commercial software requiring software licenses was used. In that sense, this functionality can be considered a success. Price, however, is always a point of contention.

Data sampling	1	All 9500+ analog and digital signals are sampled once every second with a 99.9% data accuracy confirmed by the customer. An excellent achievement.
Log file transfer	3	Collection and transfer of log files from other equipment in the control network successfully implemented.
Segmented network layout	1	Segmented network layout designed and delivered.
No VPN connections	1	No VPN connections used.
NTP time sync	N/A	Not requested initially, but this functionality was added during the commissioning of the equipment to keep the real-time clocks of the X10+ data collectors in sync with the SCADA system.

5.2 Benefits of the System

- A proven solution that can be offered to end-customers with stringent cybersecurity requirements that do not allow the use of remote support software over VPN tunnels.
- The system allows faster response time from customer support as new data is available within 5-10 minutes instead of the earlier 24-hour data transfer cycle.
- Virtually unlimited scalability of the system makes it possible to collect data from even the most extensive production facilities.
- The use of open-source software eliminates the reliance on expensive proprietary software licenses.
- Both a fully standardized version for basic data transfer and an end-customer specific, highly customizable, version is available and already in use.
- Log file collection from equipment within the network is an example of end-customer specific customization. Log files help tremendously with troubleshooting and root cause analysis of different problems, as it is known what the equipment has done and in what order.

6 CONCLUSIONS

The objective of this thesis was to create a solution for data collection and transfer without compromising cybersecurity. The transfer of data is done in real time or with a small delay depending on the quality and speed of the Internet access available to the site. As discussed in chapter 5, this objective was successfully achieved.

Even though all priority 1 features were implemented in this solution, there is always room for improvement. The future development of the system will focus on patching the Linux operating system of the X10+ collectors and updating the configuration of the X10+ collectors via the cloud. Hostname-based routing should also be developed before any large-scale use of the system to give customers flexibility when updating or changing data storage services.

A move from old-fashioned, purely file based, data transfer systems to the use of APIs is to be encouraged. Web technologies such as Representational State Transfer (REST) offer increased flexibility and faster performance due to the reliance on the use of a uniform and predefined set of stateless operations. REST APIs also offer the ability for two-way communication between client and server. Web services conforming to the REST architectural style are called RESTful Web Services.

REFERENCES

/1/ ARNON Oy web pages. Accessed 10.05.2019. http://arnon.fi/company/

/2/ Communication Technologies, Inc. October 2004. Supervisory control and data Acquisition (SCADA) system, Technical information bulletin 04-1, pages 4-9.

/3/ Technic Link webpage, Accessed 03.03.2019. http://techniclink.com/learn/electrical/scada/

/4/ Edith Cowan University 2008, Issues common to Australian critical infrastructure providers SCADA networks discovered through computer and network vulnerability analysis, pages 1-6.

/5/ Modbus Organization, Inc. web pages. Accessed 03.03.2019. http://www.modbus.org/specs.php

/6/ Siemens, Edition 12/2005, A5E00711636-01, SIMATIC Open TCP/IP Communication via Industrial Ethernet manual.

/7/ Siemens Support website. How do you configure an ISO-on-TCP connection for data exchange between S7-300 and/or S7-400 by way of Industrial Ethernet CPs? Accessed 03.03.2019. http://support.automation.siemens.com/WW/view/en/47885440

/8/ SSH Communications Security, Inc. webpage. Accessed 03.03.2019. https://www.ssh.com/ssh/sftp/

/9/ SSH Communications Security, Inc. webpage. Accessed 03.03.2019. https://www.ssh.com/ssh/protocol/

/10/ Edith Cowan University 2009, SCADA Security - Slowly Circling a Disaster Area, pages 1-7.

/11/ Department of Home Affairs, Australian Government webpage, Accessed 10.05.2019. https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/security-coordination/critical-infrastructure-resilience

/12/ Weiss, Joe. August 2008, Assuring Industrial Control System (ICS) Cyber Security. Accessed 16.05.2019. http://csis.org/files/media/csis/pubs/080825_cyber.pdf