Article

Not peer-reviewed version

GOOSE Secure: A Comprehensive Dataset for In-Depth Analysis of GOOSE Spoofing Attacks in Digital Substations

Oscar Andrés Tobar-Rosero , Omar Roa Romero , Germán Darío Rueda Carvajal , Alexander Leal , Juan Felipe Botero * , Sergio Armando Gutierrez , John Willian Branch , Germán Darío Zapata

Posted Date: 16 September 2024

doi: 10.20944/preprints202409.1170.v1

Keywords: Cybersecurity; Dataset; Digital Substation; GOOSE protocol; IEC 61850; Smart Substation; Spoofing attack

 doi:10.20944/preprints202409.1170.v

Article

GOOSE Secure: A Comprehensive Dataset for In-Depth Analysis of GOOSE Spoofing Attacks in Digital Substations

Oscar A. Tobar-Rosero [1,*,†], Omar A. Roa-Romero [1], Germán D. Rueda-Carvajal [1], Alexánder Leal-Piedrahita [2], Juan F. Botero-Vega [2], Sergio A. Gutierrez-Betancur [2], John W. Branch-Bedoya [1], and Germán D. Zapata-Madrigal [3]

[1] Universidad Nacional de Colombia - TyT Group; oaroaro@unal.edu.co (O.A.R.R.); gdruedac@unal.edu.co (G.D.R.C.); gdzapata@unal.edu.co (G.D.Z.M.)
[2] Universidad de Antioquia - GITALab; erwin.leal@udea.edu.co (A.L.P); juanf.botero@udea.edu.co (J.F.B.V.); sergio.gutierrezb@udea.edu.co (S.A.G.B)
[3] Universidad Nacional de Colombia - GIDIA Group; jwbranch@unal.edu.co (J.W.B.B.)
* Correspondence: oatobarr@unal.edu.co; Tel.: +57-317-750-6055
† Current address: Industrial Automation and Communications Laboratory - Universidad Nacional De Colombia, sede Medellín

Abstract: Cybersecurity in Critical Infrastructures, and in particular Digital Substations is a topic of wide interest for industry and academic community. One of the approaches widely used to support research on this field is the use of datasets (i.e., traffic samples collected during the operation of an infrastructure). However, generating datasets from operational electrical systems presents certain challenges: i) The generation of these datasets generally implies the operation under controlled or ideal conditions, which possible ignores the dynamics of real-world operations within a digital electrical substation; and ii) Captured data often contains sensitive information, posing challenges for publication within the research community. This paper introduces the development of a dataset for cybersecurity research, focusing on the analysis of GOOSE spoofing attacks, given the critical role of GOOSE protocol in executing operational and control actions within Digital Substations. The dataset exposes the real effect of the attack by executing unwanted operations or actions under various operational conditions, including both stable electrical system scenarios and situations where failures are present in the electrical system. The dataset was acquired from a laboratory testbed developed within a physical infrastructure that emulates the actual operation of a digital substation with two bays. Experiments allowed to observe relevant properties of the traffic associated to GOOSE protocol, and the actual vulnerability of a DS infrastructure regarding Spoofing Attacks.

Keywords: cybersecurity; dataset; digital substation; GOOSE protocol; IEC 61850; smart substation; spoofing attack

1. Introduction

Digital Substations (DS) or Smart Substations play a fundamental role in the adoption of novel monitoring, control, and supervision mechanisms in the context of electrical systems. Within a DS environment, various protocols and standards are employed to implement appropriate communication among multiple devices and technological tools [1]. However, despite the process of substation digitalization brings benefits, it also introduces cyber security vulnerabilities that must be addressed in order to guarantee secure, reliable and resilient systems [2,3].

One of the communication protocols of high interest in the operation of DSs is GOOSE (Generic Object Oriented Substation Event). This protocol is used in the transmission of messages for protection and control operations, and the monitoring of behavior of elements within DSs [4]. Considering the relevance of GOOSE in the operation of DSs, it turns to be a relevant target for several types of cyber attacks. In particular, Spoofing attacks stand out due to their potential impact [5–7]. This type of attack involves resembling legitimate devices by injecting malicious traffic with the purpose of disrupting the normal operation of a DS possibly causing failures in the overall electrical system [8].

 doi:10.20944/preprints202409.1170.v

A valid approach present in literature for the study of cyber attacks, threats and vulnerabilities is the analysis of datasets. In computer networks and cyber security research, a dataset typically consists of structured data which can be used to identify and characterize the flow of data and information exchanges among devices in a system [9,10]. In the context of DSs, by analyzing these data flow of data and information exchanges, it is possible to identify behavioral patterns, devices and messages, as well as extracting other relevant attributes in order to detect anomalies [3]. A desirable property expected in datasets used in the context of DSs is that they contain messages actually expressing the typical traffic patterns associated to actual operational events of DSs infrastructures [11]. However, to the best of our knowledge, datasets generated for DSs infrastructures are mostly derived from synthetic traffic generated through simulation tools [12,13]. In general, these datasets are created under controlled or ideal conditions, thus ignoring the dynamics of real-world operations that might appear in actual DSs.

This paper presents the construction of a dataset for in-depth analysis of GOOSE protocols in DSs in different conditions, including Spoofing Attacks. The dataset has been obtained from a laboratory testbed deployed on a physical infrastructure that accurately emulates the operation of a DS with two bays. This physical infrastructure encompasses typical communication protocols of a DS and allows to reproduce the operational dynamics of DSs. In order to complement the analysis, the testbed infrastructure was also attacked with a Spoofing attack, which evidently caused disruptions and therefore failures on the DS.

The remainder of this paper is organized as follows. Section 2 provides a theoretical background related to DSs and the type of attack of interest considered in the dataset. Next, Section 3 presents previous work related to the proposed topic. It also establishes differences in the datasets generated, highlighting the need and importance of the dataset presented here. Section 4 presents a detailed description of test scenarios as well as the methodology employed to capture data for each of them. In Section 5, we analyze the dataset and explain the detailed behavior of GOOSE messages. Finally, Section 6 concludes this paper by condensing the results and presenting the section for future work.

2. Background

This section provides a brief introduction on several key concepts considered in this paper. First, we introduce a basic definition Digital Substation (DS). Next, we present an overview of the infrastructure elements conforming Digital Substations, with particular emphasis on the communication protocols used among these elements. From these protocols, GOOSE is presented with further detail in order to understand why the spoofing attacks, finally discussed, are specially critical against this protocol.

2.1. Digital Substation in a Nutshell

The digitalization process of electrical substations is guided by the IEC 61850 standard [14], which proposes an architecture such as the one illustrated in Figure 1 [15]. This standard comprises different concepts including systems integration, interoperability, and specifications of communication technologies and protocols, aiming at simplifying supervision, control and monitoring activities in DSs [1,16]. According to the IEC 61850 standard, a DS includes three operation levels known as i) Level 0: Process; ii) Level 1: Bay; and iii) Level 2: Station. These operation levels are linked through two communication buses named station bus and process bus, which are used for information exchange among the devices of each level. For this information exchange, the DS leverages a set of communication protocols [17]. In addition to these levels mentioned, there is a Level 3 which is considered external to the DS. Level 3 comprises the control center of the utility company operating the DS, where tasks associated to remote control and supervision are carried out. There are additional standards and protocols for Layel 3 which are out of the scope of IEC 61850 such as IEC 60870 or IEEE 1815 [18,19]. These protocols are not considered in this work.

 doi:10.20944/preprints202409.1170.v

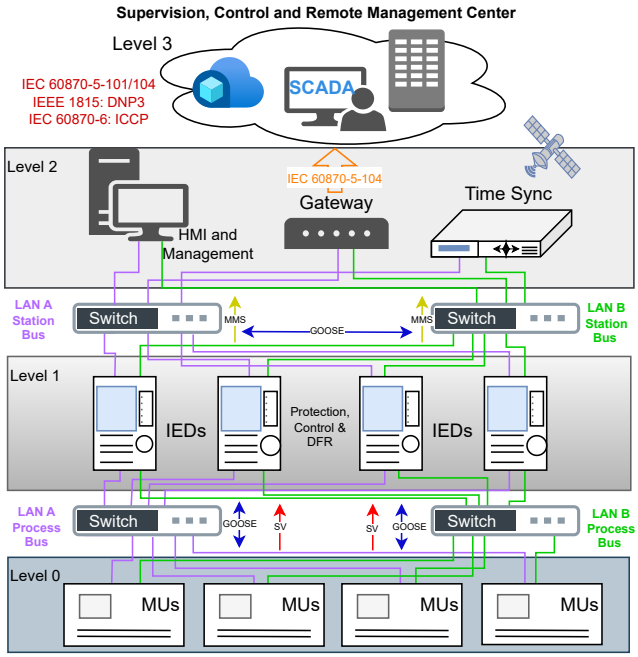

Figure 1. Digital Substation (DS) - Reference architecture

2.1.1. DS Infrastructure

In this section, we describe the main devices that appear in the infrastructure of a DS according to the reference architecture (Figure 1). We highlight some of their technical functionalities and operational features [15,20]:

Protection or Control Intelligent Electronic Device (IED): It is a class of device responsible for executing protection functions or performing control functions in DS. This device transmits messages through the communications network, associated to aspects such as general operating states, programmed protection functionalities, and signal states for supervision purposes.

Digital Fault Record (DFR) or Backup Protection IED: It is a subclass of IED device that can perform two main functions. On the one hand, it is in charge of recording and storing in a standard format [21] electrical and digital signals associated with fault events. This stored information allows further fault analysis and identification of spontaneous electrical events within DSs. On the other hand, this kind of device can also function as a backup protection IED.

Gateway: It is a device that translates the monitoring and reporting protocols used within a DS onto protocols implemented by utility companies in control centers, comprising supervision systems external to the DS.

Merging Unit (MU): It is a class of device in charge of digitizing electrical signals associated to current and voltage in order to make possible their transmission as sampled values within the DS. An MU can send state reports and subscribe to messages from IEDs in order to execute switching actions within the electrical system.

Switch: This network device connects various devices within a Local Area Network (LAN). Due to their properties, switches can implement mechanisms (e.g. IEEE 802.1Q VLANs) to isolate traffic in such a way it is delivered only to intended destination devices. It is important to remark that network switches used in DSs must ensure that the message transmission time falls within the limits defined by the IEC 61850 standard [14].

 doi:10.20944/preprints202409.1170.v

Precision Time Server: It is a device in charge of providing time synchronization for all the devices in the DS. This synchronization is particularly critical for devices such as IEDs and MUs which process digitized signals coming from electrical measurements of currents and voltages [22].

Management Workstations: These are computers used for the configuration, management and analysis of both, the devices within the DS and the information generated by these devices [1].

2.1.2. DS Communication Protocols

In this section, we will outline the communication protocols used in the operation of DSs, according to the IEC 61850 standard [15,20].

Generic Object Oriented Substation Event (GOOSE): It is a messaging model for transmitting highly critical events in a DS. It is also specified in the part 8-1 of the IEC 61850 standard. GOOSE is a data link layer protocol and it is based on publisher/subscriber model. It uses multicast Medium Access Control (MAC) addresses for the identification of devices involved in its communication [4,23].

Sampled Measured Values, or Sampled Values (SMV/SV): It is a protocol defining essential messages used to transmit current and voltage digitized signals coming from current and potential transformers into the IEDs within the DS. The structure of these messages is defined in Part 9-2 of IEC 61850 standard. Similar to GOOSE, this protocol is a data link layer protocol, and it is based on the publisher/subscriber model [1,24].

Precision Time Protocol (PTP): It is a communication protocol used to synchronize time information of devices in a network. Part 9-3 of the IEC 61850 standard defines its adoption for DSs. Similar to GOOSE and SMV/SV, it is a data link layer protocol and it is based on the publisher/subscriber model [22,25].

Manufacturing Message Specification (MMS): It is a messaging specification for industrial applications, adopted in the part 8-1 of the IEC 61850 standard. MMS operates in a Client/Server Model. It uses TCP as a transport protocol (MMS servers use to listen at port 102). It is used in the station bus for the transmission of messages from IEDs up to Supervisory Control and Data Acquisition (SCADA) systems, Human-Machine Interfaces (HMI), or Gateway devices [23,26].

Due to the relevance of GOOSE protocol among the remaining ones defined by the IEC 61850, given its criticality for the protection and control operations, in this work we focus on the GOOSE protocol and its vulnerabilities. These vulnerabilities might be exploited to compromise the security of a DS.

2.2. GOOSE Operation

GOOSE messaging plays a pivotal role in the operational infrastructure of DSs. This communication protocol enables reliable real-time exchange of data and control commands among IEDs [4,23]. GOOSE is crucial to ensure operational integrity and safe and resilient operation of the overall electrical system [27].

In a DS, GOOSE messaging is mainly used for fast and reliable transmission of information associated to critical events and control signaling. In its specification, it defines requirements in order to provide deterministic data transmission [17]. IEDs broadcast GOOSE messages over the network (i.e. publishers) which are processed by receiving IEDs (i.e. subscribers). Hence, no intermediary devices are involved in the processing of GOOSE messages, which reduces response systems while increases DS relability [4,27].

Figure 2 shows an schematic diagram of GOOSE operation. In this figure, it is displayed how the GOOSE messaging is generated whenever a state change occurs in signals (represented as Boolean fields in protocol messages). The IEC 61850 standard defines that upon these changes, a burst of messages is generated as a confirmation for the data sent. These bursts are generated due to the occurrence of *trips* (i.e. operation of circuit breakers in the DS infrastructure) associated to protection signals, or changes on control supervision signals [28].

 doi:10.20944/preprints202409.1170.v

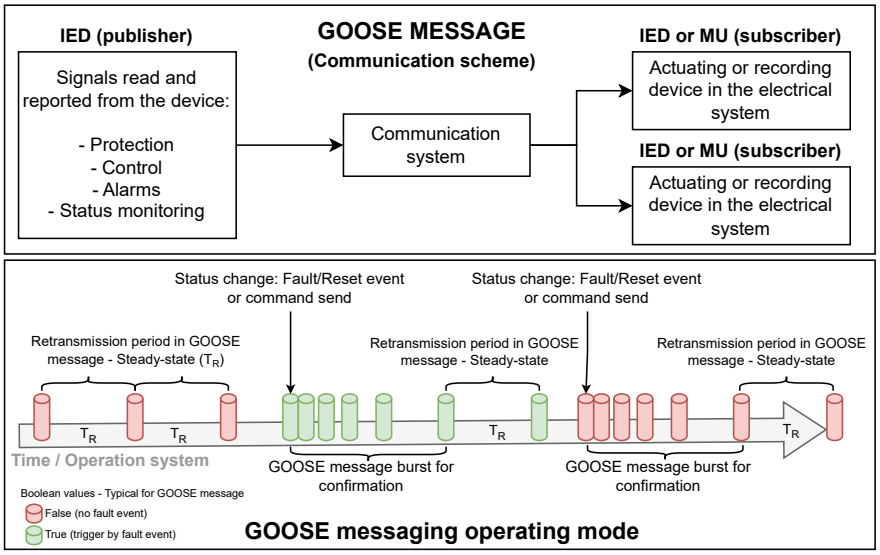

Figure 2. GOOSE Operation scheme and mode

2.3. Spoofing Attack

Spoofing attacks are those attacks where an offending device resembles a legitimate device. This resemblance is performed with the goal of injecting malicious traffic, as generated from a legitimate device, with the purpose of disrupting the normal operation of a DS. This action might possibly cause failures in the overall electrical system [8].

In a DS (see Figure 3), the spoofing of messages or devices represents a critical vulnerability that might lead to various failure scenarios. These failure scenarios represent a direct impact to the operation of the electrical infrastructure [5]. For instance, impersonating a protection device or Spoofing a specific *trip* message (change of state of a circuit breaker) might result in the omission of an activation command directed to a substation switchgear. Spoofing could lead to ignoring an operation command or fault event, potentially escalating to a large-scale event with significant effects on the safety of the electrical infrastructure. Also, due to spoofing, a subscriber device might be manipulated to execute protection actions or controlled operations that do not align with the actual state of the system. This might result in triggering failure events, often leading to cascading failures with significant impacts on the electrical system [29]. Spoofing attacks typically target IEDs, but they can also be directed against other equipment within the DSs [30]. For instance, Spoofing attacks against Precision Time Servers might cause failures in time synchronization, which is critical for IEDs and MUs [31]. In this work, we focus on analyzing the impact of an Spoofing attack against an IED.

 doi:10.20944/preprints202409.1170.v

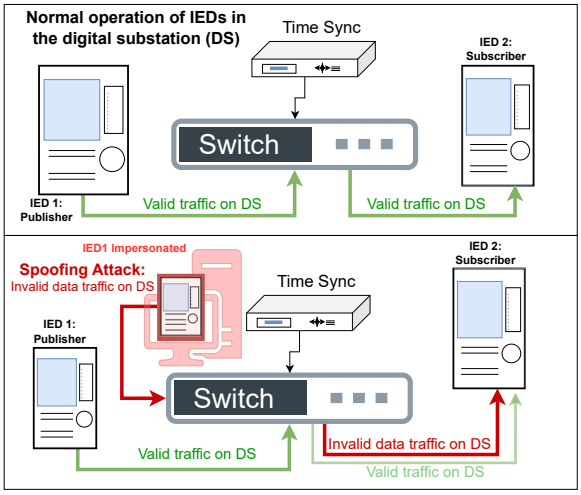

Figure 3. Spoofing Attack on Digital Substation

3. Related Work

Vulnerability assessment is a critical task in cybersecurity research [32,33], specially for the identification in advance of possible threats against networks and infrastructures. A restriction for vulnerability assessment is the fact that, in many cases, it is not possible to have access to production infrastructures in order to execute experiments. Thus, an approach commonly used by the scientific community is leveraging datasets to model the infrastructures under study, understand their dynamics, operations and interactions and propose models and solutions in different fields [34,35].

For years, researchers in the field of cybersecurity have been using datasets as input for their studies, and in many cases they have also produced datasets as contributions to the community [35]. One of the first attempts of dataset for cybersecurity research that is referenced in literature is the *KDD-Cup99* [36]. However, due to the evolution in traffic dynamics and the emergence of new and different services and applications, *KDD-Cup99* has rendered obsolete. New datasets such as *NSL-KDD and CICIDS 2017* have been presented to the community, containing large volumes of traffic samples benign and associated to a wide set of cyber attacks [37–39].

Despite the relevance of these recent datasets and their usage to produce high relevance research [40], they are intended to be general, containing samples of diverse traffic patterns and attacks. More tailored datasets are required for the assessment and analysis of particular infrastructures such as DSs [3,41]. However, dealing with such infrastructures introduces several challenges. First of all, in the context of critical infrastructures such as DSs, there are important concerns associated to information privacy. Both, client and vendor information exchanged by the devices in these infrastructures have strict requirements imposed regarding the management of this information. Also, corporate policies enforce restrictions on the disclosure of topology and descriptive information of the equipment involved in particular deployments. Hence, the generation of datasets as contribution to the academic community is more complicated than in other research fields [3,42].

Some examples can be found in literature presenting datasets for cybersecurity study in the context of critical infrastructures. The *HIL-based augmented ICS (HAI) of 2020* presents a dataset developed on a testbed based on a steam turbine and an hydroelectric system. This dataset contains different traffic samples of normal traffic, and some synthetically generated attacks [43]. *Secure Water Treatment (SWaT)* dataset is another example of dataset generated on critical infrastructures, in this case, a water treatment facility. This dataset contains also samples of normal traffic, and induced synthetic cyber attacks. [44].

There are some efforts of datasets for electrical substations. *ELECTRA* focuses on Modbus communication register [45]. *EPIC* addresses MMS but leave other protocols out of the scope of the study [46]. *IEC61850SecurityDataset* includes GOOSE protocol traffic, but the traffic is captured

 doi:10.20944/preprints202409.1170.v

by emulation of IEDs [47]. In [48] authors present a dataset coontaining traffic samples of a DS in steady state. Hence, it can be used as baseline for the development of anomaly detection solutions. A simulated dataset features a substation operating in a steady state, which implies that the flow dynamics are constant. In this case, it is analyzed using a fractional auto-regressive integrated moving average *(FARIMA)* from which characterization of traffic and anomaly detection are performed.

Quincozes et al. [3] conducted a survey of intrusion detection and prevention systems in DS, highlighting the importance of accurately characterizing the attributes of the communication protocols under analysis. This is a task that can be performed by leveraging adequate datasets. Additionally, Wang et al. [49] present an analysis of anomaly detection and attacks against IEDs on smart substations, considering different operation scenarios and working over simulated testbeds. This paper discusses the importance of having real datasets and the limitations associated with the availability of operational infrastructure datasets, access to electrical substation systems, and the inclusion of cyber attacks in the data samples.

The main contribution of our paper is presenting a public dataset for cybersecurity research in Digital Substations. This dataset is generated on a real testbed, aligned with the IEC 61850 standard, rather than on a simulated infrastructure. This dataset contains both, samples of normal traffic in different operational conditions, and attacks exploiting vulnerabilities of the GOOSE protocol. Hence, we aim at adressing two limitations of existing works. First, the dataset is based on sampling of traffic generated with real hardware. Second, the vulnerabilities of GOOSE, specially those allowing spoofing, are extensively leveraged for the generation of actual attacks, with real impact in the infrastructure.

4. Testbed and Dataset Description

In this section, we describe the testbed used in the construction of the dataset described in this work. We detail the infraestructure composing this testbed and we outline the scenarios considered, which fit real operational conditions of an actual infrastructure. Finally, we introduce some relevant properties of the dataset generated.

4.1. Infrastructure Description

As previously explained, DSs comprise different devices, which perform specific functions. The testbed built for this work aims at reproducing as precisely as possible the infrastructure of a small DS. For this reproduction, we include actual physical equipment which can reproduce all the operational procedures and conditions of a real world infrastructure. Figure 4 presents a scheme of the testbed. This is a DS with two bays, a station bus level and three devices performing Protection and Control functions. The testbed contains also a Precision Time Server and a Gateway. Given its design, the testbed incorporates the communication protocols defined in IEC 61850 standard besides complementary protocols associated to monitoring and management of the equipment.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 September 2024 doi:10.20944/preprints202409.1170.v

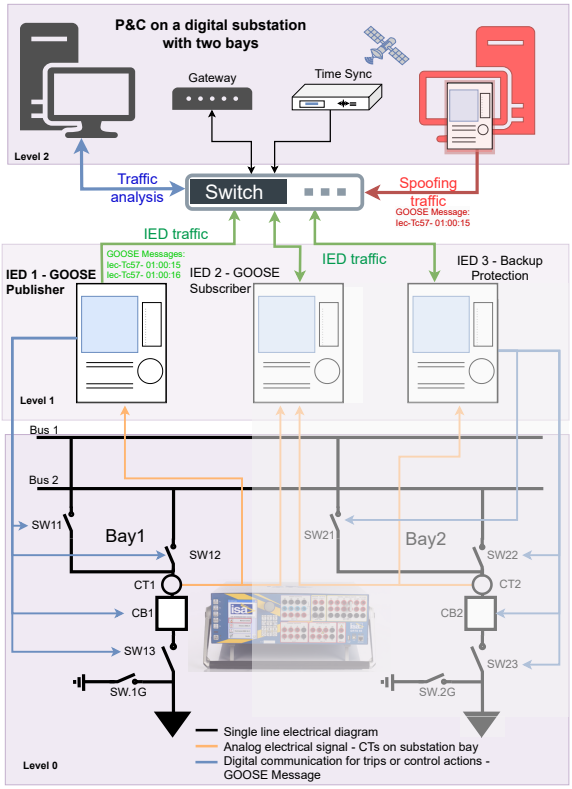

Figure 4. Testbed used

Next, we present some of the technical specifications of the equipment in the testbed.

4.1.1. **Level 2**

Gateway: It is Kalkitech brand device. It is connected to the three IEDs in the DS, for monitoring purposes

Workstations: Two high-performance workstations used for monitoring, data analysis and generation of attack traffic

Precision Time Server: It is a Meimberg GPS device. It can operate with both PTPv2 and NTP/SNTP protocols

4.1.2. **Level 1**

IEDs: These are equipments of ABB and GE brands. These devices have the following general characteristics:

- Network Interfaces: Ethernet 100BASE-FX ports (Optical Ports)
- Protection and control functions: Instantaneous overcurrent
- Configuration of GOOSE retransmission time set: 5 secs.

For visualization purposes, the IEDs were configured with a local HMI. This interface included a flag to indicate whenever a GOOSE *trip* (change of state) occurs in the electrical system. Also, the HMI contained an indication to report whenever a device suscribes to a GOOSE *trip* message.

4.1.3. **Level 0**

This level includes an ISA DRTS 66 device. This device is a signal injector, responsible for emulating the electrical network and validating switch actuation upon the occurrence of a GOOSE *trip* event, which as mentioned is associated to change of state in electrical variables.

 doi:10.20944/preprints202409.1170.v

4.1.4. **Station Bus**

This bus includes a General Electric S2024 switch. This is a Layer 2 switch, supporting LAN protocols such as IEEE 802.1Q (VLANs). It provides the connectivity among devices in the DS while enabling traffic isolation according to the required communication patterns (e.g. the different multicast groups for the communication among IEDs).

4.2. Dataset Description

In this section, we describe the general aspects of the dataset generated in this work. This dataset contains different traffic samples representing normal operations, failure events associated to electrical conditions, and the behavior of the infrastructure when facing cyber attacks. We focus on spoofing attacks, which are highly devastating and highly prevalent in the context of critical infrastructure, such as DSs [7,50].

4.2.1. Experiment Scenarios

In the development of this dataset, we considered four scenarios in order to cover different operational conditions, both in normal operation and when infrastructure is under attack. Next we describe these scenarios.

Scenario 1: Stable operation of the electrical system

In this scenario, the devices are configured for joint and synchronous operation. In this stable state, devices hey transmit operational messages. That is, the traffic reflects normal operation of the electrical system associated with the DS.

Scenario 2: Electrical system failure events

In this scenario, we considered the operation of the DS upon the occurrence of failure events. In order to reproduce this behavior, we configured general triggers associated with electrical variables monitored in the infrastructure. In this scneario, traffic varies according to the nature of the failure events. The goal of this scenario is to make evident the difference of traffic dynamics during failure events in comparison to the stable operation.

Scenario 3: Spoofing attack during stable operation of the electrical system

In this scenario, while the DS is transmitting information related to the stable operation of the electrical system, a spoofing attack is executed. This attack simluates the generation of a false report of a failure event reported by an attacker device. This spoofing causes the system responds incorrectly due the false event.

Scenario 4: Spoofing attack in the presence of electrical system failure events

In this scenario, a spooring attack is induced while the system is facing a failure event. Given the behavior of the GOOSE protocol in failure events (i.e. generating a burst of messages), this situation of spoofing upon the occurrence of attacks becomes more devastating due to the cascade effect it might introduce.

Scenarios 3 and 4 are considered analogous to scenarios 1 and 2, respectively. However, scenarios 3 and 4 include spoofing attacks. Through the design of this scnearios, we want to illustrate the behavior of the DS Infrastructure both in normal stable state conditions, and when reacting to failure events that might arise during the operation of the intrastructure.

4.2.2. Time Window for Data Collection

For the four scenarios, we considered a time window of 180 seconds. To the best of our knowledge, this time window allows to capture the traffic dynamics during operational conditions (normal, stable state and upon attacks), and the effect of spoofing attacks. This is aligned with the dynamics of the GOOSE protocol as described in literature [4,20]

 doi:10.20944/preprints202409.1170.v

4.2.3. Dataset Structure

In this section, we describe the organization of the dataset. For the sake of simplicity, traffic samples were stored in the libpcap-ng format, commonly used for traffic captures. Traffic samples were also exported to CSV format. Next we present a brief description of the variables stored in the CSV files.

No: This is a sequence number assigned by the traffic capture tool. This is a merely informational and descriptive field of the record. **Time:** This is a time mark, relative to the start of the capture process. It is also a merely informationational and descriptive field of the record. **Source:** This is the source MAC address of the frame associated to the given record.

Destination: This is the destination MAC address of the frame associated to the given record. It is worth to remark that for protocols such a GOOSE, this address represents a layer 2 multicast group.

Protocol: This is the EtherType field of the frame associated to the corresponding record.

Boolean: This is a variable that indicates whether the GOOSE message contained in the frame is reporting changes of states.

Info: This is an informational attribute containing a description of the correspoding frame associated to the record. It might contain information about devices, type of GOOSE message and error notifications. It is important to remark that this is not an actual protocol field.

Length: This value corresponds to the size of the corresponding frame in bytes.

4.2.4. Overview of Communication Protocols

The dataset built in this work contains traffic samples of the typical communication protocols present in a DS. That is, the dataset contains synchronization messages, scout messages, and redundancy protocols in addition to GOOSE, SMV/SV and PTP messages.

Table 1 presents a general overview of the distribution of packet counts for the communication protocols captured in the dataset during the experimental scenarios. It also contains information about other protocols associated to specific functions such as ARP (Address Resolution Protocol), PRP (Parallel Redundancy Protocol), HSR (High-availability Seamless Redundancy), NTP (Network Time Protocol) and PTPv2. Regarding GOOSE messaging, we discriminate GOOSE messages sent by each of the three IEDs of the testbed. Therefore, the table shows six set of messages (three pairs), associated with signals for the protection and control in each device.

The table details the number of messages (CM: Count Message) observed in each analysis scenario. Particularly, the GOOSE messages have been separated according to their multicast MAC address to highlight the distinctive behavior of the protection and control messages. In this overview, the information associated to traffic attacks is highlighted in red.

 doi:10.20944/preprints202409.1170.v

Table 1. Communication protocols in the dataset

	Message Type	Scenario 1 CM 1	Scenario 2 CM 2	Scenario 3 CM 3	Scenario 4 CM 4
Digital Substation Protocols	**GOOSE**	365	401	391	487
	IED 1: Iec-Tc57_01:00:15	**61**	**98**	**88**	**184**
	IED 1: Iec-Tc57_01:00:16	62	61	61	61
	IED 2: Iec-Tc57_01:00:40	61	61	61	61
	IED 2: Iec-Tc57_01:00:41	61	61	61	61
	IED 3: Iec-Tc57_01:00:90	60	60	60	60
	IED 3: Iec-Tc57_01:00:91	60	60	60	60
	HSR/PRP	295	294	294	295
	NTP	14	14	15	13
	PTPv2	900	900	900	900
Other Network Protocols	**ARP**	73	74	76	75
	BROWSER	0	1	0	1
	LLDP	10	11	10	10
	LLMNR	4	12	0	0
	MDNS	24	44	0	0
	NBNS	0	3	0	0
	PRES	10	10	10	10
	SSDP	10	12	8	11
	STP	150	150	151	150
	TCP	34	34	37	34
	TPKT	1	1	0	1
	UDP	42	14	0	42
	TOTAL	**1932**	**1975**	**1892**	**2029**

CM: Count Message - GOOSE Message Iec-Tc57- 01:00:15 is spoofing attack object

5. Dataset Analysis

In this section, we present a detailed description of the GOOSE messaging present in the dataset, in the different scenarios previously described. The dataset generated in this project is available at https://dx.doi.org/10.21227/jjv5-qg20

5.1. Behavior of GOOSE Messages in the Dataset

In this part, we describe the behavior of the GOOSE messaging within each scenario. In order to make clearer the understanding of each scenario, we present a figure complementing the explanation of each one. The labels used in each figure are defined as follows:

Trip Count: The number of messages associated to state changes. These events of state changes are associated to the corresponding triggers configured for each scenario.

Attack GOOSE count: The number of GOOSE messages generated as part of the spoofing attack.

GOOSE_Ctrl: It marks the generation of control messages.

GOOSE_Prot: It marks the generation of protection messages.

GOOSE_Attack: It marks the starting of the spoofing attack.

Stable state: It marks the time window of stable state operation of the DS. During this period of time, there are not either faults or anomalies. GOOSE messages associated to protection operations will indicate false, since no protection has been triggered.

Trip (Yellow shading): It marks the time window where *trip* messages are generated in respose to a fault event. GOOSE messages associated to protection operations will have a true value.

T (enclosed within a green circle): It indicates the time mark when a *trip* action is generated from the protection IED in response to an electrical fault event.

R (enclosed within a red circle): It indicates the time mark when a reset action is generated from the protection IED in response to a recovery in the electrical system.

Next, we present a detailed description of each escenario and the results obtained with the experiments performed in each case.

 doi:10.20944/preprints202409.1170.v

5.1.1. Scenario 1: Stable Operation of the Electrical System

Figure 5 illustrates the flow of GOOSE messaging during the stable state or stable operation. The figure shows the behavior in state stable of the GOOSE messages: those associated with control functions (in grey) and the protection function (in yellow), as published by one of the IED devices in the DS (e.g. the IED 1). The system starts from an operating state with no-fault events. This is indicated through a boolean value contained in GOOSE messages which is set to False.

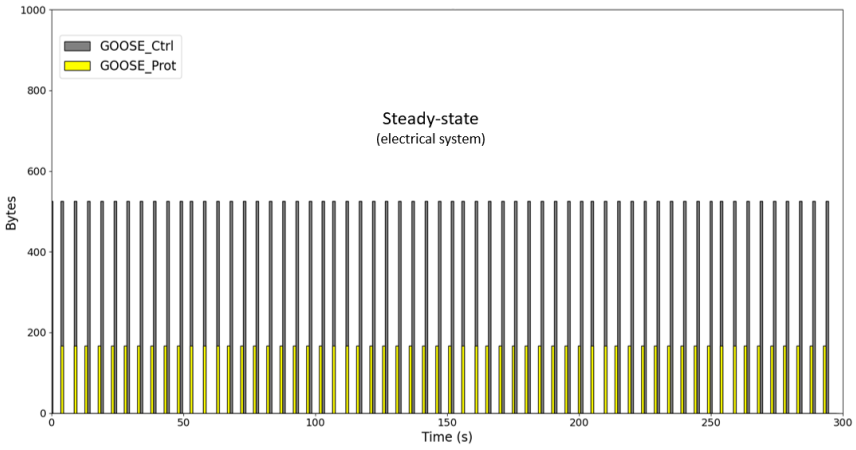

Figure 5. Scenario 1: GOOSE traffic in Stable Operation

Figure shows that frames associated to protection packets have a size of 167 bytes whereas framses associated to control packets have a size of 526 bytes. This difference lies in the amount of data contained in the GOOSE packets, which tends to be larger in control packets. The IEC 61850 standard defines protection messages as time critical. Hence, they contain only the information required for the particular operation. On the other hand, control messages have a variable length field containing information for the monitoring of the associated device, or control commands required to execute a particular operation. Variations in the size of these messages or in the frequency of their transmission might indicate changes in the behavior of the system.

5.1.2. Scenario 2: Electrical System Failure Events

In this scenario, several protection operations were engaged in the system, in response to induced failure events. Figure 6 shows different states of the electrical system, starting with a stable operation, followed by a fault event which starts with the activation of the protection function, and the resulting generation of GOOSE *trip* messages. This can be observed at time marks T at 55s, 115s, 195s, and 280s. Upon system recovery, new changes of state occur returning the system to normal condition. This can be observed at time marks at R at 76s, 143s, 222s, and 283s. Times for the generation of trigger and reset events are random. The figure shows how state changes are actually associated to protection messages. Hence, GOOSE control messages are transmitted periodically, as described in the previous scenario.

 doi:10.20944/preprints202409.1170.v

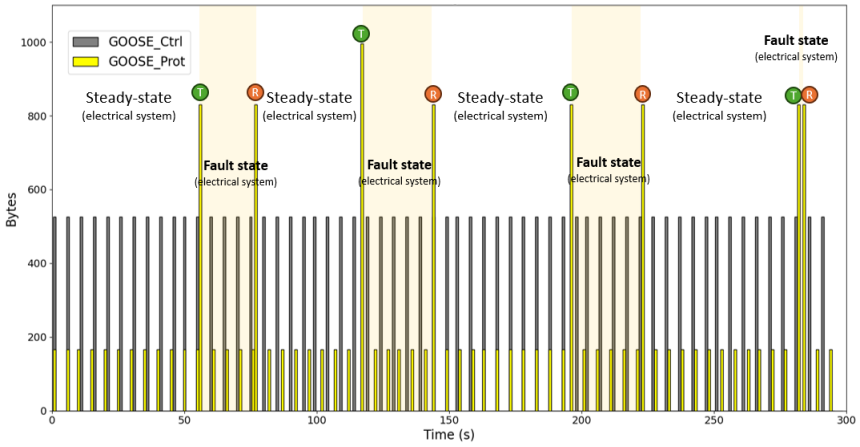

Figure 6. Scenario 2: GOOSE traffic during stable state, and upon the occurrence of some real failures in the electrical system

Figure shows eight state changes, associated with the occurrence of four fault events (T) and their corresponding resets (R). Each state change in GOOSE messaging is characterized by an increment in the number of messages retransmitted within time windows of 500ms. This is expressed with an increment in the size of the GOOSE protection messages due to the inclusion of the corresponding fields. This behavior is defined by the standard, in order to guarantee subscription, even during failures, since packet losts might occur due to congestion.

In real DS infrastructures, fault events do not have a fixed duration. We reproduced this behavior in our testbed by inducing failures with random duration. It is important to remark that this scenario does not contain attacks but failures, which can be conceived as events that might occur during the normal operation of the DS. This scenario is useful to determine a base line of the behavior of the DS when dealing with failures not attributable to cyber attacks.

5.1.3. Scenario 3: Spoofing Attack during Stable Operation of the Electrical System

In this scenario, spoofing attacks are generated while the DS is in stable operation. These spoofing attacks introduce artificial failure events due to the false messages injected. Figure 7 illustrates the behavior of the GOOSE messaging during stable operation in a particular IED, which is ultimately resembled in the spoofing attack. The figure displays with red bars the spoofed GOOSE messages. There are 25 messages associated to the attack, which are separated in two groups. A first group is generated after 120 seconds and a second group is generated after 210 seconds. The duration of the attack is random. Hence, the change state triggering varies in each case.

 doi:10.20944/preprints202409.1170.v

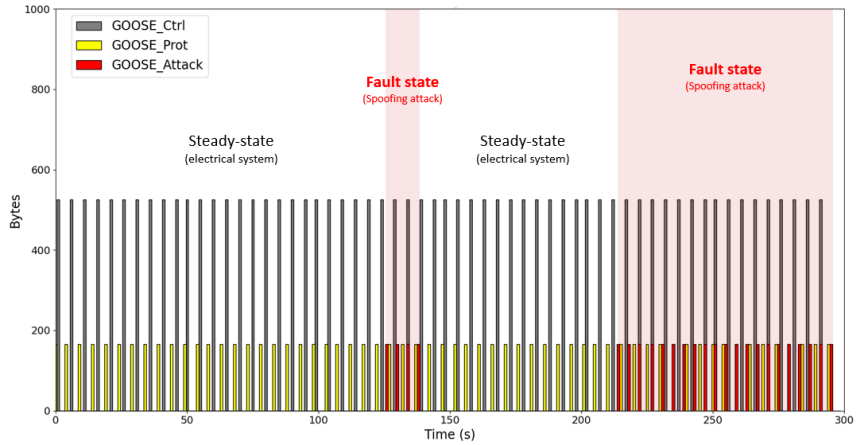

Figure 7. Scenario 3: GOOSE traffic during stable operation, and upon the occurrence of failure events due to Spoofing attacks

In this scenario, there are not real failure events, understood as those attributable to operational electrical events. The GOOSE spoofing attacks do cause the observed failure events. Hence, it can be observed that spoofing attacks might compromise the real stability of a DS.

5.1.4. Scenario 4: Spoofing Attack in the Presence of Electrical System Failure Events

This scenario shows the combination of actual failure events with spoofing attacks. Figure 8 displays this scenario which contains 14 state changes associated to real failure events associated to operational conditions. That is, seven fault events with their corresponding GOOSE *trip* messages (T) and their restoration events (R). In this scenario, 59 spoofing attacks were generated.

Similarly to the previous scenario, attacks were separated in two groups with specific characteristics. The first set of attacks introduces a change of state at time mark of 94s. This change of state induces the triggering of protection mechanisms, which remain activated during 90s, until time mark at 185s. During this time window, there were six actual state changes (*trip* and restore) not identified in the monitoring of the DS. This means that the protection system was not actually disengaged, which is a behavior consequence of the spoofing attack. The second attack scenario is started around the time mark at 240s. The attack starts shortly after the occurrence of a GOOSE *trip* message associated to a real operational event. Hence, the attack causes the system to be restored to a fake stable state operation which is kept even upon the generation of six state changes during the attack.

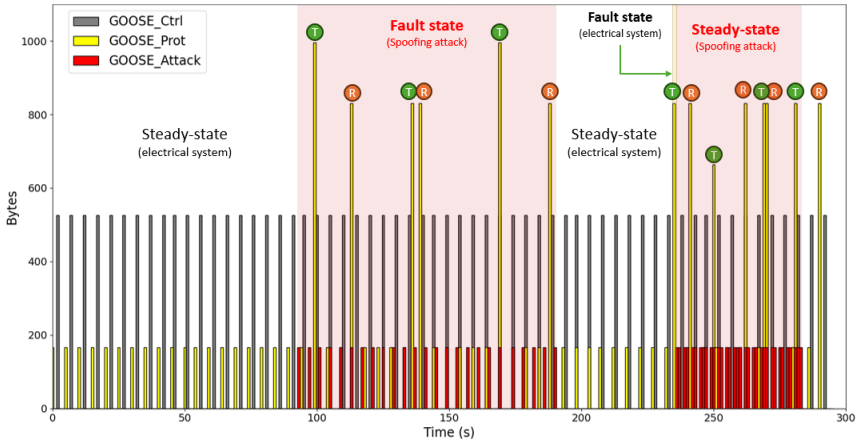

Figure 8. Scenario 4: Traffic of GOOSE protocol, Steady-State, Real Failure events, and Spoofing Failure events.

 doi:10.20944/preprints202409.1170.v

In this scenario, the Spoofing attack indeed degenerates towards a denial of service attack. The figure shows how the spoofing messages prevent the system to adequately process the messages generated by the real events. This causes that the protection and control mechanisms of the DS become isolated.

5.2. General Remarks

From the development of the dataset presented in this paper, there are several features of the traffic associated to the protocols in a DS which became evident. The identification of these features can be considered the initial step and reference point for the development of cyber security solutions to cope with the security threats that might affect critical infrastructures such as DSs. First, it is evident that a prominent property of the traffic, specially during stable operation periods is its regularity. GOOSE messages flow through the network in regular intervals, with an almost constant Inter Arrival Time. Secondly, in addition to direct features that can be extracted from traffic captures, such as the values of protocol fields indicating type of message, source and destination addresses, sequence numbers, packet sizes, among others, there are other useful features that can be derived. For instance, the packet Inter Arrival Time previously mentioned, can be used to determine the regularity of the packet transmission or the standard response when facing failure events. Behaviors that deviate from these dynamics observed in regular operational conditions might trigger alarms as possible indication of anomalies or events associated to threat exploitation.

The dataset developed in this work, collected from a real testbed, with real DS equipment and reproducing actual operational conditions, is an important avenue to understand the behavior of DS infrastructures from the study of their traffic. It allows also the understanding of protocols such as GOOSE and SV. This understanding will allow the development of different solutions, not only in the cyber security field, but also in the areas of traffic engineering and quality of service. This understanding will clearly contribute in the research of integration of novel network technologies, which will increase the resilience of performance of critical infrastructures in general, and DSs in particular.

6. Conclusions and Future work

In this paper, we have presented the development of a dataset in the context of a Digital Substation. In contrast to some of the contributions available in literature, this dataset has been developed on a testbed with actual real devices, which reproduce the operational dynamics that can be observed in real world substations. Hence, the traffic samples captured and included in the dataset really represent the behavior and dynamics of the communication protocols such as GOOSE and SMV/SV that are used in Digital Substations.

Initial observations allowed to understand the behavior of protocols in stable operation of the infrastructure, and upon the occurrence of failure events due to operational conditions. Through these observations, it was evident that the traffic in these conditions presents an evident regularity. Properties such as the packet Inter Arrival Time tend to be constant, and according to the IEC 61850 standard, there are some traffic bursts that are generated upon failure events, whose properties are predictable.

By creating different experimental scenarios, including the generation of attacks in combination with failure events due to operational conditions, we detected two important facts. First, Digital Substations are prone to Spoofing and False Data Injection attacks. Hence, and given their criticality, these are infrastructures that must be protected, specially from the point of view of their communication networks, in order to avoid further consequences. Second, these two type of attacks are not only dangerous by themselves, but they can cause Denials of Service and cascade failures which are even more catastrophic and cause incalculable effects to the users and services relying on Digital Substations.

Our literature survey has shown us that there are some related work on datasets on the context of critical infrastructures, including some examples in the field of electrical facilities. Despite the

 doi:10.20944/preprints202409.1170.v

relevance and value of these works, a noticeable drawback is that they have been developed in simulated environments, in ideal conditions and with emulated devices and infrastructures. Hence, in our work, we present a contribution based on experiments conducted on real equipment, with real operational conditions and the reproduction of actual cyber attacks, which reproduces faithfully the operation of a real world infrastructure. We consider this data set will be an important asset for further research in the topic of network traffic in the context of Digital Substations, and the development of novel solutions in fields such cyber security, quality of service and traffic engineering.

The dataset presented in this work focused mainly on the GOOSE protocol, given its relevance and criticality in the operation of Digital Substations. But GOOSE is clearly one among other relevant protocols within these infrastructures. As future work, we plan to extend this dataset with the analysis of other protocols such as SMV/SV, MMS and PTP, which according to the literature in the area, are other important sources of vulnerabilities for Digital Substations. It is more than evident the importance of protecting adequately these infrastructures, since they are clearly the most critical among the critical infrastructures, due to the existing interdependency. This research we performed is therefore a very first step, which we hope become extended by the collaborative work of the scientific community interested in these areas.

Author Contributions: In the paper, the various authors played an essential role in the culmination of the research. Although a collaborative effort was sought, each author undertook specific tasks. The principal author, Oscar Tobar-Rosero, was responsible for consolidating a significant portion of the research and article writing. Omar Roa-Romero was in charge of developing the tool implemented for the generation of attacks. Germán Rueda-Carvajal conducted corrections and made contributions to the introduction and background section. Alexander Leal-Piedrahita conducted general revisions and contributed to enhancements in the graphics and styles used in the document. Finally, Juan Botero-Vega, Sergio Gutiérrez-Betancur, John W. Branch-Bedoya, and Germán Zapata-Madrigal served as investigative leaders, overseeing, correcting, and providing necessary resources for the development of the associated research'. All authors have read and accepted the manuscript.

Funding: This research was funded by the General System of Royalties - Colombia (BPIN 2020000100381) and has been performed under the project "Fortalecimiento de las capacidades de investigación y desarrollo en gestión de riesgos cibernéticos en la infraestructura crítica del sector eléctrico desde las instituciones de educación superior públicas de Medellín". The APC was funded by the same projet.

Data Availability Statement: The availability of data generated during the research process is consolidated on the GitHub platform, which, being widely recognized in the community, provides an infrastructure enabling anyone interested to access the data remotely and without geographical restrictions. This accessibility extends the reach of the research by allowing other researchers, students, and professionals interested in the topic to access and utilize the data in their projects. To access the repository containing all the information generated in this study, click the following link: https://dx.doi.org/10.21227/jjv5-qg20.

Acknowledgments: This paper has been supported by the General System of Royalties - Colombia (BPIN 2020000100381). This work has been developed under the project "Fortalecimiento de las capacidades de investigación y desarrollo en gestión de riesgos cibernéticos en la infraestructura crítica del sector eléctrico desde las instituciones de educación superior públicas de Medellín".

Conflicts of Interest: The authors of this study have no conflicts of interest in the research results. The funding comes from public resources, which guarantees impartiality in the process. Therefore, the only interest of the participants in this article is to deepen, broaden, and facilitate the study of future research related to the topic discussed here

Abbreviations

The following abbreviations are used in this manuscript:

CI	Critical Infrastructure
DDoS	Distributed Denial of Service
DoS	Denial of Service
DS	Digital Substation
GOOSE	Generic Object Oriented Substation Event
HMI	Human Machine Interface
ICT	Information and Communication Technologies
IEC	International Electrotechnical Commission
IED	Intelligent Electronical Device
IT	Information Technologies

 doi:10.20944/preprints202409.1170.v

LAN	Local Area Network
MITM	Man-in-the-middle
MMS	Manufacturing Message Specification
MU	Merging Unit
NFV	Network Function Virtualization
OT	Operation Technologies
P&C	Protection and Control
PTP	Precision Time Protocol
QoS	Quality of Service
SCADA	Supervisory Control and Data Acquisition
SDN	Software-Defined Networking
SV/SMV	Sampled Values / Sampled Measured values

References

1. Aftab, M.A.; Hussain, S.S.; Ali, I.; Ustun, T.S. IEC 61850 based substation automation system: A survey. *International Journal of Electrical Power and Energy Systems* **2020**, *120*, 106008. doi:10.1016/J.IJEPES.2020.106008.
2. Nair, M.M.; Tyagi, A.K.; Sreenath, N. The Future with Industry 4.0 at the Core of Society 5.0: Open Issues, Future Opportunities and Challenges. 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1–7. doi:10.1109/ICCCI50826.2021.9402498.
3. Quincozes, S.E.; Albuquerque, C.; Passos, D.; Mossé, D. A survey on intrusion detection and prevention systems in digital substations. *Computer Networks* **2021**, *184*, 107679.
4. Wang, S.; Yang, F.; Yan, X.; Liu, T. Analysis of GOOSE message and the engineering application for GOOSE message in the intelligent substation. *The Journal of Engineering* **2020**, *2020*, 207–212. doi:https://doi.org/10.1049/joe.2018.5208.
5. Hoyos, J.; Dehus, M.; Brown, T.X. Exploiting the GOOSE protocol: A practical attack on cyber-infrastructure. 2012 IEEE Globecom Workshops, 2012, pp. 1508–1513. doi:10.1109/GLOCOMW.2012.6477809.
6. Roa, O.; Botero, J.F.; Gutierrez-Betancur, S.A.; Tobar-Rosero, O.A. GOOSEAttacker: Synthetic Attack Generation Tool for IEC61850. 2023 IEEE Latin-American Conference on Communications (LATINCOM), 2023, pp. 1–6. doi:10.1109/LATINCOM59467.2023.10361897.
7. Huseinović, A.; Mrdović, S.; Bicakci, K.; Uludag, S. A Survey of Denial-of-Service Attacks and Solutions in the Smart Grid. *IEEE Access* **2020**, *8*, 177447–177470. doi:10.1109/ACCESS.2020.3026923.
8. Jokar, P.; Arianpoo, N.; Leung, V.C. Spoofing detection in IEEE 802.15.4 networks based on received signal strength. *Ad Hoc Networks* **2013**, *11*, 2648–2660. doi:https://doi.org/10.1016/j.adhoc.2013.04.015.
9. Oliveira, A.d.S.; Santos, H. Continuous Industrial Sector Cybersecurity Assessment Paradigm: Proposed Model of Cybersecurity Certification. 2022 18th International Conference on the Design of Reliable Communication Networks (DRCN), 2022, pp. 1–6. doi:10.1109/DRCN53993.2022.9758022.
10. Burgetová, I.; Matoušek, P.; Ryšavý, O. Anomaly Detection of ICS Communication Using Statistical Models. 2021 17th International Conference on Network and Service Management (CNSM), 2021, pp. 166–172. doi:10.23919/CNSM52442.2021.9615510.
11. Malik, H.; Alotaibi, M.A.; Almutairi, A. Cyberattacks identification in IEC 61850 based substation using proximal support vector machine. *Journal of Intelligent and Fuzzy Systems* **2022**, *42*, 1213–1222. doi:10.3233/JIFS-189783.
12. Elmasry, A.; Albaseer, A.; Abdallah, M. OpenPLC and lib61850 Smart Grid Testbed: Performance Evaluation and Analysis of GOOSE Communication. 2023 International Symposium on Networks, Computers and Communications (ISNCC), 2023, pp. 1–6. doi:10.1109/ISNCC58260.2023.10323659.
13. Ring, M.; Wunderlich, S.; Scheuring, D.; Landes, D.; Hotho, A. A survey of network-based intrusion detection data sets. *Computers & Security* **2019**, *86*, 147–167. doi:10.1016/j.cose.2019.06.005.
14. Commission, I.E.; others. Communication networks and systems for power utility automation. *IEC Std* **2013**, *61850*.
15. Tobar Rosero, O.A.; Pérez González, E.; Botero Vega, J.F.; Zapata Madrigal, G.; Roa, O.; Candelo-Becerra, J.E.; García Sierra, R. Digital Substations and Cybersecurity in the Transformation of the Electricity Sector. 2023 IEEE Colombian Caribbean Conference (C3), 2023, pp. 1–6. doi:10.1109/C358072.2023.10436315.
16. Mesmaeker, I.D. Trends in protection and substation automation systems and feed-backs from CIGRE activities. *IET Conference Publications* **2008**, pp. 1–8. doi:10.1049/CP:20080001.

17. Apostolov, A. Impact of IEC 61850 on the interoperability and reliability of protection schemes. 2013 IEEE Power & Energy Society General Meeting, 2013, pp. 1–5. doi:10.1109/PESMG.2013.6673051.
18. Musil, P.; Mlynek, P. Overview of communication scenarios for IEC 60870-5-104 substation model. 2020 21st International Scientific Conference on Electric Power Engineering (EPE). IEEE, 2020, pp. 1–4.
19. Song, E.Y.; FitzPatrick, G.J.; Lee, K.B. Smart sensors and standard-based interoperability in smart grids. *IEEE Sensors Journal* **2017**, *17*, 7723–7730.
20. León, H.; Montez, C.; Valle, O.; Vasques, F. Real-Time Analysis of Time-Critical Messages in IEC 61850 Electrical Substation Communication Systems. *Energies* **2019**, *12*. doi:10.3390/en12122272.
21. Vahidi, S.; Ghafouri, M.; Au, M.; Kassouf, M.; Mohammadi, A.; Debbabi, M. Security of wide-area monitoring, protection, and control (WAMPAC) systems of the smart grid: A survey on challenges and opportunities. *IEEE Communications Surveys & Tutorials* **2023**, *25*, 1294–1335.
22. Hunt, R.; Dalmeny, C.; Geor, M. Time Synchronisation for IEC 61850 Systems. In *IEC 61850 Principles and Applications to Electric Power Systems*; Springer, 2023; pp. 95–130.
23. Lozano, J.C.; Koneru, K.; Ortiz, N.; Cardenas, A.A. Digital substations and iec 61850: A primer. *IEEE Communications Magazine* **2023**, *61*, 28–34.
24. Zakonjšek, J. CT/VT Sampled Value Acquisition Applied to IEC 61850. In *IEC 61850 Principles and Applications to Electric Power Systems*; Springer, 2023; pp. 301–319.
25. Balakrishnan, K.; Dhanalakshmi, R.; Sinha, B.B.; Gopalakrishnan, R. Clock synchronization in industrial Internet of Things and potential works in precision time protocol: Review, challenges and future directions. *International Journal of Cognitive Computing in Engineering* **2023**, *4*, 205–219.
26. Tightiz, L.; Yang, H. A comprehensive review on IoT protocols features in smart grid communication. *Energies* **2020**, *13*, 2762.
27. Hou, J.; Hu, C.; Lei, S.; Hou, Y. Cyber resilience of power electronics-enabled power systems: A review. *Renewable and Sustainable Energy Reviews* **2024**, *189*, 114036.
28. Ağın, A.; Demirören, A.; Usta, Ö. A Novel Approach for Power System Protection Simulation via the IEC 61850 Protocol. *IEEE Access* **2024**.
29. Rajkumar, V.S.; Tealane, M.; Ştefanov, A.; Palensky, P. Cyber Attacks on Protective Relays in Digital Substations and Impact Analysis. 2020 8th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, 2020, pp. 1–6. doi:10.1109/MSCPES49613.2020.9133698.
30. Maziku, H.; Shetty, S.; Nicol, D.M. Security risk assessment for SDN-enabled smart grids. *Computer Communications* **2019**, *133*, 1–11.
31. Akbarzadeh, A.; Erdodi, L.; Houmb, S.H.; Soltvedt, T.G.; Muggerud, H.K. Attacking IEC 61850 Substations by Targeting the PTP Protocol. *Electronics* **2023**, *12*. doi:10.3390/electronics12122596.
32. Rashid, M.T.A.; Yussof, S.; Yusoff, Y.; Ismail, R. A review of security attacks on IEC61850 substation automation system network. *Conference Proceedings - 6th International Conference on Information Technology and Multimedia at UNITEN: Cultivating Creativity and Enabling Technology Through the Internet of Things, ICIMU 2014* **2015**, pp. 5–10. doi:10.1109/ICIMU.2014.7066594.
33. Rajkumar, V.S.; Tealane, M.; Ştefanov, A.; Palensky, P. Cyber Attacks on Protective Relays in Digital Substations and Impact Analysis. 2020 8th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, 2020, pp. 1–6. doi:10.1109/MSCPES49613.2020.9133698.
34. Alshaibi, A.; Al-Ani, M.; Al-Azzawi, A.; Konev, A.; Shelupanov, A. The comparison of cybersecurity datasets. *Data* **2022**, *7*, 22.
35. Zheng, M.; Robbins, H.; Chai, Z.; Thapa, P.; Moore, T. Cybersecurity research datasets: Taxonomy and empirical analysis. 11th USENIX Workshop on Cyber Security Experimentation and Test (CSET 18), 2018.
36. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009, pp. 1–6. doi:10.1109/CISDA.2009.5356528.
37. Thakkar, A.; Lohiya, R. A review of the advancement in intrusion detection datasets. *Procedia Computer Science* **2020**, *167*, 636–645.
38. Abdulraheem, M.H.; Ibraheem, N.B. A DETAILED ANALYSIS OF NEW INTRUSION DETECTION DATASET. *Journal of Theoretical and Applied Information Technology* **2019**, *97, 17*, 4519–4537.
39. Panigrahi, R.; Borah, S. A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems. *International Journal of Engineering & Technology* **2018**, *7*, 479–482. doi:10.14419/IJET.V7I3.24.22797.

40. Mittal, M.; Kumar, K.; Behal, S. Deep learning approaches for detecting DDoS attacks: A systematic review. *Soft computing* **2023**, *27*, 13039–13075.
41. Quincozes, S.E.; Albuquerque, C.; Passos, D.; Mossé, D. ERENO: A Framework for Generating Realistic IEC–61850 Intrusion Detection Datasets for Smart Grids. *IEEE Transactions on Dependable and Secure Computing* **2023**.
42. Li, W.; Meng, W.; Kwok, L.F. Surveying trust-based collaborative intrusion detection: state-of-the-art, challenges and future directions. *IEEE Communications Surveys & Tutorials* **2021**, *24*, 280–305.
43. Shin, H.K.; Lee, W.; Yun, J.H.; Min, B.G. Two ICS Security Datasets and Anomaly Detection Contest on the HIL-based Augmented ICS Testbed. Proceedings of the 14th Cyber Security Experimentation and Test Workshop; Association for Computing Machinery: New York, NY, USA, 2021; CSET '21, p. 36–40. doi:10.1145/3474718.3474719.
44. Goh, J.; Adepu, S.; Junejo, K.N.; Mathur, A. A Dataset to Support Research in the Design of Secure Water Treatment Systems. Critical Information Infrastructures Security; Havarneanu, G.; Setola, R.; Nassopoulos, H.; Wolthusen, S., Eds.; Springer International Publishing: Cham, 2017; pp. 88–99.
45. Perales Gomez, A.L.; Fernandez Maimo, L.; Huertas Celdran, A.; Garcia Clemente, F.J.; Cadenas Sarmiento, C.; Del Canto Masa, C.J.; Mendez Nistal, R. On the Generation of Anomaly Detection Datasets in Industrial Control Systems. *IEEE Access* **2019**, *7*, 177460–177473. doi:10.1109/ACCESS.2019.2958284.
46. Adepu, S.; Kandasamy, N.K.; Mathur, A. EPIC: An Electric Power Testbed for Research and Training in Cyber Physical Systems Security. Computer Security; Katsikas, S.K.; Cuppens, F.; Cuppens, N.; Lambrinoudakis, C.; Antón, A.; Gritzalis, S.; Mylopoulos, J.; Kalloniatis, C., Eds.; Springer International Publishing: Cham, 2019; pp. 37–52.
47. Biswas, P.P.; Tan, H.C.; Zhu, Q.; Li, Y.; Mashima, D.; Chen, B. A Synthesized Dataset for Cybersecurity Study of IEC 61850 based Substation. 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2019, pp. 1–7. doi:10.1109/SmartGridComm.2019.8909783.
48. Yang, Q.; Hao, W.; Ge, L.; Ruan, W.; Chi, F. FARIMA model-based communication traffic anomaly detection in intelligent electric power substations. *IET Cyber-Physical Systems: Theory & Applications* **2019**, *4*, 22–29. doi:https://doi.org/10.1049/iet-cps.2018.5052.
49. Wang, X.; Fidge, C.; Nourbakhsh, G.; Foo, E.; Jadidi, Z.; Li, C. Anomaly Detection for Insider Attacks From Untrusted Intelligent Electronic Devices in Substation Automation Systems. *IEEE Access* **2022**, *10*, 6629–6649. doi:10.1109/ACCESS.2022.3142022.
50. Aoufi, S.; Derhab, A.; Guerroumi, M. Survey of false data injection in smart power grid: Attacks, countermeasures and challenges. *Journal of Information Security and Applications* **2020**, *54*, 102518. doi:https://doi.org/10.1016/j.jisa.2020.102518.