Modeling and Analysis of Smart Grids as Cyber-Physical Systems

Ronak Feizimirkhani

UNIVERSITÉ GRENOBLE ALPES

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Automatique Productique**

Arrêté ministériel : 7 août 2006

Présentée par
Ronak FEIZIMIRKHANI

Thèse dirigée par **Yvon BÉSANGER** et
co-encadrée par **Antoneta Iuliana BRATCU**

préparée au sein des laboratoires
GIPSA-Lab et G2ELab
dans **l'École Doctorale Electronique Electrotechnique Automatique et Traitement du Signal(EEATS)**

Modeling and analysis of the smart grid as a cyber physical system

Thèse soutenue publiquement le **4 Mars 2020**,
devant le jury composé de :

Thierry DIVOUX
Professeur, Univerités de Lorraine, Examinateur, Président du jury
Lars NORDSTÖM
Professeur, KTH Royal Institute of Technology of Stockholm, Sweden, Rapporteur
Yann LABIT
Professeur, Université de Toulouse, Rapporteur
Yvon BÉSANGER
Professeur, Grenoble INP, Directeur de thèse
Antoneta Iuliana BRATCU
Maître de Conférences, Grenoble INP, Co-encadrante

Acknowledgements

First of all, I would like to thank the members of the jury who gave me the great honor to evaluate this thesis: Mr. Thierry Divoux, professor at University of Lorraine, Mr. Lars Nordstöm, professor at KTH Royal Institute of Technology of Stockholm, Mr. Yann Labit, professor at University of Toulouse. I thank them so much for the time they have devoted to this research, despite all their responsibilities. My sincere thanks are also dedicated to my supervisor, professor Yvon Bésenger and my co-supervisor, associated professor Antoneta Iuliana BRATCU. I thank for the confidence they gave me through their agreement on my thesis supervision. Without their scientific guidance and constant feedbacks this Ph.D. would not have been achieved. I would like to thank them for their constructive advice, and their valuable ideas proposed during this work which have been very useful in my research process. I would also like to express my great thanks to the SLR team of GIPSA-Lab and SYREL of G2Elab for welcoming me to their teams. I thank them for their continuous accompanying. I am grateful to all my doctoral friends and other colleagues of GIPSA-Lab and G2ELab for their collaboration. All my thanks to my best friends for their support and encouragement and the nice moments after work and scientific exchanges during my study which helped enriching my experience. I would also like to say a heartfelt thank to my parents for always believing in me and encouraging me to fallow my dreams. And finally to my dear husband, Amer who has been by my side throughout this Ph.D., living every single minute of it, and without whom, I would not have had the courage to embark on this journey in the first place. It is obvious that without their supports, my success would not have been possible.

Contents

Table of the acronyms **xi**

General introduction **1**

1 Background **7**

1.1 Context of smart grid 8

1.2 Distribution energy generation and control 21

1.3 Conclusion 27

2 State of the art on the communication traffic modeling in smart grid applications **29**

2.1 Interaction of energy layer and communication layer in smart grids 29

2.2 Researches in smart grid communications 37

2.3 Research motivation and objectives 46

2.4 Conclusion 47

3 Introduction to the IEC 61850 standard, traffic modeling methods and transmission delay estimation **49**

3.1 Background: IEC 61850 protocol 50

3.2 Modeling methods and influence of delay factor 57

3.3 The choice of methods for traffic and delay estimation 64

3.4 Conclusion 71

4 Traffic modeling with ARIMA and application of delay estimation method **73**

4.1 Procedure to obtain an ARIMA model 74

4.2 Maximum message transmission delay 85

4.3 Conclusion 97

5 Modeling of data flows and delay estimation within illustrative case studies 99

5.1 ARIMA modeling of MATLAB®/Simulink® co-simulation data flows 100

5.2 ARIMA modeling of data flows in a real-time traffic generator 110

5.3 Test-bench validation of a delay estimation analytical model 116

5.4 Conclusion 132

Conclusion 135

Bibliographie 152

List of Figures

1.1 The hierarchical system of the present power grids [85]. 8

1.2 A smart grid perspective with different components [78]. 10

1.3 The communication infrastructure for smart grid application [160]. 11

1.4 The motivations of emerging smart grid through the communication infrastructures [132]. 12

1.5 Different domains for information security [40]. 16

1.6 A typical substation automation system and its classical communication protocols [7]. 17

1.7 A representation of the IEC 61850 standard parts [140]. 21

1.8 Energy resupply with a distributed energy generation unit [20]. 22

1.9 Object modeling of the IEC 61850 data, illustrates the relations between physical device and logical device [66]. 25

1.10 The concept of physical device with path to data attributes of logical nodes. . 26

1.11 The default Logical node LLN0 within default logical device LPHD0. 26

1.12 Multiple LNs cooperating to build functions [140]. 26

2.1 An example for interdependency effects in a typical cyber-power network [43]. In this typical network LAN stands for Local Area Network, and HMI means Human Machine Interface. 32

2.2 An example of the INEI effect in a typical cyber-power network [43]. 32

2.3 GWAC interoperability model. 34

2.4 A sample object model defined in IEC 61850 [8]. 36

3.1 Substation automation operating levels [128]. 51

3.2 OSI mapping of IEC 61850 protocol [137]. 51

3.3 IEC 61850 communication architecture [137]. 52

3.4 The TCP/IP reference model [150]. 52

3.5 Retransmission mechanism in GOOSE [128]. 53

3.6 IEC 61850 different data distributions over time [161]. 57

3.7 Model of a single queue. 59

3.8 ARIMA model as a linear filter. 65

3.9 An example of a cumulative function $F(t)$ related to a typical flow upper bounded by the arrival curve $\alpha(t)$. 69

4.1 A typical time series with an improper choice of the sampling period [158]. . 77

4.2 A time series which shows an upward trend, and related ACF diagram [25]. . 78

4.3 A seasonal time series and related ACF diagram [25]. 79

4.4 Example of ACF and PACF diagram of an ARMA(1,2) [21]. 80

4.5 Flowchart of the ARIMA model implementation and model validation steps. . 84

4.6 Overall inversion process in order to obtain the total mathematical expression. 85

4.7 A typical example of SCN port connection, including one VLAN group. . . . 86

4.8 Di-graph of the physical and logical connections. 86

4.9 Distribution path of flow 1. 91

4.10 Flowchart detailing the process of computing the message distribution matrix for flow j. 94

5.1 MATLAB®/Simulink® co-simulation model of an electrical distribution grid containing two renewable sources, and a load in interaction with its communication network. 101

5.2 Reactive power over time (kVAR) as the transmitted load reactive power profile. 101

5.3 Number of packets sampled over a sampling period of 2 ms. 102

5.4 The first-order simple difference is applied to the sampled-time series, and mean value is removed. 103

5.5 ACF coefficients decay slowly before removing the linear trend over time. . . 103

5.6 Seasonal trend is removed by seasonal period = 3 s. 103

5.7 ACF/PACF related to the series after simple difference. 104

5.8 The comparison of estimated ARIMA model against preprocessed series. . . . 104

5.9 Minimum MSE *vs.* orders p and q of ARIMA model. 105

5.10 Estimated ARIMA model by min AIC is compared to that obtained for min MSE. 105

5.11 The whiteness of residuals, showing a good estimation of data traffic behavior. 106

5.12 Estimated model is compared to the original sampled time series. 107

5.13 Estimated ARIMA model Bode plots for packet loss probability (PL) = $5\%, 10\%, 15\%$, and 20%, the nominal case corresponding to PL = 0%. 108

5.14 Estimated ARIMA model Bode plots for CHR = $0.3 \cdot 10^6$, $0.4 \cdot 10^6$, and $0.5 \cdot 10^6$ bytes/s, the nominal case corresponding to CHR = 10^6 bytes/s. 109

5.15 Experimental MV/LV substation emulation test bench. 110

5.16 Active power (kW) variation transmitted by GOOSE messages through the network. 111

5.17 Active power (kW) is entered in the Omicron Ramping module as piecewise-constant values by means of the corresponding current and voltage values. . . 111

5.18 Monitoring of the network transmission by the Wireshark packet analyzer. . . 112

5.19 Data traffic sampled each 0.2 ms. 112

5.20 Time series after simple difference application. 113

5.21 Related ACF/PACF diagram indicating existence of a cyclic behavior. 113

5.22 Time series after seasonal difference application. 114

5.23 The auto-correlation function of model residuals is calculated and graphically represented against 25 lags. 114

5.24 The estimated model is compared to the original sampled time series. 115

5.25 Original signal is estimated through all the steps back. 116

5.26 IEDs placement in the load shedding scenario. 117

5.27 Typical active power defined in Premium PLC as the load profile connected to IED3. 118

5.28 Load shedding control process in the Zenon application on PC. 119

5.29 Communication network configuration for the unperturbed scenario. 119

5.30 Di-graph of physical (P) and logical (L) connections in the scenario without perturbation. 120

5.31 Fitting the arrival curve as the upper bound for the first flow generated by IED1 and source port 1. 121

5.32 Flow1 of the scenario without perturbation is monitored in Wireshark, while sending data and receiving acknowledgement back. Time column shows RTT values in seconds. 126

5.33 Communication network configuration for the perturbed scenario. 127

5.34 Di-graph of port connections in the perturbed scenario. 128

5.35 Network architecture of two intelligent substations communicating at the distance. 138

List of Tables

1.1 Comparison between some well-known legacy communication protocols in SAS [7]. 19

3.1 Generation of different message types in SCN [158]. 56

5.1 Estimated delay (μs) corresponding to each flow while passing through different ports. 125

5.2 Estimated delay (μs) compared to the measured values. 127

5.3 Maximum delay experienced by each flow passing through different ports to reach its destination in μs. 131

5.4 Estimated delay in μs through the message distribution is compared to the measured values for the perturbed scenario. 132

Table of the acronyms

ACF	Auto Correlation Function
AC	Alternating Current
ARMA	Auto Regressive Moving Average
AR	Auto Regressive
BCU	Bay Control Unit
MA	Moving Average
ADA	Advanced Distribution Automation
AMI	Advanced Metering Infrastructure
CAGR	Compound Annual Growth Rate
CB	Circuit Breaker
CT	Current Transformer
CPS	Cyber Physical System
CDC	Common Data Classes
DER	Distributed Energy Resource
DS	Disconnect Switch
DNP3	Distributed Network Protocol, version 3
DG	Distributed Generation
DNO	Distribution Network Operator
DSO	Distribution System Operator
DSM	Demand Side Management
ENISA	European Union Agency for Network and Information Security
EPRI	Electric Power Research Institute
EMS	Energy Management System
EPA	Enhanced Performance Architecture
FIFO	First-in First-Out

FD	Fixed Driven
ED	Event Driven
FLISR	Fault Isolation, Location and System Restoration
GSE	Generic Substation Event
GFit	Goodness of Fit
GOOSE	Generic Object-Oriented Substation Event
HW	Holt-Winters
HEMS	Home Energy Management System
HMI	Human Machine Interface
IEC 61850	International Electrotechnical Commission 61850
IED	Intelligent Electronic Device
IoT	Internet of Thing
IP	Internet Protocol
ICT	Information and Communication Technology
ICD	IED Capability Description
ISA	International Society of Automation
LDC	Local Distribution Company
LTE	Long Term Evolution
LRD	Long Range Dependence
LV	Low Voltage
LN	Logical Nodes
LD	Logical Devices
LAN	Local Area Network
CDC	Common Data Classes
CAN	Control Area Network
DC	Direct Current
MAC	Media Access Control
MG	Microgrid

MPPT	Maximum Power Point Tracking
MSE	Mean Squared Error
MMS	Manufacturing Message Specification
MU	Metering Unit
MV	Medium Voltage
MSB	Swedish Civil Contingencies Agency
NIST	National Institute of Standards and Technology
NSF	U.S. National Science Foundation
OSI	Open Systems Interconnection model
OLTC	On Load Tap Changer
OPNET	Optimized Network Engineering Tool
PACF	Partial Auto Correlation Function
PV	Photovoltaic
PHD	Physical Devices
PHIL	Power-Hardware-In-the-Loop/Hardware-In-the-Loop
PLC	Power Line Communication
PMU	Phasor Measurement Unit
PVGP	PV Generation Plant
PDU	Protocol Data Unit
PHEV	Plug-in Hybrid Electric Vehicle
P2P	Point to Point
QoS	Quality of Service
RTT	Round Trip Time
RTI	Run Time Interface
RES	Renewable Energy Sources
SAS	Substation Automation System
SCADA	Supervisory, Control and Data Acquisition
SCL	Structured Control Language

SCN	Substation Communication Network
SV	Sampled measured Value
TCP	Transmission Control Protocol
TC	Technical Committee
UTP	Unshielded Twisted Pair
UCA	Utility Communication Architecture
UML	Unified Modeling Language
VLAN	Virtual Local Area Network
V-ID	VLAN Identifier
VSC	Voltage Source Converter
V2G	Vehicle-to-Grid
VT	Voltage Transformer
WPF	Wind Power Farm
WAN	Wide Area Network
XML	Extensible Markup Language

General introduction

Today, power grid architecture, planning and operation are facing numerous changes whose impact strongly the way to reach the original main aim: supplying electricity to the final consumers from the energy generators. CO_2 emission by conventional generation plants[1] (*e.g.*, coal and lignite-fired steam power plants, and gas-fired plants, *etc.*), significant increase of electricity consumption all over the world, massive insertion of renewable and non-dispatchable energy sources with modification of power flows, inefficiency of the current control-protection systems adapting to these changes, and power plants worn equipment require a viable change of the current electric power systems.

Introduction of Renewable Energy Sources (RESs) to provide a clean energy generation, and Intelligent Electronic Devices (IEDs) to define a distributed monitoring, control, and protection structure lead to a distributed generation and control system, which can efficiently fix deficiencies of the legacy power systems. All these generation-control distribution and intelligence injected into the power system result in the recent concept of smart grid, which changes the production, management, and consumption of electricity. As its first goal, smart grid aims at modernizing the current power grid through a new set of technologies and services in order to make the electricity network more reliable, efficient, secure, and environmentally friendly.

Smart grid is characterized by a two-way flow of electricity and information, creating an automated, extensively distributed energy delivery network. It incorporates into the existing power grid the benefits of distributed computing and communication to deliver real-time information. To be successful, smart grid initially requires collaboration, integration, and interoperability among a variety of technologies and disciplines. Information and Communication Technology (ICT) plays a key role to simplify the achievement of the modern power grid and its services. Existing power grid, integrating ICT, has thus be partially or totally redesigned in order to adopt RESs, and IEDs in a collaborative structure. Coupling energy and information systems in a common structure is a paramount important challenge, since the interactions between them can affect each of them.

ICT is always prone to errors, and any uncertainties of the communication network (*e.g.*, packet loss, message transmission delay, *etc.*) can affect reliability and stability of the power system. Thus, to design a new power system, it is essentially needed to know the relying cyber-physical interactions in the smart grid. A wide range of the associated research problems with smart grid focus on the co-simulation, and modeling of the electrical grid and communication network by means of a common framework to describe the existing interactions [122]. In addition, there are co-simulation, and modeling works which propose methods for performance analysis of the communication infrastructure underlying the power grid, such

[1]Nuclear plants as one of the conventional power plants do not provide CO_2, but they have other environmental impacts: heating where they are implemented (*e.g.*, sea, rivers, *etc.*), nuclear wastes treatment and storage, major failures (*e.g*, Tchernobyl, Fukushima, *etc.*)

as real-time performance, data transmission reliability, information security, optimal network configuration, *etc.* [15], [16], [101], [147], [26], [87].

The models over data traffic proposed by previous research works do not result in a mathematical model to describe both stochastic and deterministic behavior of different communication protocols in a single expression [164], [158]. Hence, the objective of this thesis is to propose methodological guidelines towards a unitary mathematical model of the data traffic through the ICT for smart grid applications as part of the solution to these problems, which may further be used as a basic model for new control paradigm over the smart grids. Towards this thesis research, the focus is on the Substation Communication Network (SCN) as an example of the ICT application for smart grids. So, modeling process is applied in case of the traffic behavior through SCN.

Substation Communication Network (SCN) traffic flow is described mathematically using time-series analysis method – ARIMA(p,d,q) – to model the stochastic behavior of traffic. Also, a series of linear equations to model the deterministic behavior of traffic. Next, model is confirmed over an SCN experimental test bench with a good estimation fit, and residuals near zero. Further, the communication channel is studied under some main parameters variation, *e.g.*, Packet Loss, using our descriptive model. Since required data for this model is the cumulative number of transmitted messages, it is applicable to all the communication protocols, no matter which type.

As a new part of thesis research, the real-time performance of the communication network is evaluated. It focuses on the message transmission delay over IEC 61850 communication protocols. Finally, delay evaluation algorithm is validated over the same intelligent substation test bench for ARIMA model. Through the literature review, among different delay estimation methods, we are interested in the model proposed by [161] which is based on the Network Calculus theorem. The estimated delay is compared to a value measured using a network analyzer.

This Ph.D. project, a joint work, has been prepared within the Grenoble Images Speech Signal and Control Laboratory (GIPSA-lab) and the Grenoble Electrical Engineering Laboratory (G2Elab), Grenoble Institute of Technology (Grenoble INP). This thesis has received its financial support from Grenoble Institute of Technology (Grenoble INP) in the framework of AGIR (Alpes Grenoble Innovation Recherche) 2016 funding program.

Outline of the thesis

This Ph.D. thesis is presented in five chapters. Following this introduction, first chapter is assigned to the topic of smart grid. Beside the motivation of ICT integration into power grids, required characteristics of communication network for smart grid applications are introduced. Also, challenges associated to a reliable communication network are presented. Characteristics of the proper communication network architecture are explained next.

As ICT integration into power grids results in their mutation, second chapter talks about the existing cyber-physical interactions. Afterwards, an interoperability model for smart grid

developed by GridWise Architecture Council (GWAC) is shown. Next, a state of the art of co-simulation, and modeling of smart grids, and communication performance analysis is presented and critically discussed. Conclusions of this analysis justify the research motivation and allow the objectives of this thesis to be formulated.

The third chapter introduces the IEC 61850 standard, and its communication protocols. As the principal application concept of the smart grid considered through this thesis, SCN structure is illustrated first. Next, intelligent nodes of the SCN, *i.e.*, IEDs, and the transmitting data in between them are represented. This chapter also overviews the literature of existing traffic modeling methods. ARIMA(p,d,q) model – applied method over the SCN traffic by this thesis – is introduced in detail. As the second objective of this thesis, the estimation of the maximum message transmission delay based on the queuing theory is studied.

The fourth chapter illustrates how ARIMA model is applied on the SCN traffic and how it is validated through a detailed procedure. Corresponding flowchart is also shown. Then, the algorithm of message transmission delay is explained through a simple example.

The fifth chapter illustrates the procedure of experimental works. Time-series ARIMA modeling is applied over the experimental data acquired from a co-simulation scenario, then validated over a real-time traffic generator. Estimated model is compared to the original data by means of some well-known validation criteria. The traffic flow calculation algorithm based on the proposed traffic flow ARIMA modeling is applied on a real-time traffic generated over a real MV/LV substation. Maximum message transmission delay is estimated. Then, estimated result is compared to the corresponding values measured using a network packet analyzer.

The manuscript body ends with a conclusion including the most relevant results, and emphasizing some of the future research works.

Contribution of the thesis

- *A new data-driven-based stochastic model* of the IEC 61850 Generic Object Oriented Substation Events (GOOSE) traffic is proposed by this thesis. The SCN traffic is well estimated over a co-simulation in the context of a simple distributed reactive power control [128], [48]. To validate, the proposed model is applied over a real-time traffic generator [49]. Modeling method is based on time-series analysis using Box-Jenkins identification method. Among all the possible models, ARIMA(p,d,q) is selected to describe the stochasticity of traffic behavior such that the global behavior is described mathematically by linear equations.
- The proposed data-driven model is integrated in an identification methodology, toward obtaining a unitary model which can effectively describe the SCN traffic behavior as the underlying paradigm of any control design approach. In addition, the effect of variations of some communication channel parameters over the identified parameters of the model are studied. This can further help to find a proper communication network configuration [48].
- As a second objective of this thesis, maximum message transmission delay is well estimated and validated over an MV/LV intelligent substation test bench. In the context of a load shedding scenario, an analytical model proposed by [161], based on queuing theory, is applied, which develops a traffic flow model. Based on this model, a traffic flow calculation

algorithm is used to obtain the distribution of traffic load and estimation of the maximum message delay. The authors suppose all the flows and network configuration are totally known. However, in this thesis, it is supposed that the information flows passing through the network are totally unknown, but network configuration is completely known, so *a new identification method* is proposed to obtain the flow characteristics.

List of publications during the thesis

A list of publications that have been written during the development of this thesis is given below. Some other still unpublished results proposed and described in this manuscript will be considered for future submission to peer-reviewed journals, and conferences.

International journal papers

- Feizimirkhani, R., Bratcu, A. I., Bésanger, Y., "Time-series Modelling of IEC 61850 GOOSE Communication Traffic between IEDs in smart grids – a parametric analysis", IFAC-PapersOnLine, ScienceDirect, ELSEVIER, Vol. 51, Issue. 28, p. 444-449, Sept. 2018, Tokyo, Japan, Conference paper published in the online international journal IFAC-PapersOnLine, ScienceDirect Journals&Books, DOI: 10.1016/j.ifacol.2018.11.743.
- Feizimirkhani, R., Bratcu, A. I., Bésanger, Y., Labonne, A., et Braconnier, T., Bamberger, J. "Evaluation of the message transmission delay over the IEC 61850 communication network – a real-time MV/LV-substation validation," (in preparation for submission to IEEE Transactions on Smart Grid).

International conference papers

- Feizimirkhani, R., Bratcu, A. I., Bésanger, Y., Labonne, A., et Braconnier, T., "Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARIMA Model", IEEE PES Innovative Smart Grid Technologies Europe (ISGT), p. 1-5, Sept. 2019, Bucharest, Romania, DOI: 10.1109/ISGTEurope.2019.8905634.

Talks

- Feizimirkhani, R., Bratcu, A. I., et Bésanger, Y. "Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARIMA Model," Groupement de Recherche "Modélisation, Analyse et Conduite des Systémes" (GdR MACS[2]) Workgroup Meeting, 15-16 Nov. 2018, Nantes, France.

Posters

- Feizimirkhani, R., Bratcu, A. I., et Bésanger, Y. "Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARIMA Model," Académie de Recherche "Alpes Grenoble Innovation Recherche" (ADR AGIR[3]), Mathématiques, Sciences et Technologies de lInformation et de la Communication (MSTIC[4]) Scientific day, 5 Feb. 2019, Grenoble,

[2]Research Group Modeling, Analysis and Systems Management
[3]Research Academy Grenoble Alps Innovation Research
[4]Mathematics, Sciences and Information and Communication Technologies

France,

- Feizimirkhani, R., Bratcu, A. I., et Bésanger, Y. "Modeling and Analysis of the Smart Grid as Cyber Physical System," Journée des Doctorants (JDD[5]), Electronique, Electrotechnique, Automatique & Traitement du signal (EEATS[6]), 14 Jun. 2017, Grenoble, France.

[5]Ph.D. Students' Day

[6]Electronics, Electromagnetic, Automatic & signal Processing

Chapter 1

Background

Contents

1.1 Context of smart grid 8

1.1.1 Smart grid concept: Energy grid and Information and Communication Technology (ICT) 8

1.1.2 Architecture of the smart grid communication network 16

1.1.3 Substation Automation System (SAS) 17

1.2 Distribution energy generation and control 21

1.2.1 Spatially Distributed Energy Resources (DERs) 21

1.2.2 Renewable Energy Sources (RESs) Integration 23

1.2.3 IEDs: distributed measuring, control and protection 24

1.3 Conclusion 27

In this chapter, a background of the topic of smart grid and fundamentals of the Information and Communication Technology (ICT) for smart grid applications is presented, together with some of the main smart grid concepts further used as basic elements throughout the considered case studies. More specifically, Section 1.1 is devoted to the context of smart grid providing a detailed explanation of the integrated ICT architecture for the smart grid applications. In addition, motivations of using communication networks for smart grid are presented, then requirements of a reliable communication network are introduced, while the challenges to be faced are mentioned next.

The communication architecture in smart grid and its standardization history is briefly presented. Next, as the core nodes of the power grid, the concept of automated substation is introduced. In Section 1.2, distributed generation and control as one of the main purposes of the integrated intelligence to the current power grid, is explained. Some concluding remarks are drawn in Section 1.3.

1.1 Context of smart grid

This section provides the structural information of the smart grid communication infrastructure; a representation of ICT for smart grids, next the related communication architecture, and finally, the communication protocols mostly used in Substation Automation System (SAS).

1.1.1 Smart grid concept: Energy grid and Information and Communication Technology (ICT)

Since decades, the transmission networks has been smart through a robust communication network. So, most of the changes through smart grid are occurring at the distribution level which needs to become smarter to cope with power systems at transmission level [97]. Structure of the electrical grid has been a strictly hierarchical, *i.e.*, top-down architecture. As shown in Fig. 1.1, large-scale power plants (*e.g.*, hydro, fossil-fuel, and nuclear power plants) were responsible for the energy production, and generated energy was transported by transmission system downwards to main consumption centers, *i.e.*, regions and cities. Through a series of top-down transformers the voltage was converted to medium and low-voltage levels, where the energy was distributed to customers at the end of the chain (*i.e.*, distribution system). Due to the strict hierarchical structure of such a system, it was characterized by a unidirectional power flow from the bulk generators to consumers [85].

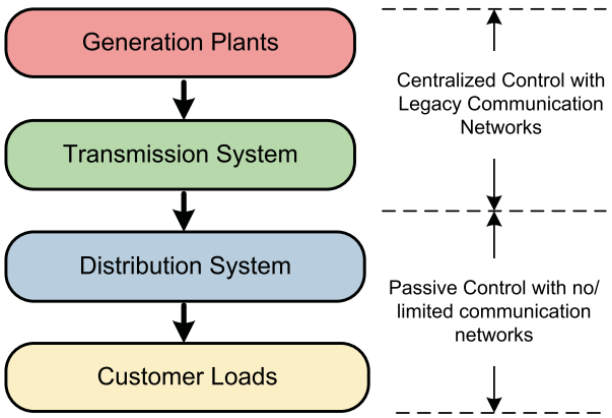


Figure 1.1 – The hierarchical system of the present power grids [85].

But, components of the hierarchical grid are moving toward the end of their lives, and meanwhile electricity demand is gradually increasing. According to the global energy statistical yearbook 2019 [27], the global electricity consumption increased by 3.5% annually over 2018. Traditional electric power distribution network is very complex and unfit for the needs of the 21^{st} century. Some of the inhibiting factors are the growing population and demand for energy, global climate change, equipment failures, energy storage problems, capacity limitations of electricity generation, one-way communication, decrease in fossil fuels, and resilience

problems [42].

In this context, smart grid, as a distributed and automated energy delivery network, enables near instantaneous balance of supply and demand by incorporating the benefits of distributed computing and communications. Smart grid integrates innovative products and services together with self-monitoring, pervasive control, two-way communication, and self-healing technologies [73].

From the conventional power grid perspective, the most vital characteristics of new intelligent grid, *i.e.*, smart grid, are [98]:

- informationization: to integrate, share and exploit real-time/non-real-time information in electric power systems,
- digitalization: to describe quantitatively the structures, states and objects of the grid,
- automation: to advance the control and protection systems that construct a dynamic and flexible grid,
- cooperation: electric generation, transmission systems and customers need to interact with each other, and they also need a coordinated operation,
- self-healing: to recover from network component failures and attacks.

To sum up, smart grid is a modern electric power grid infrastructure for improved efficiency, reliability, and safety: from large generation, through delivery systems to electricity consumers. Such a grid comprises a growing number of distributed generation – with smooth integration of renewable energy sources – and storage resources, through automated control and modern communication technologies [7]. The existing grid is in lack of communication capabilities, while a smart power grid is full of enhanced sensing, advanced communication and computing abilities.

A typical smart grid structure is illustrated in Fig. 1.2. Different components of the power system are linked together through the communication paths to provide interoperability platform between them. There are different parts of a power grid based on a communication platform: generation (renewable resources, fossil-fueled generators), distribution, transmission and other substations (*e.g.*, residential, commercial, and industrial sites).

Related changes facilitate the integration of Distribution Generation (DG) based on Renewable Energy Sources (RESs) and also energy storage technologies [129], [106], improve capacity of suppliers and consumers [112], enable the local energy demand management, and provide an Advanced Metering System (AMI) [46].

As a fundamental for smart grid[1], a robust communication infrastructure is its key enabler that differentiates it from the conventional grid.

Smart grid creates an automated, widely distributed electrical grid, over a two-way flow of electricity and information structure [45], [162]. The primary requirement for smart grid to be successful is the cooperation, integration, and interoperability among a set of technologies

[1]Through this work, smart grid considers the distribution level as the part undergoing the changes, not transmission level.

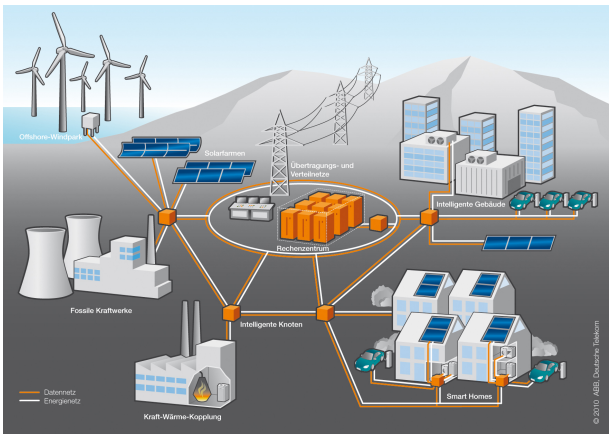


Figure 1.2 – A smart grid perspective with different components [78].

and disciplines. In order to realize smart grid, new communication capabilities are needed which do not exist in the current power grid, and their legacy control systems. Electrical communities from all over the world have started doing research and development on smart grid applications and technologies. As an example, Local Distribution Companies (LDCs) have integrated advanced metering and two-way communication, automation technologies to their distribution systems [127].

In addition to research and development projects, many electric utilities have also been taking incremental steps to realize the smart grid technology. Most of them in cooperation with telecom operators or smart meter vendors has performed real smart grid projects. Thus, global power community is now facing main challenges to find the most appropriate communication technology that can satisfy energy network future needs.

By all the movements toward the new generation of power grid technologies, *i.e.*, smart grid, some of the applications and services which have been developed over the past decades are listed below as some main research problem:

- local energy generation which led to the Distributed Generations (in a larger scale microgrids (MGs)),
- integration of RESs, storage resources and electric vehicles to decrease the effects of greenhouse gas on Earth.
- moving through a decentralized control, automation, and protection using Intelligent Electronic Devices (IEDs).

Different units of smart grid need to communicate with each other based on a reliable communication network. Any uncertainty of the communication network can impact the stability and availability of the power system. To address the challenges brought by power systems and power electronic devices, it is required to rethink and modify the existing communication network technologies for smart grid, as well as the existing control paradigm.

Communication network is the critical infrastructure for smart grid in order to connect,

configure, and control various power devices of the electrical power grid [9]. In order to properly communicate and transmit smart grid messages (*e.g.*, control signals, measured values, *etc.*) among different intelligent agents (*i.e.*, control centers, sensors/actuators), significant research should be done on the modeling, simulation and test-bed validation.

There are different types of ICT which have an important role in facilitation of modern power grid realization, and its services. Many of these communication technologies have been used by other industrial applications such as sensor networks in manufacturing or wireless networks in telecommunications. A cooperation of these communication technologies is adapted for use in new intelligent and interconnected power systems [157], and in general, they can be classified into five groups: advanced components, sensing and measurement, improved interfaces and decision support, standards and protocols, and integrated communication. Fig. 1.3 shows a general architecture of smart grid communication network which includes: Home Area Network (HAN), Neighborhood Area Network (NAN), and Business Area Network (BAN), data centers, substation automation systems [160].

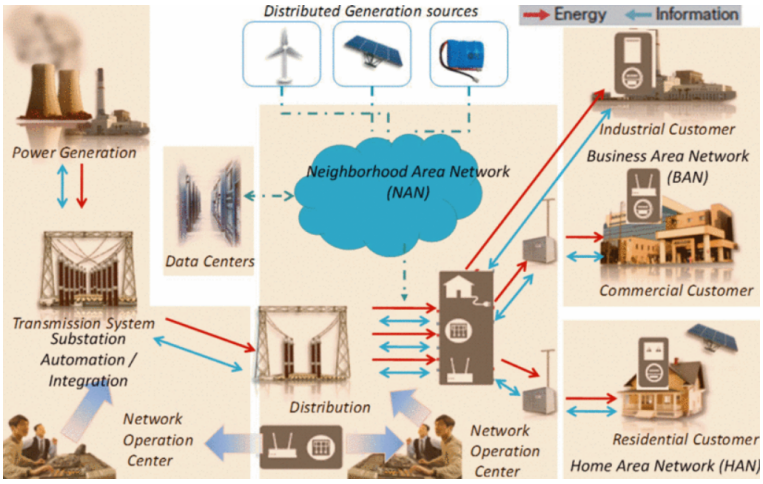


Figure 1.3 – The communication infrastructure for smart grid application [160].

Finding the best communication technology for smart grid applications is not easygoing. In the smart grid environment, the communication network has to support information exchange among a large number of smart meters, IEDs, sensors and actuators without – or very limited – human intervention. This form of communication is often known as the machine-to-machine (M2M), which is autonomous in nature and triggered either by time or events. Each of these applications has different characteristics in terms of packet arrival rate, packet size, and latency based on their deployment scenario. For example, the latency requirement of a smart meter event, and a substation event are quite different. In order to properly evaluate the existing communication technologies, architectures and protocols, it is very important to acquire detailed knowledge about the current and prospective applications of the smart grid.

As mentioned earlier, satisfying the requirements for next generation of power grid needs a better understanding of the future design process of the smart grid and identification of its key features. As it is fundamental that any smart grid should be characterized by sufficient

flexibility and resilience, here, the focus is on the information, computing and communication aspects of the future power systems. Thus, in the reminder of this section, a classification of the main motivations, requirements and challenges characterizing a communication structure for the smart grid applications is presented.

1.1.1.1 Motivations of ICT for smart grid

Some of the main motivations of the communication infrastructure for smart grid applications are briefly highlighted here. As illustrated in Fig. 1.4, the integration of ICT into smart grid is interesting for system, operation, and environment aspects.

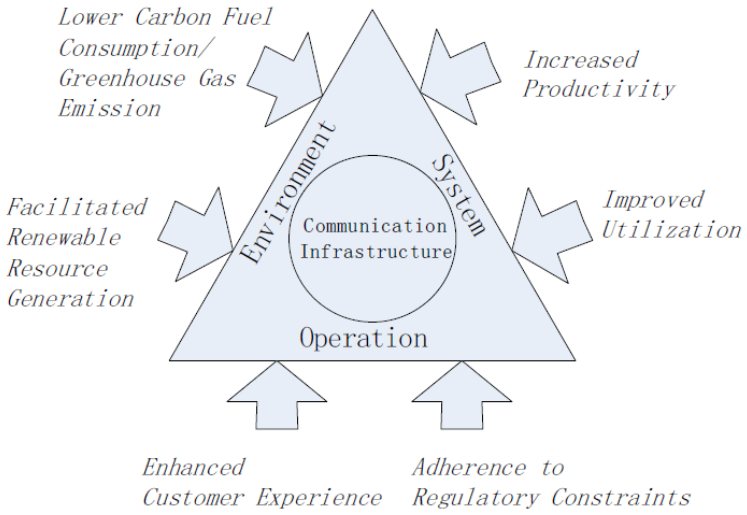


Figure 1.4 – The motivations of emerging smart grid through the communication infrastructures [132].

Enhanced customer experience: improving Quality of Service (QoS) and reliability is the main objective of smart grid, *e.g.*, reduced outage time, optimize the energy consumption, *etc.*

Increased productivity: due to the intelligence, tasks are performed more efficiently, which reduces the operational costs.

Improved utilization: provided real-time data over the distribution network improves decision making, which leads operators to find online strategies.

Lower carbon fuel consumption: using RESs, and electric vehicles decreases the power grid environmental effects. In addition, smart grid can reduce electricity losses, and reduce demands through a well monitoring system.

Facilitated renewable resource generation: integration of RESs reduces dependency from the conventional resources to reach a distributed generation system.

Adherence to regularity constraints: smart grid arrangement of a vast data traffic should guaranty a reliable and cost-effective energy system.

1.1.1.2 Requirements of the smart grid communication infrastructure

Main functional blocks of smart grids can be classified as: monitoring (metering), control, protection, and communication. Monitoring system *via* a sensory network detects any fault, and sends notifications to the control centers. To manage the reported deficiency, control block sends the proper command to protection actors installed in different points of the power grid. Also, protection units frequently send their status report to the control system. All these blocks are communicating through an underlying communication infrastructure. A robust communication network is needed to integrate the user-input data to the process. In addition, all users have to receive appropriate information from smart grid. In this part, we will discuss the major requirements of a reliable communication infrastructure for smart grid applications.

Quality of Service (QoS): realization of a practical smart grid depends strongly on the guaranteed QoS of the deployed ICT infrastructure. QoS considers a number of network evaluation factors as the following:

- latency: Communication network is characterized by how fast the interactions take place, *i.e.*, real-time performance, and the fact that information has to be transmitted in some hard time boundaries depending on the importance of transmitted information,
- bandwidth: while there are more and more intelligent devices added to the smart grid, the transmission bandwidth should increase faster than the demand of these interconnected intelligent elements in the network to be able to transmit much more messages simultaneously without serious effects on latency.

Interoperability: smart grid interoperability is the ability of different heterogen systems to work together, use compatible parts, exchange information or equipment with each other, and work cooperatively to fulfill common tasks. So, the integration of a two-way communication, and effective cooperation among many interconnected elements are possible. The National Institute for Standards and Technology (NIST) of United States of America was founded in 1901. NIST is one of the national oldest physical science laboratories and started to work as the first international coordinator for smart grid interoperability. In April 2009, NIST established a framework including protocols and standards for information management to achieve interoperability of smart grid devices and systems [58].

Scalability: smart grid communication infrastructure should be scalable, which means the capability of adopting more and more devices and services. As an effective solution for smart grid communication infrastructure, IP-based network is a proper one [105].

Security: one of the major requirements facing the smart grid development represented by is the cyber-security systems. As smart grid becomes more and more interconnected, the potential cyber attacks and incidents against this critical sector increase.

Standardization: implementing standards is a major issue to move toward advanced sensing/automation, and is also an important step to create the communication infrastructure for smart grid. Indeed, communication systems in smart grid are usually from different manufacturers/vendors with their own rules, which are proprietary and have no mutual interoperability. The identified standards which govern the necessary integration of different functions are specified by IEC. A detailed history on the standardization for smart grid communication infrastructure is presented further in Section 1.1.3.

As a pioneer, IEEE defined existing standards and wrote guidelines on how smart grid should operate using the latest technology in three main working groups: power engineering, information, and communication technology. In 1990s, it has published IEEE 1550 Report, which was further the basis for the IEC 61850 standard [143].

1.1.1.3 Challenges through a reliable communication network

To reach a reliable and robust communication infrastructure for smart grid applications, there are various type of challenges to be tackled with: widely use of the wired/wireless sensors and communication devices, necessity of specific management systems, the need of a large-scale embedded computing setup, adapting to the legacy power system, integration of IEDs, new generation of communication and cooperation systems for smart grid. Here, different challenges toward smart grid communication infrastructure are discussed:

Complexity: as a system of systems, communication infrastructure for smart grid is extremely complex. Therefore, modeling, analysis and design of a suitable communication system encounter a number of new challenges. Proposed models should consider possible uncertainties. Large-scale problems have to be solved using numerical tools in order to perform analysis. New control systems should be able to release uncertainties and inconsistencies or reduce their effects if necessary. Modeling and simulation challenges in case of the smart grid communication infrastructure complexity are summarized here:

- multi-physics approach: the complexity here is due to the discrete-time nature of the communication system, and continuous-time nature of the power system, which are tightly interconnected [108],
- multidisciplinary approach: different users working in the same scenario, and each of them are related to different aspects of the grid (*e.g.*, control centers, power flow systems, communication networks),
- re-configurable models: it is necessary to have dynamic models which may be re-configured automatically [130],
- high-level graphic visualization to support system: a proper visualization is required to synthesize "system pictures" in order to extract the necessary information over the simulation curves, and facilitate necessary monitoring,
- support for the uncertainty propagation: uncertainties in the communication network can affect the overall operation of the smart grid such as message transmission delay, packet loss, disrupted communication, *etc.*

Efficiency: an efficient communication must be able to provide an error-less and nearly instantaneous communication among all devices – ranging from individual loads to the grid wide control centers. This needs to process numerous data transactions for analysis and automation. The challenges to be meet here are listed as follows: a better telemetry, faster and more robust controls, secure communication, embedded computing capabilities, and Internet technologies to facilitate data exchange.

Reliability: to satisfy all the requirements of standards and protocols, and install the necessary analytical capabilities in the new communication infrastructure, a framework for coherent integration of reliability technologies should be defined. The challenges to meet reliability [121], also the impacts of the communication infrastructure on the reliability are discussed here:

- renewable resources: renewable resources generally bring some different challenges to the reliability of smart grid because of the following aspects: variability and low capacity, the factors which make the net demand profile steeper (*i.e.*, increasing or decreasing of electricity demand in a short period of time), low correlation with the load profile, renewable energy forecast errors increase volumes of imbalance, congestion problems at distribution level due to distributed resources, operational performance issues such as voltage, protection system, and regulation.
- demand response: demand response results in the load consumption considering the emergency and pricey conditions of smart grid – demand response process can benefit the peak load or congested operation conditions to reduce the load consumption.
- load management: emergent load rejection in order to protect the electrical grid from disruption is implemented to operate either by system operator command, or through under-frequency and/or under-voltage relays. Smart grid also provides a new method, named *soft load shedding*, which only cut off a fraction of customers' power. Hence, in case of instability or overloading, a part of each customer load is cut off, and no customer is totally in dark [109]. In a smart grid, the load shedding can be improved to function more intelligently and based on the customer participation.
- storage devices: the growth of storage systems is much smaller than the current need of them, despite various storage technologies which are appearing. Battery storage seems to be the most promising due to improvements in technology and the scale economy. Storage resources tend to make the net demand profile smoother, also are expected to improve reliability.
- electric transportation: plug-in electric vehicles (PEV, PHEV) continue to become more popular as the environmental concerns increase. They reduce significantly the green house gas emission and dependence on fossil fuels. One of the main issues for integrating electric vehicles is their long recharge time which leads to generally unacceptable level of vehicle unavailability. As PHEVs present a significant factor of load growth, it can also worsen demand variability and associated reliability problems depending on the charging schemes and consumer behavioral patterns.

Security: based on the evolution of power system communication infrastructure and cyber-security concerns, new issues have arisen in the context of smart grid:

- information security domains: since SCADA and Energy Management System (EMS) have become increasingly integrated, it becomes more difficult to treat the system structure in term of parts or subsystems. Instead, it becomes more natural to consider SCADA/EMS systems in term of different domains which were introduced by [41], for the power system applications. Related security domains are introduced in Fig. 1.5.
- de-coupling between operational SCADA/EMS and administrator IT: when the existing SCADA/EMS systems are being renovated or replaced, the information and IT security problems should be considered.
- threats: today's SCADA/EMS are interconnected and integrated into a various range of external systems, which make digital entrances, so digital vulnerabilities at the same time. The points from whom a substation can be accessed should be secured.
- vulnerability: systems in smart grid are facing security risks such as a Wireless Sensor Network (WSN) for AMI, which is vulnerable to be attacked by an intelligent adversary, or perturbed by an ordinary microwave stove.
- privacy: the privacy of terminal customers and smart metering networks is important to the eventual acceptance by the public.

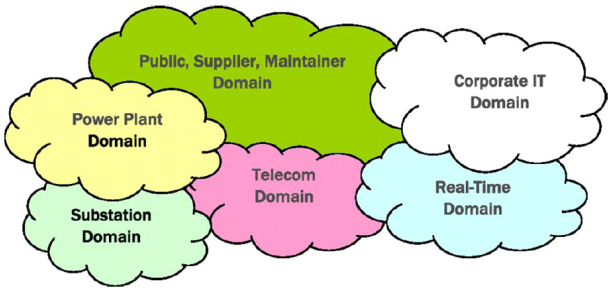


Figure 1.5 – Different domains for information security [40].

1.1.2 Architecture of the smart grid communication network

As key component of the smart grid, it is critical for power utilities to find the best communication infrastructure according to the requirements of the communication network to handle the output data and deliver a reliable, secure and cost-effective service throughout the total system.

There are two main communication media that can support different communication technologies in smart grid: wired and wireless. They are used for data transmission among different entities of the smart grid in order to give necessary services. At some points, wireless communication has some advantages over wired technologies, such as low-cost infrastructure and easy connection to difficult or unreachable areas. But, sometimes, the type of transmission path may attenuate the signals. On the other hand, wired solutions do not have interference problems and their functions are not dependent on batteries, as wireless solutions often do.

Basically, two types of information are passing through smart grid communication network. The first flow is from sensor and electrical applications to smart meters, the second

is between smart meters and the utility's data centers. As suggested in [110], the first type of data flow can be transmitted through the Power Line Communication (PLC) or wireless communication, such as ZigBee, 6LowPAN, Z-wave, *etc.* For the second information flow, cellular technologies or Internet can be used.

Nevertheless, there are important limitations that should be taken into account in the smart metering deployment process, such as time of deployment, operational costs, technology availability and rural/urban or indoor/outdoor environment, *etc.* The choice of technology that fits one environment may not be suitable for the other.

1.1.3 Substation Automation System (SAS)

Distribution substations are the main nodes of delivering electricity to industrial, commercial and residential areas. These substations follow specific requirements to make electricity cost effective for all the customers. To reach availability and scalability, power utilities involve standardized aspects, that are driven by international standards [23]. Substation automation refers to the monitoring, protection and control functions performed on substation and feeder equipment.

In the power industry, Modbus and DNP (Fig. 1.6) are well-known among the substation communication protocols. Modbus was developed by Modicon – further absorbed by Schneider Electric in 1996 – in 1979. This protocol was originally used as a control network protocol for PLC to allow process control communications. The original edition forms a master/slave environment between control devices. Through this protocol, the master initiates a request and corresponding slave(s) will respond with required action/data. Master station can initiate a broadcast message, or address only one slave station [59], [118].

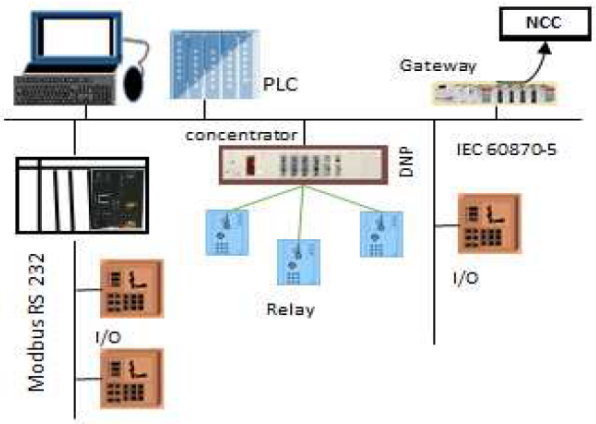


Figure 1.6 – A typical substation automation system and its classical communication protocols [7].

Physical layer of this protocol is not predefined, so manufactures can develop their own physical interface. Thus, this freedom resulted in the creation of different protocol versions, *e.g.*, Modbus RTU, ASCII and TCP. Modbus RTU and ASCII are commonly used with RS

232, RS 422 and RS 485 links with maximum baud rate between 19200 bps and 100 Kbps, while Modbus TCP supports client/server communications with different physical layers such as Ethernet Unshielded Twisted Pair (UTP) cables.

In 1993, DNP 3.0 protocol was developed by GE-Harris Canada (formerly known as Westronic, Inc.) in order to enable communications between devices of the master station level, RTUs and protective relays. This protocol was standardized based on IEC 60870-5 series. It is widely used in power, water and gas process control for SCADA connections to RTUs. DNP 3.0 uses RS232 or RS485 as serial physical layer [118].

Table. 1.1 provides a comparison between some widely used legacy protocols of substation communications.

Modern protocols in SAS

As briefly explained, many custom-built protocols were established for data exchanging and management for SAS applications. The existing protocols should be modified and converted to new protocols in order to provide interoperability among the equipment and devices using different protocols from different manufacturers. Additional efforts are required to perform these changes, which increase the cost, effort and configuration complexity [39].

Since 1986, Electric Power Research Institute (EPRI) has announced the concern of different protocols in substation installations. EPRI made some efforts that resulted in the Utility Communication Architecture (UCA 1.0) project by the end of 1991.

Pilot projects involving experimental technologies were attempting to develop a standardized approach to cover all communications from an IED up to the control center or SCADA master [39]. UCA, as the result of these projects specified the use of Manufacturing Message Specification standard (MMS) and Integrated Utility Communication (IUC). Therefore, EPRI established a forum with Northern States Power Company (NSP), about the implementation of MMS across multiple communication media. Demonstrations from the MMS forum projects have resulted in detailed specifications. These specifications addressed interoperable communications in the utility industry covering communication profiles, application services and object models for IEDs [131].

In 1999, substation implementation documents were published in the IEEE 1550 Technical Report as UCA 2.0, and further used as the basis for IEC 61850 standard [143]. A working group worked on the coordination of certain parts from UCA, that extended UCA modeling, data definitions, data types and services. IEC 61850 included these results in the related standardization parts. The parts of IEC 61850 standard are aimed to be a development of UCA.

IEC 61850 standard covers almost all aspects of SAS including real-time, high-bandwidth protection and control applications, while DNP 3.0 standard only provides communication specifications for low-bandwidth monitoring and control operations. Therefore, IEC 61850 [111] is gradually becoming the dominant protocol in this field.

Protocol	**Modbus**	**IEC 60870-5-103**	**DNP 3.0**
Release data	1979	1997 (former VDEW6, in late 1980s)	1993
Developed by	Modicon	IEC standards (TC 57 WG 03)	Harris
Standards support	Modbus organization	IEC 60870	IEEE 1815-2012, open specification
Substation use	• SCADA master to RTUs slaves, • Also as client/server with IEDs network	Interoperable connection between protection and control devices (RTUs and relays)	SCADA, RTUs and protective relays
Physical interface	• EIA (RS) 232, 422 and 485 for Modbus RTU and ASCII, • Exist also Ethernet for Modbus TCP	EIA 485, and optic fiber	• EIA (RS) 232, 485 • Exist also for Ethernet
Communication type	Master/slave, peer-to-peer and client/server	Master/slave	Master/slave, peer-to-peer and client/server
Support OSI layer	Application layer	Application layer and 3 Enhanced Performance Architecture (EPA) layers	2^{nd} layer and somehow 4^{th} and 7^th layers supporting TCP/IP
Baud rate	19200 up to 100 Kbps (EIA), and Ethernet bandwidth for Modbus TCP (up to 1 Mbps)	6900 or 19200 bps	38400 bps(some versions up to 112.5 Kbps) depends on hardware
Dominant market	worldwide	Europe	North America

Table 1.1 – Comparison between some well-known legacy communication protocols in SAS [7].

The Technical Committee (TC) 57 of the International Electrotechnical Committee (IEC) was established in 1964 to publish and expand international standards in the field of communications between the equipment and systems for electric power process, including Remote-control, Remote-protection and all other telecommunications in the electric power systems [39]. TC 57 developed the international standard IEC 61850: Communication Networks and Systems in Substations.

Utilities, suppliers and users noted that the industry have to obey a single standard for substation communication, and all of the technical problems based on UCA 2.0 application are to be resolved in the appropriate parts of IEC 61850 [10]. In 1997, a meeting between IEC TC 57 members concluded at Edinburgh that only one standard for substation automation and communication should be developed, and to merge the North American and European approaches [131].

IEC 61850 features

The primary objectives of IEC 61850 standardization were:

- interoperability: IEDs ability, from one or various manufacturers to exchange information, and use the information for their considered functions,
- free configuration: the standard should support different philosophies and allow free allocation of functions,
- long-term stability: the standard should always follow the progress in communication technology as well as evolving system requirements.

The standard has several parts that cover different domains in the field of power utility communications. In SAS applications, the standard aims at applying interoperability among devices, and integrate subsystems to build the overall SAS system. IEC-61850-based SAS should merge several devices that have certain features such as:

- data models with Logical Nodes (LN) and Common Data Classes (CDC),
- communication service interfaces,
- reporting, Generic Object-Oriented Substation Event (GOOSE) and Sampled measured Values (SV) communication services,
- interoperable protection, control, measurement and monitoring functions,
- support of eXtensible Markup Language (XML) and IED-based Capability Description (ICD) files,
- substation devices could be configured through Structured Control Language (SCL).

Different parts of IEC 61850 are illustrated in Fig. 1.7.

Standard enables the use of emerging technologies in the field of communication networks, smart protective devices and smart instrumentation and metric equipment.

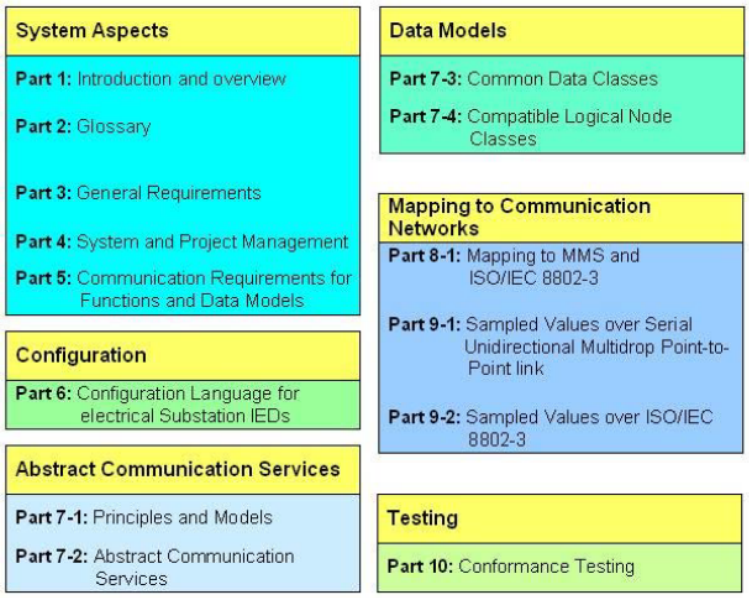


Figure 1.7 – A representation of the IEC 61850 standard parts [140].

1.2 Distribution energy generation and control

Traditional power network was designed as a centralized system such that the electric power flows unidirectional through transmission and distribution lines from power plants to the customer. The intelligence is concentrated in central locations and only partially in substations, while remote terminations (*i.e.*, loads) are almost or totally passive. The new systems would provide higher and widely distributed intelligence embedded in local electricity production, two-way electricity and information flows, thus achieving reliable, flexible, efficient, economic, and secure power delivery and use [17].

Smart grid enables, among others, new approaches in grid management by applying demand side management and providing grid energy storage, needed for load balancing and for overcoming energy fluctuations caused by the intrinsic nature of RESs, thus preventing widespread power grid cascading failures [86], and enables the integration of plug-in hybrid and electric vehicles [135], [83], [37].

In this section, a brief introduction is presented on the green distributed energy production. Thus, basic concepts used through this thesis experiments as introduced.

1.2.1 Spatially Distributed Energy Resources (DERs)

One of the main benefits of the smart grid is the introduction of DERs into the electricity grid on a large scale. These DERs can supply particular areas with electricity when they are isolated from the main power grid due to failure conditions or system and equipment failures

[94]. Actually, providing electricity for an isolated area is a very costly process for Distribution Network Operators (DNOs). The reason is the waste of major energy amount through its way to the customers. In this case, DERs represent a cheaper and more efficient solution that can deliver energy from the points closer to the consumer than the DNO's centralized power grid. Fig. 1.8 shows an area isolated and cut-off from the power supply due to a fault that occurred on a feeder, so breaker B1 is opened. The isolated area can be resupplied with electricity from generator G1, as soon as breaker B2 is closed.

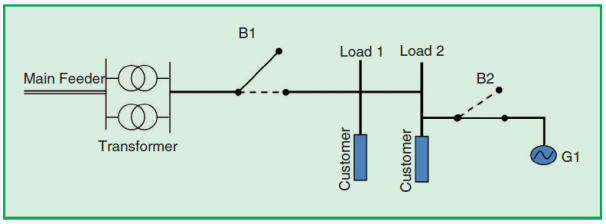


Figure 1.8 – Energy resupply with a distributed energy generation unit [20].

DERs are small sources of power generation and/or storage that are connected to the distribution grid. A DER can be from both renewable and non-renewable nature and can either be a Distributed Generation (DG) source, or a distributed storage (DS) source, or both. Solar panels, wind turbine, combustion turbine, fuel cells and battery storage systems, *etc.*, are some examples of DERs.

In general, such an area, which is able to locally supply energy in the islanded or isolated mode is named *microgrid* (MG). This is a small local electric power system having one or more DER units and loads. During normal operation, it is connected to the grid and operates in a synchronized mode. However, in case of any fault or maintenance event, it can operate autonomously in an islanded mode and capable of supporting its own load.

In a MG, loads and energy sources can be disconnected from and reconnected to the main grid with minimal disruption to the local loads. This offers potential benefits of using MGs such as reducing system load by intentionally islanding the MGs and providing uninterrupted power supply during maintenance activities [94].

The vital role of smart supervision over the operation of a MG in relation to the main grid and its proper load could be imagined. MG in smart grid is connected to the whole infrastructure using the allocated communication protocol. In this way, any necessary information from its status or any command is transmitted by the main grid controllers or a localized controller.

One of the most important concerns of the future power production is to be as clean as possible, and smart grids provide integration of the environment-friendly resources (*i.e.*, renewable resources). Thus, the first part of this thesis deals with traffic modeling under an intelligent cooperative scenario including two renewable resources equipped with IEDs. These resources follow a predefined consensus algorithm to efficiently provide electricity for a load unit.

In the following parts of this section, most used renewable resources are presented (*i.e.*, wind turbine, photovoltaic). Next, the structure of the main intelligent blocks of the related scenario, *i.e.*, IEDs, is explained. IEDs are the key devices of smart grid to apply intelligent supervision, control, and protection based on the proper ICT infrastructure. In this thesis, different scenarios with different communication protocols of IEC 61850 (Generic Object-Oriented Substation Event (GOOSE), Manufacturing Message Specification (MMS), and Sampled measured Value (SV) are considered.

1.2.2 Renewable Energy Sources (RESs) Integration

The main energy sources that are currently being used around the world are based on fuel and nuclear reactors. Chernobyl and Fukushima Daiichi nuclear catastrophes that occurred in 1986 and 2011 respectively, have warned power system designers of safety problems. This has convinced countries like Italy and Germany to abandon nuclear power plant. Germany has announced to entirely shut down nuclear power plants by 2022 [77]. In addition, global warming has become a critical issue leading governments to control CO_2 emissions by setting rules on using fossil fuel resources. These economical and political issues motivated scientists to develop new, safe and sustainable energy resources.

Hence, smart grid concept has been proposed and power companies have started building new infrastructures for renewable energy. Smart grids integrate clean energy sources (*e.g.*, solar panels, wind turbines) on a large scale for power generation whenever wind speed and solar radiation are available. Since these are not available all the time, storing excess power in storage units (*e.g.*, hydrogen, super-capacitors, electrical vehicles) is very important for the success of installing such systems. Many optimization efforts in generation, storage, and consumption have been published until now, such in [119], [92], [145], [75], [74].

Greenhouse gas emission on Earth caused by the electricity and transportation industries has become a considerable problem [136]. Renewable energy generators are a kind of technology that reduce fuel consumption and greenhouse gas emissions [51], [13]. Through a wide range of studies over RESs [146], some advances of the RES integration could be represented as reducing system losses and increasing the reliability, efficiency and security of electricity supply to customers.

It is worth to mention that the implementation of green-energy generators such as photovoltaics (PV) and wind turbines (WT) is really expensive. Thus, a hybrid-energy system is a cost-effective solution of long-term use. Since the sunlight and wind vary, some other forms of energy resources (such as gas, petroleum, hydrogen, bio-fuels) can be used at low periods to avoid blackouts. By this way, dependency on petroleum and gas is reduced significantly [67].

Since most of the renewable energy resources are connected to the distribution domain of the grid, their wide scale integration with the existing grid reverses the traditional direction of power flow. Hence, to enable bidirectional energy flow, as well as to integrate large-scale renewable energy generators, the existing grid needs advanced monitoring and control

technologies to stabilize its operating parameters (*e.g.*, voltage, current and frequency) by optimally balancing its load-supply profile.

The goal of this section was to explain the role and structure of RESs. To be mentioned, in this thesis, the example of a cooperation scenario of a single PV and WT unit communicating over IEC 61850 GOOSE protocol is considered. Next part introduces the structure of IEDs, a key grid components in transition toward smart grid.

1.2.3 IEDs: distributed measuring, control and protection

Today, protective relays become Intelligent Electronic Devices (IEDs), *i.e.*, programmable electronic-based protection and control devices with at least one communication interface. An IED is a microprocessor-based electronic device that includes input, output, memory, storage media, and communication network interface. This device is capable of doing many functions, in the same time benefiting from the processing power. IEDs contain logic programs that perform the electric power functions such as calculating reactive power, monitoring primary equipment, sending protection trip, *etc.* Generally, IEDs exchange information that can be gathered and saved locally or remotely for detailed processing and log registration. This information helps utilities to enhance reliability, and to enable asset management programs including predictive maintenance, life extensions and advanced planning [115], [1].

IEC 61850 data model: data model – IEC-61850-based modeling concept – follows a hierarchical structure where physical devices (PHD), *e.g.*, IEDs, contain Logical Devices (LD) that include predefined Logical Nodes (LN). As the smallest part of a function, LN exchanges data, and these data elements are defined by a unique name. Hence, the core of IEC 61850 standard is the information model and modeling methods. The information represented by data models and their attributes are exchanged by the communication services according to the predefined rules, and the requested performance.

Logical nodes: as mentioned, IEDs contain LDs that include LNs. An LN is an object defined by its data (*i.e.*, attributes) and its methods (*i.e.*, functions) [138]. LNs are the entities that communicate to exchange power process information, protection status, and control data. In this way, protection and control devices are made of several logical nodes. One or more LNs are integrated in different LDs, which in turn are located in different PHDs, and can cooperate to perform distributed functions. If one LN or one of the included communication links is lost, it can result in losing the functionality because of completely blocked function, or showing significant degradation.

LNs normally represent power protection tasks and related functions. IEC 61850 standard uses the object-oriented methodology to define LNs and their data regarding both content structure (syntax) and content meaning (semantic). IEDs manufacturers should follow these concepts to guarantee devices interoperability.

LNs are grouped into high-level LN groups according to their functions. For example, MMXU logical node starts with M, which represents the measurement group. Standard defines 92 different LNs classified into 13 groups in which suppliers can develop a new LN

under G group (Generic functions). Also, IEC 61850 provides names and classification of LNs according to their functions and logical location at station, bay and process levels [2].

IED, as a physical device includes a connection interface to the communication network. It has at least one network address that identifies its data set. The standard modeling starts with a physical device model that includes one or more logical devices. In this manner, the standard allows a single physical device to act as a proxy or a gateway for multiple devices (virtual devices), thus providing standard representation of a data concentrator [66], [2].

Fig. 1.9 illustrates a protection IED as a physical device, which in turn contains a logical device, *i.e.*, IED 1, which includes two logical nodes, *i.e.*, MMXU and XCBR. The average number of specific data objects provided by LNs is approximately 20. Each data, *e.g.*, circuit breaker position, comprises several details (data attributes). For instance, the circuit breaker position (called "POS") is defined in LN1 (XCBR), and data object (POS) is made up of many data attributes. For example, data attribute ctlVal represents controllable information, *i.e.*, can be set to OFF or ON. In this figure, data attribute Pos.stVal represents the position of the real breaker (could be in intermediate-state, off, on, or bad state), and data attribute q represents quality. LN2 (MMXU) has a data object representing Automatic control, A, which has two data attributes, PhA and PhB, representing the phase voltages. Data model structure is hierarchical (presented as a tree on right-hand side).

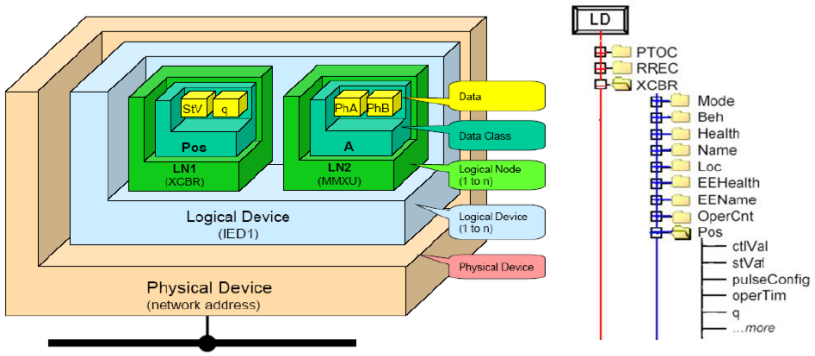


Figure 1.9 – Object modeling of the IEC 61850 data, illustrates the relations between physical device and logical device [66].

Fig. 1.10 shows the hierarchy of the IEC 61850 data model with a given example showing the LD IED1. The data attribute shall have a value that is important for exchanging the status of an equipment and related events.

Some data that refer to a physical device itself are needed, such as results of device self-supervision. Therefore, the standard introduces a default LN called as LLN0 [2]. The logical node LLN0 contains information related to the physical device (IED) data (Fig. 1.11), independent from all included logical nodes, such as device identification or nameplate, device self-supervision, *etc.*

Figure 1.10 – The concept of physical device with path to data attributes of logical nodes.

Figure 1.11 – The default Logical node LLN0 within default logical device LPHD0.

The general approach of IEC 61850 is to decompose application functions into small entities. So, every control or automation function is broken down to a collection of different LNs. These logical nodes are placed in a single IED or distributed among multiple IEDs. All the different LNs of a specific application are interconnected using logical connections. These logical connections (LC) can be over a single or multiple physical connections (PC).

Fig. 1.12 illustrates how functions are realized using LNs and LCs. In this example, two functions, F1 and F2 are shown. F1 is split into five LNs (LN1 to LN5), and F2 is split into three LNs (LN3, LN5 and LN6). These LNs are housed in three different PDs, *i.e.*, IEDs, (PHD1, PHD2 and PHD3). The LN0 is the node carrying the identification of the physical device.

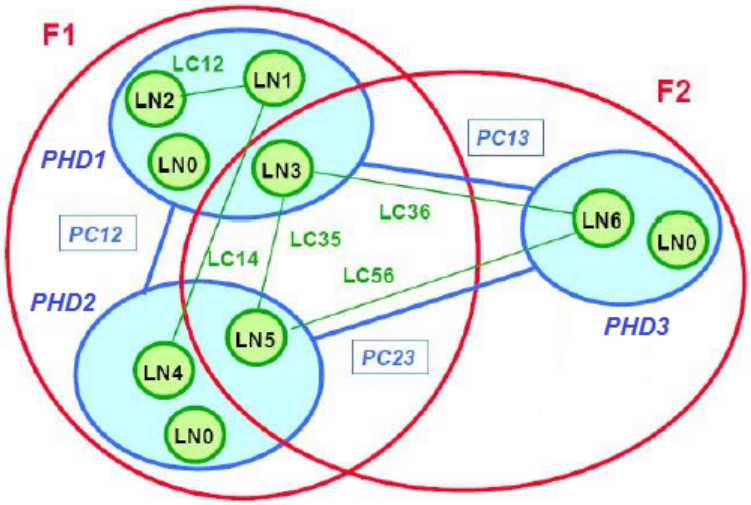


Figure 1.12 – Multiple LNs cooperating to build functions [140].

The standard defines Common Data Classes (CDC), which group LNs data object elements into specific data classes. Each LN can have a few or up to 30 data objects that belong to CDC class. Each of these data objects in turn has a few or more than 20 data attributes.

Each CDC describes the type and structure of data within the logical node, and each CDC has a defined name and set of CDC attributes with defined name, defined type and specific purpose [66].

1.3 Conclusion

In this chapter, we have presented a background over the set of required concepts used through this thesis. An introduction of smart grid concept, topics and challenges has been provided at the beginning of this chapter. The communication infrastructure as the key of transition of the future power grid toward smart grid has been represented, which explains the motivations of ICT integration, its requirements and existing challenges.

As one of the main concerns, communication standardization history for smart grid applications has been explained. The need of a common standard resulted in IEC 61850, which is now becoming the dominant standard in smart grids. In this thesis, IEC 61850, has been selected to provide communication services.

Then, distributed generation, and control have been also discussed. Two well-known types of renewable resources (PV plants and WTs) were introduced, together with their physical and communication structures. Next, the structure of intelligent electronic devices, as the intelligent agents for control and protection of the smart grid, have been presented with its, corresponding elements.

The next chapter is dedicated to an analysis of the state of the art of the communication traffic modeling for smart grid applications, performed through this thesis research.

Chapter 2

State of the art on the communication traffic modeling in smart grid applications

Contents

2.1 Interaction of energy layer and communication layer in smart grids . 29
2.1.1 Interoperability . . . 33
2.2 Researches in smart grid communications . . . 37
2.2.1 Co-simulation studies . . . 38
2.2.2 SCN traffic modeling based on IEC 61850 . . . 42
2.2.3 Delay estimation over SCN message transmission . . . 44
2.3 Research motivation and objectives . . . 46
2.4 Conclusion . . . 47

This chapter presents the interactions between electrical grid and communication network in smart grids. Then, the need of co-simulation of the smart grid as a cyber-physical system is emphasized. Modeling and analysis previously performed on the IEC 61850 communication traffic are overviewed. In addition, as a critical evaluating factor of communication infrastructure performance, research efforts on the estimation of transmission delay are discussed. Consequently, motivations and research objectives of this thesis are introduced to finish the chapter.

2.1 Interaction of energy layer and communication layer in smart grids

Growing presence of renewable and spatially distributed energy sources, as well as the intelligence embedded in increasingly sophisticated measuring and control systems, engenders a mutation of distribution energy grids. In this case, ICT has to be widely used in order to make possible an infrastructure for all different entities to communicate. Only this combination of electrical grids and communication networks will allow challenges to be overcome and smart grid to be truly smart [8].

A successful smart grid is a seamless combination and interaction of the power grid infrastructure as physical system, and ICT as cyber system. In such a system, cyber network does monitoring, protection and control over the physical system. Mainly, the interaction between power grids and cyber systems should be rethinked to be able to support flexible, available, and secure production-consumption system through the future smart grid. In addition, as cyber network is vulnerable and not failure-free, it is required to consider the characteristics of cyber network for the analysis of power system.

According to some important characteristics of the errors in cyber network, it is necessary to consider its failure risk. First, as cyber elements are more and more used, the failure risk of cyber system increases too. Second, tracking a failure of the cyber elements is more sophisticated than for the power elements. Most of the time, failures are hidden and will appear when a fault or mal-function occurs in the power system. Finally, the behavior leading to errors, faults or failures in the cyber system are so complex, that cyber-system modeling is challenging. Thus, uncertainty, unreliability and unpredictability nature of the cyber system could have negative effects on the power system.

Both cyber and power systems are in the form of a grid, containing nodes and their connections. This property makes possible to define similar structure based on the graph theory and algorithms. Beside these similarities, there are main differences that have to be considered in the modeling of a cyber-power system:

- the most important difference comes from data and electricity dissimilar distribution over time. While data in cyber network has a discrete-event behavior, electricity in power system has a continuous-time behavior. So, one of the most important issue to be considered while modeling a smart grid is to find a solution to include both natures in the same framework;
- while in the power grid, quantitative values like current, voltage, frequency are important for the performance evaluation, in the cyber network connectivity, interoperability, safety and transmission speed, *etc.*, are important;
- congestion has different effects on each network. Various events can result in a congestion: contribution of neighbors to compensate a lack of generation, triggering of a line, which will modify the power transmission on the adjacent lines by Kirshoff lows, switching on a large load when the network is already stressed, or a poor configuration of the network, *etc.* In result, temperature increases and can damage power lines. The solution here can be the redispatching of generation, and load shedding. However, congestion in cyber network occurs when there is a collision in the half-duplex connection, or at the output ports of switches, routers, bridges, and gateways. Most of the time, congestion results in delays and data loss in the ICT network;
- both of the networks – from a graph theory point of view – contain four main elements: nodes, connections, sources nodes, and sink nodes. In power system, nodes are defined as sources of energy generation, and consuming loads. Whereas, in the cyber network, each pair of nodes engage a peer-to-peer communication, so communication network is bidirectional. Each node can act as a message sender or receiver in the network;
- finally, the nature of data and electricity is different. As electricity obeys Kirchhoff's voltage and current laws (KVL, KCL), data transmission is controllable through the path definition

in the network switches.

According to [43], interdependencies between power and cyber networks can be categorized in four groups considering how they can influence each other:

- Direct Element-Element Interdependencies (DEEI): a failure in a group of elements in one network can either change the specifications of one element in the other network, or even cause failure of the element. Fig. 2.1 shows an example of this type of interdependency in which failure of the controller C1 can cause failure of the circuit breaker CB1;
- Direct Network-Element Interdependencies (DNEI): performance of each of the networks can change the specification of an element in the other network or result in a failure of the element. Fig. 2.1 contains an example for DNEI. If the connectivity between controller C1 and HMI can be established, operator can send a control command (open/close) from HMI to the circuit breaker CB1. Failure of switch S2 does not mean that there is no possibility to send this message to the CB1, because there is another path through S1, S2, and S4 to pass. Thus, a connectivity check can help to survey the impact of the cyber network performance on the power element, *i.e.*, CB1;
- Indirect Element-Element Interdependencies (IEEI): a failure in a group of elements in one network does not affect directly an element in the other network such as DEEI and DNEI, but can affect its performance against failure. Such an example is a failure of the indicators installed in a cyber network by which it monitors the power system. The indicators send an upcoming failure of the related element. Failure of any of these indicators can increase the failure risk of the power element, since the operator is not aware of the current status of the power element to make a correct protective/control decision;
- Indirect Network-Element Interdependencies (INEI): performance of each of the networks does not change the specifications of an element in the other network or cause its failure directly, but can influence its performance against failure. As shown in Fig. 2.2, there are two types of protection against faults on feeder F1: primary (*i.e.*, protective relay R1 acts as the primary protection on CB1) and backup (*i.e.*, protective relays R1, R2, R3, and R0 in communication with each other serve the backup protection) to operate the circuit breakers. It can be imagined, if connection between R2-R3 and R1-R2 fail, R2 will be isolated and will not receive the backup protection request, then CB2 cannot be opened to protect F2 from the fault on F1.

The concept of smart grid is based on such features, which requires changes in the control paradigm. Entities within a smart grid – production units, loads, circuit breakers, power stations, *etc.* – can communicate with each other and exchange information by means of a communication network that overlays – or is integrated into – the energy network. Imbrication between these two networks becomes increasingly important, leading to concepts such as "critical infrastructures". Governing phenomena have non-unitary character, *i.e.*, multi-domain, multi-physics, and take place at different time scales (multi-scale), this is a real challenge for stating a unitary control and modeling paradigm.

Power grids stability and reliability are extensively dependent on the relying communication network. Though, a robust communication infrastructure is the touchstone of smart grid that differentiates it from the conventional electrical grid by transforming it into an intelligent

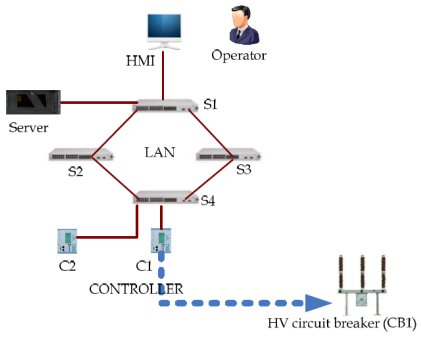


Figure 2.1 – An example for interdependency effects in a typical cyber-power network [43]. In this typical network LAN stands for Local Area Network, and HMI means Human Machine Interface.

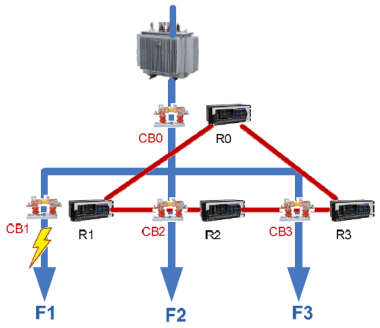


Figure 2.2 – An example of the INEI effect in a typical cyber-power network [43].

and adaptive energy delivery network.

Smart grid SAS application considering cyber-power interdependencies

As explained in the previous chapter, SAS has the task of monitoring, protecting and controlling of the related substation. SAS is a cyber network which selects required data from the power equipment (*e.g.*, incoming/outgoing feeders, transformers, circuit breakers (CBs), and disconnect switches (DSs)); data is processed in SAS, then a proper command is sent to operate the power equipment. SAS gradually replaces the conventional control systems.

IEC 61850, as a new standard for communication in SAS, provides interoperability and reliability to the communication system [89]. As SAS extends, the failure risks are increased. Any failure in SAS operation could cause failures in the power system, and may even disconnect power feeders in the substation.

Measurement, protection, and control IEDs, such as Bay Control Units (BCUs), protective relays and measuring units, are the main devices in the SAS, which are connected to each other by an Ethernet network. BCUs monitor and control each bay, and report all status and values of switchgears to other IEDs. Protection relays associated with circuit breakers

protect the substation, and all the connected lines against faults.

In addition, measuring units collect measured current and voltage, and calculate powers and energies. For each of the above-mentioned interdependencies, an example is presented below:

- DEEI: since measuring accuracy in Current Transformers (CTs) and Voltage Transformers (VTs), as well as the correct status of Circuit Breakers (CBs) and Disconnect Switches (DSs) are important concerns of SAS, any inaccuracy of them can lead to an incorrect monitoring, improper protection, and insecure control. So, DEEI can be defined between CTs, VTs, CBs, and DSs and their corresponding BCUs;
- DNEI: this type defines the interactions between cyber network and power equipment, including transformers, CBs and DSs. Any problem in the connectivity, interoperability, and synchronization in the cyber network may disturb data transmission, that, for example, are required for changing the tap of transformers, and status of CBs and DSs;
- IEEI: in the substation, there are indicators that monitor the temperature of windings and oil of transformers. If temperature exceeds the proper thresholds, related alarms are activated. Thus, if any of these indicators fails, the risk of the failure on transformers is elevated because the indicator cannot announce that the transformer is under risk;
- INEI: a peer-to-peer communication in SAS is crucial for applying protection tasks. Any connection loss or communication delay in the cyber network potentially decreases the ability of power feeders, and bus bars to defend against failures.

In this section, some of the interdependencies between communication and power grid are categorized, through which a perspective of how they are interacting, and how their stability depends on each other is explained.

2.1.1 Interoperability

Interoperability in a smart grid is the ability for IEDs from one or several manufacturers to exchange information, and to use this information for their own functions. The issue here is the interoperation of system components, which are supposed to confirm common standard specifications. Interoperability allows a network to seamlessly and autonomously integrate all the components of electric power supply, particularly monitoring and protection equipment, distribution and substation equipment, management and communication equipment.

The minimization of human intervention is an important benefit of this functionality. The challenges include the need for technical enhancement of the network, and adaptation of existing technologies, also development and implementation of comprehensive standards.

Interoperability model for smart grid

Several key features of the interoperable efficiency of a network have to be considered in relation to each other, and also independently. The GridWise Architecture Council (GWAC) was formed by the U.S. Department of Energy to promote and enable interoperability among many entities that interact with the national electric power system. GWAC developed an

eight-layer conceptual model for smart grid.

This eight-layer stack, called GWAC stack, provides a context for determining smart grid interoperability requirements and defining exchanges of information.

The layers represent the chronological processes that enable various interactions and transactions within the smart grid. Each layer depends on the lower layer, so each layer must function properly for the entire stack to be effective. Different layers of the interoperability model are shown in Fig. 2.3. The GWAC proposed some tools to achieve interoperability through covering these areas [12]:

- technical: considers the syntax, the format of the information, focusing on how the information is represented on the communication medium,
- informational: considers the semantic aspects of interoperation, paying attention to what information is exchanged and its meaning,
- operational: considers the pragmatic (business and policy) aspects of interoperation, especially those related to the electricity management.

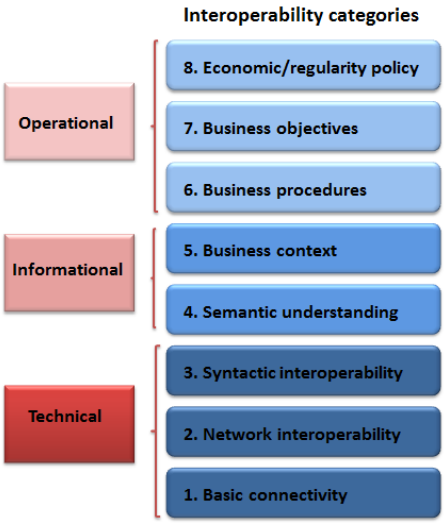


Figure 2.3 – GWAC interoperability model.

Machines (IEDs, HMI, PMUs, *etc.*) of a network need specific data and instructions in order to perform their tasks. A common language, and proper protocols, should be designed to ensure an effective communication among the machines that may be governed by the same protocol, or even further for those using different protocols. The main goal is to facilitate data transfer among different devices in an efficient and quick way.

Interoperability is not limited to the physical aspect of the network, and must also consider when two devices try to exchange data, messages have to "speak a common language" of network navigation and be properly "addressed" to reach their destination. Thus, there is a need of networking standards for the connected machines, which need to be able to communicate without interruption or disruption.

Smart grid standardization

Standards are the specifications that either establish the compatibility of a product for a particular use, or define the function and performance of a device or system. Many standards structures, including the National Institute of Standards and Technology (NIST), International Electrotechnical Commission (IEC), Institute of Electrical and Electronic Engineers (IEEE), Internet Engineering Task Force (IETF), American National Standards Institute (ANSI), North American Reliability Corporation (NERC), and the World Wide Web Consortium (W3C) are addressing interoperability issues for a broad range of industries, including power industry.

As explained in Section 1.2.1.2, the urgent need for the development of standards has led NIST to develop a plan to accelerate the identification and establishment of standards. Based on the first phase of this work, NIST published "NIST framework and road map for Smart Grid Interoperability Standards Release 1.0". In this publication, nearly 80 existing standards have been identified. The road map for interoperability by NIST includes the following applications:

- demand response and consumer energy efficiency
- wide area situation awareness
- electric storage
- electric transportation
- advanced metering management
- distribution grid management
- cyber security
- network communication

Additionally, there are several standards identified for different levels of communication, demand response, and measurement devices. As an example, in the area of power quality, the IEEE and other governing bodies have their own sets of standards. On an international level, IEC plays an important role in the power systems domain, as the development of IEDs is one of the most important steps toward an actual realization of smart grids. In this context, IEC 61850 defines a standardized way for the power utility automation and information exchange between IEDs.

Interoperability of IEC 61850 standard in power systems

The focus of IEC 61850 standard is on the interoperability between IEDs on device level, which are systems manufactured by different vendors in the power systems domain. Its first goal was to provide a new approach for the automation of substations, but further it was extended to cover the DERs components. Basically, IEC 61850 includes two main parts: power system modeling and communication between components of the electricity system.

IEC 61850 is highly focused around models, where different grid components are modeled in an object-oriented manner. Summarizing, the following three models with different objects are of high importance:

- functional model: is intended as a logical model, where the IEC 61850 objects are not necessarily connected to a real component of the power system (*i.e.*, voltage level),
- product model: provides objects for modeling IEDs and their data objects,
- communication model: provides a model of the communication network to which the IEDs are connected.

Fig. 2.4 shows the main objects provided in the previously described IEC 61850 models in Unified Modeling Language (UML) notation. In the product model, semantics for real-time and non-real-time data exchange between IEC 61850 devices (*i.e.*, IEDs), is defined. The different elements of this model are hierarchically organized: IED, server, logical device (LD), logical node (LN), and Data [7].

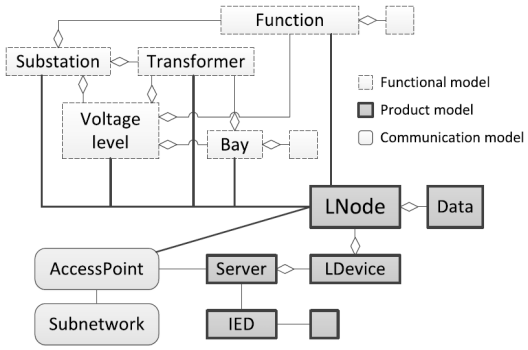


Figure 2.4 – A sample object model defined in IEC 61850 [8].

A data element is also divided into different objects: data objects, common data classes, and data attributes according to the IEC 61850 specification [8]. Moreover, the Structured Control Language (SCL) defines the concepts for configuration of IEDs, which is serialized in an Extensible Markup Language (XML) format [8]. IEC 61850 also defines communication services for the data transfer between devices (*i.e.*, IEDs). This includes client/server based but also real-time communication (*i.e.*, sampled value, Generic Substation Event (GSE) and GOOSE).

As explained in Section 2.1, considering the increasing complexity of smart grid, it is difficult to investigate research questions within one dedicated specific environment. The coupling between continuous power system and discrete-event structure of communication network engenders a mutation between two environments. As a critical issue, researchers proposed different methods to study these inter-dependencies and co-effects of two grids.

One of the practical solutions is the co-simulation of power system with communication and automation system, so the impacts and dependencies of communication on the system can be investigated. As control and automation systems in power grids rely on communication infrastructure, the properties of the communication configuration can have influence on the overall behavior of the system. For example, if the measured voltage values are not delivered in a timely manner, the response of the voltage control would lag behind and could worsen the situation.

Another example might be the impact of not received measurements due to a fail in communication, resulting in non-observability of the network state. Irrespective of the protocol used, the information flux through a communication network generally exhibits variable delays potentially responsible for unstable – or at least degraded [165] – behavior of the power system. Moreover, certain asynchronous events – erroneous transmissions, collisions, loss of data packets, *etc.* – may also take place.

Thus, a smart grid in interaction with its communication network exhibits hybrid dynamics. The energy grids have already been analyzed as complex interconnected systems and modeled as hybrid dynamic systems in a systematic context allowing to study their structural properties such as the stability [148]. Recent works have envisaged more general methods for modeling the communication protocols for large-scale systems [3]. Thus, the global goal of this thesis is to obtain a unitary model of the communication protocols that are most used for energy grids in interaction with these latter, *i.e.*, a mathematical model is needed, that can describe the traffic behavior through the communication network controlling, monitoring and protecting the power system. As a perspective, a new control paradigm may be designed based on such proposed model, while representing related interactions.

Smart grid realization, as explained, requires a vast change over the existing power system. The main challenge is to integrate a reliable and robust communication network for devices to communicate efficiently, and cooperate for the assigned tasks. Thus, to perform all these revolutionary changes, there is a need of a simulation system to validate correct operation before applying it on the real power system. In addition, it is possible to study the interactions between power grid and communication network, and according to the results, make the best decision.

Through all changes from the conventional power grid to the smart grid, the control paradigm needs to be changed too. In this case, in addition to the simulation, both of the grids should be modeled in a way to represent their interactions. So, it further becomes possible to propose new control methods on the smart grid. Next section presents a state of the art on the smart grid modeling and simulation. Finally, as one of the most important communication network evaluation factors, researches related to communication delay estimation are overviewed.

2.2 Researches in smart grid communications

The fundamental structure required to move power grid through smart grids by intelligence is the communication network. All the advantages of smart grid to be green, power-efficient, and distributed-controlled are strictly relied on ICT. So today, integration of ICT to the distribution grid is a real challenge. Based on the literature review done in the framework of this thesis, the existing studies cover principally the following issues: co-simulation, modeling and validation of the smart grid testbeds [134], [26], [15]. These issues will be detailed in next subsections.

Some critical issues for smart grid communications (*e.g.*, performance modeling, control, and optimization algorithms) are also addressed [153], [5], [163].

There are also interesting works on the modeling and analysis of the IEC 61850 protocol [158], and the security issues in Advanced Metering Infrastructure (AMI) [159].

2.2.1 Co-simulation studies

In smart grids, the bidirectional exchange of large amounts of data will create an intense interdependence between electric system and communication, automation and control systems. In order to assess the performance of communication technologies for smart grid applications and find the most appropriate one, co-simulation tools are crucial to simulate the behavior of both physical and cyber systems simultaneously, and properly consider their mutual interactions in planning and operation studies.

Transition toward smart grid involves an extensive use of ICT, and innovative control systems, enabling various intelligent and automated applications, such as building automation, distribution automation, load side management, DERs and integration of electric vehicles [38]. In smart grid, therefore, ICT is not a simple auxiliary infrastructure of the power system, but a fundamental component for its effective operation, that ensures reliable and real-time data collection from an enormous number of widely dispersed data sources and responsive data transmission for power network control.

Various types of existing communication technologies, wired or wireless, can be an opportunity for the smart grid applications; however researchers are still wondering which technology satisfies better smart grid requirements. The reason is that different smart grid applications have different communication requirements, in terms of data payloads, sampling rates, latency and reliability [93]. Thus, it is essential to have simulation tools which make possible to compare performances of different technical solutions and help utility planners do the best choice in their smart grid implementation projects. Co-simulation is a useful method to simulate power distribution system, and ICT system behavior simultaneously, including the existing relationship and interdependencies between the two systems.

Co-simulation couples the simulations of power system and communication network in order to analyze their interdependencies, *e.g.*, for control applications. Coupling of these two systems of a total different nature – discrete-event-driven and continuous-time-driven – allows a dynamic simulation of the impact of communication system on power system, and opens a wide range of mutual interaction opportunities. Computer networks and communication structures that fulfill dedicated control purposes have discrete-event-driven nature and show stochastically distributed latencies. Such multi-domain systems do not fit well into the current monolithic power system and components simulation paradigms.

One of the important challenges is that ICT and power electronic devices are developing so fast, so the behavior of each system and their interactions change quickly over time. Developments such as the Internet of Things (IoT), smart homes, and vehicle-to-grid (V2G)

communication further contribute to a data-driven power system. ICT for power systems has different applications, therefore different requirements that have to be fulfilled. For example, timing constrains in communication range from very relaxed in the case of meter reading to very strict in the case of high-speed signals for protection purposes. So, using co-simulation to investigate the mutual influences of ICT and power systems, therefore the behavior of intelligent power systems, has become significant.

Smart grids still need to be improved, and related improvements should be tested on small scale before deploying on the real-world scale [60]. Thus, their results and effects should be evaluated through simulations, and random tests. Smart grid simulators may be classified according to their modeling capabilities of power, and communication systems. Three alternative approaches have been proposed in the literature to tackle this type of study: co-simulation, comprehensive simulation, and Power-Hardware-In-the-Loop (PHIL) simulation. Co-simulation approach usually involves the integration of two or more simulators to capture the cyber-physical interdependency of a process or system.

In the co-simulation, each system is analyzed by its own dedicated simulator, and all simulators are executed simultaneously by appropriately designed Run Time Interfaces (RTI), and coordinated simulation management.

There are different solutions to implement a *co-simulation* found in literature, which vary upon their research goals on smart grids, and consequently upon the architectural choices (*e.g.*, software components, time synchronization strategy, and scalability). One of the first co-simulation tools for power systems is the Electric Power and communication synCHronizing Simulator (EPOCHS) [65], which was developed to better study the effects of communication network integration into power control systems on the stability of the electric power systems. It uses a combination of three simulators that are synchronized by a time-stepped method: a simulator for electromagnetic transient simulation (PSCAD/EMTDC), a simulator for electromechanical transient simulation (PSLF), and NS2, which is used for communication network simulation [34]. As each of these was designed as a stand-alone simulator, EPOCHS connects these simulators by an RTI method on the High-Level Architecture (HLA) in order to make them interoperable.

Another important pioneer platform for co-simulation is the Global Event-driven CO-simulation framework (GECO) [102], which applies an event-driven method for the simulation synchronization of power system (modeled with PSLF), and communication network (modeled with NS2). In this platform, each iteration of the numerical solution for power flow is an event. All events are integrated in the event scheduler of NS2, allowing a perfectly integrated simulation and the minimization of synchronization errors. In comparison to the time-stepped synchronization, event-driven synchronization reduces the simulation time, and allows the simulation of large power systems with reduced computational tasks.

Comprehensive simulation, as an alternative approach combines power system and communication network simulation in one environment. In this case, the main issue is to bring together both of the system models, and solve the routines which contributes either to the integration of power system simulation techniques into the communication network simulator

or *vice-versa*. A comprehensive simulation approach has been adopted in [117], where authors present a modular simulation environment based on discrete-event simulator (OMNeT++) using the existing models for communication network, but consciously extra models for the electrical network are developed.

The last proposed simulation approach is the Power-Hardware-In-the-Loop (PHIL), often used in power system studies for testing the control and protection strategies [28], where software simulators, and hardware components are integrated in a real test bench. The PHIL approach allows a very well correspondence with a real system, but it is more costly than a software-only simulation. A detailed state of the art review of appropriate tools for simulating both domains of power system and ICT processes in the evolution of smart grids is presented in [122], [18].

One of the other key points to be studied is the cyber security of smart grids. As a critical issue, in [147] a PHIL test bench, is proposed, whose contributions are: 1) it covers from the high-voltage to low-voltage power systems, and 2) it can simulate the distribution level of cyber attacks and check their impact on the transmission level power system. The Object linking and embedding for Process Control (OPC) communication is developed, which is widely used in industrial automation system test bench. Using the OPC server/client software, real-time data can be exchanged between cyber and physical systems in the test bench.

Among all these three methods, co-simulation is the most common method to study smart grid as a cyber-physical system, since it allows to simulate each subsystem (electrical and ICT) with a simulator specifically developed for that purpose. In comparison to the other approaches, some of the co-simulation benefits can be a stronger accordance with a real system, while the investment, and required time for a prototypical result is minimized.

A notable feature of co-simulations of intelligent power networks is the existence of different time scales that are combined in one model. ICT, and especially the controls of intelligent power grids, expose another important aspect that the simulation models have to consider: real-time guaranties (GOOSE) with no need to wait for an acknowledgment in order to reduce the transmission time, or the best-effort basis (MMS) while the message delivery is confirmed by an acknowledgment. Based on the prior discussion, it is evident that the flexibility of co-simulation potentially allows for a mixture of the ICT and power components, as well as systems of different time scales, even in real-time.

To decide for the best fitting solution among all the possible industrial applications of communication technologies for smart grid, a comparative performance evaluation is needed. In [54], authors propose a co-simulation platform, which adopts a power system simulator (OpenDSS) and an ICT network simulator (NS3), coordinated by MATLAB®. It manages the co-simulation process and implements the control logic. This platform is designed to give the chance of comparison among different ICT solutions for smart grid.

Another example of co-simulations to evaluate existing communication technologies for smart grid is presented in [53]. This work investigates the use of public mobile telecommu-

nication system 4G Long Term Evolution (LTE) applications in both radially and weakly-meshed medium voltage (MV) distribution networks. The results demonstrate that LTE can be used as communication medium for advanced fault location, extinction, and network reconfiguration in distribution networks.

As discussed previously, different natures of electrical dynamics and communication system events should be treated differently in simulations – electrical dynamics are time-driven, whereas a communication system is event-based. Because of these contrasting features, they cannot simply be combined into one single piece of software. Simulation software specialized in either analyzing electric power grids or communication networks is widespread.

[15] presents a co-simulation framework developed, over which a hybrid-state estimation application, a voltage monitoring application and a renewable energy integration application are implemented. The obtained results verify their field correctness, also validate the correctness of the co-simulation framework. OpenDSS and OMNET++ are used respectively for power systems and communication networks simulation.

In [26], authors validate that smart grid traffic can be described by a Poisson distribution and thus M/M/1 and M/G/1 queuing models are applied to analyze the delay performance. These models are based on the Kendall's notion of queuing theory standard: first letter indicates the nature of arrival curve, second letter indicates the service time distribution, and the last number indicates the number of servers. M and G signify the Poisson process, in which M stands for memoryless (*i.e.*, exponential distributed interarrival times), and G stands for general distribution (*i.e.*, not necessarily exponential). Third symbol is the number of communication channels. This paper first develops a queuing model over the communication traffic among the Distributed Grid Intelligences (DGIs) in the network with different topologies has been developed. Second, based on the proposed model, delay performance is analyzed and illustrates that the model can be used to predict grouping delays. Third, through the collection of an extensive experimental data, accuracy of the model is validated. Finally, they have implemented a HIL testbed with the capabilities of integrated real-time communication and power exchange. They report a large-scale smart grid prototype, called FREEDM[1] HIL testbed, in which power system is implemented using RTDS (Real-Time Digital Simulator, from Hydro-Quebec) and the communication network is implemented in Optimized Network Engineering Tool (OPNET) modeler.

[142] sets up a simulation model on the OPNET Modeler platform to demonstrate that Ethernet has satisfactory performance to meet real-time demands of SCN.

[139] constructed an IED model in SCN and developed research platform to be able to build SAS network models with different topologies for all kinds of substations. So, the SCN dynamic performance could be evaluated. Authors introduce the modeling technique of IED generic models using OPNET modeler.

[1]Future Renewable Electric Energy Delivery and Management system, which has been funded by U.S. National Science Foundation (NSF) since 2009 as a communication system for power grid. FREEDM proposes a novel idea of "Internet of energy" like Internet for information. It has established an energy research center with two large-scale testbeds for smart grids.

Considering the co-simulation of the communication grid event-based nature and its effects on the continuous nature of the electrical grid, there are efforts to model and co-simulate their interactions while cooperating for a common goal. In addition, considering fast enrollment of the communication networks for these functions, it is crucial to understand their impact on the performance of the electrical grid.

However, a combination of these two is not easily available. Therefore, researchers have been trying to combine popular tools, including MATLAB®/Simulink ®, with each other to ideally converge toward a single platform able to simulate both networks. This is considered very promising in giving details about the interaction between both systems. However, advanced mechanisms were needed to be developed, in order to ensure correct coordination between the two simulation tools, since they were not designed with the philosophy of having to cooperate – examples include [16], [19] and [147].

In [128], the effect of perturbations within the communication network, which is providing services for reactive power control, was modeled and implemented. A communication network, consisting of a switch and IEDs transferring data over an Ethernet wired network using the design principles of IEC 61850 GOOSE protocol, is considered. The electrical grid includes two DERs and a load, in which DERs are cooperating to compensate the load reactive power consumption. A simple decentralized reactive power control algorithm is implemented, which requires that the three buses communicate with each other in order to reach their common goal. Packet loss as a perturbation within the communication network, which is providing services for reactive power control, is analyzed. The results show a direct influence of the communication network performance on the electrical grid. It was seen that packet loss on different communication links has various effects, which cannot always be treated lightly. Communication network configuration, IEC 61850 GOOSE protocol, and its detailed model is built using blocks of the SimEvents®/Simulink® toolbox, and the electrical grid is implemented using SimPowerSystem® toolbox.

2.2.2 SCN traffic modeling based on IEC 61850

Through the evolution of smart grid, the complex and long-wired networks are replaced with Ethernet-based communication network to transmit information in SAS. So, SAS is more and more dependent on SCN for monitoring, control and protection. In this case, building a custom-designed and well-monitored SCN, which can help to maintain a fast and reliable information transmission, leads to an improved SAS operation and management. In order to achieve this goal, it is important to develop reliable SCN models and evaluate network performance under different conditions, and analyze the network performance requirements.

Hence, for a detailed understanding of both physical and cyber components of SCN, vast research effort has been done on the mathematical modeling and simulations over the SCN data flow and analyzing the influence of communication network inefficiencies (*e.g.*, transmission delays, packet loss, communication interruption, collision effects, *etc.*) on the power grid performance. Data-flow models are beneficial for the acquisition of more convincing

results to assess the network performance [141], [81], [126], [90], [151], [70], [72].

According to their goals, related researches can mainly be divided into two parts:

- improvement of communication architecture [6], [104]
- assessment of network performance [69], [80]

Although there have been much works on studying the substation communication system by means of LAN simulation, there are still several unresolved issues in order to assess the dynamic performance of SCN comprehensively and accurately. Enhancement of the previous research is required for two reasons: 1) there is no proper mathematical model to describe uniformly various messages in SAS, which is the basis for accurate evaluation of network performance, 2) comprehensive assessment of real-time performance based on the proposed data-flow models is needed, in order to acquire more convincing results.

In fact, some preliminary works have been presented in [35]. A network model of IEDs in SAS with star topology is simulated using OPNET modeler, in which different types of data flows – *i.e.*, cyclic data flow, random data flow and burst data flow – are discussed. Based on this network, the impacts of some of the communication parameters (*e.g.*, packet interval transmission, packet size, high-level protocol, and utilization of CPU background) on the end-to-end transmission real-time performance, *i.e.*, transmission delay, is analyzed. However, a specific mathematical model for the concerned three kinds of data flow has not been developed, also the traffic management technologies, such as VLAN, are not thoroughly evaluated.

A co-simulation of the power grid considering the integrated communication system is proposed in [16], which shows how a co-simulation platform can be set up using OPNET as the communication network simulator, and OPAL-RT as the power system real-time simulator. Through a case study, it is demonstrated that this co-simulation can be used for performance analysis of smart grid operation in a real-time mode.

[99] presented a 104 protocol/UDP transmission scheme based on the IEC 60870-5-104, as well as the UDP transfer protocol. The real-time performance of such protocol is evaluated in the Electric Power and Communication Synchronizing Simulator (EPOCHS) platform. Results indicate that UDP is a better transmission protocol in SAS environments in comparison to TCP protocol. However, in this work SV and GOOSE are ignored despite they are the fundamental protocols in supporting the SCN applications.

Authors of [164] propose three different models for each type of data flow within SCN, based on IEC 61850 standard (cyclic, random, burst). Based on these models, a quantitative analysis of typical data flow is carried out, and the real-time performance of a VLAN-based substation is evaluated. It can be concluded by the results that: 1) the proper configuration of network reduces the utilization of network links, and the Ethernet delay of SCN by reducing the transmission of cyclic data flow within the bay level, 2) in occurrence of system fault and FTP file transfer, the real-time performance of SCN is affected, 3) the substantial growth of stochastic and burst messages during and after fault may consume more bandwidth resources

and thus cause potential threats to the network features, 4) a comparison of ring and star topologies is performed, which shows that the ring structure is superior in both real-time and reliability performance of the network. Despite proposing mathematical models for SCN data flows, there is a lack of a unitary model which can describe all types of flows through a conclusive mathematical expression.

Considering the management challenge of a vast amount of smart grid traffic in the access network, in [4] a traffic modeling project and performance analysis based on a wireless smart grid network is proposed. This paper proposes an analytical traffic model for smart grid access networks based on a priority queuing system. The traffic is classified in Fixed Scheduling (FS) (such as smart meter readings) and Event Driven (ED) with higher priority (such as demand response). By using the proposed model, the authors drive expressions for the performance metrics, mean buffer length and queuing delay of each traffic.

A mathematical framework is developed in [88]. The framework is presented for characterizing the cyber-power system behavior, where system descriptions and associated models are selected based on the system dimension, modeling purpose, and necessary level spatio-temporal resolution or scale. State/structure interactions are expressed in a hybrid model that evolves on a continuous/discrete state/structure space, as well as mesoscopic descriptions and models that are based on Gibbsean statistical mechanics.

Paper [158] proposes a FARIMA(p, q) – Fractional Auto Regressive Integrated Moving Average – model considering all types of IEC 61850 traffic flows, and mathematically analyzes the associated parameters in order to accurately model the self-similarity behavior of SCN traffic, which describes the co-existence of short-range and long-range dependence. Finally, the model effectively predicts the traffic pattern. The model is validated over a 24-hour sampled data collected from a 110 kV substation SCADA.

2.2.3 Delay estimation over SCN message transmission

Since IEC 61850 standard has been released, different approaches were developed to evaluate the performance of the communication services defined in IEC 61850. This section represents an overview of the relevant literature, which is done with focus on performance evaluation of IEC-61850-based communication networks.

In [139], at first, three types of generic IEDs (*i.e.*, breaker IED, merging unit (MU), and combined protection and control (P&C)) are modeled for the SCN operation, next a setup is introduced by OPNET modeler as a test bench to be used for performance analysis of the communication network. The results show that the network performance satisfies time critical messages in SAS.

The studies are extended in [82], in which communication between a substation and DERs using GOOSE messages is simulated. Wired and wireless communication solutions are compared, and simulation results are validated by laboratory experiments. However, GOOSE communication is only taken into account partially. Simulation results comprise average and

worst case message delay in inter- and intra-bay communication for different sampling rates – examining both star and bus topology with 10 MBit/s and 100 MBit/s Ethernet connections.

In [91], authors present their approach of SV traffic generation. This paper features a detailed description of the SV packet format, and arising traffic when sending such packets. For simulation purposes NS3 is chosen, and results are validated by Wireshark network analyzer. Although an explicit protocol implementation for SV is provided, other message types like GOOSE or MMS need to be taken into account for a complete performance evaluation of IEC-61850-based substations.

In [55], the generic architecture of IEC 61850 is first introduced, then a modeling approach is presented for evaluating high performance and real-time capability of communication technologies for smart grid applications. To do the performance evaluation, a simulation model along with an analytical approach is developed based on the Network Calculus theorem, which enables the identification of the worst case boundaries for intra-substation communication performance. Simulation is done using OPNET modeler to examine the generation of SVs and GOOSE messages in order to evaluate the resulting overall real-time performance of the bay LAN.

Two scenarios are defined for performance evaluation: a simple scenario with the SV traffic only, and a more complex scenario that contains both SV and GOOSE traffic. In these basic examples, both simulation and analytical model have shown delays significantly less than 10 ms, which satisfies the real-time requirements defined in IEC 61850. The results show that Network Calculus theorem is able to determine appropriate upper delay bounds compared to the simulation results. Hence, the Network Calculus theorem can be used for analytical evaluation of real-time constrained communication in smart grids provided that network configuration is known.

Authors of [161] propose an analytical model based on the Network calculus theorem, which can measure the maximum end-to-end delay of the SCN message transmission. At the beginning, a set of traffic source-flow model, and traffic service-flow model of the IEC 61850 SCN traffic is established. At the next step, a traffic flow calculation algorithm is developed to obtain traffic load distribution, and calculate the maximum message delay. The accuracy of the calculation method is verified using a network analyzer.

In this thesis, we have selected the delay estimation method proposed by [161] to do performance analysis over a real-time IEC-61850-based communication network. The selected method can be applied on any types of messages through SCN while it considers the number of transmitted packets over time. All the modeling process, and delay calculation can be performed by simple functions of MATLAB®. In addition, estimated maximum delay is based on the traffic flow distribution by which an optimal network configuration can be found out to minimize delay.

2.3 Research motivation and objectives

As explained in the previous sections, communication network, as a fundamental infrastructure for smart grids, is experiencing a vast range of changes. Thus, any changes need to be evaluated through a reduced-scale test bench, before it can be introduced in the real smart grid scale.

Co-simulations and mathematical modeling are the main approaches that give the opportunity to validate new methods that describe smart grid behavior, and the existing cyber-physical interactions.

IEC 61850 standard is considered in this thesis, which provides interoperability of the IEDs through the SCN. Today, a reliable data communication among IEDs is a key issue. Modeling and analysis of the SCN traffic flow will further help to state, and solve the associated control problems of the smart power grids.

Research approach and objectives of this thesis are motivated by the need of obtaining a unitary mathematical model to describe the traffic flow behavior of SCN, and analyze the interactions of energy-communication network.

Based on the literature review, we were mainly interested in two reporting results about the SCN traffic modeling:

- [164], which models different protocols of IEC 61850, *i.e.*, GOOSE, MMS, SV, separately,
- [158], in which SCN traffic is modeled through the time-series analysis method, *i.e.*, FARIMA, which only estimates data stochastic behavior in order to predict the traffic pattern.

But, neither of these papers result in a model which can describe mathematically both stochastic and deterministic behavior of traffic flow (only stochastic part is considered). First paper models each of the IEC 61850 protocols separately, and there is a lack of a unitary model. Hence, using time-series analysis, we propose a model describing both stochastic and deterministic behaviors, which is applicable irrespective of the protocol type.

Afterwards in the second part of this thesis, as the latency can significantly affect the power grid stability, or at least degrade its performance, another objective of this thesis is to analyze the message transmission delay through SCN. By disposing of a reliable estimation of the maximum delay, it is possible to find an optimal communication network configuration or, if necessary, alternative communication paths to insure a real-time well-performing communication.

[55] and [161] analyze the transmission delay through SCN network based on the Network Calculus Theorem, and the results show network calculus theorem can accurately estimate transmission delay. We consider the model of [161], which propose a computation method to estimate the maximum end-to-end delay, with the aim of validating the proposed model in a load-shedding scenario, considered pertinent as it requires communication in between

different IEDs in order for the correct shedding decision to be taken.

2.4 Conclusion

In this chapter, a representation of smart grid as a cyber-physical system has been presented, and introduces the issues of interactions between communication network and power grid. Interoperability through the smart grid devices has been explained next, with a representation of the GridWise Architecture Council (GWAC) interoperability model. Section 2.2 has presented a state of the art, which discusses the need of co-simulation, and modeling methods for smart grid. It was discussed how these models can be used for the network performance analysis to help transforming the current power grid into a smart grid. Finally, research motivation and objectives of this thesis were identified, namely the interest of modeling an SCN traffic flow and of estimating the maximum communication delay.

Next chapter will introduce IEC 61850 protocol as the considered communication protocol through this thesis. Then, based on the need of modeling for smart grid performance analysis, next chapter will introduce the most pervasive SCN modeling methods according to the literature review by this thesis. Briefly, the selected model by this thesis is introduced. Moreover, the real-time performance of the ICT infrastructure for smart grids is analyzed IEC 61850.

Chapter 3

Introduction to the IEC 61850 standard, traffic modeling methods and transmission delay estimation

Contents

3.1 Background: IEC 61850 protocol 50

3.1.1 Substation Communication Network (SCN) 54

3.1.2 Smart supervision: Intelligent Electronic Devices (IEDs) 55

3.1.3 Types of messages in IEC 61850 55

3.2 Modeling methods and influence of delay factor 57

3.2.1 Queuing theory 58

3.2.2 Analytical methods 60

3.2.3 Time-series analysis method 61

3.3 The choice of methods for traffic and delay estimation 64

3.3.1 Time series: Auto Regressive Moving Average (ARMA) model 64

3.3.2 An analytical model for transmission delay estimation 67

3.4 Conclusion 71

All the simulation and experiments in this thesis work are based on IEC 61850 communication protocol within a Substation Communication Network (SCN). Therefore, the first section of this chapter provides an introduction of IEC 61850 structure, devices, and message types through SCN as basic notions and concepts used throughout this work. The second section presents an introduction of different modeling methods over the SCN data traffic through, among which the Box-Jenkins identification method is selected for the modeling process. Finally, the third section explains our choice of the modeling method for the SCN traffic behavior estimation. Also the analytical method to estimate the maximum message transmission delay adapted from [161] is detailed, which will further be real-time validated.

3.1 Background: IEC 61850 protocol

As explained in Section 1.2.3, among different Substation Automation System (SAS) standards, IEC 61850 is becoming the global standard for communication in substations. It enables integration of all protection, control, measurement and monitoring functions within a substation, and additionally provides the means for high-speed substation protection applications, interlocking and inter-tripping.

IEC 61850 standard was completed between 2003 and 2005 following the publications in 1999 as UCA2.0 (*i.e.*, further used as IEC 61850 fundamentals) [143]. It combines the convenience of Ethernet with the performance and security, that are essential in substations today.

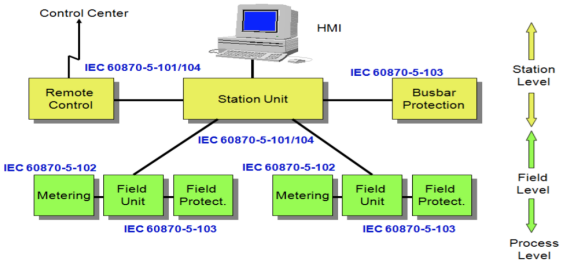
IEC 61850-enabled relays are fitted with an integral Ethernet card providing both copper and fiber Ethernet. There is no need for external adapters or data concentrators, and the only requirement is a standard Ethernet equipment such as switches and substation grade switches. While in the past the required information was transmitted over numerous type of cable in the form of binary or analogue values, today, almost all information is transmitted using Information Communication Technology (ICT).

The substation configuration language defined in the standard allows the integration of Intelligent Electronic Devices (IEDs) manufactured by different vendors, so they can operate seamlessly in a uniform communication environment. Initially, a three-leveled structure was proposed for the substation automation, including process level, field level and station level, as illustrated in Fig. 3.1 (a). Through further development of the protocol, two data buses were introduced between the levels, in order to enable the communication in a more general way. The new bus structured design can be found in Fig. 3.1 (b).

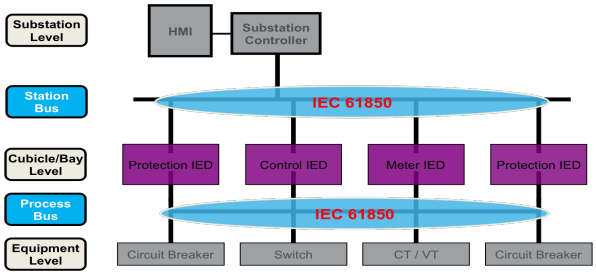
Standardized data classes and services made IEC 61850-enabled IEDs unifying the structure, requirements, and communication specifications to share data among them, and the first announcement of cooperation could be implemented. IEC 61850 architecture includes three data transmission protocols: GOOSE (Generic Object Oriented Substation Event), MMS (Manufacturing Message Specification) and SV (Sampled measured Values) along with two data buses, *i.e.*, process bus and station bus.

The OSI (Open System Interconnection) mapping of IEC 61850 protocol can be found in Fig. 3.2. Station bus interconnects the whole substation, provides connectivity between central management and individual bays and also connects different bays among themselves and bays with SCADA (Supervisory, Control and Data Acquisition system). On the other hand, process bus connects, within a bay, primary equipment to IEDs.

MMS traffic allows MMS client such as SCADA (supervision level) to access vertically all the IED objects. GOOSE traffic allows IEDs to exchange data horizontally between the bays sharing measurement and flows over station bus. SV traffic carries voltage and current samples.



(a) Operating levels before IEC 61850



(b) Operation levels of IEC 61850

Figure 3.1 – Substation automation operating levels [128].

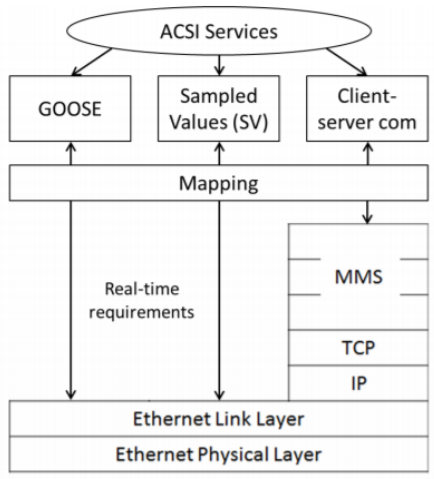


Figure 3.2 – OSI mapping of IEC 61850 protocol [137].

The communication flows are shown in Fig. 3.3. Each protocol has a transmission algorithm over time and possibly a manner to confirm the message deliverance.

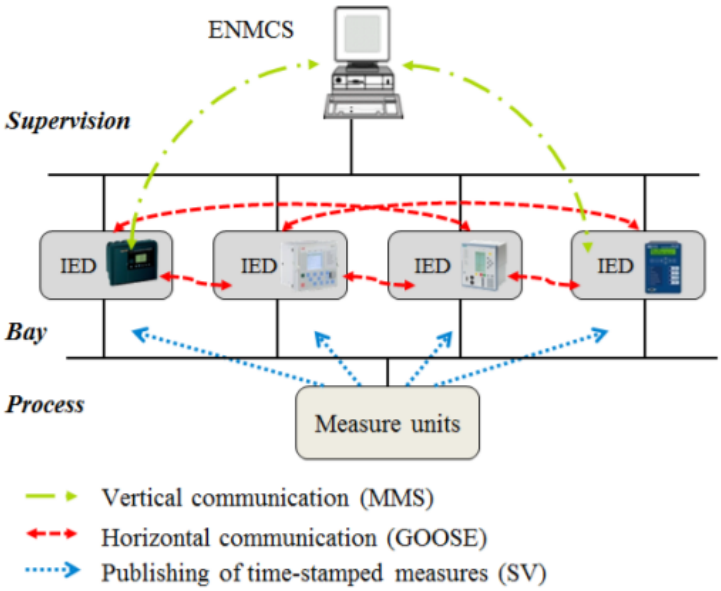


Figure 3.3 – IEC 61850 communication architecture [137].

MMS protocol is based on TCP/IP layers and allows traditional supervision, *i.e.*, read, write, and report. Considering the OSI mapping, TCP/IP has four layers: Link layer, Internet layer, Transport layer and Application layer Fig. 3.4. Link layer is the lower layer, which is responsible to move frames from one hop to another. Internet layer delivers packets from source to destination across multiple links. The intermediate transport layer is responsible for process-to-process delivery of entire message, this process is an application program running on a host. The network layer in the TCP/IP model supports the IP (Internetworking Protocol) and ICMP (Internet Control Message Protocol).

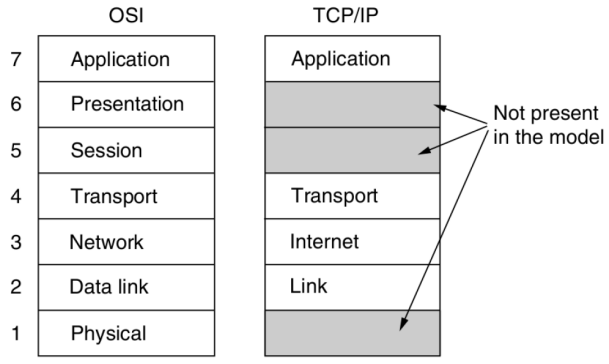


Figure 3.4 – The TCP/IP reference model [150].

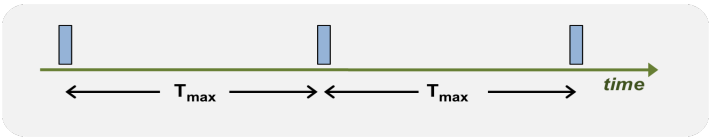
TCP is a connection-oriented transmission that requires: connection establishment, data

transfer, and connection termination. TCP uses a system of acknowledgement to confirm the reception of data. So, if the source receives back the corresponding destination acknowledgement, it is sure that data has been delivered correctly. If data is corrupted, lost or delayed, it will be retransmitted [52].

GOOSE protocol is mainly responsible for fast real-time and time-critical services. Therefore, it needs to be efficient and, with less importance, reliable. GOOSE is *multicast* so multiple IEDs can receive the same data at the same time, and allows a fast and non-redundant communication. In addition, it benefits of a *publisher/subscriber*, and *unconfirmed* communication protocol, which is a big advantage as it helps at avoiding possible delays due to acknowledgements. This features allows to discard the handshake method, used by TCP/IP based protocol, and, as a result, enables less redundant messages (beneficial for speed), potentially inducing less delay. It also makes sure that different IEDs receive the same information around the same time. Finally, it operates *directly on Ethernet*, which allows the communication once again less redundant, but also to use the built-in advanced functionalities of Ethernet. Moreover, directly operating on Ethernet allows GOOSE to evolve as Ethernet, which is prone to constant development [50].

A publisher/subscriber mechanism has many advantages and is used in an unconfirmed (without-handshake mechanism) way, so the reliability will not be ensured by acknowledgments. To increase reliability, a retransmission mechanism is used. Every frame sent by a device (*e.g*, an IED), is repeatedly sent as long as the state of that device does not change. Thus, as long as no event happens, the same data will be sent repetitively. This mechanism assumes that, as more as a frame is sent, its transmission chance through the network will increase.

This retransmission mechanism can be explained more clearly based on Fig. 3.5. The easiest way to understand this figure is to separate the mechanism into two states. System is in the normal state as nothing happens, and in the retransmission state when an event took place. In the normal state, retransmission mechanism makes sure that every T_{max} seconds the same data is sent.



(a) Normal operating condition

(b) Retranmission after event occurs

Figure 3.5 – Retransmission mechanism in GOOSE [128].

In this way, it is possible to lose some of the retransmissions or to be prone to errors while traveling through the network. In the retransmission state, as soon as an event happens, data change and new data will be sent. After a short amount of time (T_{min}, with $T_{min} < T_{max}$) data will be sent again. This happens a couple of times again, but every time with an increase of retransmission interval. The interval will increase steadily, until it reaches the same length as T_{max}, after which the system is in the first state again.

The retransmissions happening in the second state is also referred to as "fast retransmissions". A fast retransmission has the goal to send information to the receiver as fast as possible, even when something would happen with some frame along the way. In the spirit of keeping some design freedom, the length of these intervals can be chosen by the manufacturers. The retransmission system accounts for the event-based nature of the GOOSE communication protocol [44].

SV messages are published directly from the secondary converters of Merging Units (MU). MU is the general name of a device that samples conventional Current Transformer (CT) and Voltage Transformer (VT) outputs. As GOOSE messages, SVs are also transmitted directly over the Ethernet network by a MU or instrument transformer with electronic interface [71]. SVs carry the critical current and voltage values through the process bus for IEDs, and it is done periodically with a fixed sampling period.

3.1.1 Substation Communication Network (SCN)

Substations are the critical nodes in the power system network, where the system information is collected and used for the proper operation and management through SAS. Early electrical substations required manual switching or adjustment of equipment, manual data collection for load providing, energy consumption, and abnormal events. As the complexity of distribution networks grew, it became economically necessary to automate supervision, protection and control of substations from a central station, to allow overall coordination in case of emergencies and to reduce operating costs.

Early remote-controlled substations used dedicated communication wires, often accompanied with power circuits. Power-line carrier, microwave radio, fiber optic cables, as well as dedicated wired remote-control circuits have all been used within Supervisory Control and Data Acquisition (SCADA) for substations. The microprocessor development allows exponential increasing of the number of points that could be economically controlled and monitored.

Today, standardized communication protocols such as DNP3, IEC 61850 and Modbus, to list a few, are used to allow multiple IEDs to communicate with each other and with the supervisory control centers. Distributed automatic control at substations is one element of the so-called smart grid. In today's electric utilities, reliability and real-time information transmission became the key factors for trustful power delivery to the end-users, profitability of the electric utility and customer satisfaction.

The operational and commercial demands of electric utilities require a high-performance

data communication network that supports both existing functionalities and future operational requirements. An Ethernet-based logical architecture called SCN has been proposed to carry out the data transfer through the substation. A fast, reliable and deterministic communication network is required to support time-critical real-time monitoring, protection, and control functions of SAS.

The information transmission through SCN, as a key role to ensure the communication reliability and stability, makes it necessary to analyze the SCN traffic behavior through a mathematical model. In this way, the interaction between power and communication grid could be better formalized in order to find the co-effects of each to the other.

3.1.2 Smart supervision: Intelligent Electronic Devices (IEDs)

Based on the IEC 61850, all the functions in substations are decomposed into different logical nodes (LN) that may reside in one or more physical devices. The electric devices in substations evolved from the older electromechanical relays to the current IEDs as digital control and protective relays. There are four main functions in a substation: protection function, measuring/metering function, breaker control function, and voltage control function. A typical IED can contain around 5-12 protection functions, 5-8 control functions controlling separate devices, an auto-re-close function, a self-monitoring function, some communication functions, *etc.* Hence, they are properly named as Intelligent Electronic Devices.

Common types of IEDs include protective relaying devices, On Load Tap Changer (OLTC) controllers, circuit breaker controllers, capacitor bank switches, recloser controllers, voltage regulators, *etc.* IEDs receive data from sensors and power equipment and issue the control commands, such as tripping circuit breakers if they sense voltage, current, or frequency anomalies, or raise/lower voltage levels in order to maintain the voltage setpoint.

The Ethernet technology was a natural part of this evolution to provide a communication network for IEDs to talk inside the bays, with metering units and control stations. With the advent of virtual local area networks (VLANs), with an exponential increase in data communication speeds, and flow control of switched systems, Ethernet has become a reliable technology for this type of real-time application.

3.1.3 Types of messages in IEC 61850

The IEC 61850 standard allows for two groups of communication services between entities within the SAS. One group uses a client-server model, accommodating services such as Reporting and Remote Switching. The second group uses a peer-to-peer model or Generic Substation Event (GSE) services associated with time-critical activities such as fast and reliable communication between IEDs used for protection purposes.

Further in detail, IEC 61850 includes seven types of messages in SCN, where each type corresponds to a certain electrical information generated by different electrical devices. Tab. 3.1 shows this classification. Considering the traffic patterns over time, the messages could

Type	Applications
1. Fast messages	Protection command – trips/closing
2. Medium-speed messages	Measuring value, device status
3. Low-speed messages	Non-electrical parameters
4. Raw data messages	Operation parameter from transformers
5. File transfer functions	Large recording and setting files
6. Time synchronization messages	Internal clocks of IEDs and system bus
7. Command messages with access control	HMI control commands

Table 3.1 – Generation of different message types in SCN [158].

be further categorized into three groups: periodic, random and burst data flow, which are explained next.

Periodic data flow: cyclic data is a time-driven data type which is generated over the same time intervals of a fixed length. There are two types of periodic data: SV and GOOSE. SV carries time-critical information with a large amount of data flow, and has an important influence on the substation traffic. SVs are generated by MUs in the substation process level and then transmitted to the protection and control IEDs in the substation bay level. GOOSE messages transmit the meters values and breakers status information from the devices in bay level to the server in station level. These types of information are transmitted at a fixed period.

Random data flow: random data is an event-driven data flow, which means they are triggered by a fault or an unplanned event such as a trip message when a short-circuit fault occurs. In general, random messages can have a length which is either fixed or variant over time. Packets are generated in a period of time which is randomly distributed with a probability of P. There is no correlation between two consecutive packets, *i.e.*, the number of packets transmitted over two mutually exclusive time periods are independent. Therefore, the arrival of packets may be modeled by the Poisson process [120].

Stochastic data flow could be classified into two main groups. First type is smaller in size and has a short duration, so the transmission time should face the requirements of fast message type (*e.g.*, transformer tap modulation, capacitor switching, trip message, *etc*). However, the second group is larger in size and determines a sudden increase on the network flow; such messages are also considered as low-speed messages (*e.g.*, event log checking, file transfer, protection setting modification, *etc*).

Burst data flow: burst data characterizes the protection actions and the changing status of breakers, which represent events transmitted by the GOOSE flow as well. While the protection device interferes at the moment of fault occurrence, the periodic GOOSE flow changes to burst flow [30].

Generally, burst data will inject a large amount of data into the network. In a short period of time, burst data is transmitted on SAS network while the network is free for a long-time

interval, so it can be described as a series of ON/OFF micro-flows. This type of data has the characteristics of Long-Range Dependence (LRD) and self-similarity, which results in the same burstiness at different periods of time [144]. If a self-similar process has observable burst data (ON flows) distributed over all time scales, it is said to exhibit LRD characteristic through which value of any instant is typically correlated with the value at all future instants.

A time series is evaluated as an LRD process if it has correlations persisting over all time scales. Since LRD can have significant impacts on the network queuing delay and packet loss rate, it is considered as an important traffic analysis factor [29]. LRD, *i.e.*, self-similarity, is measured by parameter H, the Hurst parameter, where $H \in (0.5, 1)$, which is explained in Section 3.3.1.1. Hurst parameter expresses the decaying speed of a time-series auto-correlation function, which decays very slowly for a time series with LRD characteristic.

Different types of message distribution over time are illustrated in Fig. 3.6.

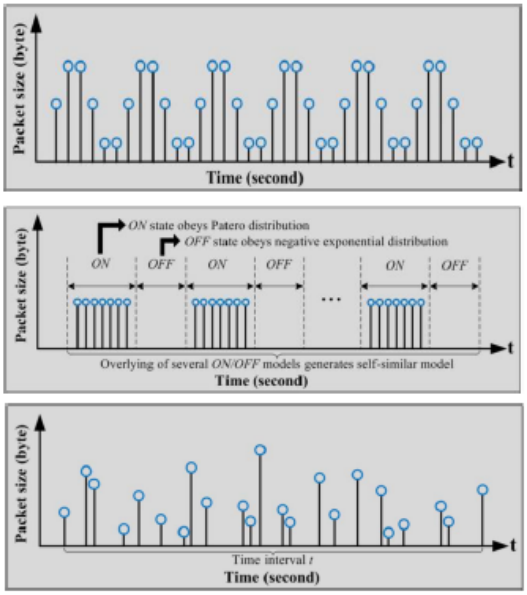


Figure 3.6 – IEC 61850 different data distributions over time [161].

3.2 Modeling methods and influence of delay factor

The spread use of ICT in the smart grid control, supervision and protection systems, while it is always prone to unexpected errors, represents a paramount important issue of the SCN traffic flow modeling. A unified model can describe the traffic flow behavior and the inter-dependencies of power and communication grids, which establishes a data base to enrich the existing control paradigm.

By now, there is a wide range of scientific and experimental researches on evaluating and co-simulating the smart grid. Different modeling methods have been proposed to investigate

the smart grid from different points of view. Among all the methods, in this section, three of the most known and widely used are introduced: queuing theory, analytical methods, time-series analysis.

3.2.1 Queuing theory

Queuing theory is the mathematical study of waiting lines, or queues. The first queuing theory problem was considered by Agner Krarup Erlang in 1908 [63], who looked at how long a telephone exchange needed to be in order to keep to a reasonable value the number of telephone calls not connected because the line is busy (lost calls). Within ten years, he had developed models to describe the Copenhagen telephone exchanges. Since then, his theory was widely used in telecommunications, traffic engineering, computing, and particularly in industrial engineering (in the design of factories, shops and hospitals, as well as project management).

Queuing theory deals with queuing problems using probabilistic methods [149]. Normally, a queuing system is considered as a group of lines (or queues) to which the customers arrive and wait to get the service and leave. It is characterized *via* the identification of probability properties of the input flow, service times and service disciplines.

The arrival process is characterized by the distribution of the inter-arrival times of input data, with the distribution function $A(t)$:

$$A(t) = P(inter - arrival_{time} < t),$$

in which the inter-arrival times are assumed to be independent and identically distributed random variables. Other random variable is the service time, and its distribution function $B(x)$ is:

$$B(x) = P(service_{time} < x)$$

Inter-arrival times and service times are supposed to be independent random variables.

According to the service structure and the service disciplines, the number of servers and the capacity of the system could be identified. The service discipline is the rule according to which the next customer is selected to be serviced. The common rules are:

- FIFO (First In First Out): the one who comes earliest leaves earliest,
- LIFO (Last In First Out): the one who comes latest, leaves earliest,
- RS (Random service): the service receiver is selected randomly,
- Priority: inputs are prioritized to get the service in order of the importance.

Use of the queuing theory is interesting because it offers the main performance probabilistic measures of the under-study system by means of the following variables: number of customers in the system, number of waiting customers, response time of a customer, waiting time of a customer, idle time of a server, busy time of a server.

In a communication network with a message transmission system, queues are the traffic flows, arrival customer is a packet of data, and service time is the average delay required to deliver a packet from origin to destination. One of the most important performance measures of a data network is the message transmission delay by which the choice and performance of network algorithms are strongly influenced, such as routing tables and flow control. For these reasons, it is paramount important to understand the nature and mechanism of delay, and the manner in which it depends on the characteristics of the network.

Queuing theory is a primary methodological framework for analyzing network delay. Its use often requires simplifying assumptions since, unfortunately, more realistic assumptions make meaningful analysis extremely difficult. For this reason, it is sometimes impossible to obtain accurate quantitative delay predictions on the basis of queuing models.

Queuing model

A queuing model is constructed in order to be able to predict the queue lengths and waiting time. Queuing systems are considered where customers arrive at random times to get service. In the context of the data network, customers represent packets assigned to the communication link for transmission. Service time corresponds to the packet transmission time and is equal to L/C, where L is the packet length in bits and C is the link transmission capacity in bits/s. In this context, the quantity estimation is of:
1. The customer arrival rate: average number of customers in the system.
2. The customer service rate: average delay per customer.

At a single queuing node, one could imagine that the packets arrive to the queue (*e.g.*, Ethernet switch), wait some time to be processed (receive service) and then depart from the queue. Fig. 3.7 shows a model of a single queue. A queue or queuing network may be studied in different ways: analysis and/or simulation.

Figure 3.7 – Model of a single queue.

The simulation is a method to simulate the system or environment in order to predict actual behavior of the system, while analysis is to decompose the system into components in order to study their interaction. Different studies have been done to propose a reliable delay estimation through the communication networks. A queuing model on a grouping network of Distributed Grid Intelligences (DGIs) has been used to predict the grouping variable communication delays [26].

Kendall's notation: There are different kinds of queuing systems, depending on how

many queues and servers exist, and how they are connected together. Various configurations of queuing systems are typically described with Kendall's notation, which gives a convenient shorthand for labeling classes of queuing systems. Different kinds of systems have very different queuing behavior, which will result in a drastically different waiting time and queue length. Kendall's notation makes it easy to be sure of what kind of system you are working with. The notation is a set of slashed-delimited letters and numbers that indicates the following:

- how arrival rates behave,
- how service time behaves,
- the number of servers,
- special parameters that are only needed in unusual cases and are usually left off. They relate to whether the system has finite capacity and might reject arrivals, how big the population of customers is, and whether the queue does anything special to select which customer gets service next, *e.g.*, order of priority.

So, a single queuing node is described by $A/S/C$, where A represents the arrival rate, S is the service time, and C is the number of servers through the queue [149]. As an example of a queuing system, $M/M/1$ describes a simple model with a single server (in the communication context, a single transmission line). First letter, M, stands for memory-less, which here means the arrival of the inputs according to the Poisson process (*i.e.*, exponentially distributed inter-arrival times), and the probability distribution of the service time is exponential too. Another example is $M/G/1$ queue, where G stands for general and indicates the arbitrary probability distribution of inter-arrival service times.

3.2.2 Analytical methods

An analytical technique (analytical method) is a procedure or a method for the analysis of some problem, status or fact. Analytical techniques are usually time-limited and task-limited. They enable researchers to examine complex relationships between variables. Generally, there are three types of analytical methods used in system modeling and identification:

- **regression analysis:** regression analysis assumes that the dependent, or outcome, variable is directly affected by one or more independent variables. Just to mention, there are four important types of regression analysis: Ordinary Least Square (OLS) regression [47], logistic regression [62], hierarchical linear modeling [11], duration models [22].
- **grouping methods:** grouping methods are techniques for classifying observations into meaningful categories. One grouping method, discriminant analysis [116], identifies characteristics that are distinctive for each group. The second grouping method, cluster analysis [61], is used to classify similar individuals together.
- **multiple equation models:** multiple equation modeling, which is an extension of regression, is used to examine the causal pathways from independent variables to the dependent variable. There are two main types of multiple equation models: path analysis [154], and structural equation modeling [124].

Analytical modeling

Analytical models are mathematical models that have a closed-form solution, *i.e.*, the solution to the equations used to describe changes in a system can be expressed as a mathematical analytic function. Analytical solutions to equations describing more complex systems can often become fairly complex. However, an analytical solution does provide a concise preview of a model behavior that is not as readily available with a numerical solution. Also implicit in the argument of an analytical model's superiority to numerical models is that graphing is tedious. One disadvantage of analytical solutions is that they are often very mathematically challenging to obtain.

In smart grids, where it is necessary to integrate the smart technologies into the existing power grids, understanding of both cyber and physical systems are a vital concern. Analytical models are applied differently to give mathematical description of this continuous-discrete integration. Some of the related works are mentioned here.

A conceptual hierarchical framework is proposed by [88] where symbol-string-based encoding, [103], made it possible to analyze complex cyber-physical systems. Authors try to address the robustness and safety by modeling smart grids as stochastic hybrid dynamical systems. This paper analyzes the cyber and physical interactions based on the analytical methods included in a stochastic hybrid model that evolves on continuous/discrete state/structure space based on Gibbsean statistical mechanics [57].

Analytical models are also applied for modeling the communication protocols over SCN. Such a model is proposed by [161], which describes the communication traffic flow based on a port connection model, and related flow source-service model. The model is designed to obtain a numerical analysis of the message transmission delay using an analytical calculation algorithm. As this model evaluates the message transmission speed, it can be used to optimize the network configuration.

3.2.3 Time-series analysis method

Time series is a sequence set of observations generated over time, and depending on the data set, it can be continuous or discrete. In this work, only discrete-time series are considered. Observations from a discrete-time series obtained at times $\tau_1, \tau_2, ..., \tau_t, ..., \tau_n$ can be denoted by $y(\tau_1), y(\tau_2), ..., y(\tau_t), ..., y(\tau_n)$, suppose that n observations on a variable over time are sampled by a fixed sampling period, from an infinite population of such samples. Let $y_1, y_2, ..., y_t, ..., y_n$ to denote observations made at equidistant time intervals $\tau_0 + h, \tau_0 + 2h, ..., \tau_0 + th, ..., \tau_0 + nh$. If we assign τ_0 as the origin and h as the sampling period, y_t represents the observation at time t. The main characteristic of time series is the stochastic dependency of adjacent observations. In order to describe this dependence, it is necessary to develop stochastic and dynamic models of time-series data [125].

Time-series analysis methods can detect and explore the existing linear relationship be-

tween the current data, historical data and the exogenous factors [123]. Time series are used vastly in scientific researches such as statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering that involves temporal measurements. Among various methods, Box-Jenkins and exponential smoothing methods are widely used to analyze the time series behavior.

Box-Jenkins method: This method searches to determine each value of time series as a function of the previous values $y_t = f(y_{t-1}, y_{t-2}, ...)$. To do so, Auto Regressive Moving Average (ARMA) models are proposed. This model was formalized first in a thesis research by Peter Whittle (1941) and further was popularized by Box and Jenkins (1970) [21]. The two main parts of ARMA model consist of Auto Regressive (AR) and Moving Average (MA) models.

AR: involves regressing the current value on its own past value:

$$y_t = c + \sum_{i=1}^{p} (\varphi_i y_{t-i}) + a_t, \tag{3.1}$$

where $\varphi_1, ..., \varphi_p$ are the parameters of the model, c is a constant, and a_t is white noise.

MA: involves modeling of the error samples as a linear combination of current occurring errors and the past errors committed by the model:

$$y_t = \mu + \sum_{i=1}^{q} (\theta_i a_{t-i}), \tag{3.2}$$

where μ is the mean of time series, $\theta_1, ..., \theta_q$ are the parameters of the model, and $a_t, a_{t-1}, ..., a_{t-q}$ are white noise samples.

Various structures are proposed by Box-Jenkins identification method such as: Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA), Auto Regressive Integrated Moving Average (ARIMA), Seasonal Auto Regressive Integrated Moving Average (SARIMA), Fractional Auto Regressive Integrated Moving Average (FARIMA), Auto Regressive Moving Average eXogenous inputs (ARMAX).

Recent researches consider modeling of the SCN traffic by the time-series-based mathematical models, which allows predicting the data traffic behavior [158], [21], [25], [152].

Exponential smoothing method: Exponential smoothing method considers data as a function of time $y = f(t)$. Models of this type are analyzed by the Fourier transform [36]. Exponential smoothing is usually employed to make short-term forecasts, as longer-term forecasts using this technique can be quite unreliable. While in simple moving average the past observations are weighted equally, exponential functions are used to assign decreasing weights over time, in which decreasing weights are assigned for the newest-to-oldest observations. In other words, the older the data is, the less priority ("weight") is given to it; newer data is seen

as more relevant and is assigned more weight. Smoothing parameters (smoothing constants) – usually denoted by α – determine the weights for observations. Considering the number of smoothing components, there are three most used models: simple, double, and triple.

Simple (single) exponential smoothing uses a weighted moving average with exponentially decreasing weights through which it is assumed the observed data varies around a stable mean without trend [79].

$$S_t = \alpha y_{t-1} + (1-\alpha) S_{t-1}, \tag{3.3}$$

where the smoothed value, S_t, is the weighted average of the current value y_t and the previous smoothed value S_{t-1}. t is the time period. Next, the forgetting parameter α is assigned a value $\in [0,1]$, which controls the decreasing of the exponential weights. The greater α is, the bigger effect has the current value on the smoothing term. Following this, the best value for α is the one that results in the smallest mean squared error (MSE). Various ways exist to do this, but a popular method is the Levenberg-Marquardt algorithm [107].

Double exponential smoothing is usually more reliable for handling data that show trends, compared to the single procedure. In addition, this is a more complicated method which adds a second equation to the procedure. Double exponential smoothing employs a level component and a trend component at each period. It uses two weights, or smoothing parameters, to update the components at each period. The double exponential smoothing equations are presented as below, through which the smoothing process is done on the "level" and "trend" parts [79]:

$$\begin{aligned} &\textbf{Trend} : b_t = \gamma(S_t - S_{t-1}) + (1-\gamma) b_{t-1} \\ &\textbf{Level} : S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1}), \end{aligned} \tag{3.4}$$

where the forgetting factors $\alpha, \gamma \in [0,1]$ respectively control the decreasing weights of level and trend components. γ, like α, can be chosen through the Levenberg-Marquardt algorithm.

Triple exponential smoothing is usually more reliable for parabolic trends or data that show trends and seasonality. In this case, there are trend and seasonal components through time series, so the model works on "level", "trend", and "seasonal" parts. In addition to the equations for single and double smoothing, a third equation is used to handle the seasonality aspect. Triple method is also known as Holt-Winters (HW) method [79]. There exists mainly two models of this kind: multiplicative seasonal and additive seasonal model.

The original data in data set is the product of the seasonal pattern and the average level in the multiplicative model assumption. In the additive model, seasonal pattern is independent of the data set average value and their sum is equal to the original data set.

The expression of the multiplicative model is as follows:

$$\begin{aligned} y_t &= (S_t + b_t t) l_t + a_t \\ \textbf{Trend} : b_t &= \gamma(S_t - S_{t-1}) + (1-\gamma) b_{t-1} \\ \textbf{Level} : S_t &= \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1}) \\ \textbf{Seasonal} : l_t &= \beta \frac{y_t}{S_t} + (1-\beta) l_{t-L}, \end{aligned} \tag{3.5}$$

where l is the seasonal index. $\alpha, \beta, \gamma \in [0, 1]$ are three forgetting parameters and L is the length of a period. β minimizes the MSE.

The additive model can be expressed as follows:

$$\begin{aligned} y_t &= S_t + b_t t + l_t + a_t \\ \textbf{Trend} : b_t &= \gamma(S_t - S_{t-1}) + (1-\gamma) b_{t-1} \\ \textbf{Level} : S_t &= \alpha(y_t - l_{t-L}) + (1-\alpha)(S_{t-1} + b_{t-1}) \\ \textbf{Seasonal} : l_t &= \beta(y_t - S_t) + (1-\beta) l_{t-L} \end{aligned} \tag{3.6}$$

The comparison of these two models shows that the additive model simply replaces seasonal multiple relationship by an additive relationship.

3.3 The choice of methods for traffic and delay estimation

In this section, it will be explained the modeling procedure employed throughout this work to describe the traffic behavior of SCN mathematically: the basics of the method application and an introduction to the modeling assumptions. As second phase of the research, one of the most important factors of the communication network evaluation, *i.e.*, message transmission delay, is also studied.

3.3.1 Time series: Auto Regressive Moving Average (ARMA) model

All the previously introduced methods have already been applied on the SCN traffic modeling over different types of messages, and for different analysis goals with varying degrees of success. Lack of a unique mathematical model regardless of the data type is the strong motivation of this thesis. The primary focus is on modeling the variation of number of data packets flowing through the communication channel over time.

The modeling process consists of a stochastic and a deterministic part. To model the SCN data flow, the number of packets passed in between the pair of communicating devices is considered as a time series, and the Box-Jenkins method has been applied based on ARMA(p,q) model to model the stochastic part, which is characterized by p, and q. p represents the AR order of a process, where p is a non-negative integer. q is a non-negative integer and represents the MA order of a process. ARMA models are widely used in identification for

control purpose. Finally, the total mathematical model describing the SCN traffic behavior is obtained by adding the deterministic part mathematical expression.

ARMA outputs a time series in response at a sequence of white Gaussian noise [21]. This is very pertinent because one looks at the system as a linear filter, likely to serve for control purposes, because all the linear analyzing methods can further be applied, as it is shown in Fig. 3.8.

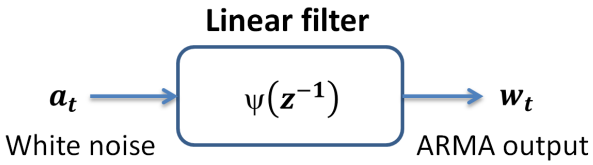


Figure 3.8 – ARIMA model as a linear filter.

To express mathematically, input a_t is multiplied by ARMA transfer function $\psi(z^{-1})$, and filter output w_t is the multiplication result, where z^{-1} is the one-sample delay operator: $z^{-1}a_t = a_{t-1}$. ARMA transfer function of a discrete-time filter is expressed as below:

$$\psi(z^{-1}) = \frac{\theta(z^{-1})}{\varphi(z^{-1})}, \tag{3.7}$$

thus w_t is obtained by:

$$w_t = \psi(z^{-1})a_t = \frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t,$$

which can be rewritten such as the following equation:

$$\varphi(z^{-1})w_t = \theta(z^{-1})a_t,$$

where w_t is the data sampled periodically and preprocessed if necessary, a_t is the white noise sequence. $\varphi(z^{-1})$ and $\theta(z^{-1})$ are polynomial functions of the delay operator z^{-1}, as below:

$$\varphi(z^{-1}) = 1 + \varphi_1 z^{-1} + \varphi_2 z^{-2} + ... + \varphi_p z^{-p},$$

$$\theta(z^{-1}) = 1 - \theta_1 z^{-1} - \theta_2 z^{-2} - ... - \theta_q z^{-q}, \tag{3.8}$$

where φ_i and θ_j are the process and noise coefficients, respectively.

Finally ARMA(p,q) polynomial expression is written as below:

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + ... + \varphi_p w_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}, \tag{3.9}$$

3.3.1.1 ARMA model versions (SARIMA, FARIMA, ARIMA, *etc.*)

According to the Box-Jenkins methodology, other versions of ARMA-based identification models are possible:

ARIMA(*p*,*d*,*q*) is built up of AR and MA terms and an extra term I which represents "Integrated". The nonseasonal order of differentiating, d, models the degree of linear nonstationarity (trending behavior) in the data. The integrated part of the model includes the differencing operator $(1-z^{-1})^d$, where d indicates how many times a differencing is needed to remove the existing linear trend (upward, downward) through data. For example, if there is a linear trend through the sampled series y_t, to remove it $(1-z^{-1})^{d=1}y_t$ is applied, which gives the simple difference of each value and its previous one: $y_t - y_{t-1}$. But if the difference needs to be applied more than once (*e.g.*, $d=2$), the difference is raised to the power d as: $(1-z^{-1})^{d=2}y_t = y_t - 2y_{t-1} + y_{t-2}$.

SARIMA$(p,d,q)(P,D,Q)_s$ is the seasonal ARIMA model. Uncapitalized letters signifies the regular nonseasonal autoregressive, integration, and moving average orders of the model. Capitalized letters represent respectively the seasonal components of the model with the seasonality order of s. As it will be explained in Chapter 4, Section 4.1, through a data preparation process, all the deterministic behavior through a time series should be removed before ARMA modeling. Thus, the seasonal nonstationarity has to be removed by a differencing process: multiplicative or additive seasonal differencing of the seasonal order s. A difference at a seasonal lag is noted as $(1-z^{-s})$, where z^{-1} is the delay operator, which results in $(1-z^{-s})y_t = y_t - y_{t-s}$. The simple application of seasonal difference obtains the additive formulation, which is simply ARMA(*p*,*d*,*q*). But if there is a multiplication of the nonseasonal and seasonal differentiating factors as $(1-z^{-d})(1-z^{-s})$, it is called Seasonal ARIMA, *i.e.*, SARIMA [156].

The multiplication of the regular and seasonal parts of SARIMA is applied on their expressions in terms of their lag operators. To clarify the procedure, a numerical example is presented below:

$$\begin{aligned}(1-0.4z^{-1})(1-0.3z^{-12})y_t &= (1-0.4z^{-1}-0.3z^{-12}+0.12z^{-13})y_t \\ &= y_t - 0.4y_{t-1} - 0.3y_{t-12} + 0.12y_{t-13},\end{aligned}$$

where there is a positive differencing interaction at lag 13. While all the other lags have negative spikes, lag 13 has a spike in the opposite direction. If the time series is not a long series, this interaction may have not a significant effect, hence it may be ignored through the identification process: $(1-0.4z^{-1}-0.12z^{-12})$ containing only the first- and 12^{th}-order regular differencing. The only difference between additive and multiplicative model is to include these interactions or not, and these interactions with opposite direction may complicate the modeling compared to the simple additive model [156], [155].

FARIMA, in which F describes the *Fractional difference* to eliminate the Long-Range Dependence (LRD) resulting from the self-similarity through data while burst traffic is present [114]. From practical modeling point of view, such processes may exhibit certain features that

could give the impression of the need for differentiating to achieve stationarity. To evaluate the self-similarity and LRD through the data, the Hurst parameter (H) is an evaluation criterion. H can be calculated by three different methods, namely R/S analysis, variance-time analysis, and periodogram analysis. If H is such that $0.5 < H < 1$, then some LRD characteristic through the traffic may be identified. Closer to 1, stronger LRD characteristic exists.

To obtain time series with only SRD characteristic, the fractional difference is applied as many times as needed to get H closer to 0.5 [76]. Nevertheless, note that the presence of H greater than 0.5 is not necessarily indicative of long-term dependence, then only the fractional difference is not a sufficient reason to be sure of the LRD characteristic. So, models of different orders (with and without fractional difference application) may be fitted to the data and if the best model fits well for not fractioned series, the conclusion is that the time series contains no long-term dependence.

In the fractional model, the power d of differencing $(1 - z^{-1})^d$ is fractional and applied to the sampled series y_t in the form of Taylor series as below:

$$\begin{aligned}(1 - z^{-1})^d &= \sum_{k=0}^{\inf} \binom{d}{k} (-z^{-1})^k = \sum_{k=0}^{\inf} \frac{\prod_{a=0}^{k-1}(d-a)(-z^{-1})^k}{k!} \\ &= 1 - dz^{-1} + \frac{d(d-1)}{2!} z^{-2} - ... \end{aligned} \tag{3.10}$$

ARMA-based identification method is widely known and used for control purpose, as it offers a sufficiently reliable model having all the advantages of the linearity. Thus, among all the possible structure models listed in 3.2.3, ARIMA is selected because:
1. the mathematical model is represented by an interconnection of linear filters,
2. the resulted model has usually sufficiently precise fitness,
3. finally, it allows simple implementation by basic MATLAB® functions.

It is good to be noted, through the experiments performed by this thesis, that FARIMA model is not selected, since any of sampled time series does not show LRD characteristics, and SARIMA is not selected in order to prevent more complications by the multiplicative seasonal difference.

3.3.2 An analytical model for transmission delay estimation

A custom-designed and well monitored SCN can maintain the fast and reliable information transmission and lead to improved operation and management of a SAS. In order to achieve this goal, it is important to develop reliable SCN models and evaluate network performance under different network conditions, and analyze network performance requirements of SCN. As one of the most important network performance evaluation, which could result in an improper information delivery and probable instability in the smart grid, the second part of this work is particularly dedicated to the maximum message transmission delay in SCN.

By delay estimation, while designing SCN, its performance could be evaluated based on the maximum message delay and the traffic load distribution under different network conditions and equipment selections, so a network designing plan could be developed according to the evaluation results. As an example, the crowded switches could be detected through the traffic load distribution, suggesting alternative configurations that can increase the reliability of the information transmission. In addition, it is possible to do a sensitivity analysis for the message delay, which can help to find out the main parameters for an optimized network performance. Delay means here the time in which a message is transmitted from its source to destination, passing through the Ethernet switches as service units.

Inside of a switch, four types of delay can be identified: packet receiving delay, processing delay, queuing delay, and transmission delay. Packet receiving and processing delays can be considered constant typically about 3 μs for each switch [70]. On the contrary, the queuing delay and transmission delay are variable and need to be estimated.

In order to achieve this, a traffic flow model is required to represent a traffic flow source model and a service model of SCN to be used in a traffic flow calculation algorithm in order to obtain the distribution of traffic load and estimation of the maximum message delay. An ARIMA-based model, proposed to represent the behavior of sources of messages in a communication grid, can be integrated to such an algorithm.

3.3.2.1 *Network calculus theorem*

The basic concepts of the network calculus theorem are *arrival curves* and *service curves* [33]. A source model is built to describe the properties of traffic flow sources, and a service model describes the service that traffic flow receives from switches. Based on the network calculus theorem, a data traffic flow is described by means of a cumulative function $F(t)$, representing the number of bits observed on the traffic flow in the time interval $[0, t]$.

An arrival curve is a model which defines constrains on the traffic flow arriving at the service unit, *e.g.*, Ethernet switch. Basically, arrival curve is used to define a guaranty to the data flow. As a flow needs some specific support in the network, *i.e.*, service curve, thus the flow sent by a source needs to be limited, *i.e.*, source curve. Traffic flow is upper bounded by an increasing function $\alpha(t)$, named arrival curve, if and only if for all $s \leqslant t$ [96]:

$$F(t) - F(s) \leqslant \alpha(t - s), \tag{3.11}$$

so $F(t)$ has $\alpha(t)$ as the arrival curve or $F(t)$ is α-smoothed. Required condition is true for a set of overlapping time intervals. One of the most important and frequently used arrival curve model is (σ,ρ) model proposed by Cruz [33] meanwhile being a simple linear model:

$$\alpha(t) = \rho t + \sigma, \tag{3.12}$$

with σ in this model representing the burstiness parameter, and ρ being an upper bound on the long-term average rate of the traffic flow.

Then traffic flow $F(t)$ is (σ, ρ) upper bounded if and only if for all $s \leqslant t$ [96]:

$$F(t) - F(s) \leqslant \rho(t - s) + \sigma \tag{3.13}$$

In order to identify related parameters of $\alpha(t)$ for each flow, a simple estimation step is applied [96]. First $\alpha(t) = \rho t$ is estimated over the measured flow, where ρ is the slope of corresponding trend, $y(t) = rt + b$, estimated for the flow. ρ means that on any time window of width τ, the number of bits for the flow is limited by $\rho\tau$, *e.g.*, $\tau = 1$ s and ρ is the in bits/s. Thus, in this case, the flow is peak-rate limited. Then, supposing an arrival curve of the form of $\alpha(t) = \sigma$, where σ is a constant value, it means that the maximum number of bits which may ever be sent on the considered flow is σ. In this thesis, σ is estimated by the maximum difference between the flow corresponding trend, $y(t)$, and the flow, itself, $F(t)$.

$$\sigma = max|F(t) - y(t)|, \tag{3.14}$$

where σ represents the maximum length of the corresponding flow. Generally, considering $\alpha(t) = \rho t + \sigma$ as the arrival curve allows a source to send σ bits at once, but not more than ρ bit/s over a long run. ρ is the burstiness parameter (in bits), and σ is the rate (in bits/s). The estimated parameter over a typical measured traffic flow is illustrated in Fig. 3.9.

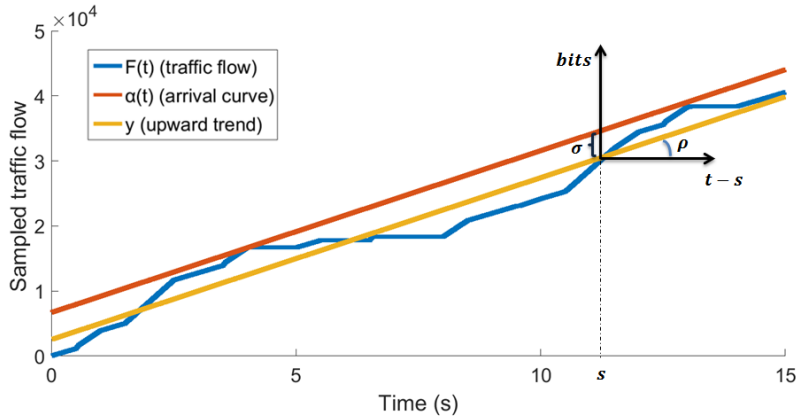


Figure 3.9 – An example of a cumulative function $F(t)$ related to a typical flow upper bounded by the arrival curve $\alpha(t)$.

A service curve is proposed to abstract the details of packet scheduling mechanisms and define lower bounds on the service applied by service units such as switches, routers and so on. The service curve model is described as a function of time that specifies the service offered by those service systems during a defined time interval. A widely used service curve type is the called rate-latency type $\beta_{R,T}$. $\beta_{R,T}$ can be represented using the linear function $\beta(t) = R[t - T]^+$, with R as the rate term and T as the latency term, where $[x]^+$ denotes $max\{x, 0\}$, [96]. If a flow receives the service of $\beta_{R,T}$ as its service curve, this means that it receives the service of rate R in T time after its arrival at the service system.

Min-plus algebra: Network calculus is established on min-plus algebra, in which it characterizes the amount of traffic and available service as functions of time [100]. Min-plus

algebra changes the most common conventional math operators on the set of integers, $\mathbb{Z}$, or reals, $\mathbb{R}$, as follows: addition to the minimum (min) computation, and multiplication to addition (plus). Min-plus algebra defines another algebraic structure, which uses the concepts of *minimum*, *infimum*, *maximum*, and *supremum* [96].

To recall, suppose S is a nonempty subset of $\mathbb{R}$. S is bounded from below if there is a value as M such that for all $s \in S$, $s \geqslant M$. Minimum is the element in a nonempty subset S of $\mathbb{R}$, which is smaller than all its other elements. In addition, there is an infimum value for every nonempty subset S, which is the greatest lower bound denoted by $inf\{S\}$. The difference between minimum and infimum can be explained by closed and open intervals like $[a, b]$ and $]a, b[$. The smallest value as the greatest lower bound for both subsets is a, which is the infimum. The minimum point does not always exist, as for the open interval, because $a \notin]a, b[$. On the other hand, if a minimum exists, it is identical to its infimum (*e.g.*, $min[a, b] = inf[a, b] = a$). Similar definition is considered for the difference between the biggest element of a nonempty subset S of $\mathbb{R}$, *i.e.*, maximum (max), and the least upper bound, *i.e.*, supremum (sup), of a subset [96].

The maximum message transmission delay is estimated, such that predefined constraints of network calculus are satisfied. As related constraints are based on the min-plus algebra, it is important to know about min-plus convolution and its dual min-plus deconvolution functions to continue the estimation procedure. Consider f and g to be two functions, then their min-plus convolution is defined as follows:

$$(f \otimes g)(t) = inf_{0 \leqslant s \leqslant t}\{f(t-s) + g(s)\}, \tag{3.15}$$

so if $t < 0$, $(f \otimes g)(t) = 0$. Min-plus deconvolution considering the same functions f and g is defined as below:

$$(f \oslash g)(t) = sup_{s \geqslant 0}\{f(t+s) - g(s)\}, \tag{3.16}$$

Based on min-plus algebra, the output flow bounds, α^*, of a service system can be computed from the min-plus deconvolution of the arrival curves and the systems service curve by replacing them in the (3.16):

$$\alpha^* = \alpha \oslash \beta = sup\{\alpha(t+s) - \beta(s)\}, \forall s \geqslant 0, \tag{3.17}$$

and the delay $D(t)$ of the flow at time t is upper bounded by:

$$D(t) \leqslant h(\alpha, \beta) = sup\{inf\{\tau \geqslant 0 : \alpha(s) \leqslant \beta(s+\tau)\}\}, \forall s \geqslant 0, \tag{3.18}$$

where $sup\{S\}$ and $inf\{S\}$ means the least upper bound and the greatest lower bound of a subset S, and α and β are the arrival curve and the service curve, respectively [96]. In particular, if α has the (σ, ρ) type and β has the $\beta_{R,T}$ type, then α^* also has the (σ^*, ρ^*) type with $\rho^* = \rho$, and $\sigma^* = \sigma + \rho T$ [96]:

$$D(t) \leqslant h(\alpha, \beta) = T + \sigma/R, \tag{3.19}$$

where the maximum message transmission delay passing through each port is obtained by the characteristics of its related arrival curve, *i.e.*, σ, and the received service from the port, *i.e.*, T and R[1].

Following the calculus theorem principles, source model and a service model are proposed in the form of matrices to illustrate the message distribution from its source passing through various switch ports (one or more switches on the path) to the destination. Then, according to the passage of each flow toward different ports, delay time is calculated for each port, and finally the total maximum message transmission delay is estimated according to (3.19).

3.4 Conclusion

Smart grid provides energy to the customers using communication protocol, *e.g.*, IEC 61850. As the possible errors through the message transmission can cause severe problems such as instability or voltage drop out, the interactions between communication and energy grid should be analyzed, which ideally requires a uniform descriptive model for all type of messages passing through SCN.

Two main objectives can thus be defined: SCN traffic modeling and maximum transmission delay estimation, respectively.

Concerning the SCN traffic modeling, a data-driven approach of modeling data traffic behavior through the communication networks within smart grids, will be investigated. A time-series-based identification procedure will be employed to this end. Evolution of the number of transmitted messages is modeled as a time series, whose parameters can be obtained by Box-Jenkins identification methods. Among all the possible model structures, ARIMA was selected, in which data is modeled as a linear filter. The detailed modeling process will be further explained in Chapter 4.

The case of a distribution grid with two renewable energy sources which exchange reactive power information for voltage regulation purposes by means of IEC 61850 GOOSE protocol will further be simulated, and coded in MATLAB®/Simulink ®. Next, a sensitivity study will be conducted to assess how the parameters of identified ARIMA models vary in relation to variations of some communication network parameters. Eventually, the identification method is applied over a real-time test-bench. In the considered scenario, it is supposed that the load active power variation is known, then an IED sends this information to all other IEDs. This scenario will be implemented on a test bench offering the possibility of real-time data packets analysis.

Concerning the maximum transmission delay estimation, one of the most important performance of the communication network will be analyzed, namely the maximum message transmission delay from its generation source to its destination. As there is a need to have

[1]The related proofs are out of the scope of this research, and they are accessible in the provided reference [96]

a message-source descriptive model, an analytical model based on the queuing theory will be used in this case. To do so, the message distribution model of each traffic flow through the Ethernet switches is obtained, then the related service model is applied to the transmission process and the waiting time at each switch port is estimated. By using these results, the total maximum delay of a message is estimated. The next chapter will explain how to build the corresponding models and through which calculation algorithm the maximum message delay is estimated.

The case studies considered in this work and implemented on a real-time traffic generator test bench are further explained in Chapter 5. For the first scenario, in order to illustrate load shedding method, a number of IEDs are communicating the energy grid critical measured values to the supervisory station PC. One of the IEDs is responsible for applying the load shedding decision as soon as it will receive the corresponding order from the control center. In the second scenario, the network is perturbed by including other traffic generators not involved in the load shedding decision, only to crowd the communication channel. The estimated delay is validated against the delay measured by a network analyzer, as discussed in Chapter 5.

Chapter 4

Traffic modeling with ARIMA and application of delay estimation method

Contents

4.1 Procedure to obtain an ARIMA model 74
4.1.1 Stochastic nature of the communication traffic flow 74
4.1.2 Correlation between observations 75
4.1.3 Data sampling 76
4.1.4 Data processing 76
4.1.5 Model identification 79
4.1.6 Parameter estimation 80
4.1.7 Model validation 81
4.1.8 Forecasting 82
4.1.9 Mathematical model 83
4.2 Maximum message transmission delay 85
4.2.1 Port connection model 86
4.2.2 Network calculus basic concepts 88
4.2.3 Message distribution algorithm 90
4.2.4 Delay estimation 95
4.3 Conclusion 97

This chapter presents ARIMA modeling method selected after a thorough inquiry over different SCN traffic modeling methods, and the analytical modeling method for maximum delay estimation over SCN. As presented in the previous chapter, we have chosen ARIMA as time-series analysis model based on Box-Jenkins identification method.

In the first section of this chapter, we will detail the modeling process that is applied in two main parts: stochastic analysis and deterministic analysis. After a proper sampling of the SCN traffic, required stationarity of the stochastic part obtained after removing the deterministic behavior is provided through a series of data preprocessing steps to be further able to model its stochastic behavior by applying ARIMA model. Illustrative flowchart of

applying the data-driven stochastic ARIMA model is presented to conclude all the modeling process. Finally, deterministic part of data behavior is added back to complete the SCN traffic model, through a global mathematical expression.

The second section presents the detailed algorithm of maximum delay estimation mentioned in Chapter 3. Here, communication network performance is evaluated by maximum message transmission delay, which has been estimated using an analytical model based on the queuing theory. The model proposed by [161] represents a source model to describe the distribution of a message from its source to its destination. Next, the estimation of maximum delay can be made based on two notions defined in calculus theorem, *i.e.*, *arrival curve* and *service curve*.

4.1 Procedure to obtain an ARIMA model

To model the SCN traffic flow, ***cumulative number of packets*** passing between pairs of IEDs are considered as our sampled time series. In this work, power grid is assumed to be stable, so no fault occurs. Hence, all devices are operating in a normal condition of monitoring, control and protection, which is considered as the nominal case.

As previously explained in Chapter 3, SCN data traffic generated by a stable power grid without fault is mostly periodic. Our proposed model will be thus applied over the periodic GOOSE traffic flow, and its results will be presented and discussed through Chapter 5. Next sections will explain the modeling process in detail.

4.1.1 Stochastic nature of the communication traffic flow

Sometimes, it is possible to describe exactly a physical phenomenon based on physical laws and using mathematical functions, which leads us to calculate the exact value of some time-dependent variables at any time instant. In this case, it is about deterministic phenomena. However, no phenomenon is totally deterministic, and random events can disturb the process. To describe this randomness, a model is required, allowing to calculate the probability distribution of the related values between two specified limits, namely a stochastic model or probability model [21].

SCN traffic shows randomness in behavior, so this is not a completely deterministic phenomenon. To model SCN traffic considering both its stochastic and deterministic behaviors, data traffic should be preprocessed. These two parts are detached, and modeled separately. ARIMA model, as a linear filter, is selected in this work to describe data stochastic behavior, and a combination of some linear filters is used to describe deterministic behavior.

Box-Jenkins method is applied as a first preprocessing step to reach an acceptable stationary level, then ARIMA is fitted through identification and parameter estimation methods. Finally, ARIMA model is evaluated by some validation criterion.

Box-Jenkins methodology is implemented within five steps as follows [125]: data preparation, model identification, estimation of the model parameters, diagnostic verification of model, possibly forecasting.

4.1.2 Correlation between observations

ARIMA analysis supposes that time-series observations are statistically dependent, so it is possible to model the relationship of an observation over its previous ones. One possibility is a graphical analysis by looking at the time-series data drawn over time, with the hope to see any pattern. But, it does not always work very well. Some data series show very clear some obvious patterns to the eye, but many do not. Even if a series shows an obvious visual pattern, estimating its exact nature by simply looking at a plot of the data would be difficult.

Hence, using the statistical concept of correlation makes it possible to judge if a data series is self-correlated or not. There are two analytical tools that can measure the statistical relationship between observations in a series, namely the Auto-Correlation Function (ACF) and the Partial Auto-Correlation Function (PACF) [21]. ACF and PACF of a data series are most useful when presented in their graphical forms. The idea of ACF and PACF is that observations from different time periods may be related to each other.

ACF: The idea of auto-correlation analysis is to calculate a correlation coefficient for each set of ordered pairs (y_t, y_{t+k}), in which y_t occurs k sampling times earlier than y_{t+k} [25]. ACF finds correlation between sets of numbers that are parts of the same series, so resulting statistic is called an auto-correlation coefficient ("auto" means "self"). After calculating auto-correlation coefficients, they are plotted graphically against sampled numbers, *i.e.*, lags, a diagram that looks something like a histogram.

PACF: It is broadly similar to ACF, and it is also a graphical representation of the statistical relationship between sets of ordered pairs (y_t, y_{t+k}) drawn from a single time series. PACF is used as a guide, along with ACF, in choosing one or more ARIMA models that might fit the available data. The idea of PACF analysis is that we want to measure how y_t and y_{t+k} are related, but with the effects of the interposing y_ts in between. For example, we want to show the relationship between the ordered pairs (y_t, y_{t+3}), but effects of both y_{t+1} and y_{t+2} on y_{t+3} are accounted for, and so forth, each time adjusting for the impact of any y_t that falls between the ordered pairs in question [25]. As PACF coefficients are calculated, they are plotted as PACF diagram.

In this work, ACF and PACF are used to justify proper sampling time, and at the identification stage, they are helpful in giving us a feel for patterns in the available data. We use ACF and PACF as guides to choose one or more ARIMA models that seem appropriate. The basic idea is: every ARIMA has a theoretical ACF and PACF associated with it. Whichever model we choose at the identification stage, we consider it only tentatively, *i.e.*, it is only a candidate for the final model. To choose a final model, we proceed to the estimation and model validation stages to see if the model was well fitted. In this stage, we will study ACF/PACF diagrams to find out the initial values for orders of ARIMA polynomials (3.7):

p and q. This will give guess values of how many previous observations in our regression part and how many previous samples of the exogenous terms should be considered in the model.

ACF and PACF coefficients for a stationary time series follow a decreasing pattern, and both are within {-1,1}. PACF of the AR(p) process is near zero from the p^{th} lagged value. ACF of the MA(q) process is near zero from the q^{th} lagged value.

4.1.3 Data sampling

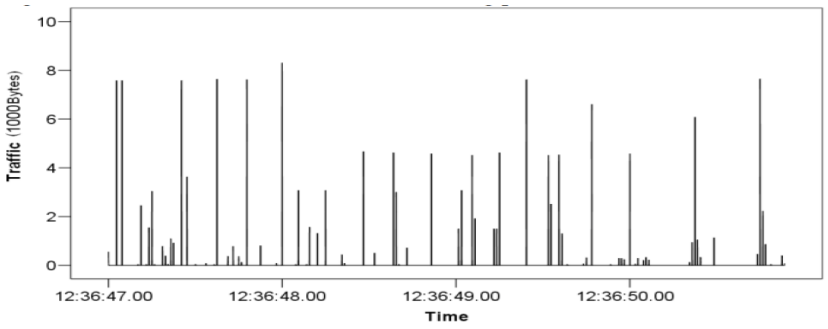
First of all, the minimum number of sampling points is important to have a well representative data set. In general, building an ARIMA model requires an adequate sample time. Box and Jenkins [21] suggest that about 50 observations is the minimum required number. In case of seasonal data, it is especially desirable to have a large number of sampling points. Next, it is important to sample data efficiently in a way to have a time series representing the data changes. A discrete time series can be obtained in two ways: by sampling a continuous time series, or by accumulating a variable over a period of time. In this work, the cumulative number of arrival packets is measured per sampling period at the receiving point of the communication link.

As this research proposes a model of SCN GOOSE traffic, in order to follow the traffic changes as much as possible, GOOSE transmission protocol should be considered to select a proper sampling period. GOOSE traffic has no handshaking topology as TCP/IP protocol, so to ensure the arrival of packets, messages are transmitted and retransmitted four times at each data change. A reasonable choice is to set the sampling time as being less than a half of the GOOSE minimum transmission period according to the Shannon's theorem [68]. In this way, sampled data follow properly – not so fast, not so slowly – the GOOSE transmission behavior.

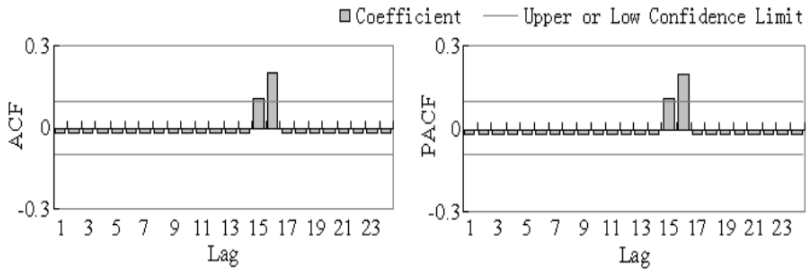
In addition to the above, looking at ACF/PACF diagrams can help to guide our choice. Visualizing these measurements gives an idea whether there is correlation through sampled data. Improper sampling period results in no correlation between observations, so ACF and PACF coefficients are near zero. As an example, a time series is considered in Fig. 4.1 (a), which is improperly sampled and obviously, the ACF/PACF coefficients are mostly near zero illustrated in Fig. 4.1 (b). So, that means there is no correlation between data to be modeled by ARIMA.

4.1.4 Data processing

ARIMA model should be applied on a special class of time series which is stationary. Stationarity supposes to be in a specific form of statistical equilibrium, and in particular, vary over time in a stable manner with constant properties (constant mean, constant variance) [21]. Since many real-world time series have nonstationary characteristics, to reach a stationary time series, any deterministic trend should be observed and removed by corresponding tools.



(a) Improper sampled time series



(b) ACF and PACF diagram of the improper sampling period selection

Figure 4.1 – A typical time series with an improper choice of the sampling period [158].

Basically, to find these trends, time-series data is plotted over time to see if any pattern exists. But as explained before, ACF and PACF graphical tools may be used at the identification stage to find more precisely different deterministic patterns over data; then these patterns are removed using related tools. Next sections include all the preprocessing steps needed to reach an acceptable stationarity level.

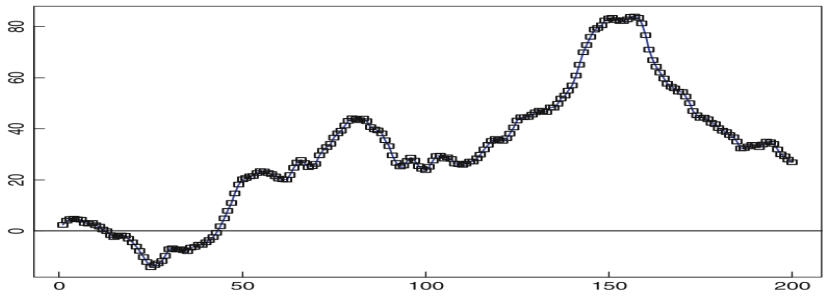
4.1.4.1 Detrending

Sometimes, there is a linear time trend which increases or decreases through time. This pattern can be found out by simply looking on the plotted time series over time. Fig. 4.2 (a) gives a simple example of a time series with an upward trend. In the case of any linear trend through data, ACF coefficients decay slowly. Related ACF diagram of time series in Fig. 4.2 (a) is presented in Fig. 4.2 (b).

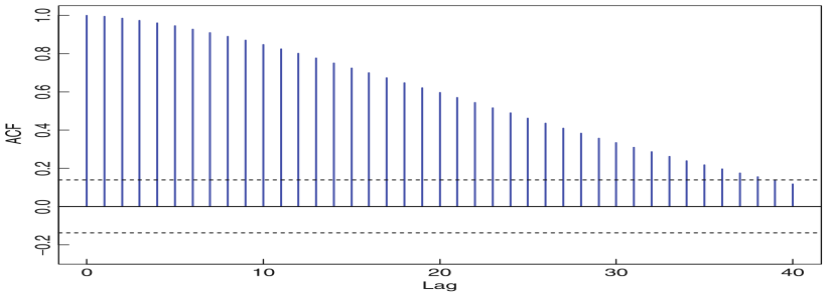
Since any dependence of time should not be present anymore, linear trend can be removed by a simple difference method. Thus, a new time series is obtained by differencing each value from its previous value, which has more stable variation around its mean. By noting y_t the original, directly measured, time series, then the simple difference is:

$$y'_t = (1 - z^{-1})^d y_t, \tag{4.1}$$

where z^{-1} is the one-sample delay operator, and d is the difference order which means how



(a) A typical time series including an upward trend through time



(b) ACF diagram of the time series with an upward trend

Figure 4.2 – A time series which shows an upward trend, and related ACF diagram [25].

many times is necessary to apply a simple difference to remove the linear trend over a time series. If $d = 1$, $y'_t = (1-z^{-1})^{d=1} y_t = y_t - y_{t-1}$, and for $d > 1$, the simple difference is repeated d times. Simple difference gives the cumulative number of packets being transmitted during each sampling period.

4.1.4.2 Zero average

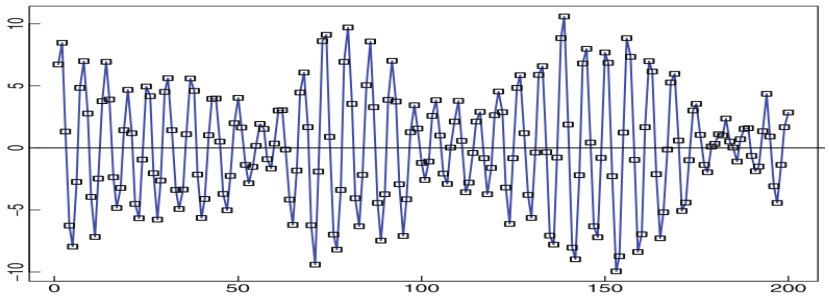
Mean value of time series is removed to have data variations around zero. Then:

$$\bar{y'}_t = y'_t - mean(y'_t), \tag{4.2}$$

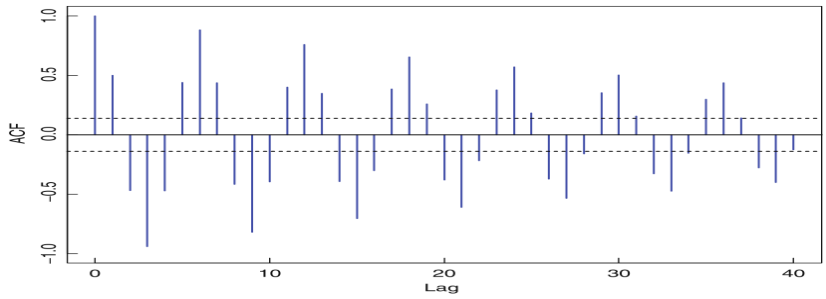
where $\bar{y'}_t$ is the zero-mean series.

4.1.4.3 Deseasonalizing

A seasonal pattern may exist through time series. These fluctuations are usually due to the measuring period and the behavior of data which is transmitted through the considered traffic. As an example, Fig. 4.3 (a) shows a seasonal time series and Fig. 4.3 (b) shows its related ACF. It can be clearly seen that the ACF diagram contains the periodic fluctuations. For series with seasonal periodicity, as for linear trend, it is necessary to remove this periodic trend



(a) A typical time series with a seasonal behavior



(b) ACF diagram of the seasonal time series

Figure 4.3 – A seasonal time series and related ACF diagram [25].

over time. Like simple difference, to smooth the seasonal behavior, a seasonal difference may be used. In a seasonal process, mean value for each season will vary from the others, which leads to a nonstationary series. Considering the sampling process, seasonal period, s, can be found out in seconds which equals to the number of samples in one period. To seasonally differentiate, we subtract from each observation the observation occurring s samples (one cycle) earlier:

$$w_t = \Delta_s^D \bar{y'}_t = (1 - z^{-s})^D \bar{y'}_t, \tag{4.3}$$

where w_t is the preprocessed series (stationary, detrended, deseasonalized), and D is the number of times that seasonal difference is applied, s is the seasonal period. If $D = 1$: $w_t = \Delta_s^1 \bar{y'}_t = \bar{y'}_t - \bar{y'}_{t-s}$. By applying the first seasonal difference, if the resulted time series still contains cyclic components, another seasonal difference should be applied, so that $D = 2$, and so forth. Generally, one seasonal difference will remove the seasonality efficiently.

4.1.5 Model identification

As previously mentioned, through the identification stage, both ACF and PACF graphical tools may help to find out the proper type of ARIMA model by measuring the correlation through the observations within series itself. The ACF and PACF plots can provide an indication about the best p and q. Fig. 4.4 shows an example of ACF/PACF diagrams as a guide to start with ARMA(1,2) that is, $p = 1$ and $q = 2$. In order to find the proper model

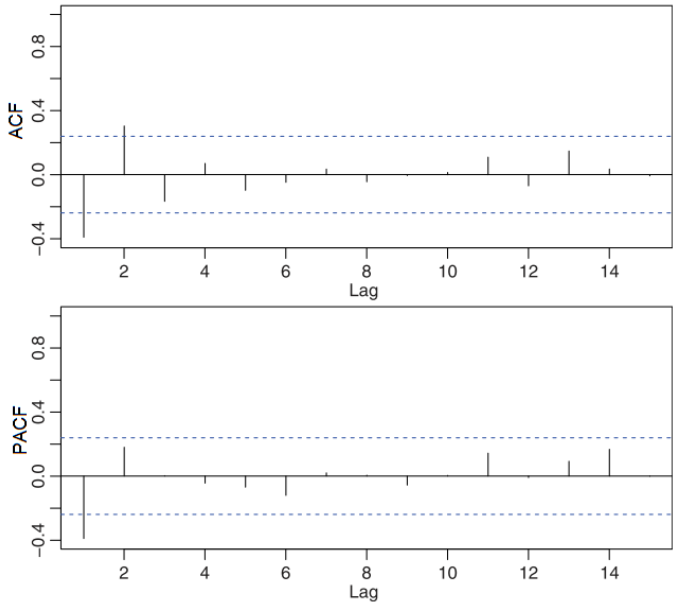


Figure 4.4 – Example of ACF and PACF diagram of an ARMA(1,2) [21].

orders, p and q, here, the Mean Squared Error (MSE) value has been considered. Almost always, increasing the orders of the model improves the goodness of fit, hence a penalty is required to avoid overfitting.

Here, the best orders are selected for which the estimated model results in MSE minimum value. So, starting with the ACF/PACF indicative initial parameters, we will have a set of choices. To reach the best estimation, their related MSE values are compared and the model that gives the minimum value is the best candidate.

4.1.6 Parameter estimation

Box-Jenkins method suggests some diagnostic checks to help determining if an estimated model is statistically adequate. A model that fails these diagnostic tests is rejected. Furthermore, the results at this stage may also indicate how a model could be improved. This leads us back to the identification stage. The cycle of identification, estimation, and diagnostic checks is repeated until a good final model is found out.

The ARIMA model coefficients are estimated and model is validated through the diagnostics concerning MSE and Goodness of Fit (GFit). Fitting a proper model results in a sequence of uncorrelated residuals. The residual ACF is plotted and if it shows significant auto-correlation in the residuals, the model should be modified to include larger auto regressive or moving average orders. A satisfactory model may be used to forecast the future behavior of series [156].

4.1.7 Model validation

Three criteria introduced for the model validation – minimum MSE, GFit, and residual ACF – are explained in the following.

Minimum MSE. Mean Squared Error (MSE) measures the average of the squares of the errors, which is the average squared difference between the estimated values by ARIMA model and the observed values. MSE is a measure of the model estimation quality, and is always non-negative. As it is closer to zero, the model is better estimated. MSE is calculated for different pairs of ARIMA orders, *i.e.*, p and q, then among all the estimations, an ARIMA(p,d,q) model which gives the minimum MSE is selected.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(w_{obs,t} - w_{model,t})^2, \tag{4.4}$$

where $w_{obs,t}$ is the observed, and preprocessed time-series value at time t, and $w_{model,t}$ is the modeled value at time t.

In the literature, there exists another criterion that provides parsimony for the number of orders, called Akaike's Information Criterion (AIC). In an estimation process, some part of information is lost. One would then choose a model among a group of candidates, which minimizes the information loss. As for minimum MSE, AIC is estimated for different pairs of p and q. By comparing their related AIC, the most accurate model is the one with minimum AIC [84]. To compare different model validation methods, AIC is also tested for different estimated ARIMA(p,d,q) models. The minimum AIC gives the most proper p and q. This thesis uses min MSE as model validation, but the obtained orders are compared to the ones selected by min AIC in the next chapter.

GFit. After fitting model(s) over a time series, the goodness of fit should be evaluated. Curve Fitting Toolbox™ of MATLAB® supports different GFit statistics among which the Root Mean Squared Error (RMSE) method is used in this thesis. RMSE method is also known as fit standard error, and with respect to the estimated value w_{model}, is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(w_{obs,t} - w_{model,t})^2}{n}} \tag{4.5}$$

In order to be free of the units, non-dimensional form of RMSE, *i.e.*, Normalized RMSE (NRMSE), is useful. There are two approaches: normalize the RMSE to the range of the observed data, and normalize to the mean value of the observed data.

$$NRMSE = \frac{RMSE}{(w_{obs,max} - w_{model,min})}, \tag{4.6}$$

$$RMSE = \frac{RMSE}{\bar{w}_{obs}}, \tag{4.7}$$

where $\bar{w}_{obs}$ is the observed data mean value. NRMSE is expressed in percentage, and results in a better GFit when close to 100%. Through this thesis, as the parameter estimation is done by $armax()$ function of MATLAB®, the NRMSE as the GFit is extracted by the function outputs. GFit of more than 80% is considered as the minimum validation level for ARIMA modeling in this thesis. For sake of parsimony in model order identification, minimum of orders are retained, beyond which GFit does not improve significantly.

Whiteness test of the residual. To validate the absolute quality of a model, it is necessary to search any possible correlation through the residuals of a fitted model (whiteness test). Mainly, residual for a fitted model is the difference between each value of fitted model and the observed data:

$$residual = data - fittedmodel \tag{4.8}$$

MATLAB® function $resid(model, data)$ is used to estimate the auto-correlations ($RN(i)$) of residuals. As the validation criterion, 25 auto-correlation lags are calculated for the whiteness test which can be changed if necessary. If the model estimation errors ($\epsilon(t)$) are perfectly white (ideal case), with a large number of samples ($N \longrightarrow \inf$) then: $RN(0) = 1$, and $RN(i) = 0$, $i \geqslant 1$[1][95]. As in real cases $RN(i)$ is never exactly zero, confidence intervals around the x-axis are considered. A good ARIMA model should have a residual auto-correlation function within the confidence intervals, indicating the residuals are uncorrelated. Confidence intervals are calculated according to the number of samples (N) as below (widely tested for a numerous applications):

$$RN(0) = 1, |RN(i)| \leqslant \frac{2.17}{\sqrt{N}}, i \geqslant 1 \tag{4.9}$$

Here, the considered intervals correspond to a signification level of 3% in case of a test hypothesis of Gaussian distribution. So as the $resid()$ function illustrates graphically, all the auto-correlations should be between the confidence levels to be uncorrelated. But if the residuals auto-correlation plot shows significant correlation in the residuals, the model should be modified to include larger auto regressive or moving average orders.

4.1.8 Forecasting

The goal in forecasting is to make statements about the likely values of future y_t's. Now, if we know the joint probability density function $p(y_1, ..., y_n, y_{n+1})$, including the relevant marginal probabilities, we could form the conditional distribution:

$$p(y_{n+1}|y_1, ..., y_n) \tag{4.10}$$

and then from the knowledge over past values $(y_1, ..., y_n)$, we could use the (4.10) expression to make a probability statement about the future value y_{n+1}. In this work, forecasting is not the main goal of obtaining the probability mathematical model of the time series.

[1]For Gaussian data, uncorrelated means independent. So, $RN(i) = 0$, $i \geqslant 1$ represents independency between $\epsilon(t)$ and $\epsilon(t-1)$, which means the residuals are white noise.

At this level, the stochastic behavior of traffic is modeled by ARIMA, and it can be written in the form of the corresponding linear transfer function, as represented by (3.7). All the model implementation steps are illustrated through a descriptive flowchart in Fig. 4.5.

4.1.9 Mathematical model

At this level, we model the deterministic behavior of time series, so finally the two modeling parts – deterministic and stochastic – are completed to give the total mathematical model. Thus, as we have ARIMA model of the stochastic data behavior, by adding all the previously removed deterministic trends, we will reconstruct the original measured traffic. To do so, inverse operations of all the differences (simple difference, mean difference and seasonal difference) are applied to the ARIMA model output. Through the inversion, the data traffic can be mathematically expressed as below, beginning from the stochastic behavior (linear filter fed by white Gaussian noise), then adding the deterministic behavior.

Estimated ARIMA output: a sequence of white Gaussian noise a_t is passed through the ARIMA(p,d,q) linear filter already identified and estimated.

$$w_t = \frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \tag{4.11}$$

in which the transfer function of ARIMA is applied on a_t, and output time series w_t expresses the stochastic behavior of data.

Add the seasonality: the periodical trend subtracted by seasonal difference is added here to the ARIMA output. According to the (4.3), the seasonal difference inversion, *i.e.*, $(\delta_s^D)^{-1} = (1 - z^{-s})^{-1}$, is applied in the form of a linear filter to w_t, and its transfer function degree is the related seasonal period, s. So, by multiplying the output of ARIMA to the seasonal transfer function inversion, *i.e.*, $\frac{1}{(1-z^{-s})}$, resulted series contains the corresponding seasonal components:

$$\bar{x'}_t = (\Delta_s^D)^{-1} w_t = \frac{1}{(1 - z^{-s})} w_t, \tag{4.12}$$

where $\bar{x'}_t$ denotes the time series including seasonal trend, with a zero mean value. $(\Delta_s^D)^{-1}$ denotes the seasonal difference inversion, in the form of a linear filter of order s.

Add the mean value: as explained in Section 4.1.4.2, through preprocessing, mean value, $mean(y'_t)$, is removed from the original simple-differenced time series. To form it back in the total description, we can add directly its corresponding value:

$$x'_t = \bar{x'}_t + mean(y'_t) = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \right) \frac{1}{(1 - z^{-s})} + mean(y'_t), \tag{4.13}$$

where x'_t is the series after adding back the mean value, which was removed in the zero-mean process. x'_t is at the same level of the original simple-differenced time series, y'_t.

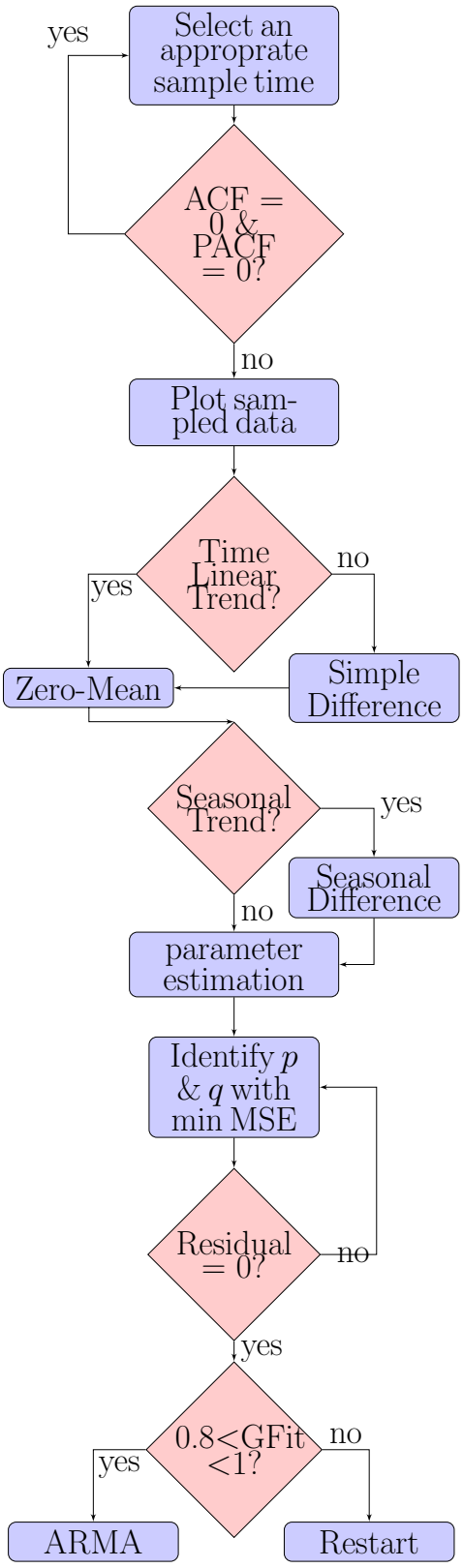


Figure 4.5 – Flowchart of the ARIMA model implementation and model validation steps.

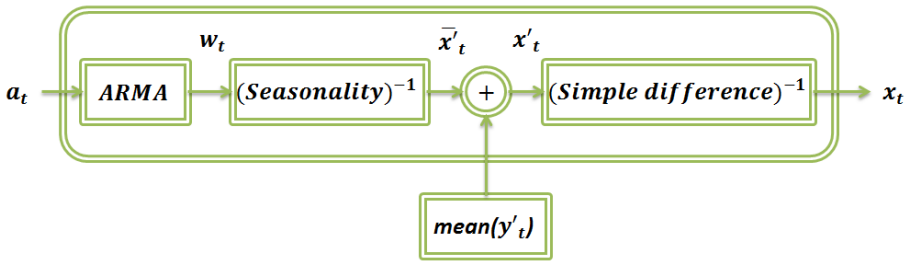


Figure 4.6 – Overall inversion process in order to obtain the total mathematical expression.

Inversion of simple difference (equivalent to the integration): same as the inversion process of seasonal difference, simple difference is applied back using a linear filter. This time, according to (4.1), the inversion of a first-order simple difference, $(1-z^{-1})^{d=1}$ is applied back as a linear filter of first degree, $(1-z^{-1})^{-1}$, to the output of reversed resulted series till now, as below:

$$x_t = \left(\frac{1}{1-z^{-1}}\right) x'_t, \tag{4.14}$$

where x_t is the estimation of original sampled time series y_t with both stochastic and deterministic parts, which is expressed mathematically below.

Total mathematical expression to estimate the original time series: At this level, all the deterministic components are added back to the model. Therefore, the total traffic behavior is expressed mathematically as below:

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + mean(y'_t)\right]\frac{1}{(1-z^{-1})}, \tag{4.15}$$

where x_t is the estimated traffic behavior through the proposed model. The inversion process is illustrated in Fig. 4.6. In Chapter 5, traffic flow is sampled over two different scenarios: over a co-simulation scenario, and a real-time traffic generation scenario respectively. The total estimated curve through ARIMA modeling and inversion process, x_t, will be compared to the original sampled data, y_t.

4.2 Maximum message transmission delay

As mentioned in Chapter 3, in this work, the maximum message transmission delay will be estimated as a network evaluation parameter, using an analytical model proposed by [161]. Here, the analytical model of Substation Communication Network (SCN) traffic flow makes use of four concepts: ***physical connection model***, ***logical connection model***, ***service model*** and ***source model***. Then, the algorithm of information flow distribution is described, whereas the communication network configuration, and parameters are completely known. In this way, it is possible to apply the related calculation procedure, and obtain the maximum transmission delay of each traffic flow from its source to destination for a given case study.

4.2.1 Port connection model

Ports of devices in SCN are independent units, so transmission path of a message can be a combination of connections between ports which either send out, forward or receive that message. In this part, a port connection model is presented to formalize the relationship between ports on a path for the information flow, including a physical connection model and a logical connection model.

A physical connection relies on physical transmission media like optical fibers or cables, and a logical connection is defined by the control logic of switches. On the arrival of a message, control logic of the switch determines which ports it should be forwarded to. The models are constructed using matrices that at first show the network message distribution, then transmission delay can be estimated while passing through each port.

Method application is explained considering a simple communication example of SCN presented in Fig. 4.7. There is only one Ethernet switch in this example, which transmits a flow from a Metering Unit (MU) to an Intelligent Electronic Device (IED).

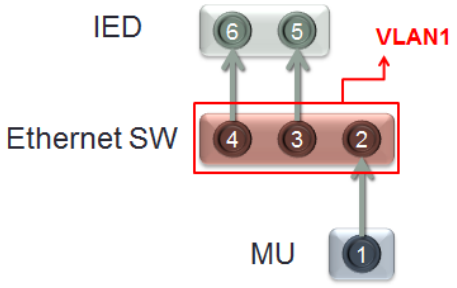


Figure 4.7 – A typical example of SCN port connection, including one VLAN group.

In order to distinguish all the port connections, first step is to draw all the connections (cabled or switch logic) as shown in Fig 4.8. All the ports are labeled by numbers from 1 to 6, and they are denoted by nodes. Physical connections and logical connections are denoted by solid arcs and dashed arcs respectively. Each arc has an arrow to show its transmission direction.

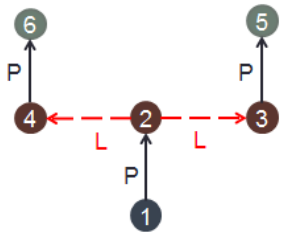


Figure 4.8 – Di-graph of the physical and logical connections.

To start the message distribution algorithm, main elements of the corresponding model have to be initialized. Physical and logical connection model (*i.e.*, matrices A and C), source flow model (*i.e.*, matrix S), and related VLAN logical connection model (*i.e.*, matrix V). Definition of each of these matrices is detailed hereafter.

Physical connection model: to build the physical connection model, a $P \times P$ matrix A is defined where P is the total number of SCN ports. A includes 0 or 1 values, in which a_{ij} is 1 if there is a solid arc from node j to node i, and equal to 0 otherwise. For the connection di-graph in Fig. 4.8, related physical connection model is defined by matrix A as below:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

where there are three solid arcs (*i.e.*, cables) connecting ports together, so A contains three values equal to 1 mentioned by a_{21} (solid arc from 1 to 2), a_{53} (solid arc from 3 to 5), and a_{64} entries (solid arc from 4 to 6).

Logical connection models: this model represents the logical connections defined by the logic inside the network switches. In case of using VLAN as a network topology in this work, there are two principal matrices for the logical connection modeling: first V for the VLAN switch ports, next C for the logical connections. A VLAN-based network consists of one or more VLAN groups, and each group can include one or more switches. Each switch port of a VLAN has a specific ID (VID) corresponding to the VLAN group that it is referred to. Suppose there are K VLANs in a SCN, then a $K \times P$ matrix V is defined to show the membership of each switch port to its related VLAN group. Our example consists of one VLAN, which in turn includes only one switch. Thus, all the switch ports, *i.e.*, 2, 3 and 4 belong to this VLAN, and related model is build by a 1×6 matrix V as follows:

$$V = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix},$$

where v_{12}, v_{13}, and v_{14} are equal to 1 as they are VLAN switch members, and all the other entries are equal to 0 as they are not a member of this VLAN group.

Likewise matrix A representing physical connections, a $P \times P$ matrix C is built denoting logical connections. c_{ij} takes 1 if there is any dashed arc directing from node j to node i in the di-graph, and 0 otherwise.

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So, matrix C is the logical connection model of the example illustrated in Fig. 4.8. As clearly there are two logical connections, *i.e.*, corresponding to the control logic of switches, 2 entries of C are equal to 1, c_{32} (dashed arc from 3 to 2) and c_{42} (dashed arc from 2 to 4).

4.2.2 Network calculus basic concepts

In this section, the concepts of service model and source model are presented. They are based on the network calculus theorem. Network calculus theorem is a methodology for performance evaluation of communication networks. With network calculus, some fundamental properties such as network services assigned to flows, window flow control, scheduling and buffer, or delay dimensioning can be studied [32], [33], [64], [24].

In this work, source port for a traffic flow is assumed to be the port that injects the flow to network. Source model is constructed to describe the properties of such source, *e.g.*, message length, message rate, *etc.* Service model describes the services received from switches by a message.

Basic concepts: the *arrival curve* and the *service curve* are the basics of network calculus theorem. Based on network calculus, each traffic flow is assumed to be a cumulative function $F(t)$, denoting the number of bits observed in a time interval $[0, t]$. As any arriving flow needs a guaranty to be served in the network, so any flow is limited by an arrival curve to receive the corresponding service curve passing through the ports [96], [56]. Upper-bound curve is defined as an arrival curve, $\alpha(t)$ that constrains the traffic flows arriving at a service system. It is said that traffic flow $F(t)$ is constrained by $\alpha(t)$, whose descriptive inequality is presented by (3.10). As previously explained in Chapter 3, the arrival curve proposed by Cruz [33] is estimated over the considered flow $\alpha(t) = \rho t + \sigma$. So that traffic flow $F(t)$ is (σ, ρ)-upper constrained.

Through a service curve, lower bounds are defined on the service provided by service systems such as switches, routers, *etc.* The service curve also abstracts the detail of packet scheduling mechanisms. Likewise the arrival curve, the service curve model is defined as a function of time during a time interval $[0, t]$. This function specifies the service provided by service systems. A linear (R, T) widely used model, $\beta(t)$, proposed by Cruz [33] will be considered as the service curve model, as follows:

$$\beta(t) = R[t - T]^+, \tag{4.16}$$

where $[x]^+$ denotes $max\{x, 0\}$ [96], R denotes the service rate, and T is the latency experienced by a message to be served after its arrival at a service system.

Source model: Assuming D as the total number of flows transmitted in the network, so a $P \times D$ matrix S is defined to express the corresponding relationship between messages and their sources. Matrix S is involved in two parts of the algorithm: first, it is used to model the distribution of a message from source to destination, and second, it includes the flow information, *i.e.*, arrival curve, passing through different ports, which is used for the

maximum delay calculation. Thus, if port i is the source of flow j, s_{ij} represents a value k_j as the length of message j, if S is intended to be used in the process of describing the traffic flow distribution. But, if S is intended to be used for the maximum message transmission delay estimation, s_{ij} is replaced with α_j which is its corresponding arrival curve function $\alpha_j(t) = \rho_j t + \sigma_j$. In both cases, s_{ij} equals to 0 if port i is not the source of message j.

Identification of ρ and σ: As explained in Section 3.3.2.1, the flow constraints are estimated for the corresponding data flow $F(t)$ as an upper bound. So, the values of ρ and σ are identified by assigning a linear upper-bound curve over the sampled flow $F(t)$. To do this, its trend $y(t)$ is estimated as a linear upward one, and its slope is considered as the slope of arrival curve which represents ρ. To fit the arrival curve on the upper bound, σ in bits is found as the maximum value of difference between the flow and its trend, $\sigma_j = max|F_j(t) - y_j(t)|$, where σ represents the maximum length of the corresponding flow j. The arrival curve estimated for a typical traffic flow is shown in Fig. 3.9. It is important to note that σ in both usages of matrix S denotes the length of the flow, so it will be replaced by k_j for the distribution algorithm.

Referring to our example in Fig. 4.7, the source model is built in case of the message distribution algorithm. It is supposed that there are two flows generated by MU (port 1) to the destination IED (port 5 or 6). Thus, the lengths of the generated flows by MU are replaced in the matrix S on the related source ports (MU port 1). As both of the flows are assumed to have the same characteristics, their lengths are the same and equal to $k_1 = k_2 = \sigma = 1216$ bits. So, $s_{11} = k_1$ and $s_{12} = k_2$, which are represented as below:

$$S = \begin{bmatrix} 1216 & 1216 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Service model: To calculate the transmission delay, service time provided at each port is considered by the service model. Service model is built as a new $P \times D$ matrix K in which k_{ij} represents the corresponding service curve $\beta_j(t) = \mathbf{R}_j[t - T_j]^+$ for flow j passing through port i.

Values of R_j and T_j for each flow are calculated according to the properties of the service providers (in our case Ethernet switches). Two important queue scheduling policies are considered in the service model of traffic flow, *i.e.*, First In First Out (FIFO) and Priority Queuing (PQ) [161]. They are used to improve the Quality of Service (QoS) and manage the flow congestion.

For a port with PQ policy, suppose packets arriving at a service system are tagged from priority 0 to n, so they are served from the highest priority to the lowest, and their arrival time is not important. In this case, a flow with the priority of $0 < k < n$ may be delayed by a flow with higher priority even arrived after, also will be a lower prioritized flow already receiving

service. But inside a flow, FIFO is the policy for packet transmission. In all the experiments performed in this work, FIFO is the policy applied for all packets of a flow and for all different flows transmitted through SCN. FIFO assigns the same priority to the packets of different transmitted flows, so they are served in order of their arrivals. Regarding to the FIFO policy, corresponding values of R_j and T_j for the flows of the same priority k are obtained as follows:

$$R_j = R - \sum_{i \in k \& i \neq j} \rho_i, \tag{4.17}$$

where R, in bps, denotes the switch transmission rate, and can be found in its datasheet specifications. This expression means that the transmission rate at each port is affected by the other flows passing simultaneously through the same port. Thus, if more than one flow pass through the same port, the transmission rate to serve flow j is reduced by the other flow rates and the experienced latency is increased accordingly:

$$T_j = T + \frac{\sum_{i \in k \& i \neq j} \sigma_i + T \times \sum_{i \in k \& i \neq j} \rho_i}{R - \sum_{i \in k \& i \neq j} \rho_i}, \tag{4.18}$$

where T, in s, denotes the minimum latency if only one message passes through a port, and as for R, it is indicated in the switch datasheet. In case of more than one flow passing through the same port simultaneously, the latency is increased by the time factor depending on the other flows lengths and transmission rates.

Next section is dedicated to the distribution algorithm, all the matrix definitions and related calculations are explained in detail.

4.2.3 Message distribution algorithm

As discussed before, a flow which is transmitted from its source to the destination is experiencing a delay through the path that is passing. Prior to delay calculation, it is necessary to obtain the distribution path of each flow. The path contains physical (cabling) and logical (switch logic) connections. At first, the distribution path from source to destination is separately described for each flow j in the form of a $P \times D$ matrix S_j. Then, a sum of all the S_j matrices results in a $P \times D$ matrix S_T using (4.19), which expresses the distribution paths of all the messages. Hence, the proposed algorithm should be executed D times, *i.e.*, equal to the total number of flows. The total distribution matrix S_T is formed by adding up all the individual distribution matrices S_j:

$$S_T = \sum_{j=1}^{D} S_j \tag{4.19}$$

The first step is to define all the basic matrices representing the models, *i.e.*, A, C, V, S. Referring to our example of Fig. 4.7, they are already defined in the previous section. For each flow distribution process, matrices S and C have to be built only for that flow related to the considered path that it passes. As in matrix S, each column is assigned to one flow j, S_j

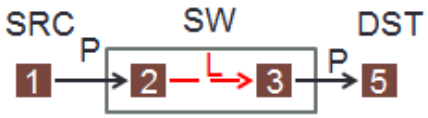


Figure 4.9 – Distribution path of flow 1.

is redefined by setting all the entries to zero except s_{ij} (representing the source of the current flow under calculation on its related column j). So, for the first flow, *i.e.*, first column $j = 1$, in our example, S_1:

$$S_1 = \begin{bmatrix} 1216 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

With the same way as S_j for each flow, matrix C_j is redefined representing only the corresponding logical connection which let the flow j passes through. Considering flow 1, Fig. 4.9 represents the distribution path of this flow. So, logical matrix C_1 for flow $j = 1$ is as below:

$$C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

As flow 1, following its transmission path, passes through one logical connection from port 2 to port 3, there is only one entry of C, *i.e.*, $c_{1,32}$, for flow 1 which is equal to 1.

Through the algorithm from source to destination, we follow the physical and logical connections alternatively. Thus, when a message passes through a cable connection, S_j it is multiplied by A (representative of the corresponding physical connection model), and if it passes through a switch, S_j it is multiplied by C_j (representative of the corresponding logical model). Since the first port is the flow generator, matrix S_j^0 is obtained as:

$$S_j^0 = -(A \times S_j), \tag{4.20}$$

where zero in S_j^0 denotes the departure of a flow from its source. Left multiplying of S_j by A means that message j is transmitted from port i to the switch port directly connected to the port i through a physical connection, *i.e.*, from the generator device to the first switch level.

Now, considering the above example, S_1^0 for the first flow is created as below:

$$S_1^0 = \begin{bmatrix} 0 & 0 \\ -1216 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

where the first flow is generated by port 1, and received by port 2 of the switch passing through cable connection (physical connection).

For the distribution algorithm, switch level are numbered. k denotes the number of switch levels where a flow enters. Starting a flow distribution model, k is equal to 1 representing the first switch level, and further if there are more than one switch for a flow to reach its destination, k is increased by one at each switch level. By setting $k = 1$, and using (4.21), a flow distribution inside the switch is modeled.

$$S_j^{3k-2} = -(C_j \times S_j^{3(k-1)}) \tag{4.21}$$

with the same assumptions, and $k = 1$ for the considered example:

$$S_1^1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ +1216 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

where flow leaves the port 3 *via* a logical connection (S_1 with an odd superscript index). Then, it is forwarded into the next level that is calculated using (4.22). The terms $3k - 2$, always gives an odd number, and $3k - 1$, always gives an even number, make the flow transmission in case of passing through a logical connection or physical connection alternatively.

$$S_j^{3k-1} = -(A \times S_j^{3k-2}) \tag{4.22}$$

and this equation applied on the example gives:

$$S_1^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1216 & 0 \\ 0 & 0 \end{bmatrix},$$

where flow arrives at the port 5 *via* a logical connection (S_1 with an even superscript index). Each entry has a negative or positive sign, which consequently denotes entering to the port or exiting from it respectively.

At this point, the algorithm requires to check if flow 1 has arrived at the destination or not. To this end, all the non-zero elements of $S_j^{3k-1} = S_1^2$ are set to zero if their corresponding port i is a non-switch port, otherwise the value remains as it is. The resulting matrix obtained through these changes is named $S_j^{3k} = S_1^3$. Hence, flow j is received at destination if all the elements of S_j^{3k} are equal to zero; if not, distribution will be continued at another switch level, so the calculation is continued with $k = k + 1$.

$$S_1^3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

In matrix S_1^2, port 5 is a non-switch port located on the IED, so it is set to zero. S_1^3 results in a matrix with all the elements equal to zero, thus indicating that flow 1 is clearly received by its destination (port 5). $n = k = 1$ is the final value of the switch counter level that is passed by flows to reach their destinations. If there are any non-zero elements, k will be incremented to denote the next level switch and the algorithm will be repeated once again for the same flow to compute the message distribution through the second switch [2]. Finally, the following addition gives the total message distribution matrix, S_j, for each flow:

$$S_j = S_j + S_j^0 + \sum_{k=1}^{n} S_j^{3k-2} + \sum_{k=1}^{n} S_j^{3k-1} \tag{4.23}$$

The distribution matrix of the first flow, *i.e.*, $j = 1$, in the considered example is represented as below:

$$S_1 = S_1 + S_1^0 + S_1^1 + S_1^2,$$

and:

$$S_1 = \begin{bmatrix} 1216 & 0 \\ -1216 & 0 \\ +1216 & 0 \\ 0 & 0 \\ -1216 & 0 \\ 0 & 0 \end{bmatrix},$$

where the ports that flow 1 passes through are shown: while it is sent by its source port 1, it is received at port 2 through a physical connection, then it is sent from port 3 through the switch logical connection, and finally it is delivered to the destination port 5 through a physical

[2] In addition, k represents the k^{th} iteration of the algorithm for each flow j, as it is incremented by each switch level, which leads to a new turn of the algorithm from the beginning to the end. The total number of iterations to obtain the total message distribution model is equal to the number of switch levels n multiplied by the total number of flows D

connection. In the considered example, the distribution algorithm should be completely applied on flow 2 to build its distribution model. As a result, S_2 given hereafter:

$$S_2 = \begin{bmatrix} 0 & 1216 \\ 0 & -1216 \\ 0 & 0 \\ 0 & +1216 \\ 0 & 0 \\ 0 & -1216 \end{bmatrix},$$

where this time the ports that flow 2 passes through can be identified: it is sent from its source port 1, then it is received at port 2 through a physical connection, then it is sent from port 4 by the switch logical connection, and finally, it is delivered to the destination port 6 through a physical connection.

Considering D as total number of messages, and P total number of ports, the total message distribution model, *i.e.*, matrix S_T, is found out as the sum of all S_j matrices. According to (4.19) for the discussed example, it is obtained by $S_T = S_1 + S_2$ represented as below:

$$S_T = \begin{bmatrix} 1216 & 1216 \\ -1216 & -1216 \\ +1216 & 0 \\ 0 & +1216 \\ -1216 & 0 \\ 0 & -1216 \end{bmatrix}$$

The message distribution algorithm is illustrated through its related flowchart in Fig. 4.10.

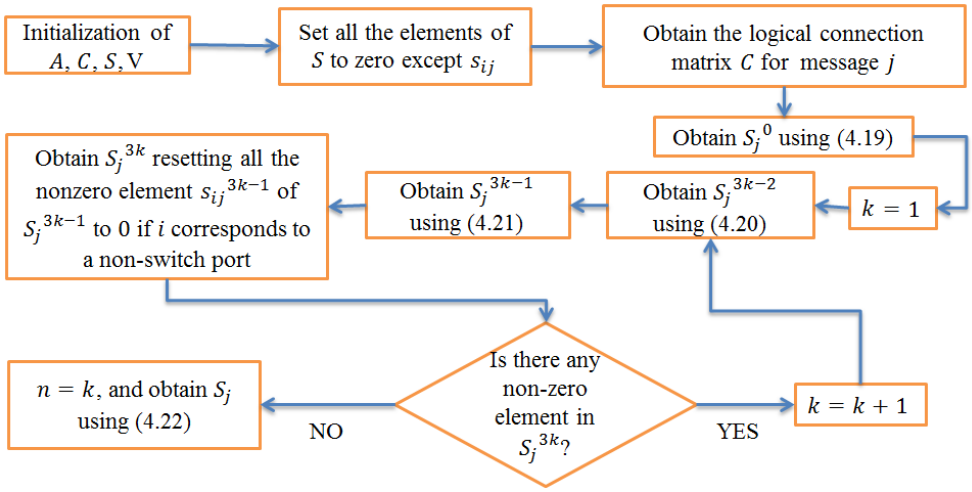


Figure 4.10 – Flowchart detailing the process of computing the message distribution matrix for flow j.

4.2.4 Delay estimation

Based on the message distribution model obtained through the previous section, all the ports that serve different flows are modeled. Now, it is possible to estimate the message transmission delay while passing through each port. Finally, by adding all the port-delay values, total delay experienced by a flow is calculated. To do so, the source and service model should be built up based upon the corresponding arrival curves (at each port) and service curves (for each flow). For the source model, as explained in Section 4.2.2, since here S is being used in the delay estimation process, s_{ij} values are filled by their proper arrival curves through the distribution algorithm. This time, instead of S_T, matrix S_D is obtained from the distribution process that identifies the message distribution matrix as the basis for delay estimation. Thus, the source model S_D related to the considered example, is written as follows:

$$S_D = \begin{bmatrix} 32 \times 10^3 t + 1216 & 32 \times 10^3 t + 1216 \\ -32 \times 10^3 t - 1216 & -32 \times 10^3 t - 1216 \\ 32 \times 10^3 t + 1216 & 0 \\ 0 & 32 \times 10^3 t + 1216 \\ -32 \times 10^3 t - 1216 & 0 \\ 0 & -32 \times 10^3 t - 1216 \end{bmatrix},$$

where, for each flow distribution, the proper parameters of their arrival curves, *i.e.*, $\rho = 32$ bps and $\sigma = 1216$ bits, are identified according to the identification process explained in Section 4.2.2.

For the service model in the form of a $P \times D$ matrix K, all the arrival curves in S_D are replaced with the corresponding service curves of each port through which the flow passes. As discussed in Section 4.2.2, while calculating R_j and T_j values, according to switch policy, all the flows passing through a switch port should be considered. If there is only one flow passing through a port, R (port transmission rate) and T (minimum latency of the port characteristics) can easily be found out from the given switch datasheet. But if there are more than one flow passing through a port, R_j and T_j are calculated based on the switch policy, because each flow is influenced by the other ones passing through the same switch port. As the policy defined in this example is FIFO, R_j and T_j for (4.17) and (4.18) are calculated for each port serving the related flow(s). Thus, service model is defined by matrix K, as follows:

$$K = \begin{bmatrix} R[t-T]^+ & R[t-T]^+ \\ R_1[t-T_1]^+ & R_2[t-T_2]^+ \\ R[t-T]^+ & 0 \\ 0 & R[t-T]^+ \\ R[t-T]^+ & 0 \\ 0 & R[t-T]^+ \end{bmatrix},$$

where two flows generated by their sources, passing through different ports of the network switch, receive services according to their presence at a given port. At port 2, these two flows are simultaneously present, so their assigned transmission rate, and the latency at port 2 is

affected by each other. This effect is calculated with new values of R_1, R_2, T_1, and T_2 using (4.17) and (4.18). But if, only one of the flows is present at a port, *e.g.*, port 3, R and T values are simply the datasheet values of the port characteristics. Assuming $R = 1$ Mbps and $T = 2.2$ μs as an example (that values are mentioned in the switch datasheet), related values are calculated for K as below:

$$K = \begin{bmatrix} (1 \times 10^6)[t - (2.2 \times 10^{-}6)]^+ & (1 \times 10^6)[t - (2.2 \times 10^{-}6)]^+ \\ (0.9 \times 10^6)[t - (2.2796 \times 10^{-6})]^+ & (0.9 \times 10^6)[t - (2.2796 \times 10^{-6})]^+ \\ (1 \times 10^6)[t - (2.2 \times 10^{-}6)]^+ & 0 \\ 0 & (1 \times 10^6)[t - (2.2 \times 10^{-6})]^+ \\ (1 \times 10^6)[t - (2.2 \times 10^{-6})]^+ & 0 \\ 0 & (1 \times 10^6)[t - (2.2 \times 10^{-6})]^+ \end{bmatrix}$$

It should be mentioned that properties of all the ports in this example – switch ports, IED's ports, and MU's ports– are assumed to have the same service values, R and T.

At this level, based on S_D and K matrices, the delay can be estimated according to the min-plus algebra. The output flow bounds of a service system can be computed from the min-plus deconvolution of arrival curve and system service curve as explained by (3.14) and delay of the flow is bounded by (3.15). Now, for delay estimation of each port by α in the (ρ, σ) type and β in the $\beta_{R,T}$ type, this is upper-bounded by (3.16) as: $D(t) \leq h(\alpha, \beta) = T + \frac{\sigma}{R}$.

Next, a $P \times D$ delay matrix Y is defined. To calculate the delay value for each flow passing through the ports, we follow (3.16). Each y_{ij} considers T and R values of the related port i, and σ value of the related flow j passing through this port. Related values of R, T, and σ are available in their corresponding entries of S_D and K matrices, *i.e.*, $S_{D,ij}$ and k_{ij}. Hence, in the case of the considered example, Y is calculated as below:

$$Y = \begin{bmatrix} T + \frac{\sigma}{R} & T + \frac{\sigma}{R} \\ T + \frac{\sigma}{R} & T + \frac{\sigma}{R} \\ T + \frac{\sigma}{R} & 0 \\ 0 & T + \frac{\sigma}{R} \\ T + \frac{\sigma}{R} & 0 \\ 0 & T + \frac{\sigma}{R} \end{bmatrix}$$

The final result of this example is obtained by replacing the corresponding values of each entry in S_D and K:

$$Y = \begin{bmatrix} 0.0012 & 0.0012 \\ 0.0014 & 0.0014 \\ 0.0012 & 0 \\ 0 & 0.0012 \\ 0.0012 & 0 \\ 0 & 0.0012 \end{bmatrix}$$

For example, y_{21} is calculated by replacing $R = 0.9 \times 10^6$ bps and $T = 2.2796 \times 10^{-6}$ s from the service curve of k_{21}, and $\sigma = 1216$ bits of the first flow from arrival curve of s_{21}^D. The calculated delay in Y represents the maximum delay at each port. To calculate the total

delay experienced by a flow from source to destination, the column sum is applied for each flow. So, we have; $D_{1,max}(t) = 0.005$ s, and $D_{2,max}(t) = 0.005$ s.

In Chapter 5, a real-world real-time communication scenario is considered for which the presented algorithm of maximum delay estimation is applied. The estimated maximum delay is further compared to the measured values using a network analyzer in order to validate the method.

4.3 Conclusion

Section 4.1 of this chapter proposed a mathematical model of the IEC 61850 GOOSE traffic sampled over SCN. Through a communication scenario in the context of decentralized control, the transmitted information among different IEDs is properly sampled. To obtain a stationary series, all the preprocessing steps are explained in detail. After removing the deterministic components – such as trends and cyclic components – the traffic stochastic behavior can be modeled by ARIMA model. ARIMA implementation procedure is illustrated through a flowchart in this section.

Finally, to find the total mathematical model that also describes the deterministic behavior of traffic, all the removed trends are added back to the ARIMA output. From a control approach point of view, the proposed model could serve as a basic mathematical description of a single communication channel, viewed as an elementary brick of the communication network structure within a smart grid. To go further, in Chapter 5, a sensitivity study will be conducted to assess how the parameters of identified ARIMA model vary in relation to the variations of some communication network parameters.

In Section 4.2, by using an analytical model proposed by [161], the maximum transmission delay for each flow is estimated while passing through different ports. To explain the delay estimation process, a simple example of SCN with a VLAN consisting of only one Ethernet switch, one IED, and a PC is defined. Considering the analytical model, delay value for each flow is estimated separately. Some limitations could restrict applicability of the proposed model for a large-scale network delay calculation, especially if this is aimed at being run on-line. For example, if there are D flows, considered algorithm should be repeated D times, which can suppose a large amount of data and running time for a large-scale system. In addition, main matrices (*e.g.* A, C, S, S_j, *etc.*) should be defined manually, which is a time-consuming task, and requires many variables to support recall tools through matrices.

Chapter 5 contains the application of the proposed ARIMA model over a simulated, and real-time scenarios, thus giving the possibility of the model validation, also emphasizing the possible limitations. The maximum delay is estimated over a real-time test bench for two communication scenarios in case of a power system that support the load shedding decision.

Chapter 5

Modeling of data flows and delay estimation within illustrative case studies

Contents

5.1 ARIMA modeling of MATLAB®/Simulink® co-simulation data flows 100
5.1.1 Case study description . . . 100
5.1.2 ARIMA modeling and mathematical model . . . 100
5.1.3 A parametric analysis of the traffic model . . . 108
5.2 ARIMA modeling of data flows in a real-time traffic generator 110
5.2.1 Reduced-scale distribution network: Communication scenario . . . 110
5.2.2 Data preprocessing and model estimation . . . 112
5.2.3 Mathematical expression . . . 114
5.3 Test-bench validation of a delay estimation analytical model 116
5.3.1 Communication in a load shedding scenario . . . 117
5.3.2 Simple scenario: without perturbation . . . 117
5.3.3 Estimated delay against measured values . . . 124
5.3.4 Traffic perturbation with background traffic . . . 127
5.4 Conclusion . . . 132

In this chapter, our proposed model detailed in Chapter 4 is applied on two different sampled-data traffics: first one over a sampled IEC 61850 GOOSE traffic in which communication scenario is simulated in Simulink®, and the second one is sampled over a real-time traffic generator of an intelligent substation test bench. Stochastic behavior is modeled by ARIMA, then deterministic descriptive equations are added to the model according to the procedure proposed in Section 4.1.8. The obtained model is submitted to the validation steps in order to reach the best estimation.

The second part of the experiments performed on the same test bench, consists in validation of a message transmission delay analytical method, as explained in Section 4.2. In this chapter, Section 5.3 presents an experimental delay estimation through the analytical method proposed by [161]. A load shedding setup is performed on two different scenarios

over a real MV/LV substation: the first scenario has no perturbation, and the second scenario is considered while communication traffic is perturbed by applying a background traffic. Delay estimation algorithm is applied on different flows for each scenario, and then estimated delays are compared to the measured delays obtained by Wireshark network analyzer. All the required codes for delay estimation process are written in MATLAB®.

5.1 ARIMA modeling of MATLAB®/Simulink® co-simulation data flows

One of the variables that different nodes could independently control, without the need of a centralized supervisor, is the reactive power absorbed or injected into the grid. Communication network in the smart grid enables implementation of some distributed and cooperative algorithms, such as Voltage Var Control (VVC), *i.e.*, voltage and/or reactive power control of a distribution grid, in a way that local energy resources cooperate to guarantee a voltage within the defined boundaries [14] , [133], [113].

Penetration of cooperative/distributed control is expected to continue; therefore the communication network, and the devices connected to it, should be able to adapt themselves to this paradigm shift.

5.1.1 Case study description

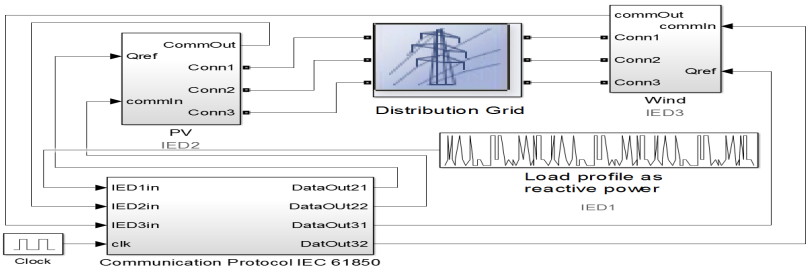
The case study considered here includes two renewable generating units, Photovoltaic (PV) and Wind Turbine (WT), committed to compensate the supposed known reactive power variation of a load unit. Each unit is equipped with an IED and communicates through a Virtual Local Area Network (VLAN).

A co-simulation model of electrical and communication grid implemented in MATLAB®/Simulink® (Electrical grid: SimPowerSystems toolbox, Communication grid: SimEvents toolbox) has been used to obtain the data traffic flow, namely the reactive power variation transmitted over IEC 61850 GOOSE [128].

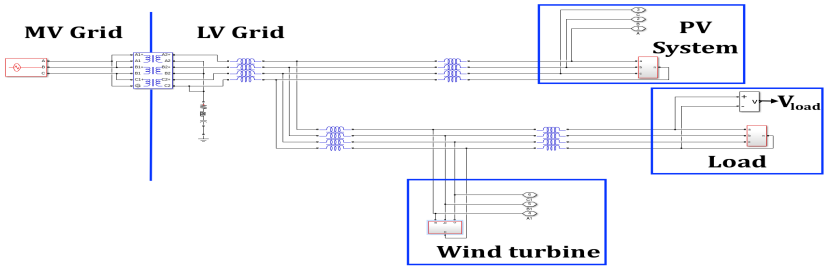
The focus is here on modeling the variation of number of data packets flowing through the communication channel over time, exchanging information about how much reactive power is able to provide each of them at a given time. Fig. 5.1 illustrates (a) the IEC 61850 communication network by which (b) a low-voltage distribution network including the PV, WT and load unit are interconnected.

5.1.2 ARIMA modeling and mathematical model

In this section, the implementation of ARIMA model over the data traffic obtained by simulation is explained. Load reactive power variations in Fig. 5.2 are sent through a communication



(a) Simulation block diagram of the communication architecture [137]



(b) Power grid configuration inside the block "Distribution Grid" [128].

Figure 5.1 – MATLAB®/Simulink® co-simulation model of an electrical distribution grid containing two renewable sources, and a load in interaction with its communication network.

channel, referred to as the nominal case. The nominal case considered in this case study refers to a traffic flow by a sampling period $T_s = 2$ ms without any packet loss. Channel speed is set at 10^6 bytes/s, and cable length is equal to 1 km.

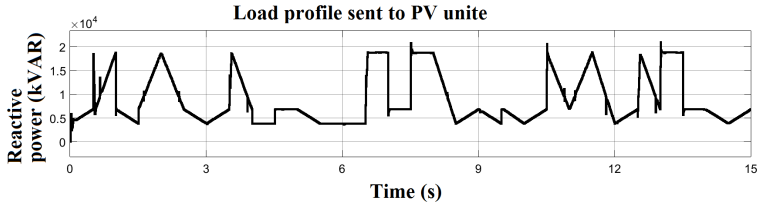


Figure 5.2 – Reactive power over time (kVAR) as the transmitted load reactive power profile.

The aim is to model the cumulative number of packets transmitted according to the reactive power curve. Data should be first prepared and smoothed over time. So, any trend (linear, cyclic) through sampled data should be removed. Then, applying a proper type of ARIMA model with the best orders (p,d,q) give the model of the stochastic part of data in a simple polynomial form. Finally, the model accuracy is evaluated by diagnostic tools: Residuals, Mean Squared Error (MSE), and Goodness of Fit (GFit). ARIMA model can also be used as a predictive model to forecast the future behavior of data [31].

Sampling period: ARIMA model is applied to the sampled data over our simulated

IEC 61850 GOOSE traffic. First step is to find a proper sampling period. GOOSE protocol has no hand-shaking topology as TCP/IP protocol, so to ensure the arrival of the packets, messages are retransmitted four times at each data change. Sampling period selection is an important step to have a series following the data changes properly.

Since here the co-simulation has a minimum retransmission period equal to 10 ms, so one can reasonably consider that the number of packets changes at least once during this time interval. Therefore, a reasonable choice for the sampling time would be less than $1/2 \times 10$ ms (according to Shannon's theorem [68]).

Different values less than 5 ms have been tested as the sampling period, among which the best one was selected as it resulted in the best GFit. Beside GFit, ACF/PACF coefficients have been checked not to be zero and show meaningful auto-correlation effects. So, sampling period has been chosen as $T_s = 2$ ms, which gives 7500 samples. Sampled time series is shown in Fig. 5.3.

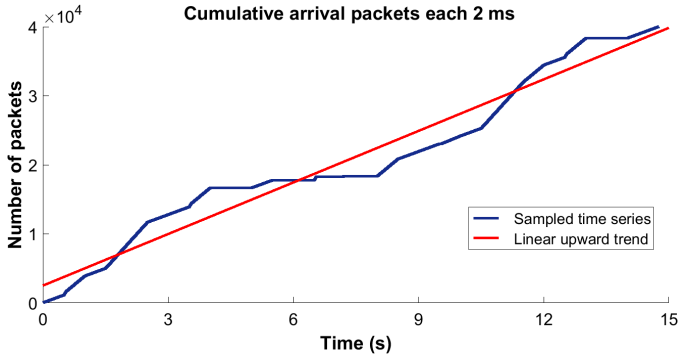


Figure 5.3 – Number of packets sampled over a sampling period of 2 ms.

Linear trend: To begin analysis, it is useful to look at time-series data graphically, so it is possible to find out some deterministic behaviors through time. On a graph, each data point represents an observation that succeeds to the previous one in time.

Data are recorded at discrete intervals: the lines connecting data points do not represent numerical values, but merely remind us that data points occur in a certain time sequence. In this case, as shown in Fig. 5.4, there is a linear upward trend over time in the original sampled data, so, to remove the trend, we have applied the first-order simple difference.

As explained in Chapter 4, ACF/PACF diagram may help to verify our steps. As shown in Fig. 5.5, ACF related to the untreated data has a slow decaying slope through time, which confirms an existing upward trend over time. This trend is also represented in Fig. 5.3 next to the original sampled data of the analyzed traffic flow.

Mean value: Next, the mean value of data has been removed to achieve a zero-mean process.

Fractional difference: As explained previously, model identification can be performed

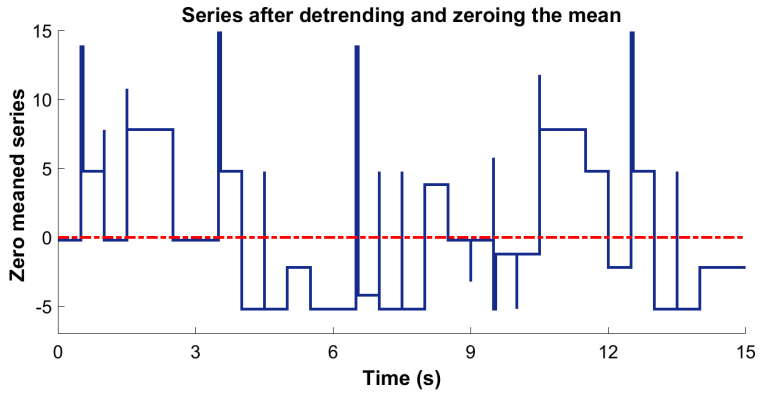


Figure 5.4 – The first-order simple difference is applied to the sampled-time series, and mean value is removed.

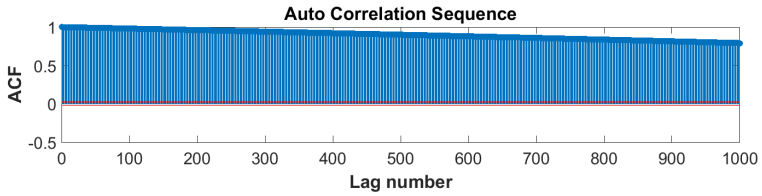


Figure 5.5 – ACF coefficients decay slowly before removing the linear trend over time.

with and without a fractional difference. The two versions of models have been evaluated by their GFit and MSE, and the version without fractional difference leads clearly to a more accurate result.

Seasonal behavior: Looking over the ACF/PACF diagrams at this level in Fig. 5.7, a periodic behavior through time appears clearly. To find the proper seasonal period, different choices are possible. First cycle is found out on 250 samples; to find the best seasonal period, different choices are applied (*e.g.*, 250, 500, 1000, 1500, *etc.*).

Then, the best one is selected for which the best GFit and minimum MSE are obtained.

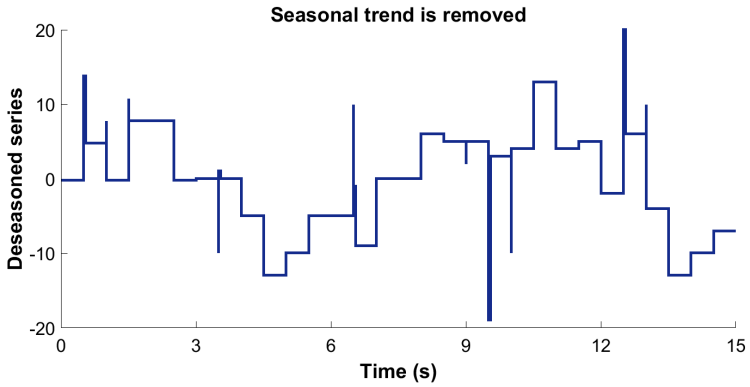


Figure 5.6 – Seasonal trend is removed by seasonal period = 3 s.

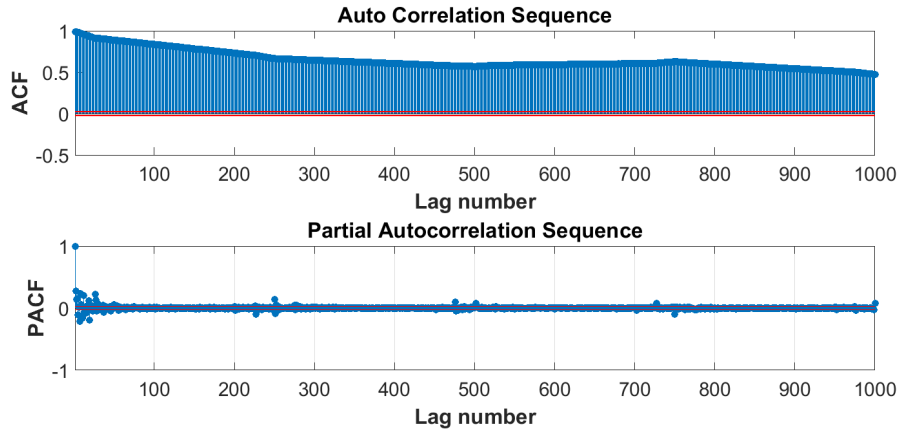


Figure 5.7 – ACF/PACF related to the series after simple difference.

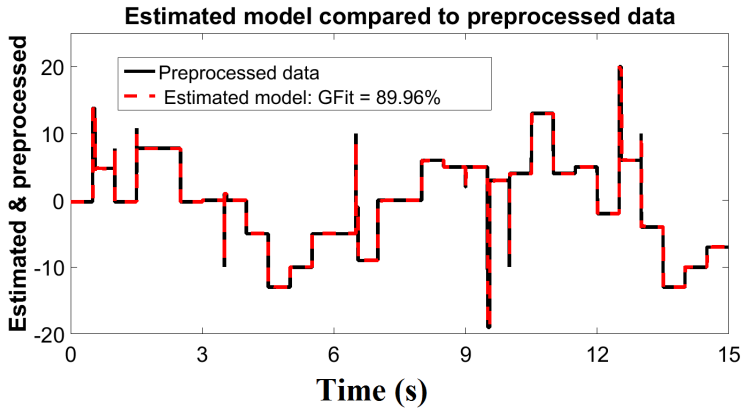


Figure 5.8 – The comparison of estimated ARIMA model against preprocessed series.

Here, the best seasonal difference is with $s = 3\text{s}/T_s = 1500$ sample times, that gives 5 cycles in total. Now, the seasonal trend is removed and data is preprocessed completely, so it can be modeled by a properly selected ARIMA model.

Parameter estimation: As all the deterministic parts are removed, thus, it is possible to identify the best orders, *i.e.*, p and q, for ARIMA model, and estimate the coefficients of the ARIMA polynomial as explained in Chapter 3:

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + ... + \varphi_p w_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}$$

The way we decided to select a proper ARIMA model was to find the minimum MSE value for all the possible pairs of p and q up to some values from where no significant increase in accuracy is obtained (*i.e.*, looking for the most parsimonious, meanwhile accurate, model). We will have the best GFit while its MSE is minimized. MSE values of different pairs tried in this model are illustrated in Fig. 5.9.

As explained previously in Chapter 3, another useful criterion, Akaike's Information Criterion (AIC), is also applied to check if the orders are properly selected. The same as MSE,

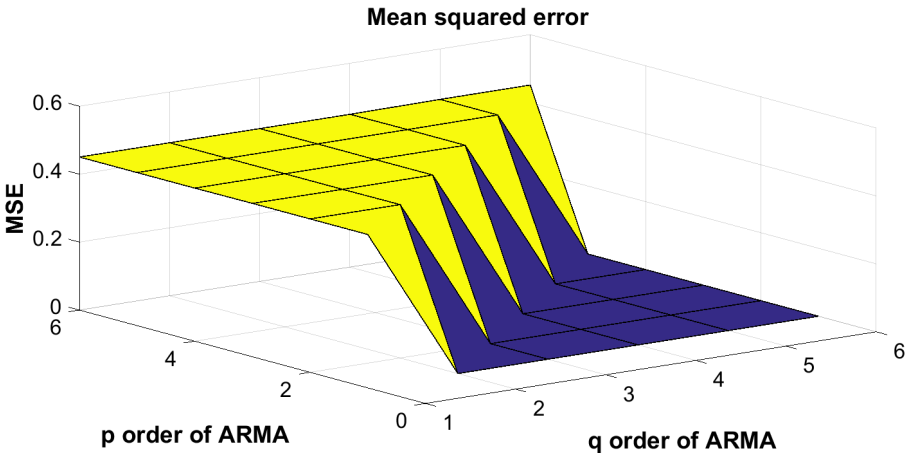


Figure 5.9 – Minimum MSE *vs.* orders p and q of ARIMA model.

ARIMA model is estimated for all the different pairs of p and q, and AIC is measured. The minimum value of AIC gives the best orders leading to a good enough GFit. AIC in comparison to the MSE gives the same Auto Regressive (AR) order, $p = 6$, but a less Moving Average (MA) order, $q = 3$. Comparing AIC with MSE, approximately the same GFit= 89.96% is obtained. Finally, Bode plots of the corresponding estimated models are compared in Fig. 5.10, where differ but insignificantly. So, we have decided to keep our choice of minimum MSE criterion for the best order selection. **Residual analysis:** Final step is to check whiteness of

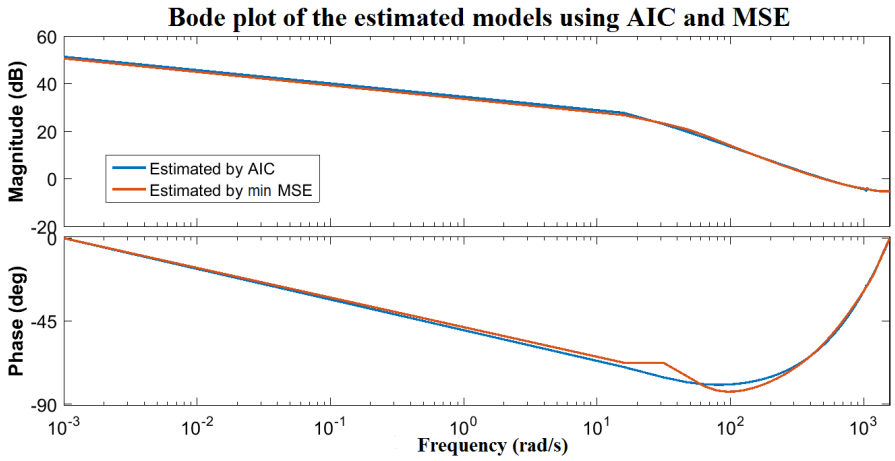


Figure 5.10 – Estimated ARIMA model by min AIC is compared to that obtained for min MSE.

model residuals. If residuals are mostly close to zero, then p and q are the best choice for the orders of ARIMA model polynomials. If not, we should increase p and q to pass the residual test. For estimated ARIMA model, whiteness of residuals is illustrated in Fig. 5.11. The confidence intervals are calculated as $|RN(i)| \leqslant \frac{2.17}{\sqrt{N=7500}} = 0.025$.

ARIMA polynomial: After selection of p and q, ARIMA polynomial coefficients will

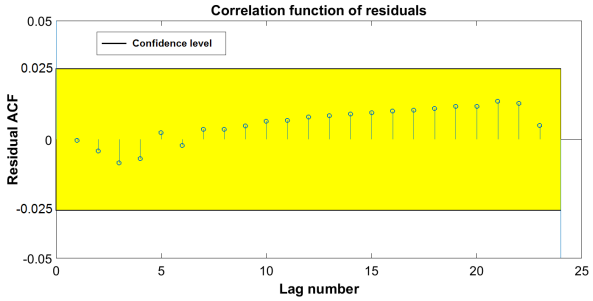


Figure 5.11 – The whiteness of residuals, showing a good estimation of data traffic behavior.

result as the final step of process. In this work, we have used the first four steps of Box-Jenkins methodology, *i.e.*, no forecasting. For the nominal case, ARIMA(6,1,4) *i.e.*, $p = 6$, $q = 4$, and first order of simple difference, the polynomial model estimated with the best estimation of GFit = 89.96% and MSE = 0.4497 is presented as below:

$$\begin{aligned} w_t &= 1.758w_{t-1} - 0.06078w_{t-2} - 0.7888w_{t-3} - 0.2517w_{t-4} \\ &\quad + 0.3021w_{t-5} + 0.04102w_{t-6} + a_t - 0.8397a_{t-1} \\ &\quad - 0.6358a_{t-2} + 0.1521a_{t-3} + 0.3345a_{t-4}, \end{aligned} \tag{5.1}$$

where a sequence of white Gaussian noise a_t is passed through the ARIMA($p = 6$,$q = 4$) linear filter, and treated by ARIMA transfer function mentioned in (3.7). So, w_t is the ARIMA output thus representing the stochastic part of the data traffic behavior.

5.1.2.1 Mathematical expression

As we have an ARIMA model of the stochastic data traffic behavior, by adding all the previously removed trends, we can reconstruct the original measured traffic. To do so, inverse operations of all the differences (simple difference, mean difference and seasonal difference) are applied to the ARIMA model output.

Through the inversion, data traffic can be mathematically expressed as below, starting from the stochastic behavior (white Gaussian noise), then adding the deterministic behavior. As all the steps are illustrated in Fig. 4.8, ARIMA output is processed here to find the overall mathematical expression of the traffic behavior.

Estimated ARIMA output: a sequence of white Gaussian noise a_t is passed through the ARIMA(p,d,q) already identified and estimated:

$$\begin{aligned} w_t &= \psi(z^{-1})a_t = \frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t \\ &= \frac{1 - 0.8397z^{-1} - 0.6358z^{-2} + 0.1521z^{-3} + 0.3345z^{-4}}{1 - 1.758z^{-1} + 0.06078z^{-2} + 0.7888z^{-3} + 0.2517z^{-4} - 0.3021z^{-5} - 0.04102z^{-6}}a_t \end{aligned} \tag{5.2}$$

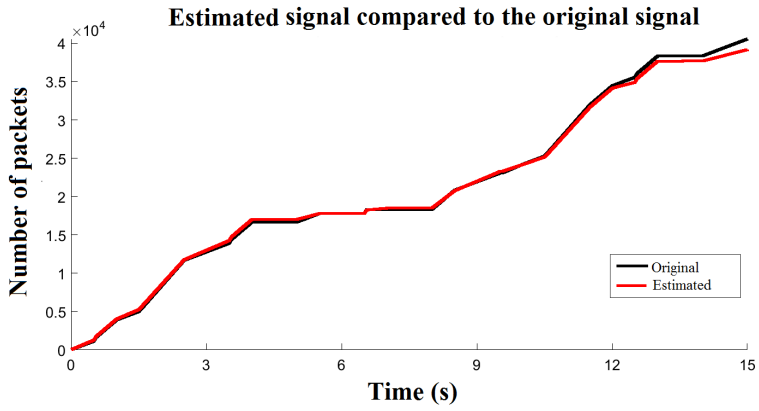


Figure 5.12 – Estimated model is compared to the original sampled time series.

Adding seasonality: periodical trend subtracted by seasonal difference is added here to the ARIMA output.

$$\bar{x'}_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})}, \tag{5.3}$$

where $\bar{x'}_t$ denotes the series including seasonal trend, and $s = 1500$.

Adding mean value:

$$x'_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + mean(y'_t), \tag{5.4}$$

where x'_t is the series after adding the mean value which was removed when zeroing the mean. Mean value is the exact value removed from the original sampled time series, y.

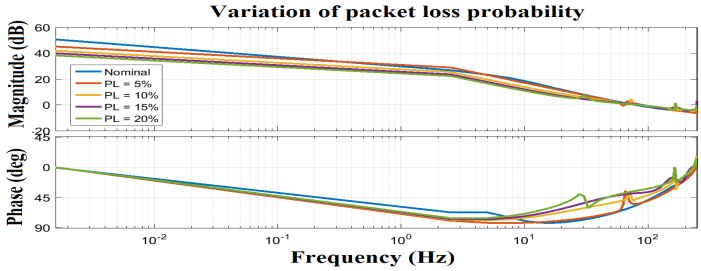
Inversion of simple difference (equivalent to integration): a discrete integrator is applied to the signal in (5.4) to account for the inversion of the simple difference.

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + mean(y'_t)\right]\frac{1}{(1-z^{-1})}, \tag{5.5}$$

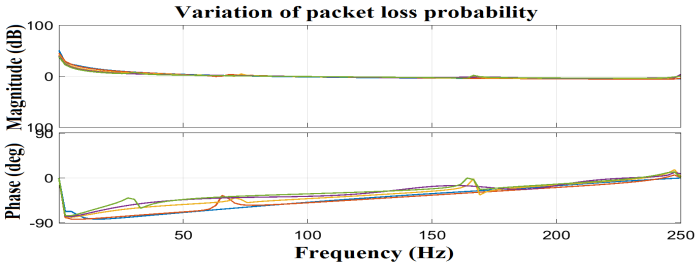
where the estimated model of the time series sampled with $T_s = 2$ ms is expressed hereafter:

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + 5.219\right]\frac{1}{(1-z^{-1})},$$

where $s = 1500$ sample times indicates the seasonal period. As all the parts are completed, *i.e.*, stochastic and deterministic, the estimated and original time series can be compared. Fig. 5.12 shows the estimated time series against the original one, that is, the number of packets transmitted over the communication channel under the considered scenario, in the nominal case.



(a) Logarithmic scale



(b) Linear scale

Figure 5.13 – Estimated ARIMA model Bode plots for packet loss probability (PL) = 5%, 10%, 15%, and 20%, the nominal case corresponding to PL = 0%.

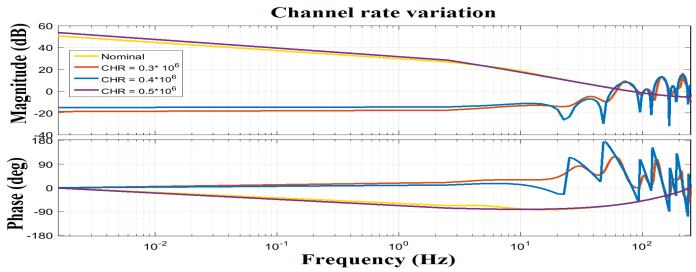
5.1.3 A parametric analysis of the traffic model

To go further, we have studied how the communication channel parameters affect the proposed traffic model. Thus, three main network channel characteristics were studied: packet loss, channel speed and cable length. For each factor, different values have been selected and set as new channel characteristics (all other parameters are constant at their nominal values), then simulation has been re-run and related traffic was sampled. So, resulted series were modeled by a proper ARIMA model. To assess the changes, we studied the identified model frequency response, so Bode plots are drawn in each case of parameter variation.

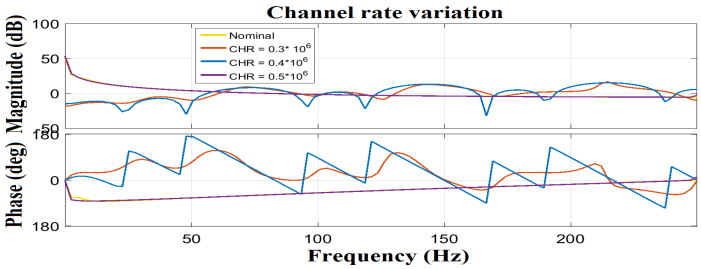
Packet loss: the probability of packet loss (PL) has been set to four different values. In the nominal case, there is no packet loss (PL = 0%). As shown in Fig. 5.13b of related Bode plots for each case, by increasing PL from 5% to 20%, the ARIMA orders increase too, so adds up the number of poles and zeros of the corresponding transfer function.

Variability exhibited in low frequency is more obvious in the Bode gain plot, whereas the Bode phase plots differs mainly in high frequency, which suggests that, the more intense the traffic is, the more sensitive the delays are to the variation of this parameter. As the considered data is discrete-time, Bode plots are also presented in the linear scale to analyze the information without accumulation in the high-frequency zone.

Channel rate or channel speed: in the same way as for PL, this time the channel transfer rate (CHR, the nominal rate in this co-simulation, is defined as 10^6 bytes/s) has been changed. Fig. 5.14b clearly shows the impact on the traffic model. If channel capacity



(a) Logarithmic scale.



(b) Linear scale.

Figure 5.14 – Estimated ARIMA model Bode plots for CHR $= 0.3 \cdot 10^6$, $0.4 \cdot 10^6$, and $0.5 \cdot 10^6$ bytes/s, the nominal case corresponding to CHR $= 10^6$ bytes/s.

decreases to half, traffic still flows close to the nominal status, but as CHR is tightened more than a half, the estimated model differs significantly from the nominal case, corresponding to somehow ill-configured communication network. In this way, limits for improper configuration of the channel capacity may be identified.

Cable Length: variation of the length of connecting cable between the IEDs indicated no considerable effect on the traffic stochastic model.

5.2 ARIMA modeling of data flows in a real-time traffic generator

This time, the proposed model is applied to a sampled series of periodic GOOSE data traffic, under stable grid conditions. So, the aim is here to model a real-time traffic by means of an ARIMA model with a good estimation fitness.

5.2.1 Reduced-scale distribution network: Communication scenario

This experiment is carried out on the test bench depicted in Fig. 5.15, where the employed devices may be identified. IEC 61850 GOOSE protocol is used as the communication protocol. It emulates a real MV/LV substation equipped with real electrical devices, like Remote Terminal Units (RTU) and protection relays. This test bench can be plugged on a real reduced-scale distribution grid, or on a current-injection box.

IEC 61850 protocol is implemented by CET850 simulator, the active power is generated by using the ramping module of an Omicron signal generator box. A typical active power signal is broadcasted over the network by a S80-Sepam PLC as the sender IED, while all the other units are listening.

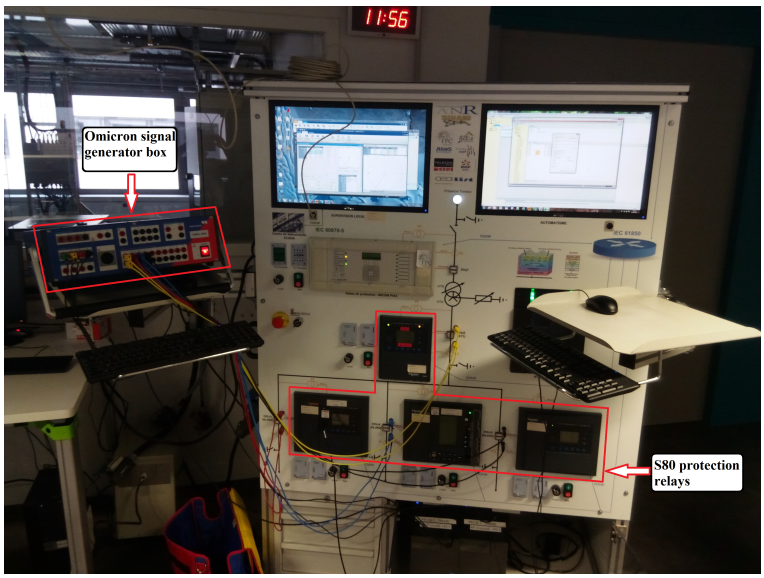


Figure 5.15 – Experimental MV/LV substation emulation test bench.

Power signal sent through the network is periodic of a total duration 144 s, its profile is presented in Fig. 5.16, which can correspond to a time-condensed variation over a day or so: power normally varies more slowly, such as 24 h instead of 37 s. To generate the active power (P) by Omicron Ramping module (Fig. 5.17), power factor is selected to be 1, therefore reactive power is zero. Each ramp (50-Hz three-phase electrical system) is constructed by the

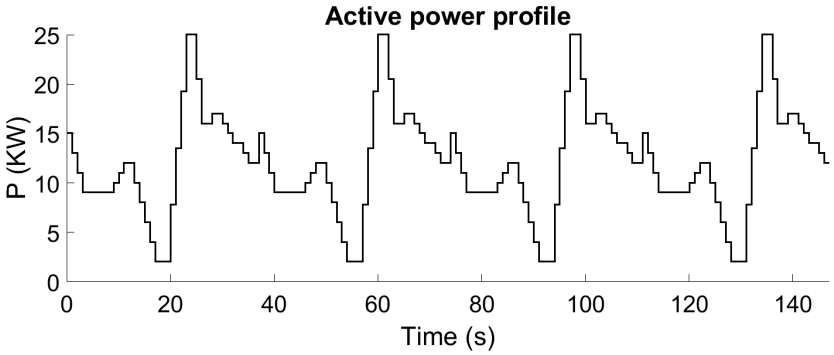


Figure 5.16 – Active power (kW) variation transmitted by GOOSE messages through the network.

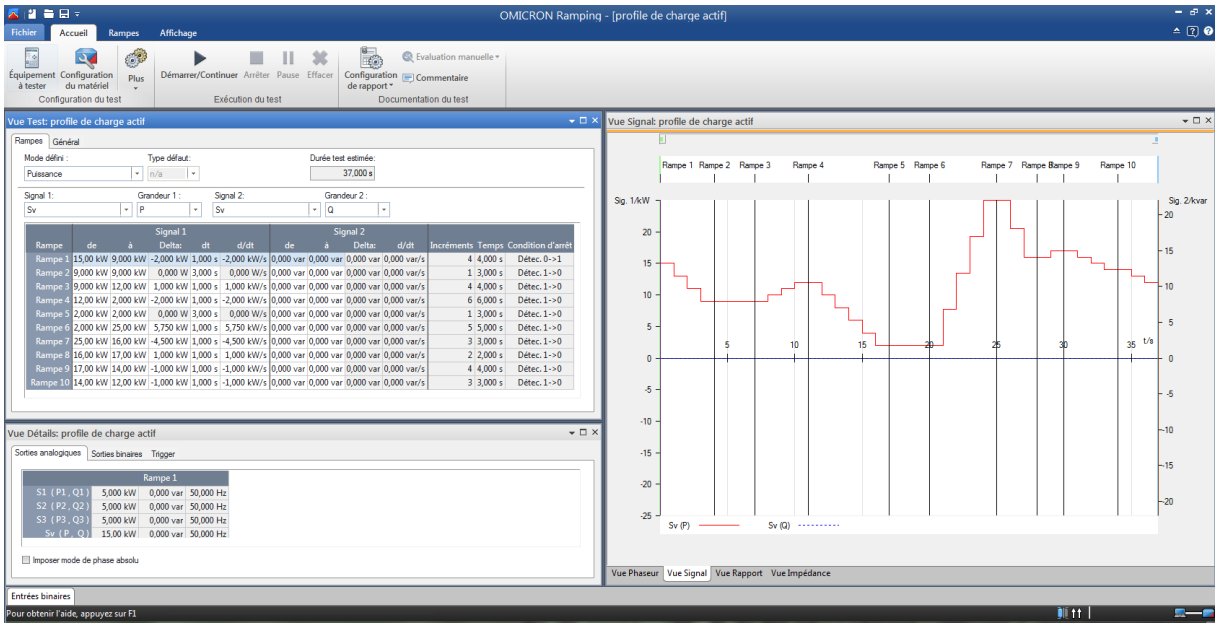


Figure 5.17 – Active power (kW) is entered in the Omicron Ramping module as piecewise-constant values by means of the corresponding current and voltage values.

related current (I) and the phase-to-neutral voltage (V) values, so $P = 3VIcos(\varphi)$.

While considered IED sends out power, number of frames is recorded by a monitoring PC *via* a network packet analyzer Wireshark (including transmission times and packet size). One of the tests is illustrated by a Wireshark analyzer window in Fig. 5.18, considering that the only traffic flow is composed of GOOSE messages transmitted from sender to receiver.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0x4607
2	0.000216000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0x4607
3	0.093935000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0x4607
4	0.093986000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0x4607
5	0.296957000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
6	0.983498000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
7	1.061138000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
8	1.825523000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
9	2.607974000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0x3a80
10	2.608193000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0x3a80
11	2.714690000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0x3a80
12	2.714740000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0x3a80
13	2.917767000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
14	3.031449000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
15	3.681894000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
16	4.446264000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
17	5.079358000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
18	5.225967000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0xb694
19	5.226257000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0xb694
20	5.335541000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0xb694
21	5.335598000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0xb694
22	5.538547000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
23	6.302626000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
24	7.067022000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>
25	7.127271000	Telemeca_d4:32:82	Iec-Tc57_01:00:00	GOOSE	151	
26	7.848206000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0xd2da
27	7.848425000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0xd2da
28	7.956216000	fe80::94f3:ae4c:7eab:f96b	ff02::1:3	LLMNR	84	Standard query 0xd2da
29	7.956275000	192.168.66.151	224.0.0.252	LLMNR	64	Standard query 0xd2da
30	8.159306000	192.168.66.151	192.168.66.255	NBNS	92	Name query NB WPAD<00>

Figure 5.18 – Monitoring of the network transmission by the Wireshark packet analyzer.

5.2.2 Data preprocessing and model estimation

Received data must be sampled at a fixed, suitably chosen sampling period. An interpolation by $T_s = 0.2$ ms is applied to data and results in 747 000 GOOSE frame samples. The sampled time series is shown in Fig. 5.19, and it is clear that there is an upward trend through time.

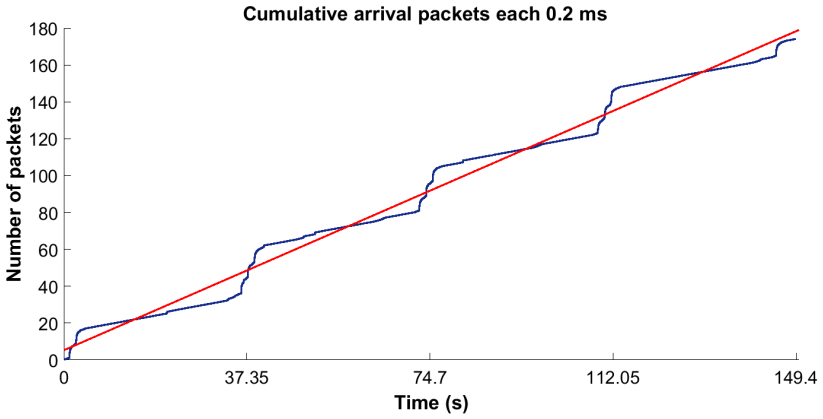


Figure 5.19 – Data traffic sampled each 0.2 ms.

Upward trend: since the number of frames is cumulative, there is a clear upward trend

over time. A first-order simple difference is applied to remove the trend, as seen in Fig. 5.20.

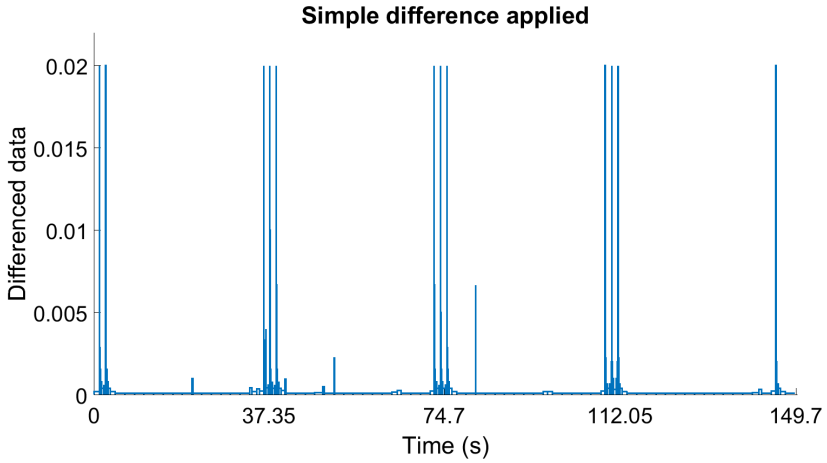


Figure 5.20 – Time series after simple difference application.

Zero mean: mean value is subtracted to result in a time series whose mean value is equal to zero.

Seasonality: at this level, by looking over the ACF/PACF or simply to the data drawn over time, the periodic components are recognized at each $s = 37.35$ s $/T_s = 186750$ samples (see Fig. 5.21). Time-series data consists of four cycles of $s = 18750$ samples. The first-order seasonal difference is applied to data which means $w_t = y_t - y_{t-186750}$, and the seasonal trend is smoothed as much as possible (Fig. 5.22). The remainder of data is mostly random, so ARIMA identification can be applied.

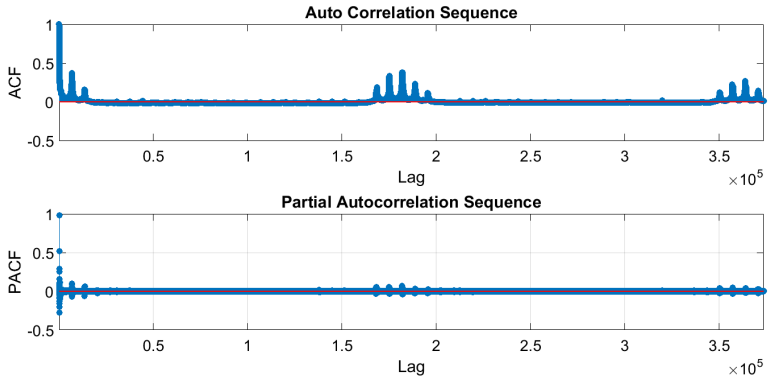


Figure 5.21 – Related ACF/PACF diagram indicating existence of a cyclic behavior.

ARIMA model best orders leading to the minimum MSE are for AR ($p = 6$) and for MA ($q = 6$), which results in an almost zero MSE (4.73×10^{-8}) and 82.48% for GFit. Finally, as shown in Fig. 5.23, since residuals are not all between the two confidence levels,

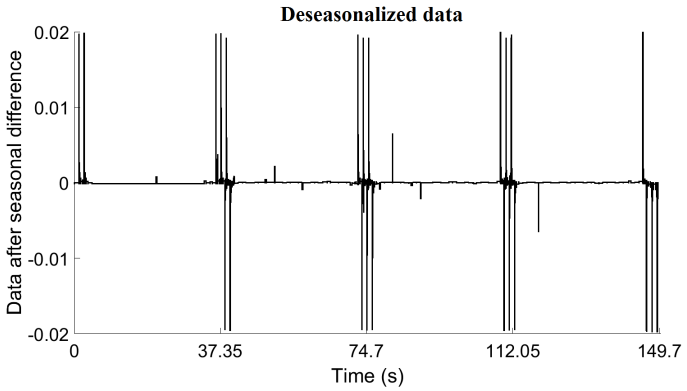


Figure 5.22 – Time series after seasonal difference application.

there still could be a solution to have a better fit, but the solution can already be declared as satisfactory, given the sufficiently high level of fit. Confidence levels are calculated by $|RN(i)| \leqslant \frac{2.17}{\sqrt{N=747000}} = 0.0025$. Fig. 5.24 shows the estimation of stochastic behavior of data. To check if our choice of the orders is the best, the criterion of minimum AIC is

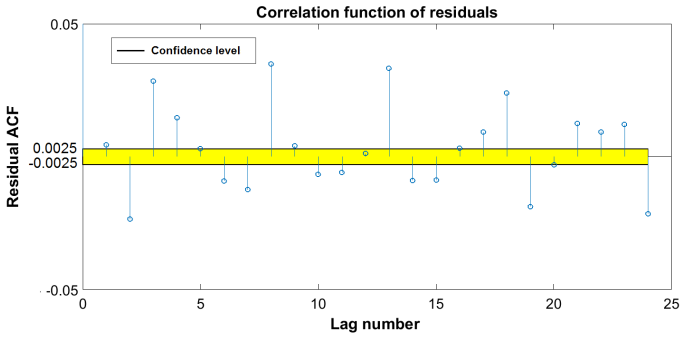
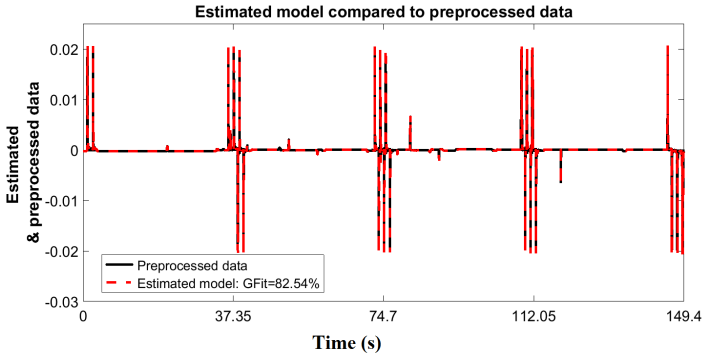


Figure 5.23 – The auto-correlation function of model residuals is calculated and graphically represented against 25 lags.

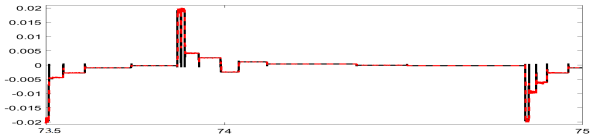
also applied. Exactly the same as for minimum-MSE criterion, $p = 6$ and $q = 6$ result as being the best orders. Hence, the minimum-AIC criterion confirms the orders selected by the minimum-MSE criterion.

5.2.3 Mathematical expression

Since estimation validation is in an acceptable range, the mathematical expression including all the deterministic components of data behavior is found out by applying all the removed trends (seasonal difference, zero mean, simple difference) *via* linear filters back to an approx-



(a) The estimated model



(b) Zoom in to a part of estimation

Figure 5.24 – The estimated model is compared to the original sampled time series.

imation of the original measured data. As the ARIMA output is as follows:

$$\begin{aligned} w_t &= \frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \\ &= \frac{1 - 0.0435z^{-1} + 0.0652z^{-2} - 0.1038z^{-3} - 0.2110z^{-4} - 0.8546z^{-5} + 0.3268z^{-6}}{1 - 0.7085z^{-1} - 0.169z^{-2} - 0.1604z^{-3} - 0.1745z^{-4} - 0.6747z^{-5} + 0.8896z^{-6}} a_t, \end{aligned} \tag{5.6}$$

a sequence of white Gaussian noise a_t is passed through the ARIMA(6,6) already identified to get w_t.

The following recurrent equation is obtained:

$$\begin{aligned} w_t = 0.7085w_{t-1} + 0.1694w_{t-2} + 0.1604w_{t-3} + 0.1745w_{t-4} + 0.6747w_{t-5} - 0.8896w_{t-6} + \\ a_t - 0.0435a_{t-1} + 0.0652a_{t-2} - 0.1038a_{t-3} - 0.2110a_{t-4} - 0.8546a_{t-5} + 0.3268a_{t-6}, \end{aligned}$$

as ARIMA(6,6) model of the stochastic part of traffic behavior.

Seasonalizing: A discrete integrator as the reversed of the seasonal difference is applied to the ARIMA output:

$$\bar{x'}_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})} a_t \right) \frac{1}{(1 - z^{-s})}, \tag{5.7}$$

where $\bar{x'}_t$ denotes the series including the seasonal trend, and $s = 186750$ denotes the seasonal period.

Adding the mean value:

$$x'_t = \left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + mean(y'_t), \tag{5.8}$$

x'_t is the series to which the removed mean value is added back.

Simple difference filtered back: finally, by applying the inverse of simple difference – that is, discrete integration – one can find the original measured data with a good precision. Fig. 5.25 compares the original sampled data to its estimation. Final model of time-series estimation, including the deterministic and the stochastic parts, is written as:

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + mean(y'_t)\right]\frac{1}{(1-z^{-1})}, \tag{5.9}$$

by replacing the corresponding estimated values, mathematical expression is written as below:

$$x_t = \left[\left(\frac{\theta(z^{-1})}{\varphi(z^{-1})}a_t\right)\frac{1}{(1-z^{-s})} + 2.3286\times10^{-4}\right]\frac{1}{(1-z^{-1})},$$

where the time-series data is sampled with $T_s = 0.2$ ms.

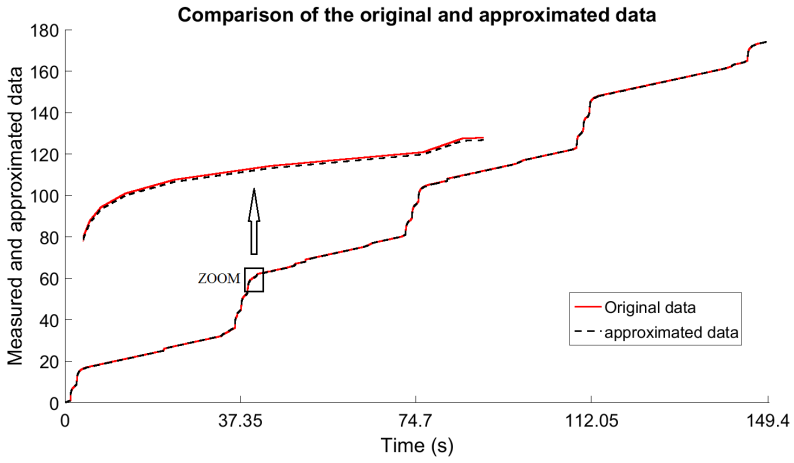


Figure 5.25 – Original signal is estimated through all the steps back.

5.3 Test-bench validation of a delay estimation analytical model

As detailed in Chapter 4, the analytical model proposed by [161] is applied over a typical example of the SCN traffic and the maximum message transmission delay is calculated based on the message distribution model. In this chapter, delay is estimated for two different scenarios with the goal of load shedding: the generated traffic by a reduced-scale SCN without perturbation, and the same traffic perturbed by some flows transmitted through the network

as background traffic. Both experiments are performed on the same real-time test bench as shown in Fig. 5.15 in Section 5.2.1, consisting of a real MV/LV substation equipped with real electrical devices. This time, the test bench plugged on a real reduced-scale distribution grid, shown on Fig. 5.26.

5.3.1 Communication in a load shedding scenario

In the load shedding setup, three loads send their typical active power (kW) consumption to the supervisory unit. IED1 and IED2 are inductive loads which are assumed to be constant, and IED3 is a typical load profile of an urban area which varies over time. For IED3, the load profile is created directly by a Premium PLC. Load units are communicating through their related S84-Sepam Power Line Communication (PLCs) as IEDs.

Data packets containing critical measured values (*e.g.*, line voltages or active power) are sent to a Zenon supervisory and control application unit, which is implemented on a PC station. As illustrated in Fig. 5.26, IEDs 1 and 3 are connected to the lines L4 and L6 respectively, which are both connected to node number 2 as the PC station placement.

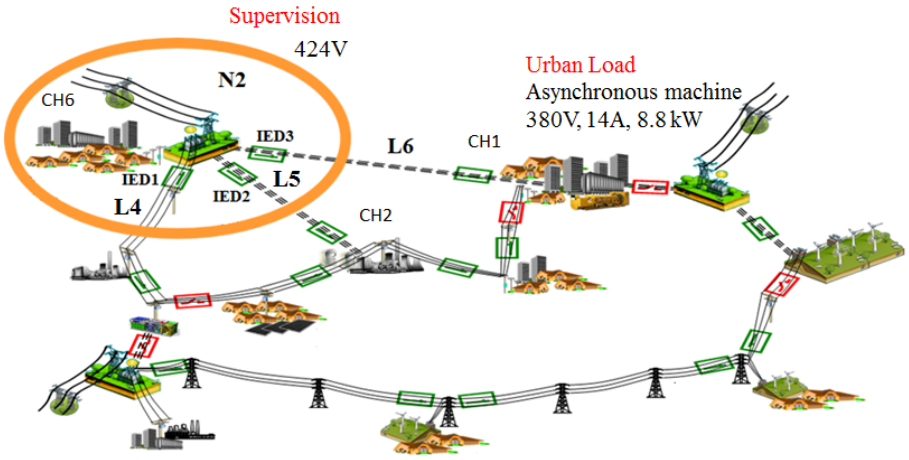


Figure 5.26 – IEDs placement in the load shedding scenario.

The communication scenario includes one VLAN group, which is realized by two switches: CISCO (2950) and NTRON (516TX). All messages passing through network are of MMS type. MMS is based on the TCP/IP protocol, and includes an acknowledgement process. So, when a message is sent, it is ensured that it will be received by a receiver through a handshaking protocol.

5.3.2 Simple scenario: without perturbation

In this scenario, there are three IEDs and a supervision PC as explained in previous section. IED1 is connected to a load on L6 which is a constant resistance , and its active power profile

is a constant value of $P = 2.6$ kW. A constant active power profile is considered for the load connected to IED2 on L5 as a motor consuming a constant value of $P = 3$kW. In case of the load connected to IED3 on L6, a variable active power (kW) profile is defined as a motor. The load profile of IED3 related load is shown in Fig. 5.27. The related load profile of IED3 is created in a Premium PLC. It contains 144 values occurring each 10s, which simulate a typical active power consumption during 24 hours, but in a reduced time scale of 24 minutes.

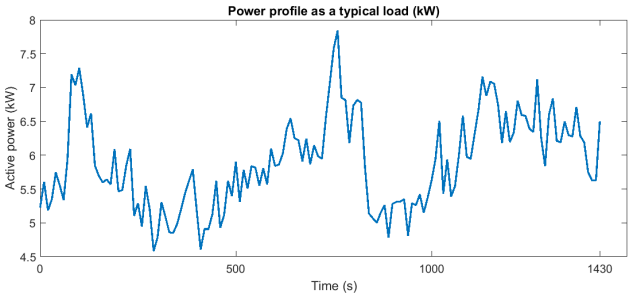


Figure 5.27 – Typical active power defined in Premium PLC as the load profile connected to IED3.

IEDs measure the critical values (*i.e.*, current and voltage) of these three load lines, and send the measured values to the PC station. An active power threshold is predefined at the PC station, so after receiving messages, PC calculates related active powers and checks the total load demand (the sum of all three loads at each time step) overpasses the threshold. If the threshold is overpassed, PC decides to shed the load with lowest priority.

The load connected to IED3 is assumed to be a residential load, and those connected to IED1 and IED2 are assumed as a factory and hospital load respectively. Thus, IED3, with the lowest priority, receives the command to open line L6, and the load on line L6 is cut out, thus the others are sure to receive electricity properly.

Threshold setting, load shedding program, and required calculations are done in the Zenon application. Related schematic of the corresponding scenario is represented in Fig. 5.28 including communication and supervision logic. As illustrated, the communication network connects three IEDs and a PC using an Ethernet switch. Measured values by IEDS are sent to the PC, and load shedding command is sent by PC to IED3. All the messages are sent using IEC 61850 MMS protocol.

In order to estimate the maximum message transmission delay for each of the flows in the represented scenario, according to the procedure presented in Section 4.2.1, the first step is to build the related port models (*i.e.*, physical connection model, and logical connection model). The communication network configuration of the unperturbed scenario is shown in Fig. 5.29, in which 4 flows pass through 2 Ethernet switches to reach their destination.

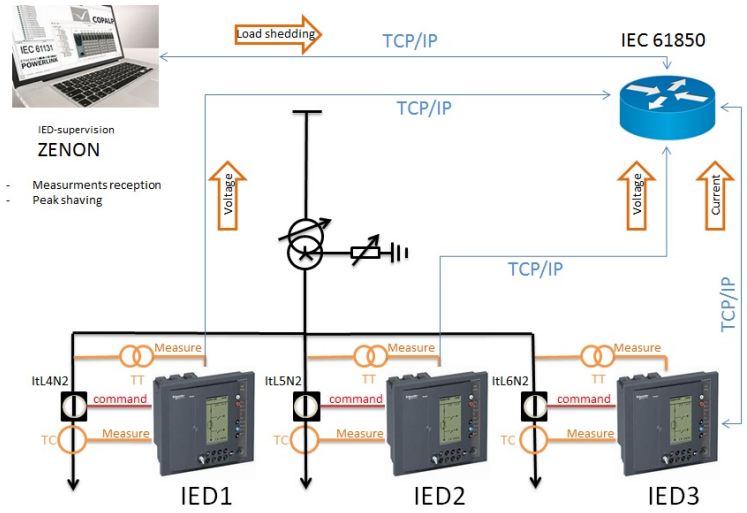


Figure 5.28 – Load shedding control process in the Zenon application on PC.

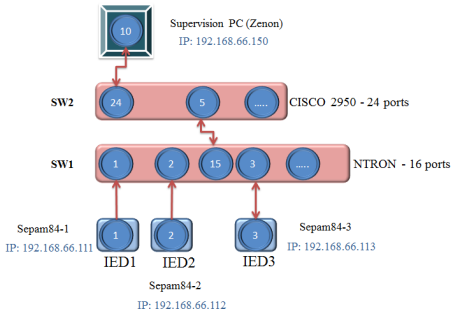


Figure 5.29 – Communication network configuration for the unperturbed scenario.

Thus, the scenario without perturbation includes 10 ports, 2 switches, and 4 flows, namely 3 flows to the PC station containing measurements, and 1 flow from PC to IED3 containing load shedding command. The IEDs are connected to the first switch (NTRON with 16 ports), and at the supervision side, PC is connected to the second switch (CISCO with 24 ports). A unique IP address in assigned to each of the devices in the network as they are marked in front of each device.

5.3.2.1 Port connection model

According to the modeling process explained in Chapter 4, the corresponding di-graph of the network configuration should be provided to build port connection models based on physical and logical connections. Physical and logical connections are distinguished by solid and dashed arcs respectively, and all the involved ports are numbered from 1 to 10 as they are represented in Fig. 5.30. Now, according to the visualized connections, it is possible to build up the physical and logical models.

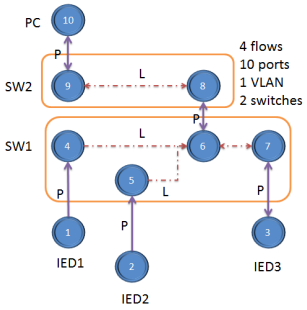


Figure 5.30 – Di-graph of physical (P) and logical (L) connections in the scenario without perturbation.

Physical model: As the unperturbed scenario contains $P = 10$ ports in total, the physical model is a 10×10 matrix A. a_{ij} is equal to 1, if there is a solid arc from port j to port i, and equal to 0 otherwise. In this way, the representing physical model for considered scenario is built as below:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

where 8 values of matrix A are equal to 1 representing 8 solid arcs (physical connections): a_{37}, a_{73}, a_{52}, a_{41}, a_{68}, a_{86}, a_{910}, and a_{109}.

Logical model: Again considering the existing 10 ports in the network configuration, a 10×10 matrix C is built, which consists of logical connections according to the logic of the switches. Each dashed arc (logical connection) is presented by 1 in the related entry of C, otherwise they are 0. Considering related di-graph in Fig. 5.30, C is presented below:

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where there are 6 logical connections, so the related entries in C are equal to one: c_{65}, c_{64}, c_{76}, c_{67}, c_{89}, and c_{98}.

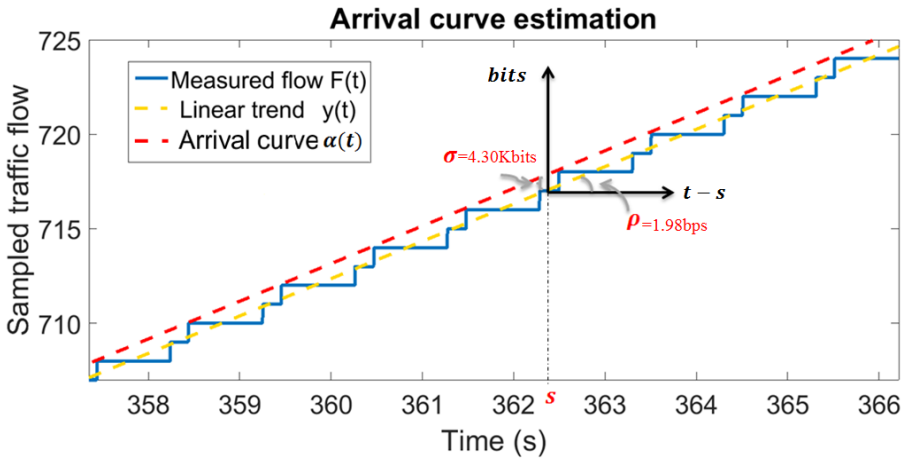


Figure 5.31 – Fitting the arrival curve as the upper bound for the first flow generated by IED1 and source port 1.

5.3.2.2 Application of network calculus theorem to the unperturbed load shedding scenario

As previously presented in Chapter 4, Section 4.2.2, according to the network calculus theorem, source and service models are constructed using the identified values based on the proposed linear models of (3.12) and (4.16).

Source model is based on the linear (ρ in kbits, σ in bits) model, $\alpha(t) = \rho t + \sigma$, where ρ in bits per second (bps) and σ in bits values result from identifying a proper arrival curve over each traffic flow emanating from a given source. Fig. 5.31 shows the estimated arrival curve for the first flow generated by IED1. As it is illustrated, $\rho = 1.98$ bps and $\sigma = 4.30$ Kbits are identified using the linear arrival curve, $\alpha(t)$. Then, the source model is constructed for two different purposes: for message distribution algorithm, and for delay estimation respectively (as explained in Section 4.2.2 of Chapter 4). As the total number of flows of the simple scenario is $D = 4$, then, for message distribution algorithm, a 10×4 matrix S is defined as follows, by indicating the length (σ in kbits) of each flow generated by its related source:

$$S = \begin{bmatrix} 4.30 & 0 & 0 & 0 \\ 0 & 4.31 & 0 & 0 \\ 0 & 0 & 4.34 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.80 \end{bmatrix},$$

whereas for delay estimation process, all the lengths are replaced with the identified arrival

curve related to each flow:

$$S = \begin{bmatrix} 1.98t + 4.30 & 0 & 0 & 0 \\ 0 & 1.98t + 4.31 & 0 & 0 \\ 0 & 0 & 1.98t + 4.34 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 106.32t + 2.80 \end{bmatrix}$$

Service model is based on the linear service curve $\beta_j(t) = R_j[t - T_j]^+$ (as explained in Section 4.2.2, in Chapter 4). (R, T) values are indicated in each of the switch datasheets. Service model, *i.e.*, a 10×4 matrix can be constructed based on the distribution matrix computed through distribution algorithm, where the service curve is replaced with the length of each flow while passing through a port (as explained in Section 4.2.4). So, prior to build the service model, message distribution algorithm is applied on the considered communication scenario.

5.3.2.3 Message distribution algorithm

According to the algorithm presented in Section 4.2.3 of Chapter 4, distribution of all messages passing through related ports is described by a 10×4 matrix S^T as follows. To build the distribution model, through the proposed algorithm, distribution of each flow is modeled separately. Here, the distribution algorithm is only applied on the first flow as follows:

- main matrices initialization (A, C, V, and S): Matrix A is already defined. C_1 for the first flow is built as below:

$$C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

  $c_{1,64} = 1$ and $c_{1,98} = 1$, since flow1 passes through these two logical connections. Matrix V contains a VLAN with 2 switches, and it is built as below:

$$V = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix},$$

  where all the switch ports of this VLAN are represented by 1 in V. Matrix S is a 10×4

matrix indicating the source model for flow1:

$$S = \begin{bmatrix} 4.30 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where port1 is indicated as the source of the first flow.

- distribution process: by applying the related algorithm of Chapter 4 for flow1, we will have matrix S_1:

$$S_1 = \begin{bmatrix} 4.30 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -4.30 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 4.30 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -4.30 & 0 & 0 & 0 \\ 4.30 & 0 & 0 & 0 \\ -4.30 & 0 & 0 & 0 \end{bmatrix},$$

where flow1 is generated by port1, received by port4 of switch1 and leaves the first switch through port6. Next, it is received by the second switch through port8, and leaves switch2 through port9. Finally, it is delivered to its destination port10.

- algorithm iteration: the distribution algorithm is applied for all the other flows, then according to (4.18): $S_T = \sum_{j=1}^{D} S_j$, the total distribution matrix S_T is built as below:

$$S_T = \begin{bmatrix} 4.30 & 0 & 0 & 0 \\ 0 & 4.31 & 0 & 0 \\ 0 & 0 & 4.34 & -2.80 \\ -4.30 & 0 & 0 & 0 \\ 0 & -4.31 & 0 & 0 \\ 4.30 & 4.31 & 4.34 & -2.80 \\ 0 & 0 & -4.34 & 2.80 \\ -4.30 & -4.31 & -4.34 & 2.80 \\ 4.30 & 4.31 & 4.34 & -2.80 \\ -4.30 & -4.31 & -4.34 & 2.80 \end{bmatrix},$$

where in this case, the algorithm is performed 4 times (equal to the number of flows). As there are 2 switch levels in the network, the switch level counter for each flow is $n = 2$, that means each flow passes through 2 switches to reach its destination.

5.3.2.4 Delay estimation process

As explained through the definition of source model in Chapter 4, in case of delay estimation, all the lengths of flows in matrix S are replaced with their corresponding arrival curves. Thus, S_D obtained for the delay estimation process, is re-written as below, containing related arrival

curves:

$$S_D = \begin{bmatrix} 1.98t+4.30 & 0 & 0 & 0 \\ 0 & 1.98t+4.31 & 0 & 0 \\ 0 & 0 & 1.98t+4.34 & 106.32t-2.80 \\ -1.98t-4.30 & 0 & 0 & 0 \\ 0 & -1.98t-4.31 & 0 & 0 \\ 1.98t+4.30 & 1.98t+4.31 & 1.98t+4.34 & -106.32t-2.80 \\ 0 & 0 & -1.98t-4.34 & 106.32t+2.80 \\ -1.98t-4.30 & -1.98t-4.31 & -1.98t-4.34 & 106.32t+2.80 \\ 1.98t+4.30 & 1.98t+4.31 & 1.98t+4.34 & -106.32t-2.80 \\ -1.98t-4.30 & -1.98t-4.31 & -1.98t-4.34 & 106.32t+2.80 \end{bmatrix}$$

Service model for the considered scenario is a 10×4 K matrix which is built according to the source model by replacing service curves assigned to each flow passing through a port. Matrix K is presented below:

$$K = \begin{bmatrix} (2.6, 2.2) & 0 & 0 & 0 \\ 0 & (2.6, 2.2) & 0 & 0 \\ 0 & 0 & (2.6, 2.2) & (2.6, 2.2) \\ (2.6, 2.2) & 0 & 0 & 0 \\ 0 & (2.6, 2.2) & 0 & 0 \\ (2.1, 3.21) & (2.0, 3.21) & (2.02, 3.22) & (1.81, 3.43) \\ 0 & 0 & (2.52, 2.24) & (2.51, 2.36) \\ (3.98, 2.45) & (3.98, 2.45) & (3.99, 2.45) & (3.85, 2.46) \\ (3.98, 2.45) & (3.98, 2.45) & (3.99, 2.45) & (3.85, 2.46) \\ (4.4, 0) & (4.4, 0) & (4.4, 0) & (4.4, 0) \end{bmatrix}$$

To obtain a service curve, (R, T) values are calculated depending on the FIFO policy and the characteristics of switches. If there is one flow passing through a port, R and T are the exact values of the switch characteristic the ones given in the datasheet. But if there are several flows passing through the same port, R in Gbps and T in μs are calculated according to (4.15), *i.e.*, $R_j = R - \sum_{i \in k \& i \neq j} \rho_i$, & (4.16), *i.e.*, $T_j = T + \frac{\sum_{i \in k \& i \neq j} \sigma_i + T \times \sum_{i \in k \& i \neq j} \rho_i}{R - \sum_{i \in k \& i \neq j} \rho_i}$. As it is shown, related values of R and T are different from the original values when there are more than one flow passing through a port simultaneously – *i.e.*, R decreases and T increases. It is important to note, (R, T) properties are assumed to be the same as those of switches for the ports of IEDs, and the latency inside the PC port is assumed to be zero.

To calculate the maximum message transmission delay experienced at each port, as detailed in Section 4.2.4, a 10×4 delay matrix Y is obtained based on the related values at each port of matrices S_T and K while satisfying (3.16), *i.e.*, $D(t) \leqslant h(\alpha, \beta) = T + \sigma/R$. Estimated delay for each port is shown in table 5.1. Now, the total delay experienced by each flow can be calculated simply by adding up all the delays at each port. According to the proposed algorithm in [161], the minimum delay for each port (original latency value T) is mentioned to be considered. In the next section, the estimation results are validated through a comparison performed against the measured values of Wireshark network analyzer.

5.3.3 Estimated delay against measured values

By now, the maximum delay experienced by each flow passing through related ports is estimated by exploiting the completely known communication network architecture and building

Port number	Flow1	Flow2	Flow3	Flow4
1	3.85	0	0	0
2	0	3.86	0	0
3	0	0	3.87	3.28
4	3.85	0	0	0
5	0	3.86	0	0
6	5.35	5.36	5.37	4.98
7	0	0	3.97	3.48
8	3.53	3.53	3.54	3.19
9	3.53	3.53	3.54	3.19
10	0.978	0.98	0.99	0.64

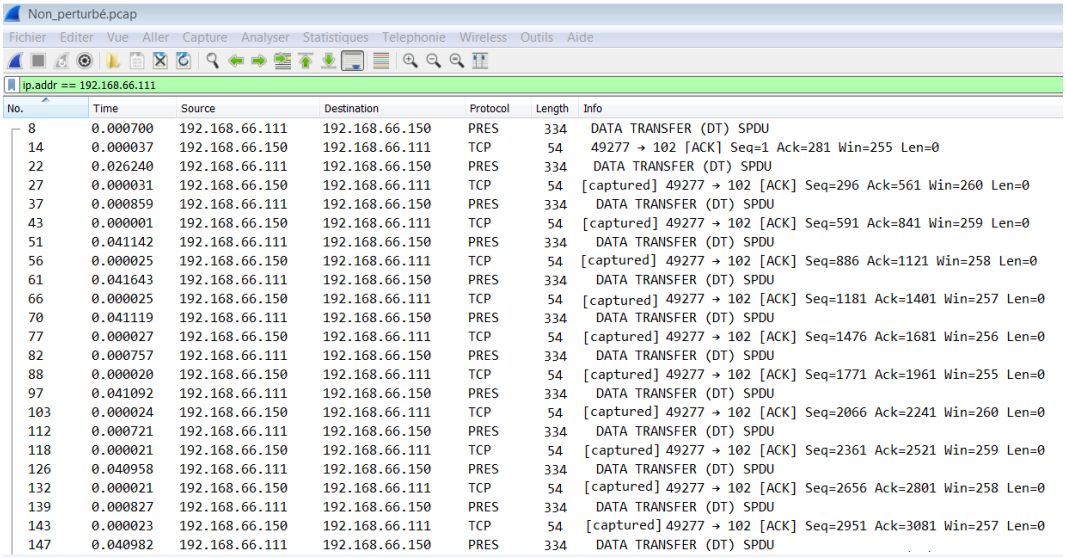
Table 5.1 – Estimated delay (μs) corresponding to each flow while passing through different ports.

up a relevant model of it. To evaluate if estimated values are reasonable, they are compared to the measured values by the Wireshark network analyzer. Wireshark monitors all the flows transmitted through the communication network under study, and gives various information including transmission delay. As a raw information, Wireshark measures the transmission time for a packet that is surely delivered to its destination. In case of the MMS protocol, a packet of data (DT) is surely delivered if the sender receives related acknowledgement message (ACK). The time measured by Wireshark is the time from sending a DT (bytes) until receiving the acknowledgement back, called Round Trip Time (RTT).

To make the RTT value comparable with the previously estimated delay, data flows monitored by Wireshark need to be prepared. First step is to sort out and recompose each flow as being the conversation between each IED and PC. For the first flow of the unperturbed scenario, the sample monitored flow1 is shown in Fig. 5.32. The two sides of the conversation are marked by their IP addresses (the same IP addresses mentioned in the communication network configuration of Fig. 5.29) in the source and destination columns. To extract related values, transmitted packets monitored by Wireshark should be filtered properly. Only packets containing data (DT) of sender, and their referred to acknowledgement (ACK) of receiver are listed in the supervision window, "Information" column.

While sorting the conversations, unnecessary information packets which were transmitted through each conversation are removed, such as Media Access Control (MAC) management packets relating to the communication protocol, and DT packets sent by PC when it was receiver in a conversation (they do not contain any information related to our communication scenario, only some packets related to the communication initialization). Thus, as it is illustrated in the Information column, there are only DT and their related ACK packets.

Through the proposed algorithm, delay is estimated per bit, while RTT in Wirshark is measured per packet in bytes and it should be converted into delay per bit. In addition, to obtain the maximum measured delay by Wireshark among all the measured delays, for each



No.	Time	Source	Destination	Protocol	Length	Info
8	0.000700	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
14	0.000037	192.168.66.150	192.168.66.111	TCP	54	49277 → 102 [ACK] Seq=1 Ack=281 Win=255 Len=0
22	0.026240	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
27	0.000031	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=296 Ack=561 Win=260 Len=0
37	0.000859	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
43	0.000001	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=591 Ack=841 Win=259 Len=0
51	0.041142	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
56	0.000025	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=886 Ack=1121 Win=258 Len=0
61	0.041643	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
66	0.000025	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=1181 Ack=1401 Win=257 Len=0
70	0.041119	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
77	0.000027	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=1476 Ack=1681 Win=256 Len=0
82	0.000757	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
88	0.000020	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=1771 Ack=1961 Win=255 Len=0
97	0.041092	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
103	0.000024	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=2066 Ack=2241 Win=260 Len=0
112	0.000721	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
118	0.000021	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=2361 Ack=2521 Win=259 Len=0
126	0.040958	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
132	0.000021	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=2656 Ack=2801 Win=258 Len=0
139	0.000827	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU
143	0.000023	192.168.66.150	192.168.66.111	TCP	54	[captured] 49277 → 102 [ACK] Seq=2951 Ack=3081 Win=257 Len=0
147	0.040982	192.168.66.111	192.168.66.150	PRES	334	DATA TRANSFER (DT) SPDU

Figure 5.32 – Flow1 of the scenario without perturbation is monitored in Wireshark, while sending data and receiving acknowledgement back. Time column shows RTT values in seconds.

conversation, the maximum RTT value is considered. Hence, the proper RTT is extracted for each flow transmission according to (5.10).

$$max(D_{measured}(t)) = max\left(\frac{RTT}{packet_{length} \times 8}\right), \tag{5.10}$$

in which $packet_{length}$ includes the sum of DT length and ACK length. Note that RTT is the same as the estimated delay, since we have considered the delay of sending a message until it is acknowledged by the receiver.

While extracting the measured maximum delay, for some of the flows there was a significant difference between the estimated value and measured value. Through the flow analysis, it was found a peak of RTT values monitored prior to the execution of the scenario, which would affect the $max(D_{measured}(t))$.

Since three different applications are cooperating to deploy the load shedding scenario – *i.e.*, Zenon, Wireshark, and Premium automate – it is reasonable to set a common start point for all of them to remove any unuseful values detected by Wireshark prior to the moment when the conversation enters a steady-state, stabilized operation. As they are activated manually, it may be difficult to start all three at the same time. Hence, to solve this, we defined a common start point for all the sampled flows by shifting our analysis window on the time axis such as to remove the initial "synchronization" phase.

At this level, measured delay can be correctly extracted through the explained data preprocessing steps. Now, to do a comparison, total estimated delay is obtained by a vertical sum of each column in matrix Y. Related estimation results are represented in the first row of Table 5.2, and measured delays for each flow is presented in the second row.

Estimated maximum delay	21.1	21.1	21.3	18.7
Measured delay	19.9	20.5	20.1	17.5
Tolerance	5.68%	2.79%	5.41%	6.45%
Exact difference	1.20	0.59	1.15	1.21

Table 5.2 – Estimated delay (μs) compared to the measured values.

As a validation criteria, the tolerance showing error value in percentage is calculated by (5.11) in the third row, and the simple difference which shows the exact value between estimated delay and the measured one in the fourth row.

$$Tolerance = \left(\frac{max(D_{estimated}(t)) - max(D_{measured}(t))}{max(D_{estimated}(t))} \right) \times 100 \tag{5.11}$$

To indicate a good estimation, tolerance is supposed to be lower than 10%. Clearly, in Table 5.2, tolerance values are in the acceptable range. Maximum transmission delay values are therefore validated with a good estimation. As all the tolerance values are positive, it can be noted that the estimated value is always larger than the measured one which is normal since it is about an upper bound of the delay.

5.3.4 Traffic perturbation with background traffic

In this section, load shedding scenario is perturbed by adding new IEDs which generate two other flows to the network. Two IEDs (S40-Sepam and P444) are added to the simple scenario in order to define a background traffic of MMS type. These messages include some measured values over the network which are not related to the load shedding goal. The new network configuration is illustrated in Fig. 5.33. This experiment contains 2 switches and one VLAN network. There are 6 flows: 5 flows to the PC station, and one flow from PC to IED3.

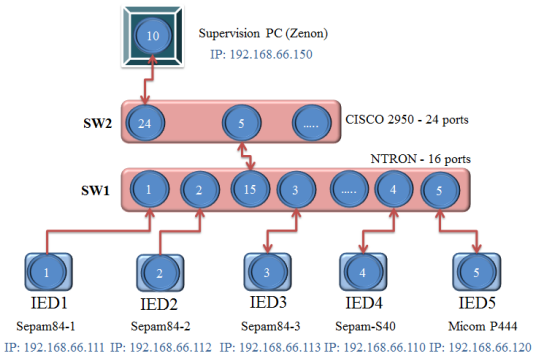


Figure 5.33 – Communication network configuration for the perturbed scenario.

5.3.4.1 Port connection model

As in the previous case, modeling process begins with the physical and logical models. Port connection di-graph is shown in Fig. 5.34. Physical connections are represented by solid arcs, and logical connections by dashed arcs. In this configuration, total number of ports are equal to $P = 14$, where each port is marked by a number.

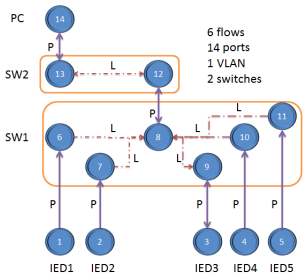


Figure 5.34 – Di-graph of port connections in the perturbed scenario.

Physical model: Considering solid arcs related to the physical connections, physical model is constructed by a 14×14 matrix A as follows:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

where any solid arc from port j to port i is represented by $a_{ij} = 1$, and the other entries are 0. In the perturbed scenario, there are 10 physical connections, so there are 10 entries of A equal to 1: a_{61}, a_{72}, a_{93}, a_{39}, a_{104}, a_{115}, a_{128}, a_{812}, a_{1314}, and a_{1413}.

Logical model: Dashed arcs represent the logical connections, which are the base of the

logical model, shown below in a 14×14 matrix C:

$$C = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},$$

where any dashed arc from port j to port i is represented by $a_{ij} = 1$, and the other entries are 0. In this scenario there are 8 logical connections, so there are 10 entries of C equal to 1: c_{86}, c_{87}, c_{89}, c_{98}, c_{810}, c_{811}, c_{1312}, and c_{1213}.

5.3.4.2 Application of network calculus theorem to the perturbed load shedding scenario

In the same way of Section 5.3.2.2, at this step, related values for arrival curves are identified by fitting a (ρ, σ) linear upper bound over each traffic flow. Then, the service curves corresponding to each port serving flows passing through them are obtained by identification of the related (R, T) parameters.

Source model is built by identifying the proper arrival curves over each of the flows, that allows (ρ, σ) values being identified over the considered linear model $\alpha(t) = \rho t + \sigma$, as illustrated in Fig. 5.29 for the unperturbed scenario following the general case in Fig. 3.9. So, in case of message distribution algorithm, source model is built by replacing the corresponding length of the flows in a 14×6 matrix S as below (note that in the perturbed scenario, there are $D = 6$ transmitted flows):

$$S = \begin{bmatrix}
4.37 & 0 & 0 & 0 & 0 & 0 \\
0 & 4.27 & 0 & 0 & 0 & 0 \\
0 & 0 & 4.37 & 0 & 0 & 0 \\
0 & 0 & 0 & 2.97 & 0 & 0 \\
0 & 0 & 0 & 0 & 1.14 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1.56
\end{bmatrix},$$

where it is mentioned which port is the source of each flow. In case of delay estimation

process, the arrival curves replace the message lengths:

$$S = \begin{bmatrix} 1.98t+4.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.98t+4.27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.98t+4.37 & 0 & 0 & 0 \\ 0 & 0 & 0 & 22.18t+2.97 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.98t+1.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.98t+1.56 \end{bmatrix}$$

Service model is a 14×6 matrix K which assigns the corresponding service curve to each flow passing through different ports. As done in the previous section for the unperturbed scenario, K is built based on the message distribution matrix S^T resulting from the related algorithm by replacing service curves with packet lengths at each port. To obtain K, first, it is necessary to apply the distribution algorithm on the total scenario to find S^T.

Message distribution model results by applying the algorithm presented in Chapter 4, Fig. 4.12. Total message distribution matrix for the perturbed scenario is obtained as follows:

$$S_T = \begin{bmatrix} 4.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4.27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.37 & -2.97 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.56 \\ -4.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & -4.27 & 0 & 0 & 0 & 0 \\ -4.37 & 4.27 & 4.37 & -2.97 & 1.14 & 1.56 \\ 0 & 0 & -4.37 & -2.97 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1.56 \\ -4.37 & -4.27 & -4.37 & 2.97 & -1.14 & -1.56 \\ 4.37 & 4.27 & 4.37 & -2.97 & 1.14 & 1.56 \\ -4.37 & -4.27 & -4.37 & 2.97 & -1.14 & -1.56 \end{bmatrix},$$

where the distribution path of the flows are detected and indicated by their related length at each port they pass through. For the purpose of delay estimation, S^T should be re-written as below:

$$S_D = \begin{bmatrix} 1.98t+4.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.98t+4.27 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.98t+4.37 & -22.18t-2.97 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.98t+1.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.98t+1.56 \\ -1.98t-4.37 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.98t-4.27 & 0 & 0 & 0 & 0 \\ -1.98t-4.30 & 1.98t+4.27 & 1.98t+4.37 & -22.18t-2.97 & 1.98t+1.14 & 1.98t+1.56 \\ 0 & 0 & -1.98t-4.37 & -22.18t-2.97 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.98t-1.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1.98t-1.56 \\ -1.98t-4.37 & -1.98t-4.27 & -1.98t-4.37 & 22.18t+2.97 & -1.98t-1.14 & -1.98t-1.56 \\ 1.98t+4.37 & 1.98t+4.27 & 1.98t+4.37 & -22.18t-2.97 & 1.98t+1.14 & 1.98t+1.56 \\ -1.98t-4.37 & -1.98t-4.27 & -1.98t-4.37 & 22.18t+2.97 & -1.98t-1.14 & -1.98t-1.56 \end{bmatrix},$$

Port number	Flow1	Flow2	Flow3	Flow4	Flow5	Flow6
1	3.88	0	0	0	0	0
2	0	3.84	0	0	0	0
3	0	0	3.88	3.34	0	0
4	0	0	0	0	2.64	0
5	0	0	0	0	0	2.80
6	3.88	0	0	0	0	0
7	0	3.84	0	0	0	0
8	7.98	7.94	7.98	4.59	9.24	9.40
9	0	0	4.68	4.09	0	0
10	0	0	0	0	2.64	0
11	0	0	0	0	0	2.80
12	6.89	6.97	6.79	3.43	4.76	4.85
13	6.89	6.97	6.79	3.43	4.76	4.85
14	0.99	0.97	0.99	0.68	0.26	0.35

Table 5.3 – Maximum delay experienced by each flow passing through different ports to reach its destination in μs.

where the arrival of the flow at each port in indicated by its related arrival curve instead of its length. Then, the service model is built based on S^D by replacing the corresponding linear service curve of each port presented by (4.15) as below:

$$K = \begin{bmatrix} (2.6,2.2) & 0 & 0 & 0 & 0 & 0 \\ 0 & (2.6,2.2) & 0 & 0 & 0 & 0 \\ 0 & 0 & (2.6,2.2) & (2.6,2.2) & 0 & 0 \\ 0 & 0 & 0 & 0 & (2.6,2.2) & 0 \\ 0 & 0 & 0 & 0 & 0 & (2.6,2.2) \\ (2.6,2.2) & 0 & 0 & 0 & 0 & 0 \\ 0 & (2.6,2.2) & 0 & 0 & 0 & 0 \\ (1.45,4.97) & (1.44,4.97) & (1.45,4.97) & (2.43,3.36) & (1.05,8.15) & (1.11,7.99) \\ 0 & 0 & (2.53,4.95) & (2.58,2.94) & 0 & 0 \\ 0 & 0 & 0 & 0 & (2.6,2.2) & 0 \\ 0 & 0 & 0 & 0 & 0 & (2.6,2.2) \\ (2.91,5.39) & (2.89,5.49) & (2.92,5.29) & (2.79,2.37) & (2.46,4.30) & (2.53,4.23) \\ (2.91,5.39) & (2.89,5.49) & (2.92,5.29) & (2.79,2.37) & (2.46,4.30) & (2.53,4.23) \\ (4.4,0) & (4.4,0) & (4.4,0) & (4.4,0) & (4.4,0) & (4.4,0) \end{bmatrix},$$

where (R, T) values are the characteristics of switches given in their respective datasheets if there is only one flow passing through a port, *e.g.*, flow1 at port1. But, if there are more than one flow passing through a port at the same time, based on the original (R, T) values, transmission rate and latency should be calculated according to (4.16) and (4.17).

Finally, using the calculated values in S_D through message distribution being served at each port in K, delay can be estimated while satisfying the required condition of (3.16). Delay is estimated for each flow passing through different ports to reach their destination, as presented in Table 5.3.

Delay estimation compared to the measurements: Now, it is possible to compare

Estimated maximum delay	30.5	30.5	31.1	19.6	24.3	25.1
Measured delay	29.9	29.7	30.0	18.5	22.8	23.8
Tolerance	2.20%	2.62%	3.76%	5.65%	6.11%	5.18%
Exact difference	0.67	0.80	1.17	1.11	1.49	1.30

Table 5.4 – Estimated delay in μs through the message distribution is compared to the measured values for the perturbed scenario.

the maximum message delay estimation to the values measured by Wireshark network analyzer. The first row of Table 5.4 shows the total estimated delay for each flow to reach their destinations. The second row is the measured delay obtained through the same data processing that was explained for the unperturbed scenario in Section 5.3.3.

First, each conversation for a pair of devices is filtered: DT packets from sender and ACK packets from receiver. It is checked if there is any nonrelated packet to the load shedding scenario as MAC Management packets or DT packet from receiver. All the involved conversations are considered at the same time by shifting the time axis at the very beginning of the test-sampled results. Finally, RTT is calculated per bit, and $max(D_{measured}(t))$ is obtained using (5.10).

Now, to compare, tolerance and exact difference of the measured and estimated delay are presented at the third and fourth row respectively. In this case also, all the tolerance values are within an acceptable range (under 10%) and all positive.

5.4 Conclusion

This chapter has presented and discussed the results of all the experiments performed at different stages of this work. Section 5.1 applies the proposed Auto Regressive Integrated Moving Average (ARIMA) modeling process on a grid scenario simulated in MATLAB®/Simulink®. The case of a distribution grid with two renewable energy sources (PV, WT) which exchange reactive power information for the voltage regulation purpose by means of IEC 61850 GOOSE protocol has been considered. Data traffic transmitted from load unit to PV unit is sampled, and modeled through the process explained in Chapter 4. ARIMA model is applied to describe the stochastic behavior of traffic, then the deterministic behavior is added to complete the modeling. Also, a sensitivity study has been performed to determine how the parameters of identified ARIMA model are affected in relation to variation of some communication network parameters.

Section 5.2 has dealt with applying the proposed ARIMA model on a IEC 61850 GOOSE traffic sampled over a real Substation Communication Network (SCN), where different IEDs exchange information in the context of decentralized control. In the considered scenario, it is supposed that the load active power variation is known, and there is an IED sending this information to all other IEDs. This scenario is implemented on a test bench offering the

possibility of real-time data packets analysis. After removing the deterministic components, stochastic behavior was modeled by ARIMA model.

If the communication architecture is known, the proposed model can be generalized from a substation to a grid. Further works may consider the potential of on-line application of proposed identification procedure, for control purposes – possibly in a linear-parameter-varying modeling and control framework – in order to formally take into account interactions between the energy layer and the communication layer in smart grids.

Section 5.3 explains the application of a model for estimating the maximum delay within communication networks in smart grids, proposed by [161] over a real-time IEC 61850 traffic generator. Two experiments were performed: the first one is a scenario where communication is needed for load shedding purposes, whereas in the second one, the same tasks are performed, but the main traffic is perturbed by a background traffic due to the presence of some IEDs not concerned by the load shedding. Message distribution was modeled according to the proposed algorithm, so the maximum message transmission delay is estimated for each flow receiving services at each port. Finally, the estimated values were compared to the effective-delay values measured by a network analyzer, Wireshark, showing acceptable estimation errors (within 10%).

It can be concluded that maximum delay estimation analytical model proposed in [161] gives a good estimation of delay, but it is limited to the cases with completely known network structure – which is the case of proprietary networks – flow characteristics and service providers (in our case Ethernet switches). Considering load shedding as an example of delay estimation scenario, it can be generally noted that any delay present within a control loop is a critical issue. If the load shedding command is delayed, the lower prioritized load cut-out is also delayed so it might result in an electricity cut out of the higher prioritized loads. In general, a delay in a control loop can result in degraded performance of electricity generation. In the worst case, control encountering transmission delay can cause instability in the power grid energy supply. Possibility to estimate the maximum delay is thus very important to be able to determine stability margins in each application case, so to prescribe these delay margins to communication network designers.

General conclusion and perspectives

General conclusion

A custom designed and well monitored substation communication network (SCN) can maintain a fast and reliable information transmission, and lead to an improved control and operation of a substation automation system (SAS). SCN traffic modeling can provide descriptive mathematical models to help changing control paradigm over SAS, and maintain the smart grid operation in a new efficient way. In order to achieve this goal, a data-driven approach of modeling over the SCN data traffic is proposed by this thesis which models the SCN traffic behavior in two parts: stochastic and deterministic. SCN traffic is analyzed as a time-series data, and modeled using ARIMA model based on the Box-Jenkins identification method. From a control approach point of view, the proposed model could serve as a basic mathematical description of a single communication channel, viewed as an elementary brick of the communication network structure within a smart grid.

From the SCN reliability point of view, a real-time message distribution is a critical issue, and delay estimation is an important factor of SCN QoS. Hence, through the second part of this thesis, a new delay estimation algorithm based on an analytical method in [161] is proposed for the SCN traffic, and validated on an intelligent substation test bench, and compared against a real-time traffic measurement provided by a network analyzer. As measured data processing appeared as not being a trivial task, then a new method is proposed for extracting the information which is relevant for communication delay measurement.

The modeling process, and delay estimation validation performed through this thesis work allow emphasizing the following results:

- As a preliminary scenario, a distribution grid of two RESs (*i.e.*, PV and WT), and a load consumption unit is simulated in Simulink®. All the units are equipped with IEDs which communicate with each other using IEC 61850 GOOSE protocol for voltage regulation purposes. Load unit sends a known reactive power profile to the sources to be provided by them through a power compensation algorithm. The stochastic part of traffic behavior is modeled by ARIMA(6,1,4), and validated by a goodness of fit (GFit) of 89.96%, minimum Mean Squared Error (MSE)= 0.4497, and whiteness of the model residuals is confirmed. Next, we have proposed a systematic approach to obtain a mathematical expression of the total traffic behavior including the deterministic part. Finally, some of the communication channel parameters are studied to check if their variation can affect the proposed model parameters: packet loss (PL), channel rate (CHR), and cable length (CL).
- The proposed model is applied over a real reduced-scale distribution grid equipped with a real-time traffic generator. An IED sends the variations of a known active power to all the other IEDs of the network. An ARIMA(6,1,6) model has been estimated with GFit= 82.48%, minimum MSE= 4.73×10^{-8}, and residuals between the confidence levels.

Final results with an acceptable GFit, MSE near zero, and residuals approximately zero confirm that our proposed model properly identifies the stochastic behavior of traffic.

- A load shedding scenario is implemented over a real-time reduced scale distribution grid for purpose of communication delay estimation under representative scenarios. The experience is applied on two communication scenarios, the simple one including main IEDs communicating with the supervision PC, and a perturbed scenario in which two other IEDs out of the load shedding goal provide a background traffic through the network. The estimated delay is compared to the measured values by the Whireshark network packet analyzer. The resulted tolerance for both of the cases is under 10%, which indicates a very good estimation precision, provided that the communication architecture is supposed completely known. A systematic method of processing the information provided by the network analyzer is here proposed, in order to extract that traffic information which is relevant for the communication delay.
- Despite the good estimation for end-to-end delay, the proposed model encounters some drawbacks and limitations:
  - the model proposed in [161] can only be applied on a communication network which is already known, *e.g.*, the specifications of the switches, network routing table, *etc.*;
  - some of the principal matrices should be defined manually at the beginning of the process, so automation of the algorithm appears to be quite difficult.

Perspectives

The results obtained within this work suggest a number of short-term and long-term perspectives.

Short-term perspectives

- Further works on the ARIMA IEC 61850 GOOSE traffic model proposed by this thesis will consider the potential of the on-line application of the proposed identification procedure, for control purposes, possibly in a linear-parameter-varying framework. To analyze the interactions between the energy and the communication grid, communication channel parameters, *e.g.*, Packet Loss (PL), transmission delay, *etc.*, can be variable over time. So, using a window function over time, the influence of parameters can be analyzed for different sampled data while testing.
- The ARIMA model that we proposed through this work is derived based on the assumption of a stable power grid without any fault, so the SCN data type is almost periodic. Although this model is applicable irrespective of the SCN data types (*i.e.*, random, burst, periodic), it is still necessary to be applied on the SCN traffic when a fault occurs (*i.e.*, unstable power grid where burst and random data interfere). In this way, the proposed model is verified on all the SCN data types and checked for its accuracy.
- Here, our proposed model is applied over smart-grid scenarios involving a single communication channel. More general and complex communication network configuration should be studied in which there are various types of IEDs, generators, consumers and for more complex goals.
- In the context of SCN traffic modeling, delay is estimated over a series of pre-sampled data, *i.e.*, off-line. One can then imagine that delay to be estimated on-line through a time window of width T seconds at any time t by applying the proposed identification procedure at each time window, provided that the required computation time is less than T. Then, variation of the estimated delay could be traced as a function of time. Thus, it is possible to monitor the transmission delay through this visualization, then analyze it under different communication status and possibly use it within machine-learning algorithms. Such information can be used to find an optimal configuration for the communication network from the point of view of the transmission delay in order to guarantee the grid operation stability under an as larger as possible set of energy-communication interaction scenario.

Long-term perspectives

- Proposed ARIMA model can be further applied on a scenario in which IEDs operate in a plug-and-play communication network, where IEDs can be freely added or removed from the network at any time. Thus, it is interesting to see how ARIMA model is applicable, and if it is reliable.
- To study the interactions between power and communication network, the proposed model can be applied on the SCN traffic in an unstable operation of the power grid. This would further allow how a fault occurrence in the energy layer affects communication model pa-

rameters.

- While monitoring information, and control signals are being transmitted, any experienced delay can cause instability or at least degrading the functionality of a co-simulation implemented over SCN. In this thesis, communication delay estimation has been studied inside of a single test bench based on the IEC 61850 protocols. A future interesting issue is to consider the variation of the communication delay in an interoperability scenario, that is, between two test benches located on distant sites, while communicating through Internet. Thus, it would be possible to determine the level of compensation to be implemented over the entire chain of delays to ensure the consistency of real-time coupled co-simulations. An example of a network set-up in this case is illustrated in Fig. 5.35.

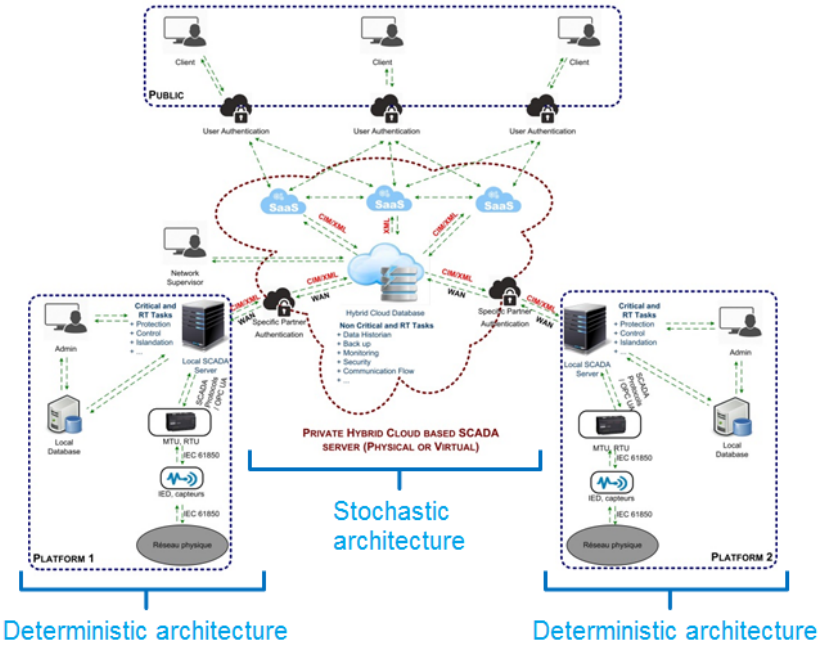


Figure 5.35 – Network architecture of two intelligent substations communicating at the distance.

Bibliography

[1] IEC TC 57. "International Standard IEC 61850-7-1: Communication networks and systems in substations Part 7-1: Basic communication structure for substation and feeder equipment Principles and models." In: 2003 (cit. on p. 24).

[2] IEC TC 57. "International Standard IEC 61850-7-4: Communication networks and systems in substations Part 7-4: Basic communication structure for substation and feeder equipment Compatible logical node classes and data classes." In: 2003 (cit. on p. 25).

[3] S. Al-Areqi, D. Grges, and S. Liu. "Stochastic event-based control and scheduling of large-scale networked control systems." In: *2014 European Control Conference (ECC), Strasbourg, 2014*, pp. 2316–2321. DOI: 10.1109/ECC.2014.6862283 (cit. on p. 37).

[4] Obada Al-Khatib, Wibowo Hardjawana, and Branka Vucetic. "Traffic modeling and performance evaluation of wireless smart grid access networks." In: *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), , Washington, DC*. IEEE. 2014, pp. 1876–1881. DOI: 10.1109/PIMRC.2014.7136476 (cit. on p. 44).

[5] Cristina Alcaraz, Javier Lopez, and Stephen Wolthusen. "Policy enforcement system for secure interoperable control in distributed Smart Grid systems." In: *Journal of Network and Computer Applications* 59 (2016), pp. 301–314. DOI: 10.1016/j.jnca.2015.05.023 (cit. on p. 38).

[6] Ikbal Ali et al. "IEC 61850 substation communication network architecture for efficient energy system automation." In: *Energy Technology & Policy* 2.1 (2015), pp. 82–91. DOI: 10.1080/23317000.2015.1043475 (cit. on p. 43).

[7] Ahmed Altaher. "Implementation of a dependability framework for smart substation automation systems : application to electric energy distribution." Theses. Université Grenoble Alpes, Feb. 2018 (cit. on pp. 9, 17, 19, 36).

[8] Filip Andrén, Roland Bründlinger, and Thomas Strasser. "IEC 61850/61499 control of distributed energy resources: Concept, guidelines, and implementation." In: *IEEE Transactions on Energy Conversion* 29.4 (2014), pp. 1008–1017 (cit. on pp. 29, 36).

[9] Nirwan Ansari and Chun-Hao Lo. "Decentralized controls and communications for autonomous distribution networks in smart grid." In: 4 (2013). US Patent, pp. 66–77. DOI: 10.1109/TSG.2012.2228282 (cit. on p. 11).

[10] Alexander Apostolov. "To GOOSE or not to GOOSE? - That is the question." In: *2015 68th Annual Conference for Protective Relay Engineers, CPRE 2015* (2015), pp. 583–596. DOI: 10.1109/CPRE.2015.7102196 (cit. on p. 20).

[11] Remo Badii and Antonio Politi. *Complexity: Hierarchical structures and scaling in physics*. Vol. 6. Cambridge University Press, 1999 (cit. on p. 60).

[12] David E Bakken, Richard E Schantz, and Richard D Tucker. "Smart grid communications: Qos stovepipes or qos interoperability?" In: *Proceedings of Grid-Interop 2009* (2009), pp. 17–19 (cit. on p. 34).

[13] Neeru Bansal, VK Srivastava, and Juzer Kheraluwala. "Rooftop Solar Power Generation: An Opportunity to Reduce Greenhouse Gas Emissions." In: *Greenhouse Gas Emissions*. Springer, 2019, pp. 141–160. DOI: 10.1007/978-981-13-3272-2_10 (cit. on p. 23).

[14] Yvon Bésanger and Tuan Tran-Quoc. "Voltage Control in Distribution Systems with Dispersed Generation." In: *Electrical Distribution Networks* (2014), pp. 237–272. DOI: 10.1109/IECON.2011.6119329 (cit. on p. 100).

[15] Dhananjay Bhor, Kavinkadhirselvan Angappan, and Krishna M Sivalingam. "A co-simulation framework for smart grid wide-area monitoring networks." In: *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), Bangalore*. IEEE. 2014, pp. 1–8. DOI: 10.1109/COMSNETS.2014.6734880 (cit. on pp. 2, 37, 41).

[16] D Bian et al. "Real-time co-simulation platform using OPAL-RT and OPNET for analyzing smart grid performance." In: *2015 IEEE Power & Energy Society General Meeting, Denver*. IEEE. 2015, pp. 1–5. DOI: 10.1109/PESGM.2015.7286238 (cit. on pp. 2, 42, 43).

[17] Frede Blaabjerg and Dan M Ionel. "Renewable energy devices and systems–state-of-the-art technology, research and development, challenges and future trends." In: *Electric Power Components and Systems* 43.12 (2015), pp. 1319–1328. DOI: 10.1080/15325008.2015.1062819 (cit. on p. 21).

[18] R. Bottura et al. "SITL and HLA co-simulation platforms: Tools for analysis of the integrated ICT and electric power system." In: *Eurocon 2013*. 2013, pp. 918–925. DOI: 10.1109/EUROCON.2013.6625092 (cit. on p. 40).

[19] Riccardo Bottura et al. "ICT-power co-simulation platform for the analysis of communication-based volt/var optimization in distribution feeders." In: *ISGT 2014*. IEEE. 2014, pp. 1–5. DOI: 10.1109/ISGT.2014.6816425 (cit. on p. 42).

[20] Faycal Bouhafs, Michael Mackay, and Madjid Merabti. "Links to the future: Communication requirements and challenges in the smart grid." In: *IEEE Power and Energy Magazine* 10.1 (2011), pp. 24–32. DOI: 10.1109/MPE.2011.943134 (cit. on p. 22).

[21] Greta Box, George, Jenkins, Gwilym, Reinsel, Gregory, Ljung. *Time series analisys: forecasting and control*. 2016. DOI: 10.1057/jors.1971.52 (cit. on pp. 62, 65, 74–76, 80).

[22] Janet M Box-Steffensmeier and Christopher Zorn. "Duration models for repeated events." In: *Journal of Politics* 64.4 (2002), pp. 1069–1094. DOI: 10.1111/1468-2508.00163 (cit. on p. 60).

[23] Klaus-Peter Brand, Volker Lohmann, and Wolfgang Wimmer. *Substation automation handbook*. Utility Automation Consulting Lohmann Bremgarten, Switzerland, 2003 (cit. on p. 17).

[24] Damien Breck, Jean-Philippe Georges, and Thierry Divoux. "Évaluation des performances de trafics agrégés 802.11n." In: *5e Journées Doctorales / Journées Nationales MACS, JD-JN-MACS 2013*. Strasbourg, France, July 2013, CDROM (cit. on p. 88).

[25] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting*. 2016 (cit. on pp. 62, 75, 78, 79).

[26] Ziyuan Cai et al. "A network model for the real-time communications of a smart grid prototype." In: *Journal of Network and Computer Applications* 59 (2016), pp. 264–273 (cit. on pp. 2, 37, 41, 59).

[27] Pascal Charriau, Bertrand Château, and Bruno Lapillonne. *Global Energy Statistical Yearbook*. 2019 (cit. on p. 8).

[28] Mehmet H Cintuglu, Tan Ma, and Osama A Mohammed. "Protection of autonomous microgrids using agent-based distributed communication." In: *IEEE Transactions on Power Delivery* 32.1 (2016), pp. 351–360. DOI: 10.1109/TPWRD.2016.2551368 (cit. on p. 40).

[29] Richard G. Clegg. "A practical guide to measuring the hurst parameter." In: *International Journal of Simulation: Systems, Science and Technology* 7.2 (2006), pp. 3–14 (cit. on p. 57).

[30] Price Code. "Communication networks and systems in substations–Part 8-1: Specific Communication Service Mapping (SCSM)–Mappings to MMS (ISO 9506-1 and ISO 9506-2)." In: (2004) (cit. on p. 56).

[31] "Coordinated Control of Distributed Energy-Storage Systems for Voltage Regulation in Distribution Networks." In: *IEEE Transactions on Power Delivery* 31.3 (2016), pp. 1132–1141. DOI: 10.1109/TPWRD.2015.2462723 (cit. on p. 101).

[32] R L Cruz. "A calculus for network delay, Part II: Network analysis." In: *IEEE Transactions on Information Theory* I (), pp. 132–141 (cit. on p. 88).

[33] Rene L Cruz et al. "A calculus for network delay, part I: Network elements in isolation." In: *IEEE Transactions on information theory* 37.1 (1991), pp. 114–131. DOI: 10.1109/18.61109 (cit. on pp. 68, 88).

[34] Gregorio D'Agostino and Antonio Scala. *Networks of networks: the last frontier of complexity*. Vol. 340. Springer, 2014. DOI: 10.1007/978-3-319-03518-5_13 (cit. on p. 39).

[35] Narottam Das, Wu Ma, and Syed Islam. "Comparison study of various factors affecting end-to-end delay in IEC 61850 substation communications using OPNET." In: *2012 22nd Australasian Universities Power Engineering Conference (AUPEC)*. IEEE. 2012, pp. 1–5 (cit. on p. 43).

[36] Jan G De Gooijer and Rob J Hyndman. "25 years of time series forecasting." In: *International Journal of Forecasting* 22.3 (2006), pp. 443–473. DOI: 10.1016/j.ijforecast.2006.01.001 (cit. on p. 62).

[37] Ruilong Deng et al. "A survey on demand response in smart grids: Mathematical models and approaches." In: *IEEE Transactions on Industrial Informatics* 11.3 (2015), pp. 570–582. DOI: 10.1109/TII.2015.2414719 (cit. on p. 21).

[38] AR Di Fazio et al. "Integration of renewable energy sources, energy storage systems, and electrical vehicles with smart power distribution networks." In: *Journal of Ambient Intelligence and Humanized Computing* 4.6 (2013), pp. 663–671. DOI: 10.1007/s12652-013-0182-y (cit. on p. 38).

[39] Dave Dolezilek. "IEC 61850: What you need to know about functionality and practical implementation." In: *2006 Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources*. IEEE. 2006, pp. 1–17. DOI: 10.1109/PSAMP.2006.285368 (cit. on pp. 18, 20).

[40] Göran N Ericsson. "Cyber security and power system communicationessential parts of a smart grid infrastructure." In: *IEEE Transactions on Power Delivery* 25.3 (2010), pp. 1501–1507. DOI: 10.1109/TPWRD.2010.2046654 (cit. on p. 16).

[41] Göran N Ericsson. "Management of information security for an electric power Utility-on security domains and use of ISO/IEC17799 standard." In: *IEEE transactions on power delivery* 20.2 (2005), pp. 683–690. DOI: 10.1109/TPWRD.2005.844318 (cit. on p. 16).

[42] Melike Erol-Kantarci and Hussein T Mouftah. "Wireless multimedia sensor and actor networks for the next generation power grid." In: *Ad Hoc Networks* 9.4 (2011), pp. 542–551. DOI: 10.1016/j.adhoc.2010.08.005 (cit. on p. 9).

[43] Bamdad Falahati and Yong Fu. "A study on interdependencies of cyber-power networks in smart grid applications." In: *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*. IEEE. 2012, pp. 1–8. DOI: 10.1109/ISGT.2012.6175593 (cit. on pp. 31, 32).

[44] Kaijun Fan et al. "Fast peer-to-peer real-time data transmission for distributed control of distribution network." In: *2014 China International Conference on Electricity Distribution (CICED)*. IEEE. 2014, pp. 1041–1045. DOI: 10.1109/CICED.2014.6991864 (cit. on p. 54).

[45] Zhong Fan et al. "Smart grid communications: Overview of research challenges, solutions, and standardization activities." In: *IEEE Communications Surveys & Tutorials* 15.1 (2012), pp. 21–38. DOI: 10.1109/SURV.2011.122211.00021 (cit. on p. 9).

[46] Xi Fang et al. "Smart grid – The new and improved power grid: A survey." In: *IEEE communications surveys & tutorials* 14.4 (2011), pp. 944–980. DOI: 10.1109/SURV.2011.101911.00087 (cit. on p. 9).

[47] Julian J Faraway. *Practical regression and ANOVA using R*. 2002 (cit. on p. 60).

[48] Ronak Feizimirkhani, Antoneta Iuliana Bratcu, and Yvon Besanger. "Time-series modelling of IEC 61850 GOOSE communication traffic between IEDs in smart grids–a parametric analysis." In: *10th Symposium on Control of Power and Energy Systems (IFAC CPES2018)*. 2018. DOI: 10.1016/j.ifacol.2018.11.743 (cit. on p. 3).

[49] Ronak Feizimirkhani et al. "Modeling of IEC 61850 GOOSE Substation Communication Traffic Using ARMA Model." In: 2019. DOI: 10.1109/ISGTEurope.2019.8905634 (cit. on p. 3).

[50] Chilton Fernandes, Samarth Borkar, and Jignesh Gohil. "Testing of GOOSE protocol of IEC61850 standard in protection IED." In: *International Journal of Computer Applications* 93.16 (2014). DOI: 10.5120/16301-6112 (cit. on p. 53).

[51] Paul Feron et al. "Towards Zero Emissions from Fossil Fuel Power Stations." In: *International Journal of Greenhouse Gas Control* 87 (2019), pp. 188–202. DOI: 10.1016/j.ijggc.2019.05.018 (cit. on p. 23).

[52] A Behrouz Forouzan. *Data communications & networking (sie)*. Tata McGraw-Hill Education, 2007 (cit. on p. 53).

[53] Michele Garau et al. "Co-simulation of smart distribution network fault management and reconfiguration with lte communication." In: *Energies* 11.6 (2018), p. 1332. DOI: 10.1109/MWC.2017.1600214 (cit. on p. 40).

[54] Michele Garau et al. "Evaluation of smart grid communication technologies with a co-simulation platform." In: *IEEE Wireless Communications* 24.2 (2017), pp. 42–49. DOI: 10.3390/en11061332 (cit. on p. 40).

[55] Hanno Georg et al. "Performance evaluation of time-critical communication networks for smart grids based on IEC 61850." In: *2013 Proceedings IEEE INFOCOM, Turin*. IEEE. 2013, pp. 3417–3422. DOI: 10.1109/INFCOM.2013.6567174 (cit. on pp. 45, 46).

[56] Jean-Philippe Georges et al. "A design process of switched Ethernet architectures according to real-time application constraints." In: *Engineering Applications of Artificial Intelligence* 19.3 (2006), pp. 335 –344. DOI: 10.1016/j.engappai.2005.09.004 (cit. on p. 88).

[57] J Willard Gibbs. *Elementary principles in statistical mechanics*. Courier Corporation, 2014, pp. 291–292. DOI: 10.1038/066291a0 (cit. on p. 61).

[58] Christopher Greer et al. *Nist framework and roadmap for smart grid interoperability standards, release 3.0*. Tech. rep. 2014. DOI: 10.6028/NIST.SP.1108r3 (cit. on p. 13).

[59] V. Cagri Gungor et al. "A Survey on smart grid potential applications and communication requirements." In: *IEEE Transactions on Industrial Informatics* 9.1 (2013), pp. 28–42. DOI: 10.1109/TII.2012.2218253. arXiv: arXiv:1011.1669v3 (cit. on p. 17).

[60] Gilbert Habib, Thierry Divoux, and Jean-Franois Ptin. "Estimating maximum and minimum delays for wireless discrete networked control systems." In: May 2009, pp. 1 –5. DOI: 10.1109/WTS.2009.5068966 (cit. on p. 39).

[61] Jiawei Han, Micheline Kamber, and Anthony KH Tung. "Spatial clustering methods in data mining." In: *Geographic data mining and knowledge discovery* (2001), pp. 188–217 (cit. on p. 60).

[62] Frank E Harrell Jr. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer International Publishing Switzerland 2015, 2015. DOI: 10.1007/978-3-319-19425-7 (cit. on p. 60).

[63] CC Heyde. "Agner Krarup Erlang." In: *Statisticians of the Centuries*. Springer, New York, NY, 2001, pp. 328–330. DOI: 10.1007/978-1-4613-0179-0_70 (cit. on p. 58).

[64] Nicholas Honeth et al. "Development of the IEC 61850 software merging unit IED test and training platform." In: June 2013, pp. 1–6. DOI: 10.1109/PTC.2013.6652385 (cit. on p. 88).

[65] Kenneth Hopkinson et al. "EPOCHS: a platform for agent-based electric power and communication simulation built from commercial off-the-shelf components." In: *IEEE Transactions on Power Systems* 21.2 (2006), pp. 548–558. DOI: 10.1109/TPWRS.2006.873129 (cit. on p. 39).

[66] Wei Huang. "Learn IEC 61850 configuration in 30 minutes." In: *2018 71st Annual Conference for Protective Relay Engineers (CPRE)*. IEEE. 2018, pp. 1–5. DOI: 10.1109/CPRE.2018.8349803 (cit. on pp. 25, 27).

[67] Akhtar Hussain, Syed Muhammad Arif, and Muhammad Aslam. "Emerging renewable and sustainable energy technologies: State of the art." In: *Renewable and Sustainable Energy Reviews* 71 (2017), pp. 12–28. DOI: 10.1016/j.rser.2016.12.033 (cit. on p. 23).

[68] Robert J Marks Ii. *Introduction to Shannon Sampling and Interpolation Theory*. 1991 (cit. on pp. 76, 102).

[69] David ME Ingram, Pascal Schaub, and Duncan A Campbell. "Multicast traffic filtering for sampled value process bus networks." In: *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2011, pp. 4710–4715. DOI: 10.1109/IECON.2011.6120087 (cit. on p. 43).

[70] David ME Ingram et al. "Direct evaluation of IEC 61850-9-2 process bus network performance." In: *IEEE Transactions on Smart Grid* 3.4 (2012), pp. 1853–1854. DOI: 10.1109/TSG.2012.2205637 (cit. on pp. 43, 68).

[71] David ME Ingram et al. "Performance analysis of IEC 61850 sampled value process bus networks." In: *IEEE Transactions on industrial informatics* 9.3 (2012), pp. 1445–1454. DOI: 10.1109/TII.2012.2228874 (cit. on p. 54).

[72] David ME Ingram et al. "Performance analysis of PTP components for IEC 61850 process bus applications." In: *IEEE Transactions on Instrumentation and Measurement* 62.4 (2013), pp. 710–719. DOI: 10.1109/TIM.2013.2245188 (cit. on p. 43).

[73] Ali Ipakchi and Farrokh Albuyeh. "Grid of the future." In: *IEEE power and energy magazine* 7.2 (2009), pp. 52–62. DOI: 10.1109/MPE.2008.931384 (cit. on p. 9).

[74] Manar Jaradat et al. "Integration of renewable energy in demand-side management for home appliances." In: *2014 International Renewable and Sustainable Energy Conference (IRSEC)*. IEEE. 2014, pp. 571–576. DOI: 10.1109/IRSEC.2014.7059843 (cit. on p. 23).

[75] Moath Jarrah et al. "A hierarchical optimization model for energy data flow in smart grid power systems." In: *Information Systems* 53 (2015), pp. 190–200 (cit. on p. 23).

[76] Andreas Noack Jensen and Morten Ørregaard Nielsen. "A fast fractional difference algorithm." In: *Journal of Time Series Analysis* 35.5 (2014), pp. 428–436. DOI: 10.1111/jtsa.12074 (cit. on p. 67).

[77] Paul L Joskow and John E Parsons. "The future of nuclear power after Fukushima." In: *Economics of Energy & Environmental Policy* 1.2 (2012), pp. 99–114. DOI: 10.5547/2160-5890.1.2.7 (cit. on p. 23).

[78] Yasin Kabalci. "A survey on smart metering and smart grid communication." In: *Renewable and Sustainable Energy Reviews* 57 (2016), pp. 302–318 (cit. on p. 10).

[79] Prajakta S Kalekar. "Time series forecasting using Holt-Winters exponential smoothing." In: *Kanwal Rekhi School of Information Technology* 4329008.13 (2004). DOI: 10.1063/1.5137999 (cit. on p. 63).

[80] Mitalkumar G Kanabar and Tarlochan S Sidhu. "Performance of IEC 61850-9-2 process bus and corrective measure for digital relaying." In: *IEEE Transactions on Power Delivery* 26.2 (2010), pp. 725–735. DOI: 10.1109/TPWRD.2009.2038702 (cit. on p. 43).

[81] Mitalkumar G Kanabar and Tarlochan S Sidhu. "Reliability and availability analysis of IEC 61850 based substation communication architectures." In: *2009 IEEE Power & Energy Society General Meeting*. IEEE. 2009, pp. 1–8. DOI: 10.1109/PES.2009.5276001 (cit. on p. 43).

[82] Palak M Kanabar et al. "Evaluation of communication technologies for IEC 61850 based distribution automation system with distributed energy resources." In: *2009 IEEE Power & Energy Society General Meeting*. IEEE. 2009, pp. 1–8. DOI: 10.1109/PES.2009.5275787 (cit. on p. 44).

[83] Hristiyan Kanchev et al. "Energy management and operational planning of a microgrid with a PV-based active generator for smart grid applications." In: *IEEE Transactions on Industrial Electronics* 58.10 (2011), pp. 4583–4592. DOI: 10.1109/TIE.2011.2119451 (cit. on p. 21).

[84] Kathryn P Huyvaert Kenneth P Burnham David R Anderson. "AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons." In: *Second. NY: Springer-Verlag* 65 (2011), pp. 23–35. DOI: 10.1007/s00265-010-1029-6 (cit. on p. 81).

[85] Reduan H Khan and Jamil Y Khan. "A comprehensive review of the application characteristics and traffic requirements of a smart grid communications network." In: *Computer Networks* 57.3 (2013), pp. 825–845 (cit. on p. 8).

[86] Iman Kiaei and Saeed Lotfifard. "Fault Section Identification in Smart Distribution Systems Using Multi-source Data Based on Fuzzy Petri Nets." In: *IEEE Transactions on Smart Grid* (2019). DOI: 10.1109/TSG.2019.2917506 (cit. on p. 21).

[87] J. Knig et al. "An extended framework for reliability analysis of ict for power systems." In: *2011 IEEE Trondheim PowerTech*. 2011, pp. 1–6. DOI: 10.1109/PTC.2011.6019343 (cit. on p. 2).

[88] Richard M Kolacinski and Kenneth A Loparo. "A mathematic framework for analysis of complex cyber-physical power systems." In: *2012 IEEE Power and Energy Society General Meeting*. IEEE. 2012, pp. 1–8 (cit. on pp. 44, 61).

[89] J. Konig et al. "Mapping the Substation Configuration Language of IEC 61850 to ArchiMate." In: *2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops*. 2010, pp. 60–68. DOI: 10.1109/EDOCW.2010.35 (cit. on p. 32).

[90] Johan Konig, Lars Nordstrom, and Magnus Osterlind. "Reliability analysis of substation automation system functions using PRMs." In: *IEEE Transactions on Smart Grid* 4.1 (2013), pp. 206–213. DOI: 10.1109/TSG.2012.2225452 (cit. on p. 43).

[91] Jakub W Konka et al. "Traffic generation of IEC 61850 sampled values." In: *2011 IEEE First International Workshop on Smart Grid Modeling and Simulation (SGMS)*. IEEE. 2011, pp. 43–48. DOI: 10.1109/SGMS.2011.6089025 (cit. on p. 45).

[92] Iordanis Koutsopoulos and Leandros Tassiulas. "Control and optimization meet the smart power grid: Scheduling of power demands for optimal energy management." In: *Proceedings of the 2nd International Conference on Energy-efficient Computing and Networking*. Vol. 57. ACM. 2011, pp. 41–50. DOI: 10.1145/2318716.2318723 (cit. on p. 23).

[93] M Kuzlu, M Pipattanasomporn, and S Rahman. "Review of communication technologies for smart homes/building applications." In: *2015 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA)*. IEEE. 2015, pp. 1–6. DOI: 10.1109/ISGT-Asia.2015.7437036 (cit. on p. 38).

[94] Quang Linh Lam. "Advanced control of microgrids for frequency and voltage stability : robust control co-design and real-time validation." Ph.D thesis. Université Grenoble Alpes, Jan. 2018 (cit. on p. 22).

[95] Ioan Doré Landau and Gianluca Zito. *Digital control systems: design, identification and implementation*. Springer Science & Business Media, 2007 (cit. on p. 82).

[96] Jean-Yves Le Boudec and Patrick Thiran. "Network Calculus a Theory of Deterministic Queuing Systems for the Internet." In: *Online* 2050 (2004), pp. xix –274. DOI: 10.1007/3-540-45318-0. arXiv: 9780201398298 (cit. on pp. 68–71, 88).

[97] Fangxing Li et al. "Smart transmission grid: Vision and framework." In: *IEEE transactions on Smart Grid* 1.2 (2010), pp. 168–177 (cit. on p. 8).

[98] Qilin Li and Mingtian Zhou. "The future-oriented grid-smart grid." In: *Journal of Computers* 6.1 (2011), pp. 98–105. DOI: 10.4304/jcp.6.1.98-105 (cit. on p. 9).

[99] Guodong Liao et al. "A simulation study on the ethernet communication of a substation automation system based on EPOCHS." In: *2006 International Conference on Power System Technology, Chongqing*. IEEE. 2006, pp. 1–7. DOI: 10.1109/ICPST.2006.321691 (cit. on p. 43).

[100] Jörg Liebeherr et al. "Duality of the max-plus and min-plus network calculus." In: *Foundations and Trends® in Networking* 11.3-4 (2017), pp. 139–282. DOI: 10.1561/1300000059 (cit. on p. 69).

[101] Hua Lin et al. "Cyber security impacts on all-PMU state estimator-a case study on co-simulation platform GECO." In: *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*. IEEE. 2012, pp. 587–592. DOI: 10.1109/SmartGridComm.2012.6486049 (cit. on p. 2).

[102] Hua Lin et al. "GECO: Global event-driven co-simulation framework for interconnected power system and communication network." In: *IEEE Transactions on Smart Grid* 3.3 (2012), pp. 1444–1456. DOI: 10.1109/TSG.2012.2191805 (cit. on p. 39).

[103] Douglas Lind et al. *An introduction to symbolic dynamics and coding*. Cambridge university press, 1995 (cit. on p. 61).

[104] Xiaosheng Liu et al. "A high-reliability and determinacy architecture for smart substation process-level network based on cobweb topology." In: *IEEE Transactions on Power Delivery* 29.2 (2014), pp. 842–850. DOI: 10.1109/TPWRD.2013.2280763 (cit. on p. 43).

[105] Francisco Lobo et al. "Distribution network as communication system." In: *CIRED Seminar 2008: SmartGrids for Distribution*. IET. 2008, pp. 1–4. DOI: 10.1049/ic:20080448 (cit. on p. 13).

[106] JA Pecas Lopes et al. "Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities." In: *Electric power systems research* 77.9 (2007), pp. 1189–1203. DOI: 10.1016/j.epsr.2006.08.016 (cit. on p. 9).

[107] Manolis IA Lourakis et al. "A brief description of the Levenberg-Marquardt algorithm implemented by levmar." In: *Foundation of Research and Technology* 4.1 (2005), pp. 1–6 (cit. on p. 63).

[108] T Lovett et al. "A multilanguage environment for interactive simulation and development controls for power electronics." In: *2001 IEEE 32nd Annual Power Electronics Specialists Conference (IEEE Cat. No. 01CH37230)*. Vol. 3. IEEE. 2001, pp. 1725–1729. DOI: 10.1109/PESC.2001.954368 (cit. on p. 14).

[109] Wei Lu. "Le délestage optimal pour la prévention des grandes pannes d'électricité." PhD thesis. 2009 (cit. on p. 15).

[110] Wenpeng Luan, Duncan Sharp, and Sol Lancashire. "Smart grid communication network capacity planning for power utilities." In: *IEEE PES T&D 2010*. IEEE. 2010, pp. 1–4. DOI: 10.1109/TDC.2010.5484223 (cit. on p. 17).

[111] Ralph E Mackiewicz. "Overview of IEC 61850 and Benefits." In: *2006 IEEE Power Engineering Society General Meeting*. IEEE. 2006, pp. 623–630. DOI: 10.1109/PSCE.2006.296392 (cit. on p. 18).

[112] Alireza Majzoobi and Amin Khodaei. "Application of microgrids in supporting distribution grid flexibility." In: *IEEE Transactions on Power Systems* 32.5 (2016), pp. 3660–3669. DOI: 10.1109/TPWRS.2016.2635024 (cit. on p. 9).

[113] Moein Manbachi et al. "Real-time communication platform for Smart Grid adaptive Volt-VAR Optimization of distribution networks." In: *International Conference on Smart Energy Grid Engineering, SEGE 2015* (2015). DOI: 10.1109/SEGE.2015.7324592 (cit. on p. 100).

[114] Mihalis G Markakis, Eytan H Modiano, and John N Tsitsiklis. "Max-Weight Scheduling in Networks with Heavy-Tailed Traffic." In: 22.1 (2012), pp. 2318–2326. DOI: 10.1109/TNET.2013.2246869 (cit. on p. 66).

[115] John McDonald. "Substation Automation Basics-The Next Generation." In: *Electric Energy T&D Magazine* (2007) (cit. on p. 24).

[116] Geoffrey McLachlan. *Discriminant analysis and statistical pattern recognition*. Vol. 544. John Wiley & Sons, 2004 (cit. on p. 60).

[117] Kevin Mets et al. "Integrated simulation of power and communication networks for smart grid applications." In: *2011 IEEE 16th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE. 2011, pp. 61–65. DOI: 10.1109/CAMAD.2011.5941119 (cit. on p. 40).

[118] S Mohagheghi, J Stoupis, and Z Wang. "Communication protocols and networks for power systems-current status and future trends." In: *2009 IEEE/PES Power Systems Conference and Exposition*. IEEE. 2009, pp. 1–9. DOI: 10.1109/PSCE.2009.4840174 (cit. on pp. 17, 18).

[119] Amir-Hamed Mohsenian-Rad and Alberto Leon-Garcia. "Optimal residential load control with price prediction in real-time electricity pricing environments." In: *IEEE Trans. Smart Grid* 1.2 (2010), pp. 120–133. DOI: 10.1109/TSG.2010.2055903 (cit. on p. 23).

[120] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015 (cit. on p. 56).

[121] Khosrow Moslehi, Ranjit Kumar, et al. "A reliability perspective of the smart grid." In: *IEEE Trans. Smart Grid* 1.1 (2010), pp. 57–64. DOI: 10.1109/TSG.2010.2046346 (cit. on p. 15).

[122] Sven Christian Müller et al. "Interfacing power system and ict simulators: Challenges, state-of-the-art, and case studies." In: *IEEE Transactions on Smart Grid* 9.1 (2016), pp. 14–24. DOI: 10.1109/TSG.2016.2542824 (cit. on pp. 1, 40).

[123] Ni Ding. "Load models for operation and planning of electricity distribution networks with smart metering data." PhD thesis. 2017 (cit. on p. 62).

[124] Joseph A Olsen and David A Kenny. "Structural equation modeling with interchangeable dyads." In: *Psychological methods* 11.2 (2006), p. 127. DOI: 10.1037/1082-989X.11.2.127 (cit. on p. 60).

[125] Alan Pankratz. *Forecasting with Univariate Box-Jenkins Models*. 1983. DOI: 10.1002/9780470316566 (cit. on pp. 61, 75).

[126] Palak P Parikh, Tarlochan S Sidhu, and Abdallah Shami. "A comprehensive investigation of wireless LAN for IEC 61850–based smart distribution substation applications." In: *IEEE Transactions on Industrial Informatics* 9.3 (2012), pp. 1466–1476. DOI: 10.1109/TII.2012.2223225 (cit. on p. 43).

[127] Sumit Paudyal, Claudio A Canizares, and Kankar Bhattacharya. "Optimal operation of distribution feeders in smart grids." In: *IEEE Transactions on Industrial Electronics* 58.10 (2011), pp. 4495–4503. DOI: 10.1109/TIE.2011.2112314 (cit. on p. 10).

[128] Thijs Peirelinck, A. Bratcu, and Y. Besenger. "Modelling of the interaction between the energy layer and communication layer in smart energy microgrids." In: (2016) (cit. on pp. 3, 42, 51, 53, 100, 101).

[129] Guido Pepermans et al. "Distributed generation: definition, benefits and issues." In: *Energy policy* 33.6 (2005), pp. 787–798. DOI: 10.1016/j.enpol.2003.10.004 (cit. on p. 9).

[130] Ferdinanda Ponci, Enrico Santi, and Antonello Monti. "Discrete-time multi-resolution modeling of switching power converters using wavelets." In: *Simulation* 85.2 (2009), pp. 69–88. DOI: 10.1177/0037549708100412 (cit. on p. 14).

[131] IEEE PSRC. "Protection Systems Relay Committee H6 SPECIAL REPORT Application Considerations of IEC 61850." In: 2005 (cit. on pp. 18, 20).

[132] Farrokh Rahimi and Ali Ipakchi. "Overview of demand response under the smart grid and market paradigms." In: *2010 Innovative Smart Grid Technologies (ISGT)*. IEEE. 2010, pp. 1–7. DOI: 10.1109/ISGT.2010.5434754 (cit. on p. 12).

[133] Olivier Richardot et al. "Coordinated voltage control in distribution networks using distributed generation." In: *Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES*. Vol. 1. 2006, pp. 1196–1201. DOI: 10.1109/TDC.2006.1668675 (cit. on p. 100).

[134] Danilo Antonio Sbordone et al. "Reactive power control for an energy storage system: A real implementation in a Micro-Grid." In: *Journal of Network and Computer Applications* 59 (2016), pp. 250–263. DOI: 10.1016/j.jnca.2015.05.006 (cit. on p. 37).

[135] Rakibuzzaman Shah et al. "A review of key power system stability challenges for large-scale PV integration." In: *Renewable and Sustainable Energy Reviews* 41 (2015), pp. 1423–1436. DOI: 10.1016/j.rser.2014.09.027 (cit. on p. 21).

[136] Mehdi Shamshirband, Javad Salehi, and Farhad Samadi Gazijahani. "Decentralized trading of plug-in electric vehicle aggregation agents for optimal energy management of smart renewable penetrated microgrids with the aim of CO2 emission reduction." In: *Journal of Cleaner Production* 200 (2018), pp. 622–640. DOI: 10.1016/j.jclepro.2018.07.315 (cit. on p. 23).

[137] EKANKI SHARMA, A. Bratcu, and Y. Besanger. "Cyber-physical modelling of energy microgrids for decentralized and cooperative control." In: (2015) (cit. on pp. 51, 52, 101).

[138] Baiqiang Shen et al. "Reasch on IED configurator automatic modeling based on IEC 61850." In: *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. IEEE. 2017, pp. 1–5. DOI: 10.1109/EI2.2017.8245427 (cit. on p. 24).

[139] Tarlochan S Sidhu and Yujie Yin. "Modelling and simulation for performance evaluation of IEC61850-based substation communication systems." In: *IEEE transactions on power delivery* 22.3 (2007), pp. 1482–1489. DOI: 10.1109/TPWRD.2006.886788 (cit. on pp. 41, 44).

[140] TS Sidhu and Pradeep K Gangadharan. "Control and automation of power system substation using IEC61850 communication." In: *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005.* IEEE. 2005, pp. 1331–1336. DOI: 10.1109/CCA.2005.1507316 (cit. on pp. 21, 26).

[141] TS Sidhu and Yujie Yin. "IED modelling for IEC61850 based substation automation system performance simulation." In: *2006 IEEE Power Engineering Society General Meeting.* IEEE. 2006, 7–pp. DOI: 10.1109/PES.2006.1708970 (cit. on p. 43).

[142] Tor Skeie, Svein Johannessen, and Christoph Brunner. "Ethernet in substation automation." In: *IEEE control systems magazine* 22.3 (2002), pp. 43–51. DOI: 10.1109/MCS.2002.1003998 (cit. on p. 41).

[143] Veselin Skendzic and Armando Guzma. "Enhancing power system automation through the use of real-time ethernet." In: *2006 Power Systems Conference: Advanced Metering, Protection, Control, Communication, and Distributed Resources.* IEEE. 2006, pp. 480–495. DOI: 10.1109/PSAMP.2006.285416 (cit. on pp. 14, 18, 50).

[144] Shibin Song, JK-Y Ng, and Bihai Tang. "Some results on the self-similarity property in communication networks." In: *IEEE Transactions on Communications* 52.10 (2004), pp. 1636–1642 (cit. on p. 57).

[145] Kin Cheong Sou et al. "Scheduling smart home appliances using mixed integer linear programming." In: *50th IEEE Conference on Decsion and Control and European Control Conference.* IEEE. 2011. DOI: 10.1109/CDC.2011.6161081 (cit. on p. 23).

[146] Lydia Stougie et al. "Environmental, economic and exergetic sustainability assessment of power generation from fossil and renewable energy sources." In: *International Journal of Energy Research* 42.9 (2018), pp. 2916–2926. DOI: 10.1002/er.4037 (cit. on p. 23).

[147] Chih-Che Sun, Junho Hong, and Chen-Ching Liu. "A co-simulation environment for integrated cyber and power systems." In: *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm).* IEEE. 2015, pp. 133–138. DOI: 10.1109/SmartGridComm.2015.7436289 (cit. on pp. 2, 40, 42).

[148] Yoshihiko Susuki et al. "A hybrid system approach to the analysis and design of power grid dynamic performance." In: *Proceedings of the IEEE* 100.1 (2011), pp. 225–239 (cit. on p. 37).

[149] Janos Sztrik. "Basic Queueing Theory." In: (2010), p. 193. DOI: 10.2307/2983006 (cit. on pp. 58, 60).

[150] Andrew S Tanenbaum and David J Wetherall. *Computer networks.* Pearson, 2011 (cit. on p. 52).

[151] Chee-Wooi Ten, Junho Hong, and Chen-Ching Liu. "Anomaly detection for cybersecurity of the substations." In: *IEEE Transactions on Smart Grid* 2.4 (2011), pp. 865–873. DOI: 10.1109/TSG.2011.2159406 (cit. on p. 43).

[152] Jin Wang. "A process level network traffic prediction algorithm based on ARIMA model in smart substation." In: *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)*. IEEE. 2013, pp. 1–5. DOI: 10.1109/ICSPCC.2013.6663896 (cit. on p. 62).

[153] Martin Wolkerstorfer et al. "Measurement and simulation framework for throughput evaluation of narrowband power line communication links in low-voltage grids." In: *Journal of Network and Computer Applications* 59 (2016), pp. 285–300. DOI: 10.1016/j.jnca.2015.05.022 (cit. on p. 38).

[154] Sewall Wright. "Correlation and causation." In: *Journal of Agricultural Research* 20.7 (1921), pp. 557–585 (cit. on p. 60).

[155] M. Xie et al. "A seasonal ARIMA model with exogenous variables for elspot electricity prices in Sweden." In: *2013 10th International Conference on the European Energy Market (EEM)*. 2013, pp. 1–4. DOI: 10.1109/EEM.2013.6607293 (cit. on p. 66).

[156] Robert Alan Yaffee and Monnie McGee. *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier, 2000 (cit. on pp. 66, 80).

[157] Ye Yan et al. "A survey on smart grid communication infrastructures: Motivations, requirements and challenges." In: *IEEE communications surveys & tutorials* 15.1 (2012), pp. 5–20. DOI: 10.1109/SURV.2012.021312.00034 (cit. on p. 11).

[158] Ting Yang et al. "On the Modeling and Analysis of Communication Traffic in Intelligent Electric Power Substations." In: *IEEE Transactions on Power Delivery* 8977.m (2016), pp. 1–1. DOI: 10.1109/TPWRD.2016.2573320 (cit. on pp. 2, 38, 44, 46, 56, 62, 77).

[159] Ping Yi et al. "Puppet attack: A denial of service attack in advanced metering infrastructure network." In: *Journal of Network and Computer Applications* 59 (2016), pp. 325–332. DOI: 10.1109/ICC.2014.6883456 (cit. on p. 38).

[160] Rong Yu et al. "Cognitive radio based hierarchical communications infrastructure for smart grid." In: *IEEE network* 25.5 (2011), pp. 6–14. DOI: 10.1109/MNET.2011.6033030 (cit. on p. 11).

[161] Yanxu Zhang et al. "Analytical Modeling of Traffic Flow in the Substation Communication Network." In: *IEEE Transactions on Power Delivery* 30.5 (2015), pp. 2119–2127. DOI: 10.1109/TPWRD.2014.2377475 (cit. on pp. 2, 3, 45, 46, 49, 57, 61, 74, 85, 89, 97, 99, 116, 124, 133, 135, 136).

[162] Yichi Zhang et al. "Distributed intrusion detection system in a multi-layer network architecture of smart grids." In: *IEEE Transactions on Smart Grid* 2.4 (2011), pp. 796–808. DOI: 10.1109/TSG.2011.2159818 (cit. on p. 9).

[163] Yuan Zhang, Navid Rahbari-Asr, and Mo-Yuen Chow. "A robust distributed system incremental cost estimation algorithm for smart grid economic dispatch with communications information losses." In: *Journal of Network and Computer Applications* 59 (2016), pp. 315–324. DOI: 10.1016/j.jnca.2015.05.014 (cit. on p. 38).

[164] Zhidan Zhang et al. “Modeling and Simulation of Data Flow for VLAN-Based Communication in Substations.” In: (2015), pp. 1–12. DOI: 10.1109/JSYST.2015.2428058 (cit. on pp. 2, 43, 46).

[165] K. Zhu, M. Chenine, and L. Nordstrom. “ICT Architecture Impact on Wide Area Monitoring and Control Systems’ Reliability.” In: *IEEE Transactions on Power Delivery* 26.4 (2011), pp. 2801–2808. DOI: 10.1109/TPWRD.2011.2160879 (cit. on p. 37).

Résumé – Aujourd'hui, les réseaux électriques deviennent de plus en plus intelligents grâce à l'intégration des technologies de l'information et de la communication (TIC) ce qui a donn naissance au concept de *smart grid*. Différents Dispositifs Électroniques Intelligents (DEIs) interopèrent pour faire supervision, commande et protection sur le Réseau de Communication de Sous-station (RCS). Compte tenu des interactions entre réseau d'energie et de communication, il est nécessaire de développer les modéles mathmathiques pour les décrire. Nous proposons un nouveau modèle mathématique qui décrit le comportement du trafic SCN, qui se basé sur le modèle d'identification de Box-Jenkins. Un modèle stochastique Auto Regressive Integrated Moving Average (ARIMA) basé sur les données est développé dans cette thèse, en utilisant les méthodes d'analyse des séries temporelles. D'abord, le modéle est appliqué sur les données issues d'une co-simulation d'un scénario de communication contenant deux sources renouvelables qui se partagent la fourniture de puissance réactive afin de réguler la tension. Dans un deuxième scénario, l'application de ce modèle a été faite sur un banc de test d'une sous-station intelligente, où un signal de puissance active est à travers dans le réseau de communication. Finalement, le modèle mathèmatique global du comportement de trafic est obtenu en rajoutant la partie déterministe à la partie stochastique identifiée par ARIMA. En plus, les effects de certains paramètres du canal de communication sur le modèle ARIMA sont étudiés, *e.g.*, la perte de paquets, le taux de canal, et la longueur de câble. Ensuite, comme la performance en temps réel d'un réseau de communication est primordiale pour un réseau électrique fiable, la dernière partie de cette thèse inclut une estimation du délai maximum de transmission des messages de RCS. Sur la base du théorème de calcul du réseau, un algorithme d'estimation de la distribution, puis du délai, est appliqué sur le trafic échantillonné de données. Sur la base de notre choix d'un modèle analytique proposé dans la littérature, le délai maximum est estimé dans le cas où un réseau électrique supporte la décision de délestage en utilisant un poste MT/BT réel. L'algorithme a été appliqué pour un réseau de communication et des flux de trafic totalement connus. Dans ce travail de thèse, nous proposons une méthode d'identification basée sur le Théorème de Calcul du Réseau pour appliquer l'algorithme indépendamment des caractéristiques des flux. L'environnement logiciel pour la modélisation, la co-simulation et l'identification est supporté par les toolboxes spécialisés de sur MATLAB®/Simulink®.

Mots clés: réseau électrique, série temporelle, Auto Regressive Integrated Moving Average (ARIMA) modèle, IEC 61850 GOOSE, Théorème de Calcul du Réseau, délai maximal

Abstract – Nowadays power grids are becoming more and more intelligent due to the integration of Information and Communication Technology (ICT), that has justified the use of "smart grid" concept. Different Intelligent Electronic Devices (IEDs) interoperate to perform supervision, control and protection functions. Due to the interactions between power and communication networks, the power grid stability can be affected by the related ICT infrastructure. Therefore, it is necessary to develop modeling and co-simulation methods to describe these interactions. As the changes in smart grid mainly occur at the distribution level, design and maintain of a reliable Substation Communication Network (SCN) is paramount important. To this end, we propose a new mathematical model that describes the SCN traffic behavior. Based on the Box-Jenkins identification model, using time-series analysis methods, a data-driven-based stochastic model Auto Regressive Moving Average (ARIMA) is proposed in this thesis. The first application scenario is a co-simulation one, namely that of an electrical distribution grid containing two renewable sources in the context of a simple distributed reactive power control. Next, the real-time performance of a communication network is studied by estimation of the maximum message transmission delay through SCN. Based on the Network Calculus Theorem, a message distribution is modeled, and delay is then estimated for each traffic flow. The communication scenario considers a power system that supports the load shedding decision on an intelligent MV/LV substation test bench. A new identification method based on the Network Calculus Theorem is proposed in this thesis, that does not need any *a priori* information about the flow information. All the modeling, co-simulation and identification computations are performed by using MATLAB®/Simulink® dedicated toolboxes.

Keywords: smart grid, time series, ARIMA model, IEC 61850 GOOSE, Network Calculus Theorem, maximum delay

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab/G2ELab
Grenoble