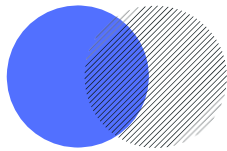


# ADVANCE'2020

8<sup>th</sup> International Workshop on ADVANCES in  
ICT Infrastructures and Services

Cancún  
27-29 January 2020



# ADVANCE'2020

8<sup>th</sup> International Workshop  
on ADVANCES in ICT  
Infrastructures and  
Services

January 27th-29th 2020, Cancún, Mexico

Universidad del Caribe

Universidad Autónoma De Yucatan



978-2-9561128-3-9 9782956112839

## Preface

Welcome to ADVANCE'2020: the 8th International Workshop on ADVANCEs in ICT Infrastructures and Services, held this year (2020) from 27 to 29 January in Cancún, Mexico.

ICT technologies and more particularly novel networking, computing and service infrastructures are drastically changing our society in all its dimensions. These advances have not only an impact on the way people are working but also on the way they are interacting, learning, educating, playing – on all human activities. It has become critical to understand and discuss how these technologies impact our society and how they should evolve to meet future needs. The focus of the ADVANCE series of workshops is to provide a forum for the publication, presentation and discussion of relevant efforts of the worldwide scientific community, practitioners, researchers, engineers from both academia and industry on the latest theoretical and technological advances in ICT.

After the successful organization of the 1st ADVANCE workshop in 2012 in Canoa Quebrada (Brazil) with the support of IFCE Aracati, the 2nd edition was held in the city of Morro de Sao Paulo (Brazil) in 2013 with the support of IFCE, the 3rd edition was held in Miami (USA) in 2014 with the support of IFU, the 4th edition was held in Recife (Brazil) in 2015 with the support of UFPE, the 5th edition was held in the city of Evry Val d'Essonne (France) in 2017 with the support of UEVE/Paris Saclay, the 6th edition of the workshop was held in the beautiful city of Santiago de Chile (Chile) with the support of the Universidad De Chile (UC), the 7th edition was held in Cape Verde Islands with the support of the Universidad de Cabo Verde and finally this 8th edition is being held in the city of Cancun (Mexico) with the support of the Universidad del Caribe and the Universidad Autonoma de Yucatan.

ADVANCE'2020 consists of seven Technical Sessions, plus Invited Talks, Tutorials, and Panels. The technical sessions are on currently hot topics, including Software Defined Networking (SDN), Cloud & Fog Computing, Data Profiling & Integration, e-Health, Education, Network Monitoring, Machine Learning, Blockchain, Internet of Things (IoT) and Security. The 7 Technical Sessions consist of 12 full papers and 10 short papers. The first invited talk about "Network traffic management" is given by invited speaker Prof Juio César Ramirez Pachero from the University of Caribe. The second invited talk is given by Prof invited speaker Prof Abdelhakim Hafid from the University of Montreal. Finally, the programme contains also 3 tutorials, the first one is an "Introduction to IoT and MEC (Mobile Edge Computing)" given by Prof Nazim Agoulmine from the University of Evry - Paris-Saclay University, the second one is on "Intelligent Network Resource Allocation and Applications" is given by Prof Joberto Martins from UNIFACS and finally the third is an "Introduction to BlockChain Technology: Concepts and Applications" given by Prof Abdelhakim Hafid.

We thank the 62 authors that submitted papers to ADVANCE'2020. Our deep gratitude also goes to the 35 members of the Technical Program Committee for their hard work reviewing papers. Finally, we thank our colleagues from Universidad del Caribe and Universidad Autonoma de Yucatan for the organization and making it possible to have ADVANCE'2020 in Cancun, Mexico.

Enjoy ADVANCE'2020 and Cancun!

Francisco Moo-Mena, TPC Co-Chair  
Elias P. Duarte Jr., TPC Co-Chair

## Program Committee

### General Chair:

Candy E. Sansores

Universidad del Caribe, Mexico

### TPC Co-Chairs:

Elias Duarte

Federal University of Parana, Brazil

Francisco Moo Mena

Universidad Autonoma De Yucatan, Mexico

### Scientific Committee:

Wassila Aggoune-Mtalaa

Luxembourg Institute of Science and Technology, Luxembourg

Nazim Agoulmine

University of Evry Val d'Essonne, France

Mustapha Ait-Idir

Banque du Canada, Canada

Rossana Andrade

Federal University of Cear, Brazil

Javier Baliosian

Universidad de la Repblica, Uruguay

Luis Basto Daz

Universidad Autnoma de Yucatn, Mexico

Sonia Ben Rejeb

Mediatron-Supcom, ISI, Tunisia

Reinaldo Bezerra Braga

LAR / Federal Institute of Technology in Cear (IFCE), Brazil

Abdelmadjid Bouabdallah

UTC, France

Karima Boudaoud

Laboratoire I3S - CNRS / Uni. of Nice Sophia Antipoli, France

Carlos Andr Batista De Carvalho

Federal University of Piaui, Brazil

Olf Chabbouh

Supcom, Tunisia

Nada Chendeb

Lebanese university, Lebanon

Elhadi Cherkaoui

Beamap, France

Emanuel Coutinho

Federal University of Cear, Brazil

Paulo Cunha

Federal University of Pernambuco - UFPE, Brazil

Antonio Wendell Rodrigues

Instituto Federal do Cear - IFCE, Brazil

Jose De Souza

Federal University of Cear, Brazil

Elias Duarte

Federal University of Parana, Brazil

Ahmed Elmisery

Nottingham Trent University, United Kingdom

Farid Alilat

USTHB, Algeria

Miguel Franklin De Castro

Federal University of Cear, Brazil

Rafael Freitas Reale

Instituto Federal da Bahia - IFBA / UFB, Brazil

Mirbella Gallareta

Universidad del Caribe, Mexico

Juan Garcilazo Ortiz

Universidad Autnoma de Yucatn, Mexico

Jorge Gmez Montalvo

Universidad Autnoma de Yucatn, Mexico

Sergio Manuel Serra Da Cruz

UFRRJ, Mexico

Francisco Moo-Mena

Universidad Autonoma de Yucatan, Mexico

Hassine Moun gla

Paris Descartes University, France

Jonice Oliveira

UFRJ, Brazil

Sergio Rajsbaum

Instituto de Matematicas, UNAM, Mexico

Julio Csar Ramirez Pacheco

Universidad del Caribe, Mexico

Paulo Sampaio

Universidade Salvador (UNIFACS), Brazil

Marcelo Santos

UFPE, Brazil

Carina Teixeira De Oliveira

Federal Institute of Cear (IFCE), Brazil

Rafael Tolosana-Calasanz

Universidad de Zaragoza, Spain

Julio Weissman Vilanova

Universidad de Sonora, Mexico

Marco Winckler

IRIT, University Paul Sabatier, France

## Author Index

Agoulmine, Nazim	9, 57, 73
Almeida, Kílvia L. De A.	81
Alves de Oliveira, Renato	33
Ammi, Mehdi	101
Amroun, Hamdi	101
Araujo, Mateus	49
Badaró Neto, Francisco	41
Barreto, Ivana	108
Basto-Diaz, Luis	89
Batista, Bruno	130
Bezerra Braga, Reinaldo	33, 65
Boudaoud, Karima	126
Castellanos-Bolaños, María-Enriqueta	138
Celestino Júnior, Joaquim	81
Celestino, Joaquim	130
Chendeb, Nada	57, 73
Costa Filho, Raimundo Valter	108
El-Assaad, Mohammad	57
F. Ramos, Ronaldo	108
Fievet, Sadry	126
Freitas, Renato	114
Garcia-Garcia, Michel	89
Gomes, Francisco	97
Gomes, Rafael L.	81
Gonzalez-Segura, Cinhtia	89
Gonzalez-Segura, Sergio	89
Guerra, César	134
Guerrero-Sosa, Jared D.T.	138
Holanda, Ivana	114
Imeri, Adnan	9
Jacob Miguel, Constantino	41
Khadraoui, Djamel	9
Khaled, Nour	73

Lessa, Lucília	81
Lopes Pereira, Silas Santiago	108
Macedo, José	97
Maia, Marcio	49
Marques, Paulo	17
Menendez-Dominguez, Victor	134
Menéndez-Domínguez, Víctor Hugo	138
Monteiro de Andrade, Luiz Odorico	108
Monteiro, Odorico	114, 118
Moo-Mena, Francisco	138, 142
Moscovits, Yanny	25
N. M. Sampaio, Paulo	41
Nascimento, Leonardo	114
Neuman, José	49, 130
Nunes Barbosa, Vinícius	65
Ojeda-Pat, Allan	142
Olavo, Cesar	114
Oliveira, Mauro	108, 114, 118
Oliveira, Wendell	118
Peixoto, Anny B. S.	81
Perez-Gonzalez, Hector	134
Ramírez, Reyes Juárez	134
Rego, Paulo	49, 97
Rios-Martinez, Jorge	89
S. B. Martins, Joberto	25
Sampaio, Paulo Nazareno Maia	1
Silva, Robson Gonzaga	1
Souza, José	97
Talbur, John	134
Teixeira de Oliveira, Carina	65
Teixeira De Oliveira, Carina	33
Torres, Eliseu	25
Trinta, Fernando	97
Valter Filho, Raimundo	118
Viana, David	118
Viana, Windson	97

# Table of Contents

Mitigating Man in the Middle attacks within Context-based SDNs . . . . .	1
<i>Robson Gonzaga Silva and Paulo Nazareno Maia Sampaio</i>	
Smart Contract modeling and verification techniques: A survey . . . . .	9
<i>Adnan Imeri, Nazim Agoulmine and Djamel Khadraoui</i>	
Machine Learning models for the prediction of Wi-Fi links performance using a CityLab testbed . . .	17
<i>Paulo Marques</i>	
Managing IEC 61850 Message Exchange for SDN-Controlled Cognitive Communication Resource Allocation in the Smart Grid . . . . .	25
<i>Yanny Moscovits, Eliseu Torres and Joberto S. B. Martins</i>	
Plante: An Intelligent Agent-based Platform for Monitoring and Controlling of Agricultural Environments . . . . .	33
<i>Renato Alves de Oliveira, Reinaldo Bezerra Braga and Carina Teixeira De Oliveira</i>	
The CAARF approach towards Monitoring and analysis of Contextual data within SDN networks . .	41
<i>Constantino Jacob Miguel, Francisco Badaró Neto and Paulo N. M. Sampaio</i>	
Performance analysis of computational offloading on embedded platforms using the gRPC framework . . . . .	48
<i>Mateus Araujo, Marcio Maia, Paulo Rego and José Neuman</i>	
Inspiring from SDN to Efficiently Deploy IoT Applications in the Cloud/Fog/IoT ecosystem . . . . .	57
<i>Nada Chendeb, Nazim Agoulmine and Mohammad El-Assaad</i>	
AuFa - Automatic Detection and Classification of Fake News Using Neural networks . . . . .	65
<i>Vinícius Nunes Barbosa, Carina Teixeira de Oliveira and Reinaldo Bezerra Braga</i>	
Integrating Blockchain with IoT for a Secure Healthcare Digital System . . . . .	73
<i>Nada Chendeb, Nour Khaled and Nazim Agoulmine</i>	
Kidney Failure Detection Using Machine Learning Techniques . . . . .	81
<i>Kílvia L. De A. Almeida, Lucília Lessa, Anny B. S. Peixoto, Rafael L. Gomes and Joaquim Celestino Júnior</i>	
Educational robotics at K-12 schools in the southeast of Mexico . . . . .	92
<i>Cinthia Gonzalez-Segura, Michel Garcia-Garcia, Jorge Rios-Martinez, Sergio Gonzalez-Segura and Luis Basto-Diaz</i>	
Supporting encryption in a MCC framework . . . . .	100
<i>Francisco Gomes, Paulo Rego, Fernando Trinta, Windson Viana, José Macedo and José Souza</i>	
Deploying-in-Production of a connected object-user identification approach by recognizing physical activity in a Big Data environment . . . . .	104
<i>Hamdi Amroun and Mehdi Ammi</i>	
Machine Learning Supporting Brazilian Public Health Care policies . . . . .	111
<i>Raimundo Valter Costa Filho, Silas Santiago Lopes Pereira, Ronaldo F. Ramos, Luiz Odorico Monteiro de Andrade, Ivana Barreto and Mauro Oliveira</i>	

Analysis of Interoperability in Public Health Systems .....	117
<i>Leonardo Nascimento, Renato Freitas, Cesar Olavo, Ivana Holanda, Odorico Monteiro and Mauro Oliveira</i>	
Quality of Health Service, an architecture to optimizing an IoT solution with Diffserv and health protocol EWS.....	121
<i>David Viana, Raimundo Valter Filho, Wendell Oliveira, Odorico Monteiro and Mauro Oliveira</i>	
SECURITY ISSUES OF HEALTHCARE IOT DEVICES .....	129
<i>Sadry Fievet and Karima Boudaoud</i>	
Heimdall: An Authorization Framework Based on Blockchain for Sensitive Data Access .....	133
<i>Bruno Batista, José Neuman and Joaquim Celestino</i>	
A proposal methodological to Assure Quality using Data Profiling Techniques .....	137
<i>César Guerra, Hector Perez-Gonzalez, Victor Menendez-Dominguez, Reyes Juárez Ramírez and John Talburt</i>	
Document Database for Scientific Production .....	141
<i>Jared D.T. Guerrero-Sosa, Víctor Hugo Menéndez-Domínguez, María-Enriqueta Castellanos-Bolaños and Francisco Moo-Mena</i>	
Gesture Recognition in Sign Languages: Methods and Approaches.....	145
<i>Allan Ojeda-Pat and Francisco Moo-Mena</i>	

## Keyword Index

Abstraction	57
Architecture	57
Authorization	130
Big Data	101
Blockchain	9, 73, 130
Brazilian health data	108
CAARF-SDN	1
Cloud	57, 73
Cognitive Resource Allocation	25
Compromise of personal and sensitive data	126
Confirmation of the strategy of the Medjack attacks	126
Contextual Analysis	41
Convolutional neural network	142
data mining	108
data profiling techniques	134
data quality dimensions	134
data quality methodology	134
DDoS	1
Document database	138
educational robotics	89
Embedded Systems	49
Encryption	97
EWS	118
Fake News.	65
Flexibility	57
Formal Verification	9
Gesture recognition	142
GISSA	114
Glomerular Filtration Rate	81
GOOSE	25
Health Information Systems	114
Healthcare	73, 118
Human Activity Recognition	101
ICT in education	89
IEC 61850	25



IEC Communication Management	25
IEC Messages	25
information availability	108
Information Security Network Security	1
Intelligent Agriculture	33
Internet of Things	130
Interoperability	114
IoT	57, 73, 101, 118
Kidney Failure	81
Machine learning	101
Machine Learning	17, 81
Mobile Application	33
Model Checking	9
MongoDB	138
Natural Language Processing.	65
Network Monitoring	41
Neural networks.	65
New cyber threat for health and medical facilities	126
Offloading	49, 97
ONOS	1
optmization	118
Performance Evaluation	49, 97
process improvement	134
Quality of Devices	41
Quality of Experience	41
Quality of Service	41
Scalability	57, 73
scientific production	138
Scopus	138
SDN	1, 57
SDN/ OpenFlow	25
Security of healthcare and medical IoT devices	126
Sensor Network	33
Sign language	142
Smart Cities	17
Smart Contract	9, 130
Smart Grid	25
Software-Defined Networks	41
STEM	89
Substation Communication	25
SV	25



# Mitigating Man In The Middle attacks within Context-based SDNs

Robson Gonzaga Silva and Paulo Maia Sampaio Nazareno

UNIFACS - University of Salvador, Salvador, Brazil.

Professor.robsongonzaga@gmail.com, pnms.funchal@gmail.com

## Abstract

The application of experimental networks through hybrid proposals such as the combination of Context-Sensitive Networks and Software-Defined Networks (SDNs) is a promising approach due to the combination of its key benefits such as scalability, dynamism, flexibility, easy management, function control, programming and others. However, despite the many advantages of hybrid approaches and new network paradigms, the study of the literature reveals some existing security challenges, since in addition to introducing new vulnerabilities to the context of computer networks, they end up potentially leveraging existing vulnerabilities, which are more critical, increasing the risks of exploiting vulnerabilities by malicious users and cybercriminals. This paper aims at presenting a study about potential man-in-the-middle attacks in the context of hybrid networks, based on the CAARF-SDN project, with the objective of identifying and assessing risks related to Information Security aspects (Confidentiality, Integrity and Availability). Moreover, solutions are also conceived through the implementation and validation of security mechanisms in order to mitigate these attacks, enhancing network security and guaranteeing its features and services.

**Keywords–** Software Defined Networks - SDN, Open Network Operating System - ONOS, Information Security, Network Security, CAARF-SDN.

## 1 Introduction

The demand for optimized management of the available resources of the network infrastructure is growing exponentially in order to cope with the co-existence of the also growing heterogeneous traffic in the Internet. Therefore, since the Internet was not designed to meet this demand, issues such as vulnerability, instability, scalability and incompatibilities are more evident as well. The success and growth of the Internet is undeniable, however some of its limitations are being unveiled, thus it is important to reconsider its architecture and main protocols with new experimental networks.

The implementation of context-aware networks can be helpful in order to improve user's satisfaction when accessing network resources and to enrich traffic management since it considers users, network and end-user devices requirements, providing a generic and cutting-edge approach for traffic optimization.

In this context, another important paradigm is the Software Defined Networks (SDN) which provide the required mechanisms for the implementation of a dynamic control architecture and management of network resources in order to deliver heterogeneous traffic due to the decoupling of the control plan and routing plan (Kim; Feamster, 2013). Nevertheless, flowtable configuration within SDNs controllers is still carried out statically, which does not allow the description of the dynamic nature of context-based networks.

In order to provide the dynamic configuration of SDN networks, in this work we propose the application of a user-centric (context-based) optimization solution to SDNs called *Context-Aware*

*Adaptive Routing Framework* - applied to SDN networks (also called CAARF-SDN) (Spinola, 2015). Hybrid networks such as CAARF-SDN rely on the integration of the concepts of Quality of Service (QoS), Quality of Experience (QoE) and Quality of Device (QoD) in order to provide a more proactive and dynamic approach for time-sensitive traffic delivery (such as VoIP and video), while aiming at the improvement of user perception over a conventional and experimental IP network.

By the combination of Context-based and Software Defined Networks (SDNs) in a hybrid model it is possible to sum the benefits (Yoon et al., 2017) of both paradigms, such as scalability, dynamism, flexibility, easy management and function control, programming, etc (Shin et al., 2014).

Nevertheless, despite the several advantages of these hybrid approaches, the CAARF-SDN paradigm also introduces new vulnerabilities and threats. According to Porras et al. (2012) SDNs, in particular, present new challenges for security in computer networks affecting authenticity, integrity and availability. Different contributions in the literature aim at unveiling security aspects within SDN networks (Secci et al., 2017). However, none of these works addresses security aspects within the CAARF-SDN hybrid approaches. Therefore, this paper aims at studying and discussing the impact of Man-in-the-Middle attacks within CAARF-SDN, proposing some possible solutions for these limitations, in order to improve confidentiality and integrity within these networks.

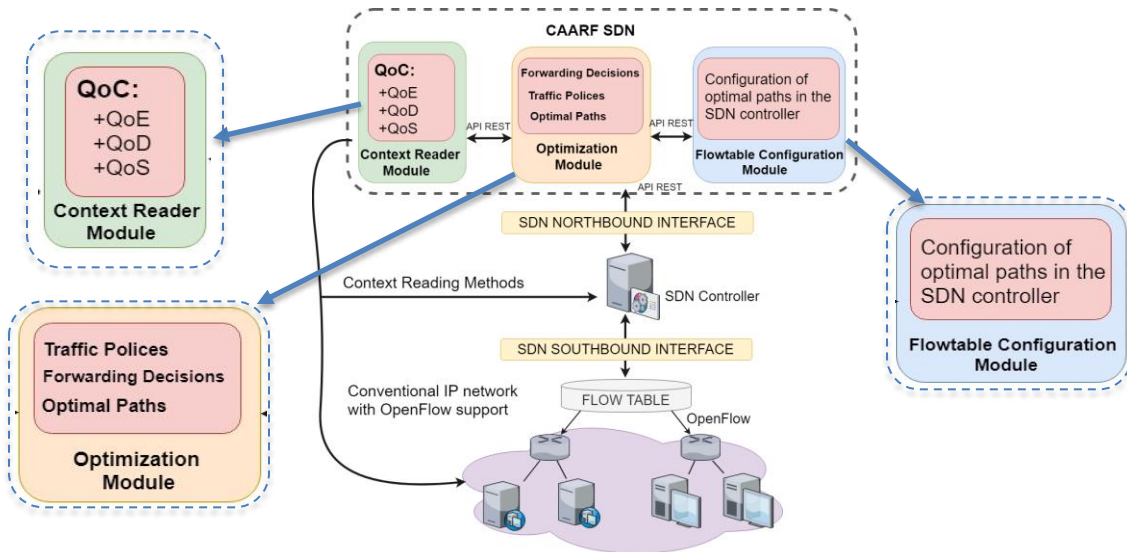
This paper is organized as follows: Section II introduces context-sensitive systems and the CAARF-SDN architecture; Section III discusses some security issues related to hybrid networks; Section IV presents some solutions to overcome vulnerabilities within CAAR-SDN; Section V discusses some lessons learned and section VI presents the conclusions of this paper and some perspectives for future work.

## 2 Context-Aware Adaptative Routing Framework (CAARF-SDN)

The **Context-Aware Adaptive Routing Framework** applied to SDN Networks (also called CAARF-SDN) (Oliveira, 2015) is a conceptual context-based solution proposed for traffic optimization within SDN networks. The conceptual architecture of CAARF-SDN is illustrated in Figure 1 and it is composed of the following modules: *Context Reader*, *Optimization* and *Flowtable Configuration* (Silva, 2015).

The **Context Reader** module aims at collecting the Quality of Service (QoS), Quality of Device (QoD) and Quality of Experience (QoE) notifications from their respective sources (network devices, end-user devices and users) and process them in order to verify the global context of the system (Quality of Context - QoC) (Muakad, 2015). These data are used by the other modules of CAARF-SDN to support traffic optimization decisions. If a relevant context modification is verified a notification for the *Optimization Module* is issued.

The **Optimization Module** aims at automatically selecting the existent optimal paths based on contextual information generated by the *Context Reader Module*. The path selection relies on a set of pre-defined policies, built upon data analysis and performance indexes also generated by the *Context Reader Module*. At last, the *Optimization Module* sends to the *Flowtable Configuration* module the configuration directives that have to be applied on the SDN controller's flowtable (Spinola, 2015).

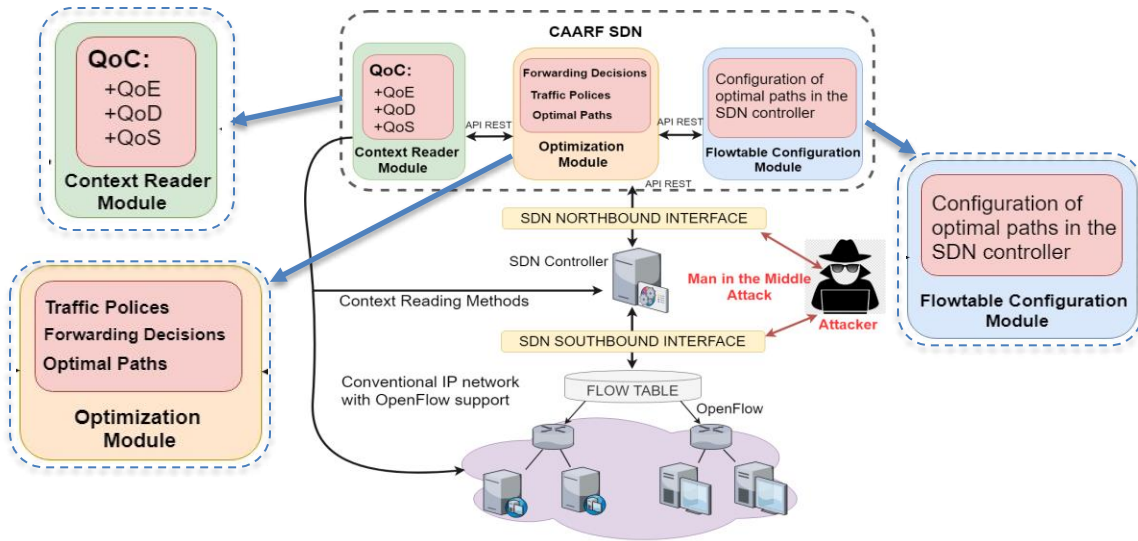


**Figure 1:** CAARF-SDN Conceptual Architecture

The main goal of the *Flowtable Configuration* module is to set the SDN controller's flowtable dynamically based on the configuration directives issued by the *optimization module*. This configuration is carried out through the SDN controller API functions, using the controller's Northbound interface. Once the controller's flowtable is reconfigured, it updates the SDN switches flowtable using its *Southbound* interface.

### 3 Man in the Middle Attacks within Context-based SDNs

When a Man-In-The-Middle (MITM) attack occurs within SDN networks, damage can significantly increase the risks of network components, since a malicious user can be connected between the data layer and the capture control layer all traffic from northbound and southbound interfaces. Thus, a malicious user in addition to capturing information is also able to modify such information being from the system or context. Figure 2 illustrates a Man-In-The-Middle attack within a hybrid network described by CAARF-SDN, targeting the SDN controller and its southbound interface. The Man-in-the-Middle attack affects integrity within CAARF-SDN in different ways, the main ones being: (i) modifying context notifications and (ii) falsifying OpenFlow flow rules (Benton et al., 2013; ONF, 2016; Hayward, 2015b).



**Figure 2:** Illustration of a *Man In The Middle* attack within CAARF-SDN.

Figure 2 illustrates these attacks and, in a first attempt, the attacker is able to intercept and modify the context notifications sent to CAARF-SDN. Therefore, with forged context messages the CAARF-SDN mechanisms will not be able to correctly optimize the traffic delivery.

Furthermore, the attacker would also be able to explore the interface between the *Application Layer* and *Control Layer*, capturing the configuration directives sent from CAARF-SDN to the SDN Controller through a vulnerable communication channel of the Controller's Northbound interface. Through this attack the controller's configuration directives can be inserted, the existing ones modified or excluded, affecting the correct behavior of the SDN devices (Young et al. 2017; Migault, 2016; ONF, 2013). Furthermore, the behavior of the SDN devices can also be affected if the attacker is able to access and reconfigure the controller's flowtable, forging false and incorrect switching rules.

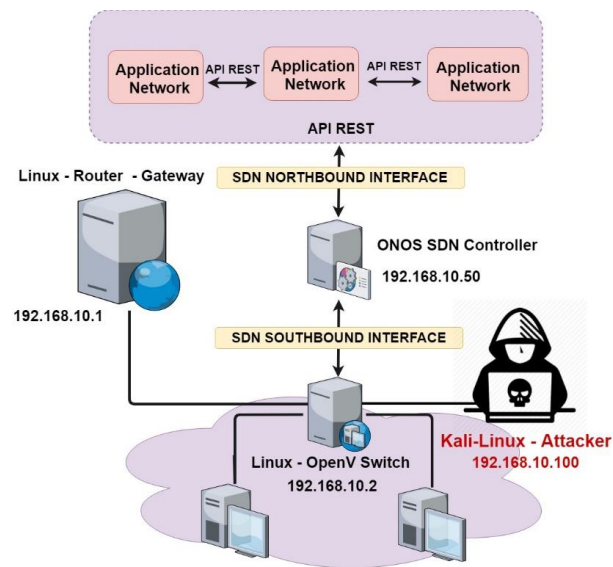
At last, another vulnerability is also present in the communication channel between the *Control Layer* and the *Data Layer* determined by the controller's Southbound interface. This attack would also affect the integrity of the flowtable content leading to the incorrect configuration of the SDN devices.

Some strategies and measures are proposed in the literature in order to increase the security level against this type of attack (Corrêa et al., 2016). For instance, the implementation of cryptography mechanisms within SDN networks is broadly discussed in the literature since due the lack (and complexity) of implementation of TLS/SSH on communication channels (Northbound and Southbound) between the *Application* and *Control layers* data remain vulnerable (Shu et al., 2016; Yoon et al., 2017).

## 4 Mitigating MITM Attacks within Hybrid Networks

The experiment carried out in this work focus on a common existing vulnerability within known SDN Network Operating Systems. The absence of cryptography solutions related to the local or remote access to the SDN Controller's GUI (Graphical User Interface) which is highly vulnerable to MITM attacks. This vulnerability is associated with the HTTP communication which exposes the requested text, including authentication information with the Controller itself (Shin et al., 2014). In this scenario, any malicious user could easily hijack the SDN Controller, consequently, gaining all the network control, compromising all the services including CAARF-SDN.

The Network Operating System (NOS) chosen for this experiment was *ONOS* (*Open Network Operating System*), one of the most used and documented NOS currently. ONOS is a JAVA-based system that provides a consistent control plan for Software-defined Networks, allowing the management of components, such as links and switches, executing network applications and services available to all the hosts and close networks (Adenuga; Heydari, 2016). A general perspective of the proposed topology is depicted in Figure 3.



**Figure 3:** SDN Laboratory with vulnerability tests

Once the scenario is configured and operational, a network scanning tool (NMAP) is executed within the malicious host, aiming at mapping all the IPs addresses of the network and identify the SDN Controller IP. For this, an in-depth scan is performed in the IP range identified previously through the command “nmap p 192.168.10.0/24”, in search of the default port access to the ONOS controller GUI, which is port 8181. Figure 5 presents the scanning result with detailed information about the IP of the SDN controller.

```
Nmap scan report for 192.168.10.50
Host is up (0.00090s latency).
Not shown: 995 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
80/tcp    open  http
443/tcp   open  https
8181/tcp  open  unknown
9876/tcp  open  sd
MAC Address: 00:0C:29:3D:DB:01 (VMw
```

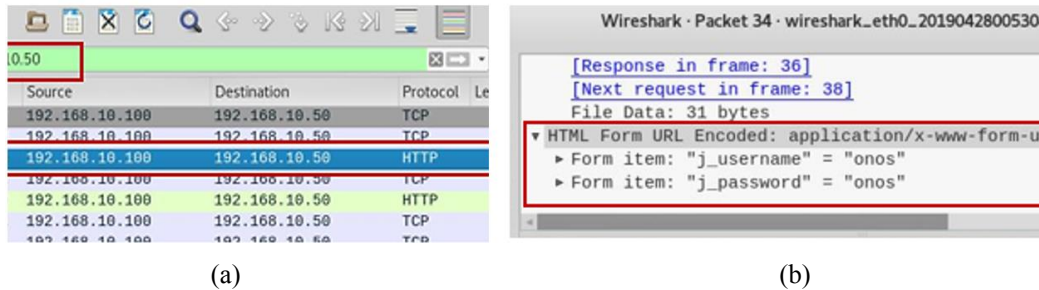
**Figure 5:** Result of the Nmap report with relevant information concerning the services related to the controller IP

For instance, it is possible to observe that IP 192.168.10.50 is the address executing an HTTP request at ONOS through port 8181. After having identified the IP of the controller, Wireshark is executed at the attacker machine as a sniffer to scan and “listen” to all the traffic addressed to the SDN controller. At this moment, the goal is to intercept the login and password information to the ONOS system. After

that, the controller can be accessed directly at the controller host, or remotely from a machine inside or outside the network.

Since all the communication with the ONOS controller Web interface is being carried out through the HTTP protocol using the port 8181, it is rather simple to a malicious user or to a hacker to eavesdrop traffic and to obtain controller's authentication information.

Figure 6(a) presents the result obtained after intercepting all the network traffic addressed to the IP 192.168.10.50 using Wireshark. In turn, Figure 6(b) presents the controller's authentication information being intercepted as raw text.



**Figure 6(a):** Controller's Traffic being intercepted using Wireshark.

**Figure 6(b):** Controller's authentication information captured and presented as raw text.

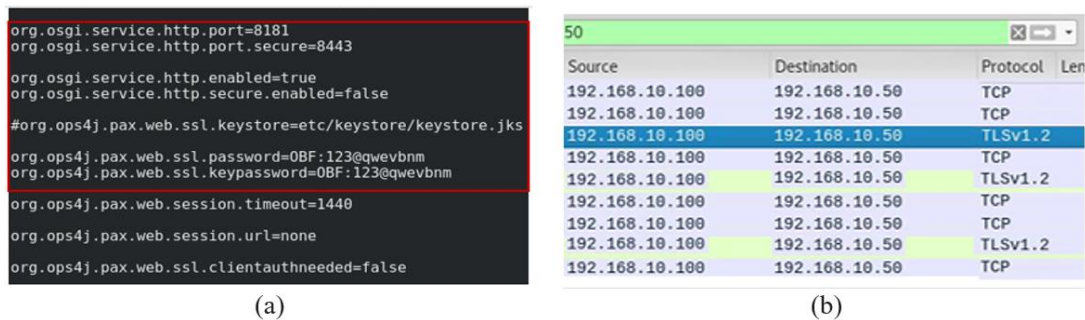
It is possible to observe that the lack of secure access to the SDN controller facilitates the authentication information to easily eavesdrop. Since the main component of an SDN network can be easily accessed all the services and resources relying on the controller turn out to be compromised. It is important to note that this vulnerability is not related only to the ONOS controller, but instead to most of the controllers commercially available since their authentication relies mostly on HTTP based insecure access (CORRÊA et al., 2016).

## 5 Discussion and Proposed Solution

The process of eavesdropping and capturing authentication information of network packets through MITM techniques, as previously illustrated, is rather effective and of great risk to hybrid networks, in particular to those logically centralized such as SDN. Nevertheless, it is possible to mitigate this type of attack through the implementation of cryptography and authentication mechanisms (CORRÊA et al., 2016).

In the case of the scenario previously presented in this paper, in order to overcome the vulnerabilities related to the ONO SDN controller's web GUI, some configuration procedures were carried out to enable web secure access through HTTPS. Nevertheless, this configuration process presented some complexity and, therefore, required some technical skills since the creation and signature of private and public keys within the controller were necessary. Thus, after the creation of both keys, it was possible to configure the secure access to the ONOS controller's web GUI using HTTPS on the port 8442 instead of using port 8181. Figures 7(a) e 7(b) present respectively the configuration process and the result obtained.





**Figure 7(a):** Java Karaf configuration file to enable HTTPS communication with the ONOS controller's Web GUI.

**Figure 7(b):** Result of Wireshark traffic capture addressed to the controller's IP with HTTPs communication presenting cryptographic information.

As we can observe in Figure 7, through the configuration of secure access to the controller, it was possible to ensure integrity, confidentiality and authenticity effectively. Therefore, eavesdropping traffic in order to capture authentication information to the ONOS controller turned out to be a hard task to malicious users since this information will no longer be available to non-authorized users. The solutions presented in this paper are, in a first glance, rather obvious as security mechanisms in order to ensure integrity, confidentiality and authenticity. However, considering that in hybrid networks, based on the such on the SDN paradigm and context-sensitive networks, control is logically centralized, and the lack of mechanisms to mitigate vulnerabilities in these networks allows the emergence of several security issues.

## 6 Conclusions and Future Works

One cannot deny the advances and advantages the new technological paradigms propose to the information society, as in the case of hybrid networks, allowing the development of Research and Innovation in the era of Industrial Revolution 4.0. Nevertheless, the same technologies and paradigms also introduce new security concerns, increasing the number of vulnerabilities.

This paper is the result of further studies that allowed to unveil some vulnerabilities related to SND networks, in particular, due to its centralized logic facilitating the exploitation of some attacks such as Man-In-The-Middle (MITM). Therefore, it was important to propose solutions to mitigate these attacks through an appropriate configuration allowing secure access to SDN controllers using the protocol HTTPs. The proposed solution was relevant in order to guarantee security aspects such as authenticity and integrity. As for future works, further vulnerabilities and solutions should also be investigated and implemented in order to propose more secured hybrid networks.

## References

- Abdelsalam, A. M.; El-Sisi, A. B. and Reddy, V. Mitigating ARP Spoofing Attacks in Software-Defined Networks. **ICCTA 2015, At Alexandria, Egypt**, n. October, 2015.
- Adenuga-Taiwo, O. and Heydari, S. Security Analysis of ONOS Software-Defined Network Platform. 2016.

Muakad, C.F.J. Context-based Dynamic and Adaptive Forwarding Management. **UNIFACS Universidade Salvador**, 2015.

Corrêa, J. H. G.; Nigam, V.; Ribeiro, M.; Mafioletti, D. and Fonseca, I. E. SHADE: Uma estratégia seletiva para mitigar ataques DDoS na camada de aplica ao em redes definidas por software. **XXXIV Simpósio Brasileiro de Telecomunicações - SBrT2016**, p. 964–968, 2016.

Kim, H.; Feamster, N. Improving network management with software defined networking. **IEEE Communications Magazine**, v. 51, n. 2, p. 114–119, 2013.

Migault, D.; Pourzandi, M. Identifying and addressing the vulnerabilities and security issues of SDN. n. January 2016, 2015.

Oliveira, A. L. C. D. Context-based Notification Mechanism for Adaptive Forwarding. In: **UNIFACS Universidade Salvador**. Salvador, Brasil, 2015.

ONF. SDN Security Considerations in the Data Center. **ONF Solution Brief**, p. 1–12, 2013.

Porras, P.; Shin, S.; Yegneswaran, V.; Fong, M.; Tyson, M. and Gui, G. A security enforcement kernel for OpenFlow networks. **Proceedings of the first workshop on Hot topics in software defined networks - HotSDN '12**, p. 121, 2012.

Secci, S.; Scoot-Hayward, S.; Wang, Y.; Van, Q. P.; Verchere, D.; Sow, A.; Basquin, C.; Smyth, D.; Attou, K.; Timmaraju, K. and Campanella, A. **ONOS Security and Performance Analysis: Report No. 1**. n. 2, p. 93, 2017.

Shin, S.; Song, Y.; Lee, T.; Lee, S.; Chung, J.; Porras, P.; Yegneswaran, V.; Noh, J. and Kang, B. B. **Rosemary: A Robust, Secure, and High-Performance Network Operating System**. 2014.

Shu, Z.; Wan, J.; Li, D.; Lin, J. and Vasilakos, A. V. **Security in Software-Defined Networking: Threats and Countermeasures**. *Mobile Networks and Applications*, v. 21, n. 5, p. 764–776, 2016.

Shubh, T.; Sharma, S. Man-In-The-Middle-Attack Prevention Using HTTPS and SSL. **Ijcsmc**, v. 5, n. 6, p. 569–579, 2016.

Silva, J. P. S. D. Implementation of an Architecture for the Context-Aware Adaptative Routing Framework (CAARF). In: **UNIFACS Un ed.** Salvador, Brasil, 2015.

Spinola, S. S. Context Management applied to Adaptive Forwarding within Convergent Solutions. **UNIFACS Universidade Salvador**, 2015.

Yoon, C.; Lee, S.; Kang, H.; Park, T.; Shin, S.; Yegneswaran, V.; Porras, P. and Gu, G. Flow Wars: Systemizing the Attack Surface and Defenses in Software-Defined Networks. **IEEE/ACM Transactions on Networking**, v. 25, n. 6, p. 3514–3530, 2017.

# Smart Contract modeling and verification techniques: A survey

Adnan Imeri<sup>1,2</sup>, Nazim Agoulmine<sup>2</sup>, and Djamel Khadraoui<sup>1</sup>

<sup>1</sup> Luxembourg Institute of Science and Technology  
adnan.imeri@list.lu

<sup>2</sup> Université of Èvry Val d'Essonne - Paris Saclay University  
nazim.agoulmine@univ-evry.fr

## Abstract

The capabilities of smart contracts for supporting and enhancing business processes in distributed-decentralized environments have affected the technological transformation of numerous industries. Designing and developing blockchain-based solutions requires model checking and verification of the components of the system such as smart contracts, for well-behave, correct execution and fulfilling of the business process requirements. Certainly, there are concerns about the execution of smart contracts in such distributed environments. This study shows the research results about model checking of smart contracts, performing a deep analysis of current approaches on modeling and verifying smart contracts and reviewing available tools for such practices. Modeling and verifying smart contracts are addressed at the levels of programming and run time execution.

## 1 Introduction

Blockchain maintains a distributed decentralized and shared ledger, that allows securely exchanging transactions between users. The network of blockchain nodes follows a peer-to-peer communication protocol, and the involved nodes contain the same ledger. The transaction data are governed by consensus protocol, that guarantees the trust and reliability of users (end-user, miners). These transactions are stored into blocks, besides the block characteristics (blockhead, nonce, transactions root), it contains also the hash of the previous block, thus forming the link of the blocks. This presents a fundamental characteristic of blockchain which in addition it supports the properties of immutability, data integrity, and non-repudiation properties [36] [34].

*Smart contract (SC)* is an autonomous computer programming code, that runs on the blockchain, and it is executed when a certain event happens, based on specified parameters [4]. For the SC that is deployed on the blockchain, a unique address is assigned, that identified the SC. The blockchain users can invoke the SC, by sending a transaction to the SC address [4] [26]. The logic implemented by SC is based on domain-specific and the source of the SC can be a natural language law, scope of any agreement between parties, and other possible sources depend on the business process requirements [29]. For the transaction that is accepted on the blockchain, and if they contain the contract address as a message received, all the miners will execute the code of the SC and react according to the specific tasks given on SC. SC is self-executed programs, moreover, it can invoke another SC, call external service, for fulfilling given task and they have the ability to automatize and implement a wide range of applications of domain-specific [26] [4] [23].

## 1.1 Formal representation of SC

The mathematical model for the formal representation of the SC in the state-transition system is a quintuple set of elements [2]:

$$M = (Q, \Sigma, \delta, s_0, F), \text{ where}$$

- $Q$ , is finite set of all possible states of the SC;
- $\Sigma$ , is the set of all input event on the SC;
- $\delta$ , is the set of transition-function of the SC ,  $\delta: Q \times \Sigma \rightarrow Q$ ;
- $F$ , is final state the SC,  $F \in Q$ ;
- $s_0$ , is the initial state of SC,  $s_0 \in Q$ ;

If the blockchain state is noted by  $\gamma$ , for any successful transaction executed by the SC, the blockchain state will be updated into  $\gamma'$  [26]:

$$\gamma \xrightarrow{Tx} \gamma'$$

The new state  $\gamma'$ , might impact many user accounts, or other SC, that might have their impact on the empirical data on the blockchain.

Simultaneously with the advantageous technological characteristics of the SC, still, relevant questions are rising when designing the SC: “Does the SC is behaving as is intended?”; “Does SC fulfill the shared intention of the parties?”; “Are the SC reliable enough to perform complex tasks, e.g., financial transactions?”

For responding to these questions and other SC vulnerabilities, the formal proof is required to ensure the SC is behaving as intended and fulfilling user requirements.

## 1.2 The vulnerabilities of SC

The security issues of SC, need high attention since they contain millions of dollars in virtual coins, or they run the business process tasks daily. Primarily, high attention is required before deploying SC. The risks stand on the fact that once deploying SC into the blockchain, they remain immutable (impossible to patch) and there is no way of stopping them [21] [25]. The well-know case of SC vulnerability is *theDAO* attack, which causes the loss of more than sixty million dollars in ethers [21][3].

Amongst the security bugs and other SC issues discovered are [25] [26] [16] :

- *Dependence on transaction-ordering*: Presents security issues where an event (of function) of SC is depended on the other previous event in order to behave correctly.
- *Timestamp*: The dependence of the SC for performing an event.
- *Throwing an uncontrollable exception*: This is the situation when a SC calls another SC, or some function of the SC by using *someThing.send(someValues)*. In case there is an exception for a certain reasons, and the called SC or function returns *false*, the process value “*someValues*”, will not reach the destination. In this situation, there is required that the SC that calls any other SC or function should check priory if the calls are made properly.
- *Reentrancy*: This is a security flaw when a SC function calls another entrusted SC function, on the SC. The called SC, can take control of data flow and make changes over data. The *theDAO* attack sourced from the reentrancy issues [7].

- *Linearizability*<sup>1</sup>: This is a security issue raised in SC when an off-chain service is called.
- *SC misbehaviour*: Present an issue when a SC is not behaving as it was intended.

## 2 Research methodology and research results

Blockchain and SC are currently in the highest level of interest as research topics from scholars and market researchers. The results in this study signify that the research field of SC and model checking and verification is relatively new based on the research articles published. For achieving significant results on this survey paper, we define a research method for selecting, classifying, and analyzing the most significant research results. Initially, we define main research questions that are the core of this research: *Model-checking techniques for SC?*; *How to verify the SC is running as it intended?*; *How to confirm the correctness of SC with natural laws and regulation?*; *Tools and best practices on verifying SC?*

From these main question, there are formalized set of queries that are used on main research libraries, such as Web of Science, IEEE, ACM, Scopus, Google Scholar, etc. The research method is composed of the following steps:

- S1: Query definition by using keywords: Smart contract & (or) formal modeling, model checking, security, verification, contract validate tools, compliant smart contract, consistency, correctness.
- S2: Search (using S1) for research articles in main research Libraries;
- S3: Formalizing the corpus of articles;
- S4: Analysis of the abstract and details of articles (selected on S3), moreover, classifying the articles based on the research topic;
- S5: Eliminating the non-relevant articles and reformulation queries (S1), if necessary after discovering other possible research challenges;
- S6: Repeating steps S2-S5, until the same results are shown again;

The method presented above allows us to formulate a systematic study of SC modeling and verification. There are retrieved a significant number of articles from our research method. The research tendency for modeling and verification of the SC is rising. For practical reasons, we do now show here results and graphs.

## 3 SC model checking and verification approaches

This section introduces an overview of model checking techniques and further, we highlight the most relevant scientific approaches for model checking and verification of the SC, for responding SC vulnerabilities presented in section 1.2. Table 1, summary the current most relevant tools, frameworks and approaches for a secure and well-behaved SC.

### 3.1 Overview of model checking and verification techniques

The formal method allows expressing a complex model for a computer system based on mathematical expressions. For obtaining the correct behavior of the model, formal methods use mathematical proofs to ensure the correctness of the model [6]. Further, the model checking techniques allow verifying all the states that are provided by the model. Initially, there is a required specification of the model, mainly by using temporal logic<sup>2</sup> and then systematically

<sup>1</sup>A classic example of linearizability is that all the users involved in the concurrent process should see the same state of data. Source: <https://jepsen.io/consistency/models/linearizable>

<sup>2</sup>[Temporal Logic Model Checking](#)

performing verification over all the specifications defined [9] [6]. This means that all possible “theorems” defined on the specification, need to be examined for all possible states of the model [6]. The model would be possible to be implemented when the previous stages “specification” and “verification” are successfully completed [6]. Model-checking and verification is a way to determine the behavior of the SC. For designing SC that is intended to run correctly and securely on the blockchain, model checking and verification is necessary. Mainly a model checking for SC will provide the necessary proof, to avoid the well-known vulnerabilities of SC 1.2, and possibly to discover new SC security and misbehavior issues.

### 3.2 The scientific approach for model checking and verification of SC

The *theDAO* attack raised the attention for the researches and scholars to improve and avoid the vulnerabilities and security issues of the SC. Research in [31], uses NuSMV for the expression of the blockchain and SC model. The model is composed of three main parts, highlighting, first, the Ethereum (kernel layer) as a distributed system for managing transactions between users. Secondly, it uses the SC (application layer) that is expressed on Solidity [10], to represent them in model checking language, i.e., in NuSMV, and the third part determines the execution environment for the application [31]. This research is to verify if the SC is behaving as they are expected. For achieving this, the expected properties need to be formalized into temporary logic (Computation Tree Logic (CTL) ) [31]. In case the property does not behave as requested, the model-checker produces a counterexample, that allows determining the problem and its genesis [31]. The research in [17] use SPIN [30] for formal verification of the properties of the SC. This research contributes by formally defining SC and providing a model for SC based on PROMELA/SPIN [17]. A formal verification of SC based on user and blockchain behaviors is proposed by research [14]. The author highlights the fact that the previous efforts for capturing the SC vulnerabilities by documenting them and by using formal verification fail because of not considering user and blockchain behaviors. The authors from [19] use the non-cooperative game theory to model the transaction performed by two players. This is possible since the terms of the contract are agreed and the players act independently. A finite-state machine (FSM) based tool, named FSolidM [1], is presented in [28], for designing secured Ethereum based SC. Further, a formal verification for SC behaviour, by using F\* [5], is showed in research [18]. The security of SC is an extremely difficult task due to the openness of the blockchain frameworks [18]. The research focuses on the behavior of the SC, and proposes a framework for analyzing and verifying the functional correctness and the run time safety of the SC by using F\* [18]. Initially, there is given a clear guide for translating Solidity and bytecode generated from the SC, into F\*. Then a detection of vulnerabilities of SC is presented. Besides, verification of the functional correctness of SC by using the Solidity subset into F\*, further, the framework proposed in [18] analysis the byte code generated for given SC and intend to prove the equivalent running of SC in solidity level (functional level) and bytecode (runtime level). In [25], the authors present a tool that intends to find the run time errors of SC, in the class of bugs of event-ordering. Basically, the idea behind this research is to see if the output from SC differs when the input order of the event (functions) is changed. In [24], researchers present a a framework for “correctness” at the level of programming aspects and the business process “validity” of SC. The formal verification of SC is performed by using abstract interpretation and symbolic model checking, where SC is taken as input, while the output in XACML style [13] is the generation of correctness or fairness [24]. Also, an intermediate-level programming language for SC is presented on [32], and the intention behind this research is to verify the high-level language programming language, e.g., Solidity, before deploying it into the blockchain. Symbolic verification of the SC is showed in

[26]. The values of program variables are represented by the symbolic parameters. The symbolic paths are formulas over the symbolic input, which these inputs should satisfy [26]. Also, the authors from [26] implemented a tool that verifies the correctness of the SC by using the SC byte code. This tool is able also to catch the famous DAO bug (reentrancy) [3] on the SC. Ethereum is proposing Vyper environment [11] to prevent the reentrancy attack [8]. Another symbolic approach based on the dependency graph, that verifies the SC behavior in the report with given properties and classify it as safe/unsafe is presented in [33].

Besides being focused on verifying the SC, on the programming level, other research highlight the verification of the SC at the runtime level, i.e., bytecode. In [20] a framework for verification of the SC is proposed by combining SC and its specification. The misbehavior of the SC is identified when a specification is violated. The research from [35] uses Coq proof assistant [12] for formal symbolic development and verification of processes of virtual machine (VM). The intention behind this study is to prove the reliability and security of the Ethereum-based SC [35]. The K framework has been used to build a tool that allows formal specification and analysis of the Ethereum VM bytecode of SC [22]. Another approach that applies formal verification of SC at the bytecode level by using Isabelle/HOL, is explained in [15]. The bytecode is structured in a block of code, and further creating a logic for reasoning this code [15].

Model Checking Tools	Main Characteristics	Operation over SC sources	Limitations
NuSMV	model checking-functional correctness	solidity	It does not support the complete expression of a blockchian environment [31]
F*	functional and runtime checking	solidity; bytecode	The presented tool for model-checking SC based on F*, does not support entirely the syntax features of Solidity, e.g., loops [18]
BIP Framework	component based and statistical model checking	solidity and blockchain	The current model does not support entirely the blockchian components, e.g., mining process, block, etc. [14]
Scilla	intermediate checking	solidity	Explicit exception are not covered on this version of Scilla [32]
EthRacer	runtime checking	bytecode	Focused only on event-order bugs by suing notions of linearizability and synchronisation [25]
ZEUS	runtime checking	solidity and bitcode	It requires to add the policy specification of the SC [24]
Oyente	pre-deployment SC checking	bytecode	Limited only on the bytecode, and thus losing the contextual information e.g. types, integer underflow (or overflow) [24]

Table 1: Summery of the main approaches related to modeling and verification of SC

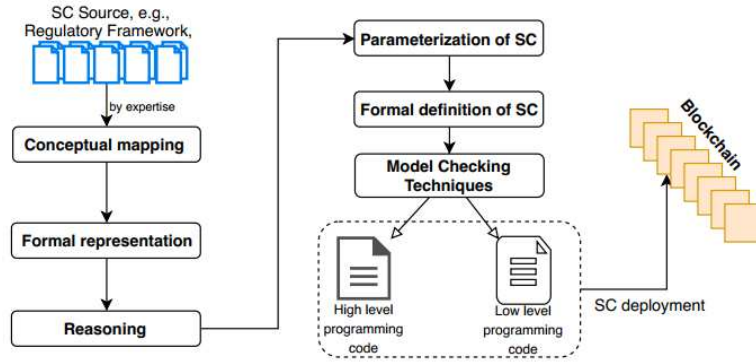


Figure 1: The schema for deriving a secured and well-behaved SC

## 4 Perspectives, conclusions and future works

We consider that the security and auditing aspects of the SC is one of the key steps before deploying and blockchain-based solution. The model checking methods that are showed in this research have essential elements to fulfill the model checking for SC. Considering, the SC are emerging from different sources, based on domain-specific requirements, the above-mentioned model checking methods do not fulfill entirely the model checking components for each SC. For the SC that are emerging from natural language, it is necessary to checker their validity in terms of if natural contract are correctly transformed into SC, and if they are running as they are intended [27]. Thus, an adaptation of the methods (or tools) is necessary for considering all possible gaps in behavior and security aspects of the SC.

For a wide examination of the behavior and security aspects of SC, a strict design method is required by considering in advance the source of SC. In terms of a method (techniques) for model checking of the SC, we propose, initially, formal reasoning over the source of SC, further at the SC level, in code level and run time level. In the context of our proposed approach, initially we select the source of the SC, which is from the perspective of regulatory frameworks, and further, we apply reasoning technique (over ontology's, e.g., by using LIKA) over the concepts that emerge from this regulatory document, e.g., legal or procedural text. Once the behavior of the conceptual model is verified, further we extract the necessary parameters for defining SC. Moreover, we express formally SC, and then we apply the relevant model checking techniques. The outline steps of our proposed approach for a secured and well-behaved SC are showed in Figure 1.

*Conclusions:* This paper summarizes the model checking techniques for SC, and we propose a new perspective on the way of modeling and verifying SC. To the best of our knowledge, there is not any model that encounter all the components of blockchain and models and verifies them. This is more due to the perspectives of the use case, which means some of the SC need to verify the financial instruments, some of them the interaction between stakeholders, and some of the fulfillment of a given task in the appropriate way.

*Future works:* We intend to apply our proposed approach, to model and verify the SC in case of anomalies, e.g., accidents or new ad-hoc decisions, on the business process. Meaning that we intend to respond to questions on adapting SC, which allows system running normally without any long disturbance.



## References

- [1] anmavrid/smart-contracts. <https://github.com/anmavrid/smart-contracts>. (Accessed on 10/31/2019).
- [2] automata2.pdf. <https://www3.cs.stonybrook.edu/~cse350/slides/automata2.pdf>. (Accessed on 10/28/2019).
- [3] The dao attacked: Code issue leads to \$60 million ether theft - coindesk. <https://www.coindesk.com/dao-attacked-code-issue-leads-60-million-ether-theft>. (Accessed on 10/09/2019).
- [4] Ethereum\_white\_paper-a\_next\_generation\_smart\_contract\_and\_decentralized\_application\_platform-vitalik-buterin.pdf. [http://blockchainlab.com/pdf/Ethereum\\_white\\_paper-a\\_next\\_generation\\_smart\\_contract\\_and\\_decentralized\\_application\\_platform-vitalik-buterin.pdf](http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf). (Accessed on 10/11/2019).
- [5] F\*: A higher-order effectful language designed for program verification. <https://www.fstar-lang.org/>. (Accessed on 09/28/2019).
- [6] Formal methods. [https://users.ece.cmu.edu/~koopman/des\\_s99/formal\\_methods/#targetText=Formal%20methods%20are%20system%20design,order%20to%20ensure%20correct%20behavior](https://users.ece.cmu.edu/~koopman/des_s99/formal_methods/#targetText=Formal%20methods%20are%20system%20design,order%20to%20ensure%20correct%20behavior). (Accessed on 10/16/2019).
- [7] Known attacks - ethereum smart contract best practices. [https://consensys.github.io/smart-contract-best-practices/known\\_attacks/](https://consensys.github.io/smart-contract-best-practices/known_attacks/). (Accessed on 10/11/2019).
- [8] Learn vyper in y minutes. <https://learnxinyminutes.com/docs/vyper/#targetText=Vyper%20lets%20you%20program%20on,requiring%20centralized%20or%20trusted%20parties.&targetText=Like%20objects%20in%2000P%2C%20each,function%20and%20common%20data%20types>. (Accessed on 10/31/2019).
- [9] Model checking overview [read-only]. <http://www.cs.cmu.edu/~emc/15-398/lectures/overview.pdf>. (Accessed on 10/16/2019).
- [10] Solidity — solidity 0.5.11 documentation. <https://solidity.readthedocs.io/en/v0.5.11/>. (Accessed on 09/16/2019).
- [11] Vyper — vyper documentation. <https://vyper.readthedocs.io/en/v0.1.0-beta.13/>. (Accessed on 10/31/2019).
- [12] Welcome! — the coq proof assistant. <https://coq.inria.fr/>. (Accessed on 11/01/2019).
- [13] Xacml 3.0 xacml:policy - complete documentation and samples. [http://www.datypic.com/sc/xacml30/e-xacml\\_Policy.html](http://www.datypic.com/sc/xacml30/e-xacml_Policy.html). (Accessed on 10/16/2019).
- [14] Tesnim Abdellatif and Kei-Leo Brousmiche. Formal verification of smart contracts based on users and blockchain behaviors models. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5. IEEE.
- [15] Sidney Amani, Myriam Bégel, Maksym Bortin, and Mark Staples. Towards verifying ethereum smart contract bytecode in isabelle/hol. In *Proceedings of the 7th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 66–77. ACM, 2018.
- [16] Nicola Atzei, Massimo Bartoletti, and Tiziana Cimoli. A survey of attacks on ethereum smart contracts. *IACR Cryptology ePrint Archive*, 2016:1007, 2016.
- [17] Xiaomin Bai, Zijing Cheng, Zhangbo Duan, and Kai Hu. Formal modeling and verification of smart contracts. In *Proceedings of the 2018 7th International Conference on Software and Computer Applications*, pages 322–326. ACM, 2018.
- [18] Karthikeyan Bhargavan, Nikhil Swamy, Santiago Zanella-Béguélin, Antoine Delignat-Lavaud, Cédric Fournet, Anitha Gollamudi, Georges Gonthier, Nadim Kobeissi, Natalia Kulatova, Aseem Rastogi, and Thomas Sibut-Pinote. Formal verification of smart contracts: Short paper. In *Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security - PLAS'16*, pages 91–96. ACM Press.
- [19] Giancarlo Bigi, Andrea Bracciali, Giovanni Meacci, and Emilio Tuosto. Validation of decentralised smart contracts through game theory and formal methods. In Chiara Bodei, Gianluigi Ferrari,

- and Corrado Priami, editors, *Programming Languages with Applications to Biology and Security*, volume 9465, pages 142–161. Springer International Publishing.
- [20] Joshua Ellul and Gordon J Pace. Runtime verification of ethereum smart contracts. In *2018 14th European Dependable Computing Conference (EDCC)*, pages 158–163. IEEE, 2018.
  - [21] Shelly Grossman, Ittai Abraham, Guy Golan-Gueta, Yan Michalevsky, Noam Rinetzky, Mooly Sagiv, and Yoni Zohar. Online detection of effectively callback free objects with applications to smart contracts. *Proceedings of the ACM on Programming Languages*, 2(POPL):48, 2017.
  - [22] Everett Hildenbrandt, Manasvi Saxena, Nishant Rodrigues, Xiaoran Zhu, Philip Daian, Dwight Guth, Brandon Moore, Daejun Park, Yi Zhang, Andrei Stefanescu, et al. Kevm: A complete formal semantics of the ethereum virtual machine. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, pages 204–217. IEEE, 2018.
  - [23] Adnan Imeri, Nazim Agoulmine, and Djamel Khadraoui. A secure and smart environment for the transportation of dangerous goods by using blockchain and iot devices. 2019.
  - [24] Sukrit Kalra, Seep Goel, Mohan Dhawan, and Subodh Sharma. Zeus: Analyzing safety of smart contracts. In *NDSS*, 2018.
  - [25] Aashish Kolluri, Ivica Nikolic, Ilya Sergey, Aquinas Hobor, and Prateek Saxena. Exploiting the laws of order in smart contracts. In *Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis*, pages 363–373. ACM, 2019.
  - [26] Loi Luu, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena, and Aquinas Hobor. Making smart contracts smarter. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 254–269. ACM, 2016.
  - [27] Daniele Magazzeni, Peter McBurney, and William Nash. Validation and verification of smart contracts: A research agenda. 50(9):50–57.
  - [28] Anastasia Mavridou and Aron Laszka. Designing secure ethereum smart contracts: A finite state machine based approach. In *International Conference on Financial Cryptography and Data Security*, pages 523–540. Springer, 2018.
  - [29] Eliza Mik. Smart contracts: terminology, technical limitations and real world complexity. *Law, Innovation and Technology*, 9(2):269–300, 2017.
  - [30] Erich Mikk, Yassine Lakhnech, Michael Siegel, and Gerard J Holzmann. Implementing statecharts in promela/spin. In *Proceedings. 2nd IEEE Workshop on Industrial Strength Formal Specification Techniques*, pages 90–101. IEEE, 1998.
  - [31] Zeinab Nehai, Pierre-Yves Piriou, and Frederic Daumas. Model-checking of smart contracts.
  - [32] Ilya Sergey, Amrit Kumar, and Aquinas Hobor. Scilla: a smart contract intermediate-level language. *arXiv preprint arXiv:1801.00687*, 2018.
  - [33] Petar Tsankov, Andrei Dan, Dana Drachler-Cohen, Arthur Gervais, Florian Buenzli, and Martin Vechev. Securify: Practical security analysis of smart contracts. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 67–82. ACM, 2018.
  - [34] Xiwei Xu, Ingo Weber, Mark Staples, Liming Zhu, Jan Bosch, Len Bass, Cesare Pautasso, and Paul Rimba. A taxonomy of blockchain-based systems for architecture design. In *Software Architecture (ICSA), 2017 IEEE International Conference on*, pages 243–252. IEEE, 2017.
  - [35] Zheng Yang and Hang Lei. Formal process virtual machine for smart contracts verification. *arXiv preprint arXiv:1805.00808*, 2018.
  - [36] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, and Huaimin Wang. Blockchain challenges and opportunities: A survey. *Work Pap.-2016*, 2016.

# Machine Learning models for the prediction of Wi-Fi links performance using a CityLab testbed

Paulo Marques<sup>1</sup>

<sup>1</sup> Instituto Politécnico de Castelo Branco, Portugal  
[paulomarques@ipcb.pt](mailto:paulomarques@ipcb.pt)

## Abstract

The Wi-Fi links performance depends in a highly complex way on the actual topology, channel qualities, spectral configurations, etc. Existing Wi-Fi radio link performance models usually adopt explicit and bottom-up approaches in order to predict throughput figures based on Markov chains and SINR levels. In this work we have validated a new approach for predicting the performance of Wi-Fi networks. Based on data measurements from the outdoor Wi-Fi CityLab testbed in Antwerp we have tested four different supervised learning algorithms. We observed that abstract “black box” models built using supervised machine learning techniques – without any deep knowledge of the complex interference dynamics of IEEE 802.11 networks – can estimate the link throughput with very good accuracy, reaching a value of R2-score of 90% for the case of the Gradient Boosting Regressor.

## 1 Introduction

Accurate prediction of wireless performance links can be very useful to optimize the Wi-Fi radio planning and resources allocation. However, the vast variety of possible wireless configurations and propagation scenarios make it hard to design explicit/theoretical models to forecast the performance of a specific link. Wi-Fi networks are notoriously hard to model in multi node scenarios. They exhibit several performance intricacies due to complex interactions between the PHY and MAC layers, which manifest themselves in frequency, spatial and time domains.

Existing radio link performance models for Wi-Fi networks, such as the model proposed in [1], usually adopt explicit and bottom-up approaches; they model the actual mechanics of the protocol (for example, the CSMA/CA procedure of the MAC layer) in order to predict throughput figures based on Markov chains.

Due to the difficulty of predicting performance in the presence of complex interference patterns, most works proposing models or optimizations for the PHY layer (e.g., [2],[3]) are reduced to using SINR-based models and ideal AWGN channels. Although SINR models can provide a characterization

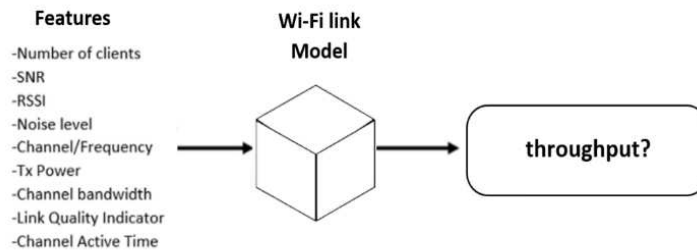
of the Shannon capacity at the PHY layer, they are not meant to capture IEEE 802.11 performance and they can fail to capture important CSMA/CA performance patterns.

In this experiment we did an experimental validation of a different approach for predicting the performance of Wi-Fi radio links. Rather than manually fitting analytical models to capture complex dependencies, we have directly learned the models themselves, using Machine Learning techniques with a limited set of observed measurements. In fact, we do not attempt to seed a pre-existing model (such as SINR based or Markov-based) with measurements. Rather, we learn and build the model itself from a limited set of measurements (state parameters) as illustrated in Fig. 1.

We treat Wi-Fi links as black boxes with potentially unknown internal mechanics. Such a black box takes some input parameters and it outputs the estimated throughput value.

The main objective of this work is the experimental validation of machine learning algorithms for predicting the performance of Wi-Fi radio links in multi node scenarios.

This paper is organized as follows: Section II describes the setup of this experiment, Section III describes the collected measurements and do a correlation analysis, Section IV proposes four Machine Learning algorithms to forecast the Wi-Fi link throughput, Section V shows the performance analysis and finally section VI concludes the paper and hints at future work.



**Fig. 1.** Prediction of a link throughput based on a “black-box” model.

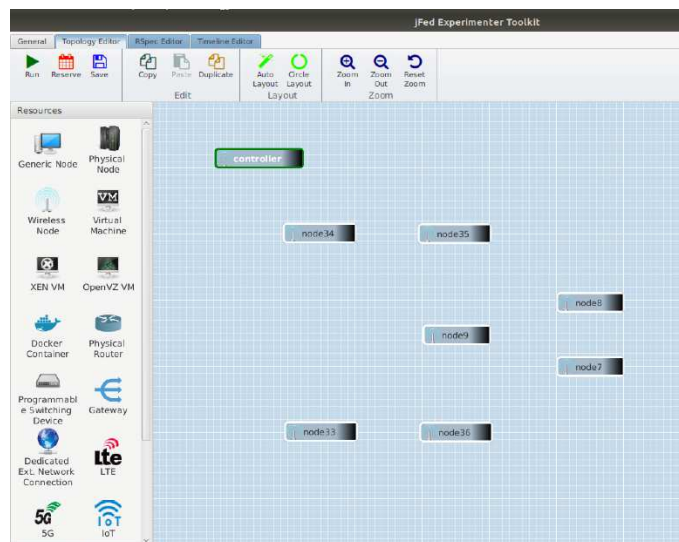
## 2 Setup of the Experiment

In this experiment we have used the CityLab (part of City of Things) testbed which is a smart cities FIRE testbed federated through the Fed4FIRE federation, operated by imec [4]. It is intended for large-scale wireless networking experimentation at a city neighbourhood level in the unlicensed spectrum. CityLab is in the city center of Antwerp, Belgium. The testbed can be found in the streets in and around the city campus of the University of Antwerp, in an area of about 0.5km by 0.5km. This testbed is a realistic environment where experiments typically face a lot of external radio interference from nearby equipment (e.g. Wi-Fi networks, IoT devices, ...). Hardware is installed at 50 locations, each with its own gateway attached to houses in the street or installed on a pole on a roof. Each gateway houses multiple radios with full low-level access for experimenters, including Wi-Fi at 2.4GHz and 5GHz.

Fig. 2 illustrates two outdoor nodes from the CityLab testbed and Fig. 4 shows the area of the CityLab testbed where this experiment was remotely carried through the jFed toolkit (Fig. 3). In order to test different deployment scenarios and configurations, a gateway acts as experiment’s controller which can change the configuration of all the nodes on the fly.



**Fig. 2.** Example of gateway deployment in the city of Antwerp available for remotely wireless experimentation.



**Fig. 3.** jFed toolkit used to remotely setup the experiment.

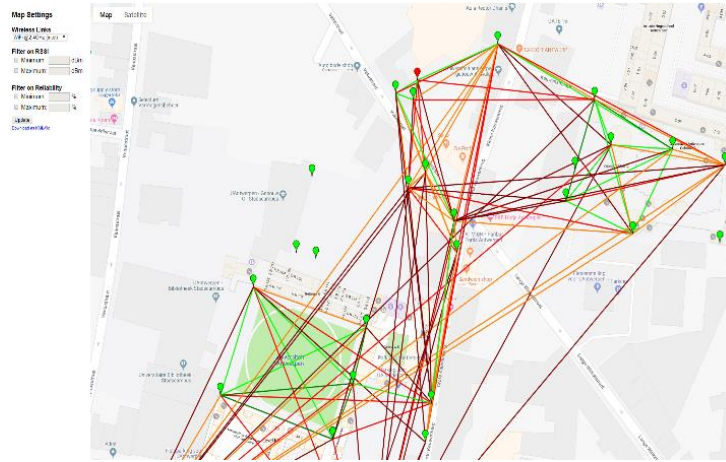


Fig. 4. Layout of the CityLab wireless testbed used in this experiment.

### 3 Measurements and Correlation Analysis

In this work we have performed  $N$  short-duration controlled experiments in the CityLab Wi-Fi outdoor testbed. Considering the “black box” representation of Fig. 1, each experiment consists in measuring the throughput ( $t$ ) of a given link ( $l$ ), for each combination of features. Those features are: number of clients, SNR, RSSI, noise level, channel, txPower, and the link quality in percentage. The goal is to expose the learning procedure to a wide variety of possible configurations. In total we did 3851 different tests.

In this experiment the throughput prediction is a multivariable regression problem with seven input features and one output to be estimated. Priority to build the Machine Learning models is important to understand the variables interdependencies and therefore the correlation level between them was computed according to the equation 1.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{cov(x,y)}{\sqrt{var(x) \cdot var(y)}} \quad (1)$$

Fig. 5 shows the measured correlation matrix where  $\rho=1$  means a perfect positive correlation between the variables;  $\rho=-1$  means a perfect negative correlation between the variables and  $\rho=0$  indicates that the variables don't have linear dependencies between them. Based on these results we can see that there is a strong positive correlation between the txPower and the throughput and a strong negative correlation between the number of clients and the throughput. These dependence between variables indicate that linear regression models can be used in the throughput estimation process.

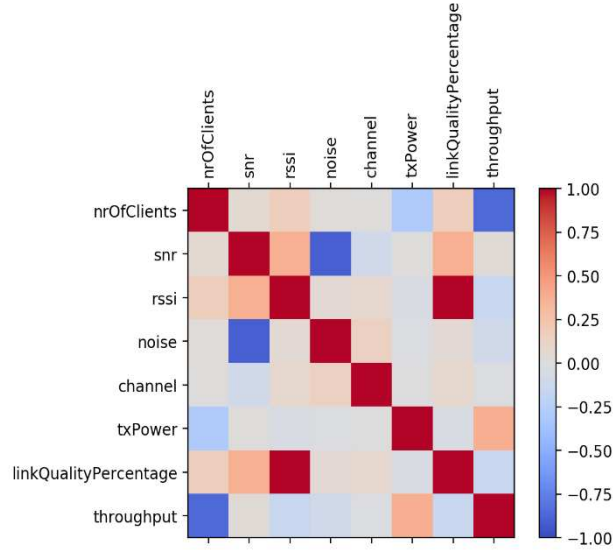


Fig. 5. Correlation matrix between all the measured features of the Wi-Fi links.

## 4 Machine Learning Models

Let us consider  $(X, t)_{i=1}^N$  the set of  $N$  measurements.  $X$  is a matrix of multiple independent input variables  $(x_1, x_2, \dots, x_7)_{i=1}^N$ , i.e., number of clients, SNR, RSSI, noise, channel, txPower and link Quality Percentage. The goal is to find a function  $f: X \rightarrow t$  that maps  $x_i$  to a value close to  $t_i$  for each measurement  $i$ . This is an instance of a regression problem where the function  $f$  is learned directly from the observed data.

The estimate  $\hat{f}(X)$ , minimizes the loss function  $\Psi(t, f)$  given by equation (2):

$$\hat{f}(X) = t \Leftrightarrow \hat{f}(X) = \arg \min_{f(X)} \Psi(t, f(X)) \quad (2)$$

There are several supervised learning methods in the literature to solve multiple regression problems (e.g. [5]). In this experiment we are going to test the following four Machine Learning algorithms: Gradient Boosting Regressor, Linear Regression, kNN (k-Nearest Neighbors) and Decision Tree. We have used the Python machine learning package scikit-learn [6] to implement the various models.

## 5 Performance Analysis

The objective of this experiment is to test the performance of the predictive algorithms of Wi-Fi throughput with unknown combinations of features. As such, we only predict throughputs for data points that do not appear in the  $N$  measurements used for learning (or training). To this end, we split our total set of measurements into a training set and a test set. The training set consists in the actual  $N$  measurements used for learning the models and their parameters, whereas the test set is used only once, for measuring the final accuracy. We compute the root mean squared error (RMSE) for each algorithm:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - \hat{t}_i)^2} \quad (3)$$

where  $t_i$  is the actual measured throughput and  $\hat{t}_i$  is the estimated value. We also compute the  $R^2$ -score given by:

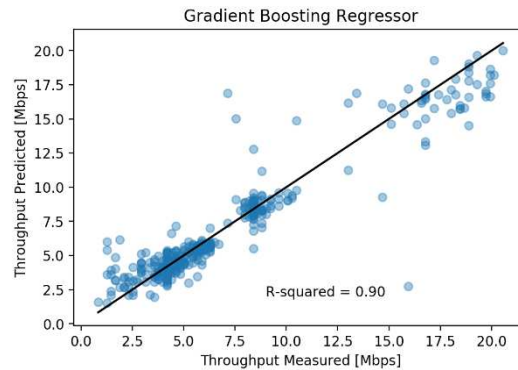
$$R^2 = 1 - \frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2} \quad (4)$$

where  $\bar{t}$  is the average throughput. Concretely, the  $R^2$ -score quantifies how well a predictor does, compared to the simplest baseline strategy, which always predicts the mean throughput. It is equal to 1 if there is a perfect match between predicted and measured throughputs.

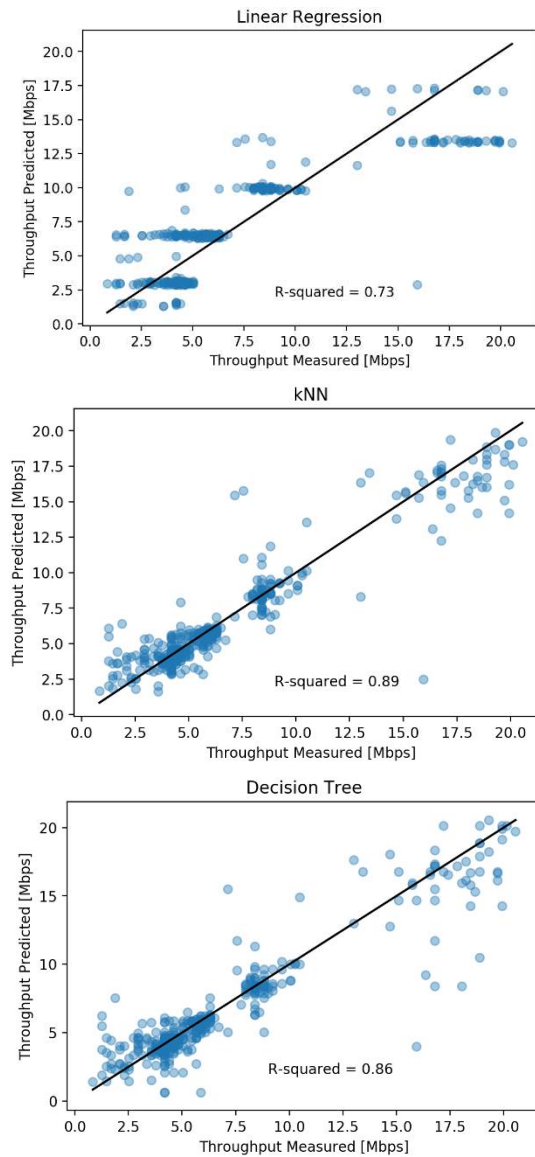
For each algorithm we have computed the RMSE and the  $R^2$ -score (Table 1), moreover, in order to visualize the actual predictions in detail, we also show a scatter plot of the predicted throughputs, against the actual measured throughputs as illustrated in Fig. 6. On these plots, the closer the points are to the diagonal, the better the prediction accuracy. The Gradient Boosting Regressor outperforms the other methods and produce fewer outlying predictions.

Method	RMSE (Mbps)	$R^2$ -score [%]
Gradient Boosting Regressor	1.48	90%
Linear Regression	2.39	73%
kNN	1.53	89%
Decision Tree	1.73	86%

**Table 1.** Performance analysis of the machine learning algorithms







**Fig. 6.** Predicted versus measured throughput for 4 Machine Learning algorithms.

## 6 Conclusions and future work

The Wi-Fi links performance depends in a highly complex way on the actual topology, channel qualities, spectral configurations, etc. It is especially hard to predict in quantitative terms how a given configuration will perform.

In this experiment we advocate an approach of “learning by observation” that can remove the need for designing explicit and complex performance models. We use machine learning techniques to learn implicit performance models, from a limited number of real-world measurements. These models do not

require to know the internal mechanics of interfering Wi-Fi links.

In this work we investigated and validated a different approach for predicting the performance of Wi-Fi links. Rather than manually fitting complex models to capture complex dependencies, we have shown that it is possible to directly learn the models themselves, from a limited set of observed measurements. This approach bypasses the usual analytical modelling process, which requires deep knowledge, and yet often yields models that are either too restricted or too inaccurate [7].

Based on data measurements from the outdoor Wi-Fi CityLab testbed in Antwerp (imec) we have tested four different supervised learning algorithms. Using supervised machine learning techniques, it is possible to generalize the observations made on this limited subset of measurements, while still capturing the complex relationships between the inputs. We build such implicit models using real-world measurements and we test them systematically, by asking them to predict the throughput for links and configurations that have never been observed during the initial measurement phase

We observed that abstract “black box” models built using supervised machine learning techniques – without any deep knowledge of the complex interference dynamics of IEEE 802.11 networks – can estimate the link throughput with very good accuracy, reaching a value of R2-score of 90% for the case of the Gradient Boosting Regressor.

A scientific level, the results obtained on the modelling of multi-node Wi-Fi networks have potential to help on the developing of better resource management algorithms and help provide guidance to radio network planners.

A possible follow-up of this work is the extension of the “black box” approach to forecast the QoE (Quality of Experience) delivered by the Wi-Fi link for specific applications such as video or web browsing, taking as inputs QoS parameters. Another interesting follow-up is the extension of the Machine Learning models to the forecast the capacity of LTE radio links without using active transmission over the mobile network.

## Acknowledgment

This work was funded from the Fundo Europeu de Desenvolvimento Regional (FEDER) through the Programa Operacional Regional do Centro (CENTRO2020) [Project Nr. 17711].

## References

1. G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*.
2. S. Rayanchu, V. Shrivastava, S. Banerjee, and R. Chandra. FLUID: Improving through-puts in enterprise wireless LANS through flexible channelization. In *ACM MobiCom*, 2011.
3. V. Mhatre, K. Papagiannaki, and F. Baccelli. Interference mitigation through power control in high density 802.11 WLANs. In *IEEE INFOCOM*, 2007.
4. J. Struye, B. Braem, S. Latré and J. Marquez-Barja, The CityLab testbed — Large-scale multi-technology wireless experimentation in a city environment: Neural network-based interference prediction in a smart city, *IEEE INFOCOM 2018*
5. Friedman, J. (2001). Greedy boosting approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
6. F. Pedregosa, G. Varoquaux, A. Gramfort, and al. Scikit-learn: Machine learning in Py-thon. *Journal of Machine Learning Research*, 12:2825– 2830, 2011.
7. J. Herzen, H. Lundgren and N. Hegde “Learning Wi-Fi Performance”, 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 2015.

# Managing IEC 61850 Message Exchange for SDN-Controlled Cognitive Communication Resource Allocation in the Smart Grid

Yanny Moscovits<sup>1\*</sup>, Eliseu Torres<sup>1†</sup>, and Joberto S. B. Martins<sup>1‡</sup>

Salvador University, Salvador, Brazil

yannymoscovits@gmail.com, eliseutorres.gx@gmail.com, joberto.martins@gmail.com

## Abstract

The IEC 61850 standard is being largely used in the Smart Grid (SG) context mainly due to its ability to address communication, interoperability and migration issues. IEC 61850 currently aims at internal substation communication. Nevertheless, there is a demand to generalize its use for distributed SG systems like Home Energy Management Systems (HEMS) and Advanced Monitoring Infrastructure (AMI) Communication which potentially involves distributed substations or distributed SG components. IEC 61850-based systems require constrained timing requirements for communication and the common approach is to allocate static link bandwidth resources leading in some cases to over dimensioning. This paper presents the Substation Cognitive Communication Resource Management (IC<sup>2</sup>RM), that aims the management of bandwidth allocation for IEC messages using a cognitive approach for its provisioning and the SDN/OpenFlow for its deployment. By dynamically deploying bandwidth for IEC messages, IC<sup>2</sup>RM optimizes links between SG substations and systems and potentially reduces the operational costs (OPEX).

**Keywords:** IEC 61850, Smart Grid, Substation Communication, IEC Communication Management, SDN/ OpenFlow, Cognitive Resource Allocation, IEC Messages, GOOSE, SV.

## 1 Introduction

The IEC 61850 standard is being largely used in the Smart Grid (SG) context. It addresses and standardizes important aspects of the grid operation and management such as the communication messages required, topologies and services. The IEC 61850 also proposes complete models, describing everything from physical devices to data and attributes [1] [2].

In the Smart Grid a robust and largely distributed network communication capability is essential to guarantee the grid operation and management [3]. The grid structure and operation can be segmented in various ways. This fundamentally depends on the problem perspective being focused. The most basic and generic structured segmentation consists of 3 general grid components that must communicate: i) generation; ii) transmission; and iii) distribution [2].

In regards to network communication supporting the main SG components, there are grid elements that have a functionality and communicate as part of their operation and management processes. Considering this perspective, the main functional elements communicating within the SG are: i) the Home Energy Management Systems (HEMS); ii) the Substation Automation Systems (SAS); iii) the Grid Energy Management System (GEMS); and iv) the Advanced Monitoring Infrastructure (AMI) Communication [4].

\*Moscovits, Y. is with UNIFACS IPQoS research group

†Torres, E. is with FIBRE UNIFACS IPQoS research group

‡Prof. Dr. Martins, J. is with UNIFACS IPQoS and NUPERC research groups

The Home Energy Management Systems (HEMS) is part of the Smart Grid on the consumption side where home appliances (e.g., air conditioner, dishwasher, dryer, refrigerator, kitchen stove, and washing machine) data are collected using smart meters. This data will be used to optimize power source and distribution. HEMS allows the end user to track consumption and optimize it, reducing energy costs [5].

The Substation Automation Systems (SAS) enables control through physical elements without the need for human interference, increasing reliability and reducing the duration of disturbances or failures. For this, communication protocols between the IEDs (Intelligent Electronic Device) are used. There are approximately 150 different communication protocols for data transmission and 20 different communication protocols in specific equipment used in utility companies. This makes it difficult to interconnect equipment from different manufacturers and gateways introduce delays in messaging can lead to improper operation. In this specific SG functional component, IEC 61850 plays a fundamental role and is relevant.

The Grid Energy Management System aims the overall management of the entities involved in the SG such as distributed energy sources, microgrids, energy storage, smart buildings, smart homes and electric vehicles, among others.

The Advanced Monitoring Infrastructure (AMI) is a bi-directional communication network integrated with sensors, intelligent meters and monitoring systems that enable the collection and distribution of information between meters and utilities [6].

Although the IEC 61850 was initially developed aiming to support internal substation communications addressing its problems and issues (SAS), the standard is being applied and extended for other SG communication scenarios like HEMS, GEMS and AIM [7] [8].

The main advantages of using IEC 61850 in substations are its lower installation cost and its capability to support new features and advanced services such as the ones required in the Smart Grid. IEC 61850 defines external visible aspects of the devices beyond data encoding on the wire and consequently enables interoperability, eases programming and lower equipment migration cost [2].

A problem concerning the adoption of IEC 61850 for communication either inside a substation or between substations is the need to have dedicated high speed capacity to support the timing requirement of priority messages. Inside a substation, dedicated switch ports are often allocated for IEC 61850 exchange of priority messages. For exchange of priority messages between substations, a dedicated overdimensioned link capacity is often used and this results in a relevant operational cost for the majority of the deployments. That being said, this paper proposes the cognitive allocation of communication resources, like link bandwidth and port capacity, in such a way that they can be optimized and shared among IEC 61850 messages.

This paper presents the Substation Cognitive Communication Resource Management (IC<sup>2</sup>RM). IC<sup>2</sup>RM objective is to allow the cognitive control of communication resources allocated for groups of IEC 61850 messages inside substation's network or between substations and functional elements of the Smart Grid. IC<sup>2</sup>RM addresses the issue of optimizing network resources deployed for communication among Smart Grid functional elements.

In the next part of this article, Section 2 discusses IEC 61850 message types, their scope and related requirements. Section 3 indicates the relevant work being done related to SG components communication with IEC 61850. Section 5 describes the IC<sup>2</sup>RM architecture and 61850 data modeling approach used to control and manage the IEC messages. Section 6 presents the implementation and section 7 presents the final considerations.

## 2 IEC Messages Types, Scope and Requirements

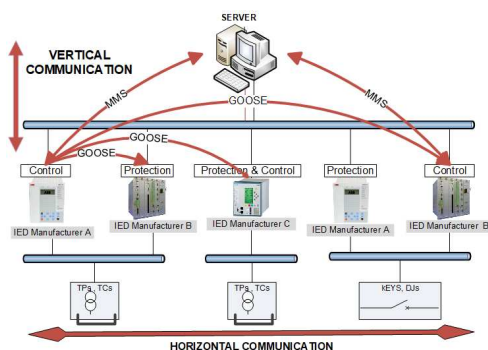
The Substation Cognitive Communication Resource Management (IC<sup>2</sup>RM) objective is to allow the cognitive control of communication resources allocated for groups of IEC 61850 messages and, in this perspective, its necessary to identify the messages types and scope alongside with their format and requirements.

There are 4 basic IEC 61850 messages types: i) Sampled Values (SV) messages; ii) PTP/S-NTP (Precision Time Protocol/ Synchronous Network Time Protocol) messages; iii) GOOSE (Generic Object Oriented Substation Event) messages; and iv) MMS (Manufacturing Message Specification) messages.

SV, PTP/SNTP and GOOSE IEC 61850 messages are transported over the SG network using either UDP/IP (User Datagram Protocol/ Internet Protocol) or straight in the Ethernet frame payload. SV messages are intended to support efficient monitoring and control of substation and SG equipment. PTP/SNTP messages are intended to support time synchronization among IEDs (Intelligent Electronic Device) or IEC-based equipment. GOOSE messages are mainly intended to support hard real-time control applications. The Manufacturing Message Specification (MMS) is a client-server based communication protocol. The client is a network application or device that requests data and actions from a server. The server contains a Virtual Manufacturing Device (VMD) in which it allocates objects and contents [9].

The Smart Grid (SG) requires communication resources between components at various levels and uses various network and communication technologies [2]. A relevant question concerning IEC 61850 in the SG is: What is the IEC 61850 main utilization scope and what are its deployment communication issues?

The scope of the IEC 61850 in the Smart Grid has been primarily defined on supporting message exchanges internally and between substations. In transmission and distribution substations, the IEC 61850 supports two group of messages that provide communication with different functionality and timing requirements: i) Horizontal communication; and ii) Vertical communication (Figure 1) [2]



Message Type	Time Constraint
Fast messages, Trip	$\leq 3ms$
Fast messages, Other	$\leq 20ms$
Medium speed messages	$\leq 100ms$
Low speed messages	$\leq 500ms$
Raw data messages	$\leq 3ms$
File transfer functions	$\leq 1000ms$
Time synchronization messages	none

Table 1: IEC 61850 Message Constraints

Figure 1: IEC 61850 Horizontal and Vertical Communication Messages and Timings [2]

Horizontal communication uses typically GOOSE messages and are intended to support critical protection and control applications with real-time transmission delay requirements. Vertical communication uses typically MMS messages that are intended to support non-critical supervision applications. An example of typical timing requirements for the utilization of IEC 61850 message exchanges is illustrated in Table 1 (Figure 1) [10].

It is a fact that the main IEC 61850 communication approach to support the exchange of

messages in substations is to over dimension network links in such a way that IEC messages always get all bandwidth they need and pass through with the required delay. The basic IC<sup>2</sup>RM motivation is then to propose a new approach for link resource allocation looking for answers to the following research question: Is it possible to deploy IEC 61850 messages communication using links with shareable and limited resources?

The relevance of this approach is based on the following aspects: i) an IEC 61850 network with shareable resources do represent a more economical and efficient use of network technologies and resources in substations and among substations. In effect, the over dimensioning of substation's communication resources does represent an investment (CAPEX). On the other hand, the over dimensioning of communication resources among substations (typically wide area telecommunication links) does represent an operational cost (OPEX) and, as such, its reduction is relevant; and ii) SG uses multiple systems at substation level that require communication with heterogeneous requirements and the deployment of shareable resources lead to a potentially more efficient solution.

The next rationale involving IEC 61850 message exchanges in the Smart Grid is: Can machine learning (ML) be used to support efficient and adequate communication resource allocation for IEC messages and other application and systems in substations and among substations?

The IC<sup>2</sup>RM is a cognitive communication approach based on SDN/OpenFlow for IEC 61850's communication resources allocation in the Smart Grid considering message exchanges inside substation and between substations and SG systems. IC<sup>2</sup>RM architecture is illustrated in Figure 2 and, in summary, it aims to allow the utilization of shareable communication resources by critical and non-critical IEC messages for functional components of the Smart Grid.

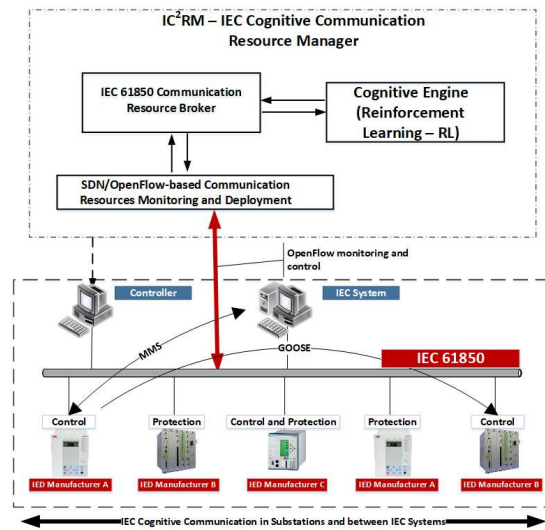


Figure 2: IC<sup>2</sup>RM Architecture and Operation Scope

### 3 Related Work

Ustun (2019) in [8], presents recent application of the IEC 61850 to support communication in various Smart Grid application scenarios like substation communication, microgrid communication and home network communication, among others. Ustun (2011) in [11] applies extensive

communication capabilities of the IEC 61850 in microgrid distributed energy resource (DER) deployments with fault current limiters IEC modeling. Solar Home System (SHS) and Smart Meter (SM) are modeled in [12] with IEC 61850 communication performance evaluation for different network technologies.

IEC 61850 communication-based coordinated operation of distributed energy resources (DER) and locally controlled distribution static compensators (DSTATCOM) is presented in [7] where MMS type IEC 61850 messages are mapped onto the XMPP (eXtensible Message Presence Protocol) web protocol to provide microgrid communication.

In all the above references, no resource allocation scheme is presented for optimize IEC message communication and guarantee its time constraints. To the extent of our knowledge, IEC 61850 message time guarantees deployed using SDN and machine learning support for managing real time message constraints have not yet been proposed.

## 4 IC<sup>2</sup>RM Architecture and Link Management

The IC<sup>2</sup>RM architecture is illustrated in Figure 2. IC<sup>2</sup>RM is composed by 3 modules: i) The IEC 61850 communication broker; ii) a cognitive engine; and iii) a SDN-based network monitoring and communication resource deployment module.

The IEC 61850 communication broker manages the constrained and shareable communication resources either inside or between substations and systems. It decides how much bandwidth is allocated for each specific set of messages exchange.

The cognitive module processes all new message exchanges and, dynamically, suggests to the broker what resource allocation action should be executed upon the set of switches used in the substation or between substations. The main purpose of the cognitive engine is to provide feedback to the broker in terms of what is the best resource allocation action to be executed based on the learning process involved in the system operation.

The SDN-based network monitoring and communication resources deployment module is basically an SDN interface between the broker and the set of OpenFlow-based switches. It monitors new incoming IEC 61850 message traffic by the use of OpenFlow *Packet-In* messages and sets-up the flow-tables in the switches of the target communication grid system.

The IC<sup>2</sup>RM's operation (dataflow) is modeled using the following principles: i) all IEC 61850 message exchanges are, at least during the first message exchange between control or supervisory equipment, inspected by the IC<sup>2</sup>RM; ii) there is a message priority scheme defined by the manager that basically defines messages that do have stringent bandwidth reservation allocation and messages that do not need it; and iii) there is a message communication identification model in a way that messages can be identified and detected for on-the-fly processing.

## 5 Message Identification Modeling for SDN-based Link Management supporting IEC 61850-based Systems

IEC 61850-based systems are modeled using typically the following steps: i) An information model is proposed with the modules that communicate in the IEC 61850-based system; ii) A set of messages is created supporting the expected service or functionality for the system; and iii) These messages are mapped onto the basic set of IEC messages. As an example, Kikusato in [8] defines an IEC-based information module with a set of messages for an EV (Electric Vehicle) charging system.

IC<sup>2</sup>RM focuses on managing basic IEC 61850 messages and, as such, IC<sup>2</sup>RM message identification modeling has to do with identifying and allocating link bandwidth for these messages. This approach guarantees that IC<sup>2</sup>RM Message would work with any IEC 61850-based system development.

In terms of the IC<sup>2</sup>RM implementation, the following priority message modeling is defined: i) GOOSE and SV messages have an assigned minimum private bandwidth allocated; ii) MMS and PTP/SNTP share a maximum limited amount of link bandwidth; and iii) MMS and PTP/SNTP maximum configured bandwidth can be, dynamically, re-allocated to GOOSE/ SV messages to guarantee its timing requirements. This priority modeling approach reflects the main objective of the IC<sup>2</sup>RM which is to provide enough bandwidth to priority messages (GOOSE/ SV) while keeping some room to allow less priority messages (MMS and PTP/SNTP) over a restrained link with limited bandwidth resources.

From the operational point of view, IEC 61850 message resource allocation requires the following actions to be executed by the IC<sup>2</sup>RM: i) GOOSE and SV messages exchanged by IEC-based equipment are processed by the broker when communication starts (1st message) to allocate bandwidth resource; ii) IC<sup>2</sup>RM implements soft-state control of critical and non-critical message exchanges among all IEC-based equipment; and iii) Messages are identified using SDN/OpenFlow *PacketIn* message and other OpenFlow protocol resources.

IC<sup>2</sup>RM message identification modeling uses the following OpenFlow flowtable parameters:

- GOOSE messages (Figure 3) [13]: i) Source/ Destination MAC addresses (equipment identification); ii) EtherType (GOOSE message); and iii) APPID (Application ID).
- SV messages: i) Source/ Destination MAC addresses (equipment identification); ii) Ether-Type (SV message); and iii) APPID (Application ID).

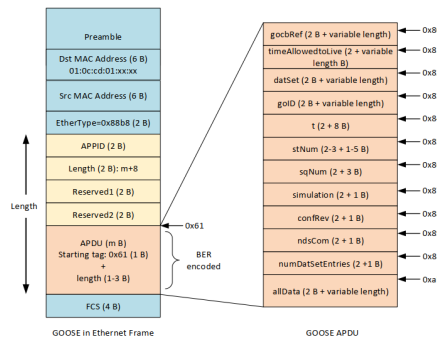


Figure 3: IEC 61850 GOOSE Message Structure [13]

## 6 IC<sup>2</sup>RM Prototype Implementation

The IC<sup>2</sup>RM prototype implementation focuses initially on managing inter-substation IEC 61850 message exchanges as illustrated in Figure 4.

In this experimental setup 2 substations are interconnected using a link (100 Mbps) and there is one ethernet OpenFlow-capable switch connecting IEC-capable and non-capable equipment per substation. The IC<sup>2</sup>RM runs on a server (controller) located in one of the substations.



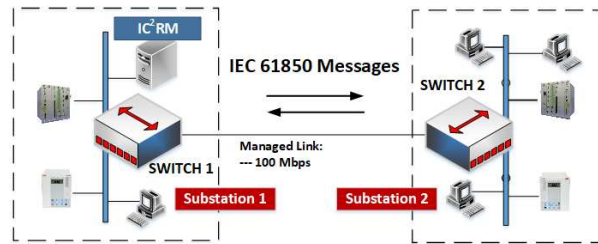


Figure 4: Inter-Substation Communication Experimental Setup

Its location in either substation 1 or 2 does not affect the overall operation because OpenFlow *PacketIn* messages generated when new IEC 61850 message traffic is generated are forwarded to the controller independently of its position and both switches use the same controller.

This IEC 61850 application scenario is relevant to control and supervision applications that must operate beyond substations limits. This is, for instance, the case of microgrid's communication in wide area for functionalities like the ones required by distributed energy resources (DER) which are highly dynamic in nature. Current IEC 61850 standard does not fully support this kind of setup and is oriented for internal substation communication.

The IC<sup>2</sup>RM prototype uses the following software components: i) Ubuntu Server 14.04.4 as the operating system; iii) Mininet Version 2.2.2 to deploy the inter-substation communication setup of hosts and switches indicated in Figure 4 iii) Oracle VM VirtualBox Version 4.1.18 r78361 as the virtualization software and iv) Pox controller; and v) OpenVSwitch (OVS) as the SDN/Openflow basic components.

In terms of the IC<sup>2</sup>RM's architecture module deployment, the IEC 61850 communication resource broker and the reinforcement learning (RL) cognitive engine are user applications running as a SDN-OpenFlow controller module in one of the hosts configured with the Mininet.

Link bandwidth allocation per prioritized (GOOSE and SV) and non-prioritized IEC message exchanges is realized by using the Queue Manager <sup>1</sup> [14]. The QueueManager is a module developed to allow the OpenFlow controller to manage dynamically the bandwidth allocation directly on OpenVSwitch. With this module, when the controller is setting a new flow, it can specify the maximum rate of bandwidth in a output port of OVS, and split the bandwidth between priority queues. Each queue receives then a specific tag that indicates the bandwidth rate. This solution solves a typical problem in controllers that can set queues, but the configurations of bandwidth allocation is intended to be done externally to the controller, preventing in some case the deployment of a dynamically configured switch.

## 7 Final Considerations

IC<sup>2</sup>RM proposes a new scheme to allow priority and non-priority IEC 61850 messages to be exchanged in substations and between substations and SG functional components keeping their time constraint requirements. The IC<sup>2</sup>RM uses a SDN-based broker that dynamically manages the bandwidth allocation for all messages flows. This allows link optimization and, consequently, the utilization of links that do not need anymore to be over dimensioned, leading to the utilization of IEC 61850 standard in a larger set of distributed applications and systems in the Smart Grid context.

<sup>1</sup><https://github.com/EliseuTorres/QueueManager>

IC<sup>2</sup>RM current prototype implementation runs on a Mininet emulated experimental setup with a POX SDN controller. The deployed prototype is able to detect new incoming IEC 61850 message flows that require bandwidth allocation using the OpenFlow *PacketIn* resource. The broker allocates then the required bandwidth per message type to comply with its standard timing requirements. In current version of the IC<sup>2</sup>RM prototype, a straightforward resource allocation method is used with the continuous and static allocation of bandwidth per message flow. At this prototype stage, the cognitive allocation approach based on the RL is not yet implemented since the current objective is to have a proof of concept of the prototype basic operation.

The next steps in the prototype implementation will include the use of reinforcement learning to learn about message traffic patterns that allow the dynamic management of under dimensioned links either inside or between substations.

## References

- [1] N. Honeth, Wu Yiming, , and L. Nordström. Application of the IEC 61850-7-420 Data Model on a Hybrid Renewable Energy System. In *IEEE Trondheim PowerTech*, pages 1–6, June 2011.
- [2] Yona Lopes, R. Franco, D. Molano, M. Santos, Flávio Calhau, C. Bastos, Joberto Martins, and N. Fernandes. Smart Grid e IEC 61850: Novos Desafios em Redes e Telecomunicações para o Sistema Elétrico. In *Brazilian Symposium on Telecommunications*, pages 1–44. September 2012.
- [3] Romildo Bezerra, F. Calhau, F. Nascimento, and Joberto Martins. A Framework to Support Smart Grid Solutions with Ubiquitous, Autonomic and Real-Time Features Targeting the Sustainable Use of Renewable Power. *Journal of Selected Areas in Renewable Energy*, 3(10):1–6, October 2013.
- [4] N. Dorsch, H. Georg, and C. Wietfeld. Analysing the Real-Time-Capability of Wide Area Communication in Smart Grids. In *2014 INFOCOM WKSHP*, pages 682–687, April 2014.
- [5] Dusit Niyato, Lu Xiao, and Ping Wang. Machine-to-machine communications for home energy management system in smart grid. *IEEE Communications Magazine*, 49(4):53–59, April 2011.
- [6] F. Bouhafs, M. Mackay, and M. Merabti. Links to the Future: Communication Requirements and Challenges in the Smart Grid. *IEEE Power and Energy Magazine*, 10(1):24–32, January 2012.
- [7] S. Hussain and I. Ali. IEC 61850 Modeling of DSTATCOM and XMPP Communication for Reactive Power Management in Microgrids. *IEEE Syst Jour*, 12(4):3215–3225, December 2018.
- [8] T. Ustun, S. Hussain, and H. Kikusato. IEC 61850-Based Communication Modeling of EV Charge-Discharge Management for Maximum PV Generation. *IEEE Access*, 7:4219–4231, 2019.
- [9] C. Ozansoy, A. Zayegh, and A. Kalam. Time Synchronisation in a IEC 61850 Based Substation Automation System. In *Proc. of the Australasian University Power Engin. Conf*, pages 1–7, 2008.
- [10] H. León, C. Montez, O. Valle, and F. Vasques. Real-Time Analysis of Time-Critical Messages in IEC 61850 Electrical Substation Communication Systems. *Energies*, 12(12):2272, January 2019.
- [11] T. Ustun, C. Ozansoy, and A. Zayegh. Extending Iec 61850-7-420 for Distributed Generators with Fault Current Limiters. In *2011 IEEE PES*, pages 1–8, November 2011.
- [12] S. Hussain, A. Tak, T. Ustun, and I. Ali. Communication Modeling of Solar Home System and Smart Meter in Smart Grids. *IEEE Access*, 6:16985–16996, 2018.
- [13] Petr Matousek. Description of IEC 61850 Communication. Technical Report FIT-TR-2018-01, Brno University of Technology, February 2019.
- [14] E Torres, R Reale, L. Sampaio, and J. Martins. BAMSND: Uma Ferramenta para a Exploração Dinâmica e Flexível de Recursos Baseada em Modelo de Alocação de Banda e SDN/OpenFlow. In *Simpósio Brasileiro de Redes de Computadores*, pages 1–8, May 2018.

# Supporting encryption in a MCC framework

Francisco A. A. Gomes, Paulo A. L. Rego, Fernando A. M. Trinta, Windson Viana, José A. F. de Macêdo, and José N. de Souza

Group of Computer Networks, Software Engineering and Systems (GREat)  
Federal University of Ceará (UFC), Fortaleza, CE, Brazil  
almada@crateus.ufc.br, pauloalr@ufc.br, fernando.trinta@dc.ufc.br,  
windson@virtual.ufc.br, jose.macedo@ufc.br, and neuman@ufc.br

## Abstract

Mobile Cloud Computing (MCC) unites two complementary paradigms by allowing the migration of tasks and data from resource-constrained devices into remote servers with higher processing capabilities in an approach known as offloading. An essential aspect of any offloading solution is the privacy support of information transferred between mobile devices and remote servers. A common solution to address privacy issues in data transmission is the use of encryption. Nevertheless, encryption algorithms impose additional processing tasks that impact both the offloading performance and the power consumption of mobile devices. This paper discusses how we extended the frameworks CAOS and CAOS D2D to supporting encryption and presents initial results on the impact caused by encryption algorithms on the execution time of offloaded methods.

## 1 Introduction

According to [1], Mobile Cloud Computing (MCC) is a novel approach for mobile applications (apps) aiming at providing a range of services, equivalents to the cloud, adapted to the capacity of resource-constrained devices, besides performing improvements of telecommunications infrastructure to improve the service provisioning.

MCC addresses applications that are very sensitive to high network delays due to the communication overhead between mobile devices and data centers resources located in the core of the current Internet infrastructure, and far away from the network edge. Some examples include real-time mobile games, crowdsensing systems, and augmented reality applications. In MCC, the most common research topic is offloading, which represents the idea of moving data and processes from mobile devices with scarce resources to more powerful machines [2]. Many research has been done on the offloading topic, and several frameworks have been proposed to provide offloading features in mobile apps [3]. One of these solutions is CAOS (Context Acquisition and Offloading System) [4], a software infrastructure to support the development of mobile context-aware applications based on the Android platform. CAOS provides offloading features to enable the processing of contextual data from mobile devices into cloud platforms. CAOS has also an version called CAOS Device-to-Device (D2D), which supports offloading between mobile devices [5]. Both CAOS and CAOS D2D allows Android programmers to mark which methods should be offloaded to remote servers using Java annotation, and the frameworks use a hybrid decision-making strategy to decide if it is worthy to offload.

Despite its potential, MCC has several challenges, such as the privacy and security of sensitive data used on offloadable processes [6]. Protecting user privacy enforces consumers' trust in a mobile or cloud platform. However, it is challenging to achieve privacy on MCC systems once the data transferred between mobile devices and remote nodes (such as methods parameters) may include user's sensitive data [7]. Thus, this paper discusses how we extended the frameworks CAOS and CAOS D2D to supporting encryption and presents initial results on the impact caused by encryption algorithms on the execution time of offloaded methods.

## 2 CAOS

Both CAOS and CAOS D2D are based on a client/server architecture as shown in Figure 1. The CAOS API runs on client mobile devices and is composed by 6 (six) components: *Discovery and Deployment Client*, *Profile Monitor*, *Authentication Client*, *Security Service*, *Offloading Client*, and *Context Client*.

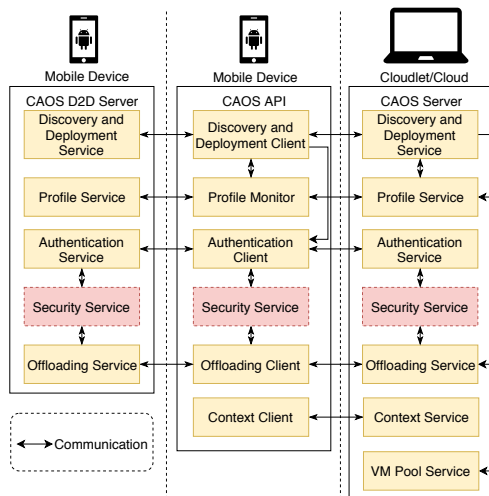


Figure 1: Overview of the CAOS/CAOS D2D Architecture.

The *Discovery Client* module uses a mechanism based on UDP/Multicast to discover CAOS Servers running in the user's local network (i.e., CAOS D2D server or cloudlets). The *Deployment Client* module injects dependencies into the CAOS Server, while the *Authentication Client* is responsible for sending the device's data to the server-side to keep the list of devices attached to a specific CAOS server through the *Authentication Service* and is connected to the *Security service*, ensuring security in the framework. (see Section 2.1). CAOS monitors the mobile application life-cycle and intercepts its execution flow whenever an annotated method is called, then decides whether to start the offloading process or not as presented in [8].

All context information of each mobile device connected to the CAOS is sent by the *Context Client* to the *Context Service* to keep a database of contextual information history. The idea is to explore the global context (i.e., the context of all mobile devices) to provide more accurate and rich context information. CAOS provides two classes of filters: one to be performed locally (on the mobile device) and other that runs in the global context repository when the method is offloaded to the cloud. The choice of which filter will be executed is performed automatically by CAOS. The components of the contextual data do not exist in CAOS D2D. The CAOS server tier has two versions: CAOS D2D that runs on mobile devices and the traditional CAOS for cloudlet/cloud. Both versions present five components: *Discovery and Deployment Service*, *Profile Service*, *Authentication Service*, *Security Service* and *Offloading Service*. The *Discovery Service* provides the correct endpoints for clients access to the CAOS Services. The *Deployment Service* receives dependencies from client applications, and store them in the server file system, to enable offloading for those applications. The *Authentication Service* is responsible for saving device data and controlling which devices are currently connected to the CAOS services, besides ensuring secure authentication by means of the *Security Service* (see Section 2.1). The *Profile*

*Service* is a set of services which receive device data related to connectivity quality and local execution time of offloadable methods from *Profile Monitor*, to keep an historical evaluation of the elapsed time for these methods. These records may be used to decide whether a method should be offloadable or not. The *Offloading Service* receives offloading requests directly from the *Offloading Client* and redirects to the *VM Pool Service*, in cloudlet/cloud tier, or directly to a mobile server. When the offloading process finishes, the *Offloading Service* returns the result to the *Offloading Client* and persists offloading information. The *VM Pool Service*, in cloudlet/cloud tier, is responsible for providing an environment that redirects offloading requests to a proper Android Virtual Machine where the offloading execution happens.

## 2.1 Security Service

In order to protect offloaded data, the *Security Service* component has been added to the CAOS framework. The module exists on both client and server sides and is responsible for confidentiality to the entire migration process. The module uses a hybrid encryption solution, which consists of combining symmetric (AES) and asymmetric (RSA) encryption, popular algorithms in this process. The following is the step-by-step for the offloading between tiers: (i) When the server is started, public and private keys (asymmetric encryption) are generated. When the client is started, the secret key (symmetric encryption) is generated; (ii) The generated public key is sent to the client through *Authentication Service*. On the client side, we use the server's public key to encrypt the secret key, then the encrypted secret key is sent to the server; (iii) Before offloading a method, the client encrypts the request object (the method and its parameters) using the client secret key; (iv) The server decrypts the request object using the secret key, which is decrypted using the server's private key, then the method is executed; and (v) After executing the method, the server encrypts the result object using the secret key and sends it to the client. In the client, the result is decrypted using the secret key.

## 3 Initial Results

We performed a initial experiment to evaluate the impact of using encryption on CAOS when offloading a method of an image processing application that applies a red tone filter to pictures. During the experiment, the method was offloaded from a mobile device (*LG X Style with Android 6.0.1, 1.5 GB of RAM memory and processor Cortex-A7 Qualcomm Snapdragon 210 MSM8909*) to a cloudlet (*laptop running Linux Mint 17.2 64 bit operating system, with 8 GB RAM and processor Core i5-4200U*), which were connected through a dedicated 802.11n wireless network. We executed the method 30 times using pictures with different resolutions (1MP, 2MP, and 4MP) and using a 128 bit AES encryption key.

Figure 2 presents, for each picture resolution, the total offloading time and the respective encryption time. As expected, the use of encryption causes overhead in the offloading process. Large images demand more time for transferring between the mobile device and the cloudlet, besides more time for applying the filter to the picture as well as for encrypting/decrypting data. The time spent with encryption procedures was approximately 82 ms, 138 ms, and 182 ms for, respectively, the 1, 2 and 4 MP pictures. The results show an increase of up to 6.57% in the total offloading time, but they also indicate that the larger the offloading time (4MP image), the less the impact of encryption procedures on such time.

## 4 Conclusion and Future Work

Offloading is a relevant research topic recently, and many different solutions have been proposed to offer offloading features. Being able to protect the communication between clients and

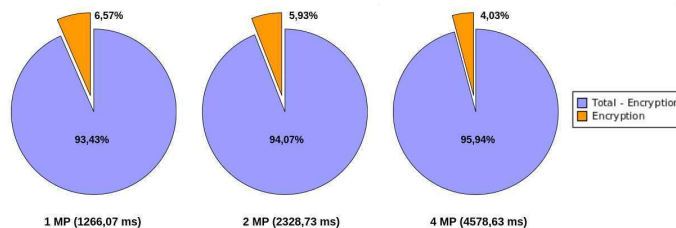


Figure 2: Initial results for encryption time

servers is an important feature of offloading solutions, but it is important to highlight that it comes with a price. The overhead of supporting encryption might impact the decision of when to offloading, so it must be investigated in greater depth. As future work, we intend to expand the experimentation and evaluate different applications, mobile devices, and wireless technologies (e.g., 4G, 5G), besides using public cloud instances as remote servers and test other cryptographic algorithms.

## Acknowledgments

The authors would like to thank The Ceará State Foundation for the Support of Scientific and Technological Development (FUNCAP) for the financial support (grant number 6945087/2019).

## References

- [1] Luis M Vaquero and Luis Rodero-Merino. Finding your way in the fog: Towards a comprehensive definition of fog computing. *ACM SIGCOMM Computer Communication Review*, 44(5):27–32, 2014.
- [2] Niroshinie Fernando, Seng W. Loke, and Wenny Rahayu. Mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(1):84 – 106, 2013.
- [3] Paulo A L Rego, Philipp B Costa, Emanuel F Coutinho, Lincoln S Rocha, Fernando AM Trinta, and Jose N de Souza. Performing computation offloading on multiple platforms. *Computer Communications*, 2016.
- [4] Francisco AA Gomes, Paulo AL Rego, Lincoln Rocha, José N de Souza, and Fernando Trinta. CAOS: A context acquisition and offloading system. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, pages 957–966. IEEE, 2017.
- [5] Gabriel B Dos Santos, Fernando AM Trinta, Paulo AL Rego, Francisco A Silva, and José N De Souza. Performance and energy consumption evaluation of computation offloading using caos d2d. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2018.
- [6] B. B. Gupta, Shingo Yamaguchi, and Dharma P. Agrawal. Advances in security and privacy of multimedia big data in mobile and cloud computing. *Multimedia Tools and Applications*, 77(7):9203–9208, Apr 2018.
- [7] Muhammad Baqer Mollah, Md Abul Kalam Azad, and Athanasios Vasilakos. Security and privacy challenges in mobile cloud computing: Survey and way ahead. *Journal of Network and Computer Applications*, 84:38–54, 2017.
- [8] Paulo AL Rego, Fernando AM Trinta, Masum Z Hasan, and Jose N de Souza. Enhancing offloading systems with smart decisions, adaptive monitoring, and mobility support. *Wireless Communications and Mobile Computing*, 2019.

# Plante: An Intelligent Platform for Monitoring and Controlling of Agricultural Environments

Renato Oliveira, Reinaldo Braga, and Carina Oliveira

Federal Institute of Education, Science and Technology of Ceará (IFCE)  
renatoalves97.123@gmail.com, (reinaldo,carina)@lar.ifce.edu.br

## Abstract

Several studies estimate a population increase for the coming decades that will result in a rise in global food demand. In this scenario, the primary sector, especially agriculture, must take a series of measures to modernize its production processes, aiming at reducing waste and increasing production levels. To effect this modernization, the sector must overcome a series of challenges related to its production processes, ranging from planning to quality control. In this context, this paper introduces *Plante* as an intelligent platform composed by three main parts: (i) a hardware module responsible for collecting environmental data and irrigating agricultural environments; (ii) a mobile application, responsible for monitoring climate aspects, providing weather forecast, and alert the farmer when something is nonstandard with the plantation; and (iii) a network layer to address storage, communication and data processing issues. We present a prototype of *Plante* to validate the proposal.

## 1 Introduction

In recent decades, society has gone through a series of transformations. One of them is the growth of the world's population. The 2019 revision of the World Population Prospects from the United Nations (UN) [9] estimates that there are 7.7 billion people worldwide and that this number could grow to around 8.5 billion in 2030, 9.7 billion in 2050, and 10.9 billion in 2100. This phenomenon deserves great attention from the government as well as the general population. If this issue is not properly addressed, it would result in a series of damages, such as the reduction of general well-being of individuals. In particular, one of the expected consequences of population growth is the increased demand for basic services, such as health, education, security and food. To meet this demand, the productive sectors will have to go through modernization processes.

The World Resources Institute (WRI), through the publication of the report *Creating a Sustainable Food Future* [10], addresses the issue of demographic growth and the quest for food. The study estimates that a 56% higher food production will be required in 2050 in relation to 2010. Also, the WRI report outlines recommendations for guiding the productive sector to deal with this issue in a sustainable way. For instance, (1) To increase food production without expanding agricultural land; (2) To protect and restore natural ecosystems; and (3) To improve soil and water management.

Nowadays, the agricultural industry worldwide produces 2,609 million tons of grain per year [6]. It corresponds to the area that produces most food in the world. Despite this large volume, a considerable increase will be necessary to cope with the future food demand. Thus, it is fundamental that the sector goes through a process of renewal and technological advances, aiming to increase the crop productivity within the recommendations pointed out by the WRI.

In this renewal process, the industry will have to solve a series of recurring impasses and shortcomings. One of them is that the increase in crop yields is intrinsically related to the

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

increase in cultivation area. For instance, the Brazilian Institute of Geography and Statistics (IBGE) states that between 2010 and 2012 the branch was responsible for the deforestation of 77,520 square kilometers of forest in Brazil [3]. Only the agricultural sector accounted for 68% of the deforestation observed in this period.

Irrigation management deficiencies are also observed in the field. It is well known that agriculture is responsible for a huge consumption of water around the globe. For example, the Brazilian National Water Agency (ANA) points out that in 2015 the volume of water used in the irrigation process in Brazil was 745 thousand liters per second; by 2030 an increase of 46% is expected, resulting in 1,055 million liters per second [4]. It is worth noting that not only the agricultural sector will have a higher demand for water resources, but also the population and several industrial sectors. This data reinforces the urgency to improve water management as adverted by the WRI.

In this sense, the low utilization of plantation monitoring and control technologies is a major drawback, as the adoption of automated solutions offers greater accuracy in data collection. Besides that, automated solutions can be configured to be always available to agricultural decision makers. Due to this factor, farmers often do not have access to accurate information regarding environmental variables such as temperature, air and soil humidity, solar incidence rate etc. This lack of accurate information negatively affects the entire decision-making process, resulting in inappropriate decisions and leading to a number of inconveniences, such as crop losses, waste of crop inputs (i.e., fertilizers, seed, etc) and reduction of profits.

Given the importance of agriculture for food supply worldwide and the benefits of Information and Communication Technologies (ICTs) for its improvement, this paper presents *Plante*, an intelligent platform for monitoring and controlling of agricultural environments. The platform is divided into three main parts: a hardware module called *Plante Box*, a mobile application called *Plante App*, and a network layer to allow the communication between the *Plante Box* and the *Plante App*. The solution provides a set of features to help the decision maker (farmer), such as the visualization of data collected by sensors, the reception of alerts and suggestions for preventive and corrective actions, the visualization of weather forecasts and the controlling of actuators. We present a prototype to validate the proposal.

## 2 Related Works

This section presents some related works that, like *Plante*, aim to combine the primary sector and ICT to achieve sustainable development and efficient production.

In the work of Celso et al. [8] proposes an architecture that uses IoT and Business Process Model and Notation (BPMN) for application in agricultural scenarios. Business Process Model and Notation is a notation developed for the construction of flowcharts and it is linked to the methodology of business process management. The work shows an IoT device integrated with sensors and actuators, which enables monitoring and control via web application, allowing part of the agricultural management to be performed remotely. The proposed web system allows the user himself, through BPMN, to model business rules for the cultivation of different plant species, automating the use of actuators. The authors demonstrated the use of architecture using lettuce cultivation in a greenhouse as a use case.

Jordano et al. [11] highlight the great relevance of three points for obtaining good results in agricultural harvests: 1) monitoring parameters such as temperature, humidity and lighting, which are the main factors in product yield and quality; 2) monitoring data analysis and decision making for optimization; and 3) applying control mechanisms. According to these points, the authors proposed a low cost agricultural monitoring system, containing a set of wireless sensors



interconnected via Wireless Sensor Network (WSN). The system also has an actuator to remote drip irrigation control. The two main disadvantages of using wireless sensors are the need to recharge batteries (requiring a continuous maintenance process) and the need for a gateway (dedicated to the sensors), increasing the cost of the product.

The authors Hamouda and Elhabil et al. [7] present a system developed for greenhouse monitoring and control. The Greenhouse Smart Management System (GSMS) has a network of WSN-connected sensors, fans and irrigators that are triggered when the system detects that humidity or temperature is not in accordance with the needs of the cultivation. In this work only two sensors are used, one of temperature and other of relative humidity. In this work, Bluetooth is used to synchronize the data collected with the mobile application, which limits the communication distance between the application and the hardware to 100 meters.

Some advantages of *Plante* over the above works are the presence of: 1) An integrated database with information about plant families, genera and species and their ideal cultivation characteristics (temperature, type and composition of soil, moisture and light level, type of nutrition, etc.); 2) Integrated weather forecasting; and 3) Alert system (in-app) that informs if the weather conditions of the cultivation are appropriate or not. These are features that add value to the product and facilitate the decision making process.

### 3 Plante Overview

The *Plante* platform provides an autonomous environment is not in accordance wthat aims at the careith of plants, reducing the human intervention, whether in large areas (such as agricultural areas, parks and squares) or small ones (such as community gardens, indoor gardens and terrariums).

Figure 1 shows the *Plante* Platform Overview, which is composed by three main parts: 1) Application, so that the farmer can monitor the data collected by the sensors and control the available actuators; 2) Network, responsible for storage, data processing and communication management between the Application and Devices; and 3) Devices, composed of a micro-controller, sensors and actuators. A case study (prototype) of the proposed solution is presented in the next section.

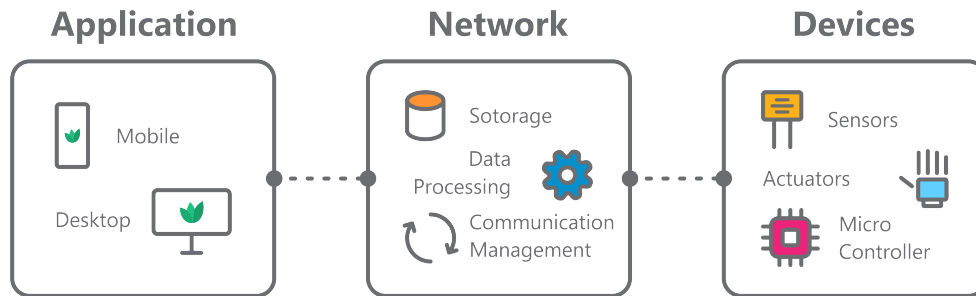


Figure 1: *Plante* Platform Overview.

### 4 Prototype

Figure 2 shows the prototype implementation realized based on the *Plante* Platform Overview presented in Figure 1. Next section details the prototype, explaining the technologies used in

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

each part of the proposal.

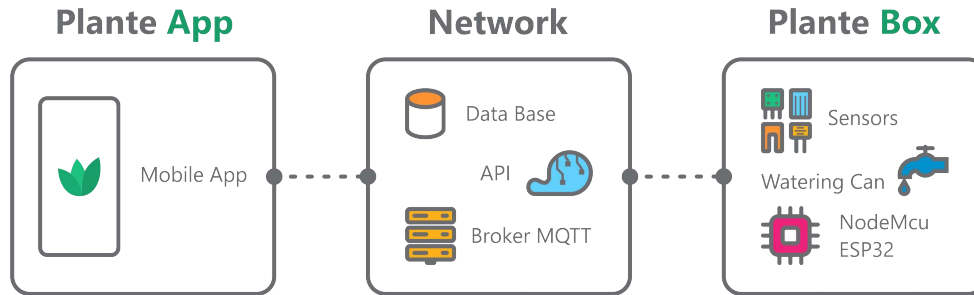


Figure 2: *Plante* Prototype.

## 4.1 Plante Box

As illustrated in Figure 2, the hardware module is called *Plante Box*. It is responsible for acquiring weather information and watering the soil when needed. It is composed by three parts: Sensors, Watering Can and NodeMCU ESP32.

### 4.1.1 Sensors

The *Plante Box* uses four sensors: DHT11, to measure air humidity and ambient temperature; YL69, which measures the soil moisture level; LDR photoresist to measure the percentage of light; and the YL83, which measures the intensity of rain.

The source code developed for *Plante Box* and written to ESP32 is based on two pillars: energy efficiency and component life. As a result, the code turns off the sensors most of the day. Each sensor receives electric current during 1.5 seconds. This is time enough to capture the environment data and send the information to the micro-controller. The frequency of data collection is 5 minutes<sup>1</sup>, for example: each sensor is turned on every 5 minutes, for 1.5 seconds of data capturing and then it is turned off until the cycle repetition. This behavior allows the sensors to remain up to 98% of the time off, ie 2% of the time on (this is 72 seconds per hour).

### 4.1.2 Watering can

Watering can consists of a solenoid valve (adapted for plumbing) and a relay module. The solenoid acts as a two-state motor, turning off (preventing water passing) and on (allowing water passing). The relay is used to activate or deactivate the solenoid. Similarly to the sensors, the electrical connection between the NodeMCU and the watering can remain off by default and is only established when requested by the user of the application or by the *Plante* Application Programming Interface (API).

### 4.1.3 NodeMCU ESP32

We have adopted NodeMCU ESP32 [13] for controlling the watering can and sensors. NodeMCU ESP32 is an open source Internet of Things (IoT) platform that uses the micro-controller ESP32. It is a low cost and low power chip developed by Espressif [5]. The hardware, which uses the

<sup>1</sup>except when watering can is on

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

Tensilica Xtensa LX6 microprocessor, has Wi-Fi and Bluetooth chips built natively into the board.

NodeMCU also acts as a synchronizer of the collected data (sending it to the Broker MQTT) and receiver of some commands sent from the application. Two of the main advantages of using this equipment are related to its low power consumption (it works with 3.3v) and the native integration with network technologies, eliminating the need to purchase extra components.

## 4.2 Network

This layer consists of three parts: 1) Data Base, responsible for storing all data; 2) API, responsible for processing data and decision-making (such as the right time to water the plantation); and the Broker Message Queuing Telemetry Transport (MQTT). The MQTT has the function to quickly enable communication between the *Plante* modules, everywhere and anytime. One of its advantages is the need for low bandwidth to perform [12] communication. The broker used in this use case is Eclipse Mosquitto™ [1].

The communication between the Broker and the ESP32 is performed in three situations: 1) When sensors capture data and send the information to the *Plante App*; 2) When the user turns the watering can on or off through the app; and 3) When the API detects the time to start or finish watering the plantation.

## 4.3 Plante App

Following the platform presented in Figure 2, the software module is called *Plante App*. Basically, *Plante App* is divided into two modes: 1) Administrative, aiming to generate a database of plants and ideal cultivation characteristics. It associates plant families, genera and species with the most appropriate types of climate, soil, light and nutrients; and 2) Commercial, where users can track and manage their plantations. This second mode was developed to assist farmers in the planning, monitoring and controlling of their plantations. The Commercial mode, focus of this article, is divided into three main sections which are described in detail below.

### 4.3.1 Sensors and Watering Can

This functionality gives users full and real-time access to the climate characteristics of their plantations. Through the tab *Sensors* the user has access to five types of data: temperature, soil moisture, luminosity, air humidity and rainfall. All this data is obtained from the *Plante Box* through the Broker MQTT.

In Figure 3(A) it is possible to observe the temperature and soil moisture levels. The red circle, at the top right of each card, indicates an alert related to the acquired data. In Figure 3(B) we can perceive the alerts generated by the API, informing when the environmental characteristics are incompatible with the plant species requirements. In the case presented in Figure 3(B), the temperature is 11.4° C above ideal, while the soil moisture is 22% above ideal.

In the *Watering* tab, according to Figure 3(C), the user can trigger the plantation watering from anywhere. This is possible due to the integration of *Plante Box* with MQTT. It is important to highlight that irrigation, by default, is done automatically, dispensing with the user's intervention in order to trigger it. The *Plante* API is responsible for turning the watering can on/off at a time when soil moisture is below/above ideal for the controlled specie. In Figure 3 there is also a summary of the weather forecast (12 hours and 5 days) displaying the predicted temperature.

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

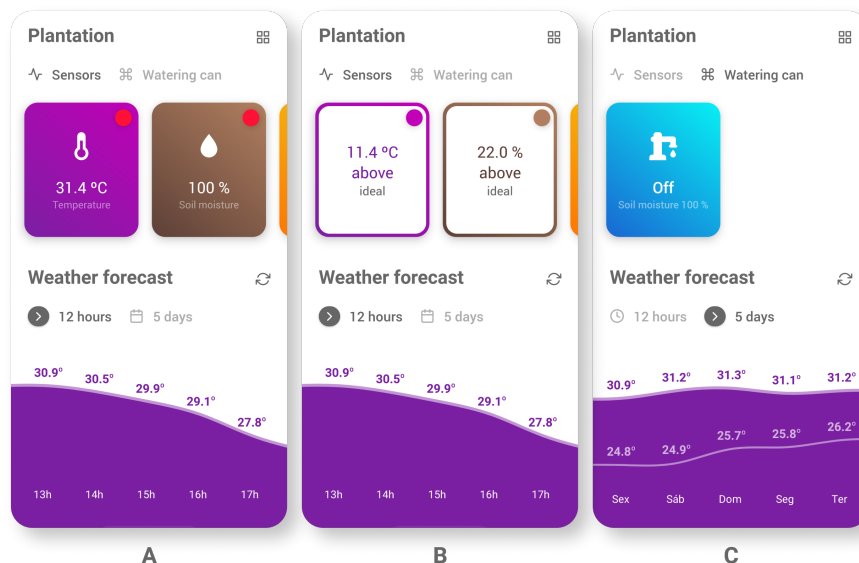


Figure 3: Dashboard

### 4.3.2 Plantations

In this module farmers can register plantations, informing the name, location and type of cultivated plantation. After registration, the information is crossed with the *Plante* Data Base and the plantation is linked to the ideal cultivation characteristics of the cultivated species. Therefore, the user has complete access to ideal cultivation characteristics such as climate type, soil, light and nutrients. It is worth noting that the Data Base is populated on demand when new users (with new plant species) hire the *Plante*. Data collection takes place through consultations with agronomists, through the analysis of scientific articles dealing with the species or through academic databases available on the Internet.

Figure 4(A) exemplifies the visualization of information from the *Plante* Data Base for a Cherry Tomato plantation. This screen presents the ideal climate characteristics, such as climate type, minimum and maximum air temperatures and humidity; the soil with its minimum and maximum levels of acidity and humidity, as well as the composition ingredients; the type of luminosity; and the types of nutrients with their respective quantity. This screen uses fictitious data.

Based on this information, it is expected that the farmer will make progress in the decision making process, because he has accurate information about the ideal conditions of cultivation for the species of plant. There is a number of decisions the farmer can make with this information about his plantation, decisions ranging from the planting planning process to the execution and control processes.

In this context, the farmer may decide to: i) establish his planting location in an area with mild or intense sunlight; choose land in a region that floods easily during winter or in a region with low soil moisture, even in rainy periods; ii) opt for a planting site that is affected by winds most of the year or not; iii) and select the region according to the type of soil composition, which has the best root development. During the execution and control steps he can decide which nutrient types to use, and in what proportion each should be mixed.

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

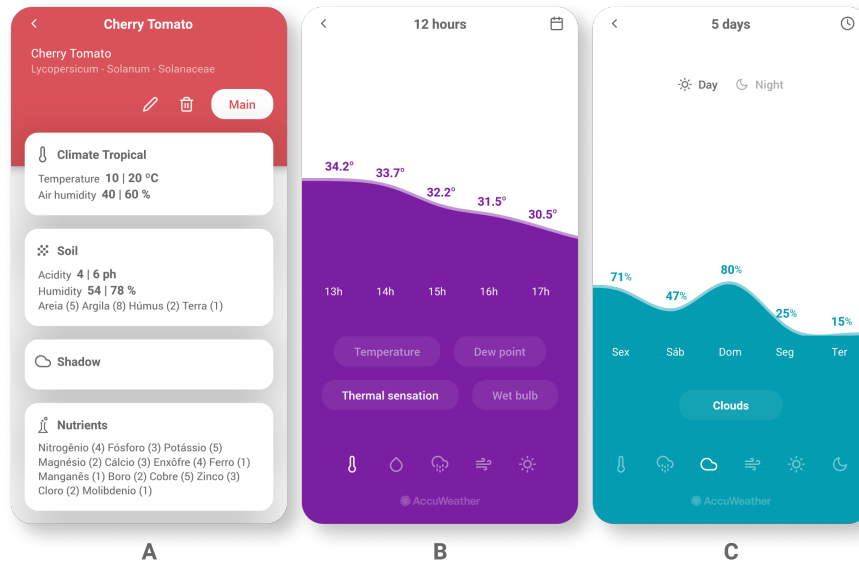


Figure 4: Planting Preview and Weather Forecast

### 4.3.3 Weather forecast

*Plante* has a set of screens that inform the prediction of a series of climatic characteristics. There are two types of views: one that shows the *12 hour* weather forecast and one that shows *5 days*. All data is obtained through the AccuWeather API [2], a platform that provides global weather data. In the app, all weather data is displayed as graphs, making it easier and faster to view and understand the weather situation.

The 12 hour forecast data is presented as follows: Temperature (actual, thermal sensation, dew point and wet bulb); Air humidity; Rain (amount and probability); Wind (speed and direction); and Ultraviolet radiation. In parallel, information for the 5 day forecast is: Temperature (actual temperature, thermal sensation in sun and shade); Rain (amount and probability); Clouds (percentage of coverage); Wind (speed and direction); Sun (sunrise and sunset times and daily duration); and Moon (sunrise and sunset times). Rain, cloud and wind data can be filtered by day or night.

Figure 4(B) shows the twelve hour forecast of the thermal sensation. Finally, Figure 4(C) shows the percentage of cloud coverage during the day, for five days of the week.

## 5 Conclusions

*Plante* brings a proposal for integrating agriculture with Information and Communication Technology (ICT) in order to increase the efficiency of agricultural management and control processes. In addition, *Plante* allows to improve water management, becoming a green solution for smart cultivations, watering only at the exact moment that is needed, reducing overuse. We can also notice that the solution presented is energy-efficient, since it remains up to 98% of the time off.

Another important improvement is the reduction in crop losses, once the platform provides complete information about the ideal cultivation conditions for each plant species. Finally, it

Plante

Renato Oliveira, Reinaldo Braga, Carina Oliveira

is important to stress that the farmer has access to collected information of the plantation anytime and anywhere.

As future works, we intend to integrate *Plante App* with user notification and alert mechanisms, alerting the user what actions he should take, without the need to open the application. In addition, we intend to use artificial intelligence to ascertain whether the *Plante* Data Base, which contains optimal cultivation information. This will be done by crossing the existing data in the database with the data obtained by the sensors and the degree of effectiveness pointed out by farmers. Finally, the application, intelligently and autonomously, will make modifications to the database based on known experiences, which will result in much more accurate information.

## References

- [1] Roger A Light. Mosquitto: server and client implementation of the MQTT protocol. *The Journal of Open Source Software*, 2(13):265, 5 2017.
- [2] AccuWeather. Accuweather api, 2019. <https://developer.accuweather.com/> (last accessed 20/12/2019).
- [3] Instituto Brasileiro de Geografia e Estatística. Levantamento e classificação de uso da terra. IBGE - Instituto Brasileiro de Geografia e Estatística, 2010.
- [4] Agência Nacional de Águas. Atlas irrigação. In Superintendência de Planejamento de Recursos Hídricos (SPR), editor, *Uso da Água na Agricultura Irrigada*. Ministério do Meio Ambiente (MMA), 2017.
- [5] LTD ESPRESSIF SYSTEMS (SHANGHAI) CO. Espressif, 2019. <https://www.espressif.com/en> (last accessed 20/12/2019).
- [6] Food and Agriculture Organization of the United Nations. Crop prospects and food situation. In *Global Information and Early Warning System on Food and Agriculture*. FAO - Food and Agriculture Organization of the United Nations, 2019.
- [7] Y. E. M. Hamouda and B. H. Y. Elhabil. Precision agriculture for greenhouses using a wireless sensor network. In *2017 Palestinian International Conference on Information and Communication Technology (PICICT)*, pages 78–83, May 2017.
- [8] Estêvão B. Saleme Celso A. S. Santos José G. Pereira Filho Jordano R. Celestrini, Renato N. Rocha and Rodrigo V. Andreão. An architecture and its tools for integrating iot and bpmn in agriculture scenarios. *The 34th ACM/SIGAPP Symposium on Applied Computing*, 2019.
- [9] United Nations. World population prospects 2019. In Population Division, editor, *Highlights*. United Nations - Department of Economic and Social Affairs, 2019.
- [10] World Resources Report. Creating a sustainable food future. In Emily Matthews, editor, *A Menu of Solutions to Feed Nearly 10 Billion People by 2050*. World Resources Report, 2018.
- [11] Mare Srbinovska, Cvetan Gavrovski, Vladimir Dimcev, Aleksandra Krkoleva, and Vesna Borozan. Environmental parameters monitoring in precision agriculture using wireless sensor networks. *Journal of Cleaner Production*, 88, 05 2014.
- [12] OASIS Standard. MQTT Version 3.1.1, 2014. <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.pdf> (last accessed 20/12/2019).
- [13] NodeMcu Team. nodemcu@nodemcu.com, 2019. [https://www.nodemcu.com/index\\_en.html](https://www.nodemcu.com/index_en.html) (last accessed 20/12/2019).

# *The CAARF approach towards Monitoring and analysis of Contextual data within SDN networks*

Constantino J. Miguel, Francisco Badaró Neto, Paulo N. M. Sampaio  
Computing and Systems Graduate Program (PPGcomp),  
Universidade Salvador (Unifacs)  
Salvador, Bahia, Brasil  
{constantino.jacob, fjbvneto, pnms.funchal}@gmail.com

*Abstract* — Software-Defined Networks (SDN) modified the way network traffic can be monitored and analyzed, allowing for new approaches for network optimization. This paper presents an approach for monitoring and analyzing context-based data flow considering the global state of the computational environment (users, presentation/communication devices, and network infra-structure). The proposed solution relays on the development and deployment of an API called *Context-Aware Adaptive Routing Framework (CAARF-SDN)*. CAARF-SDN is a context-based service that provides the monitoring and optimization of the SDN traffic. In this paper we present the CAARF-SDN based monitoring through the implementation of a virtual network using Mininet in order to analyze contextual information such as Quality of Service (QoS), Quality of Device (QoD) and Quality of Experience (QoE). For this purpose, different scenarios are proposed in order to validate the so-called network Quality of Context (QoC) which is related to the global state of the computational environment and allows for the optimization of the network.

**Keywords:** Quality of Service, Quality of Experience, Quality of Devices, Software-Defined Networks, Network Monitoring, Contextual Analysis.

## I. INTRODUCTION

In 2017 the global data traffic on the Internet reached 1,5 ZB (Zettabytes), with an estimate to climb up to 4,8 ZB in 2022 [1]. With the growing demand of bandwidth, different contributions are required in the traffic engineering domain in order to optimize the use of network resources. Therefore, new proposals for the network optimization need to be developed, and in this context, monitoring network data is crucial for the correct implementation of optimization strategies.

The framework *Context-Aware Adaptive Routing Framework applied to SDN (CAARF-SDN)* [2] [3] was proposed as a solution for the automatic analysis and optimization of the network resources. This framework proposes the analysis of different metrics (Key Performance Indexes - KPI), which can be Quality of Service (QoS), Quality of Device (QoD) and Quality of Experience (QoE). QoS is related to the traffic delivery under the infrastructure's perspective. QoD is related to presentation and communication device's performance and features. QoE describes the quality of presentation under the user's perception of the service delivered. The cutting-edge aspect about CAARF-SDN is related to the integrated deployment of these three metrics in order to provide the optimization of SDN traffic to improve user's perception of the service delivered.

The architecture of CAARF-SDN is composed of four main modules, depicted in Figure 1: MONITORING, OPTIMIZATION, DEPLOYMENT AND INTEGRATION. The MONITORING module is responsible for coordinating the monitoring activities of the required data (KPIs) in order to determine the Quality of Context (QoC) and notify the OPTIMIZATION module of a context modification. The OPTIMIZATION module is responsible to determine the optimal path for traffic delivery based on the QoC. The DEPLOYMENT module is responsible to generate the configuration directives of the optimal path. The INTEGRATION module is responsible for adapting the DEPLOYMENT module's configuration directives to the format suitable to the adopted SDN's controller. Therefore, the former module makes CAARF-SDN agnostic to any type of network paradigm such as SDN, GMPLS, ASON, among others [2]. This paper details the MONITORING module and its interactions with all the CAARF-SDN modules. Therefore, the main contributions of this paper are:

- 1) The presentation of a dynamic and scalable solution to provide the context-based traffic optimization through the proposal of CAARF-SDN networks;
- 2) The proposal of a context model based on the concepts of Quality of Service (QoS), Quality of Device (QoD) and Quality of Experience (QoE);
- 3) The deployment of a generic context model based on a JSON notation to describe the context notification provided by the user, presentation and communication devices;
- 4) The introduction of the implemented architecture of CAARF-SDN identifying its main modules and their interactions;
- 5) At last, a further presentation of the MONITORING module discussing and illustrating how it works using Mininet [4].

This paper is organized through the following sections: Section II presents some related works; Section III discusses how the QoE is obtained and how the KPIs are applied to determine the QoC; Section IV introduces the main aspects entailing the monitoring of contextual information; Section V describes the experiment carried out, the emulation of an SDN network and the CAARF-SDN monitoring using Mininet in order to validate the concepts proposed, and; Section VI presents some conclusions and future perspectives.

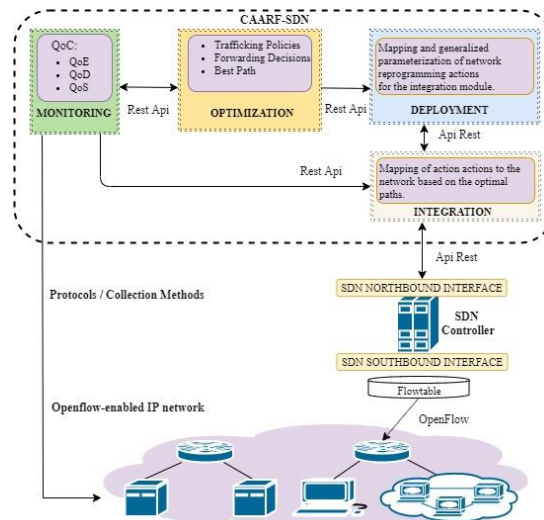


Fig. 1 Conceptual Architecture of CAARF-SDN.

## II. RELATED WORKS

The monitoring of data within SDN networks is based on the combination of information managed by the SDN controller concerning the status of the network. A monitoring tool should meet the following requirements: scalability, non-intrusiveness, interoperability, robustness and



fault-tolerance. Nevertheless, these requirements are not fully implemented within several solutions due their implementation complexity [5].

Therefore, different monitoring tools were studied and compared. One of these contributions identified in the literature is work of Tambourine [6] that introduces a software of reverse proxy using the HTTP protocol to communicate through an API REST with monitored devices using the SNMP [7] protocol. The main goal is to come up with a monitored information cache in order to prevent overloading the Network Management System (NMS) server.

Similarly, PayLess [8] presents a framework for monitoring SDN networks through its Northbound interface. This solution offers a RESTful API to support other types of applications compatible with this interface. Nevertheless, it generates a JSON-like information as an output, which narrows down the integration with other tools.

MonSamp [9] introduces a SDN monitoring application for QoS data based on its traffic. Two main aspects are outlined: (1) It is possible to monitor flows without affecting performance, and; (2) SDN switches of different providers and virtual switches can present different behaviors, even if they execute the same version of OpenFlow, which requires a further tuning of the monitoring process.

Compared to the previous contributions, the CAARF-SDN MONITORING module provides: (i) Monitoring user information (QoE), network environment (QoS) and presentation and communication devices (QoD); (ii) Mapping the network topology in real-time, and; (iii) Analysis of the monitored information, verification of the global state of the computational environment (QoC) and notification of contextual modification towards the OPTIMIZATION module. Table I outlines the main differences among the solutions studied.

TABLE I  
COMPARATIVE AMONG DATA MONITORING STRATEGIES

Solution	Access SNMP	Access API SDN Controller	QoS Monitoring	QoD Monitoring	QoE Monitoring
Tambourine	X		X	X	
PayLess		X	X		
MonSamp		X	X		
CAARF- SDN	X	X	X	X	X

Next section discusses further aspects about contextual information monitored.

### III. MONITORING CONTEXTUAL INFORMATION

As introduced in [2] [3], the notion of QoC was proposed in order to evaluate accuracy of contextual information related to QoS, QoD and QoE. Therefore, to determine QoC the Key Performance Indexes (KPIs) were proposed, such as [2]:

- QoS: delay, jitter, packets loss, bandwidth, throughput, etc.;
- QoD: CPU usage, Memory usage, Battery level, GPS positioning, etc., and;
- QoE: estimated MOS and factor R.

As a reminder, a *Mean Opinion Score (MOS)* [10] is a numerical measure of the human-judged overall quality of an event or experience. In this approach, MOS is derived, as an adaptation of factor R, using metrics such as delay, jitter and packet loss, as presented in [2], and illustrated in Figure 2.

Estimated derivation of MOS	$\text{EffectiveLatency} = (\text{AverageLatency} + \text{Jitter} * 2 + 10)$ $\text{if } (\text{EffectiveLatency} < 160) \text{ then}$ $R = 93.2 - (\text{EffectiveLatency} / 40)$ $\text{else}$ $R = 93.2 - (\text{EffectiveLatency} - 120 / 10)$ $R = R - (\text{PacketLoss} * 2.5)$ $\text{MOS} = 1 + (0.035) * R + (.000007) * R * (R-60) * (100 - R)$
-----------------------------------	--

Fig. 2 Estimated derivation of MOS

The proposed KPIs in [2] are extensible to any traffic metrics supported by the monitoring protocol, as for instance some details of the traffic features that can be monitored using Netflow/IPFIX. In other words, any metrics monitored using ICMP, SNMP, Neflow v9/IPFIX [11].

#### IV. MONITORING MODULE

The CAARF-SDN MONITORING module was proposed to support several types of technologies and network topologies. The architecture of the MONITORING module is depicted in Figure 3. This architecture is composed of the following submodules:

- *Network Management System (NMS)*: this module is supported by Prometheus [12], which supports SNMP communication and provides an API that facilitates integration to different solutions;
- *Integration API Reader*: this module is responsible for the access to the *Integration* module, which in turn is configured to access the SDN controller;
- *Topology Mapping*: module that processes the information from *NMS* and *Integration API reader* modules. This module aims at storing network topology in a database belonging to the *Context Management* module;
- *Context Management*: module that stores the information about the current state of the network. This module is composed of the following components: (i) *Network Logging Records*, responsible for storing all the monitored information; (ii) *QoC Analysis*, which is responsible for the QoC derivation from all the monitored ports; (iii) *Notification*, when the QoC values fall under certain acceptable limits a notification of context modification should be sent to the OPTIMIZATION module, and;
- *API Reader Service*: responsible to communicate with the API service agent running on the user presentation device (client). The main goal is to extend the scope of information monitored in order to improve the accuracy of the QoC calculus.

Next section describes the derivation of the QoC within a simulated environment.

#### V. EXPERIMENT AND ANALYSIS

In order to illustrate the application of CAARF-SDN this paper applies scenarios using audio and video stream applications. Due to the nature of these applications, traffic conditions variation causes a loss on quality perception providing a poor user experience. One of the goals of this work is to illustrate how user experience is improved through the implementation of the proposed model based on a network scenario within a controlled domain.

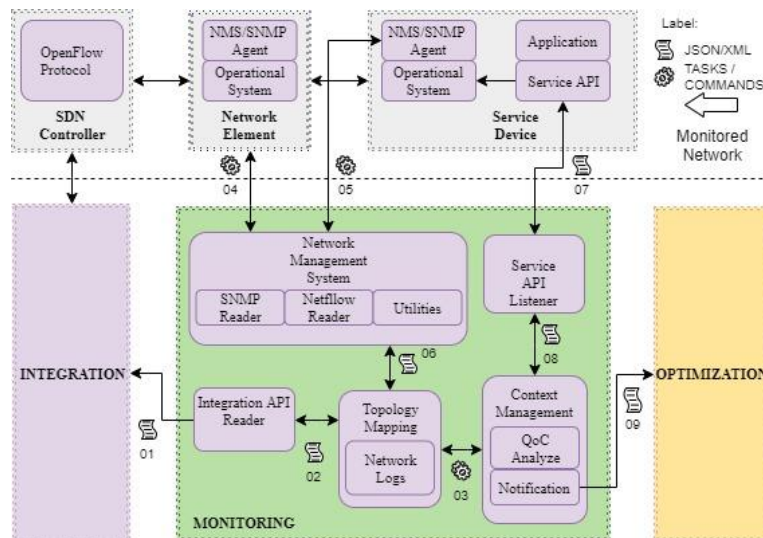


Fig. 3 Architecture of the MONITORING module.

The experiment aims at unveiling how the QoC is calculated in each part of the components of the network, relating this information with user experience determined by the estimated QoE, leading to a traffic optimization.

The experiment was carried out in an environment with virtual machines with support to Mininet. Mininet creates a testing environment that enables the development of programs for SDN networks, allowing the experiment to be easily replicated if needed.

The environment was simulated in a computer with Intel Core i7 processor with 16 GB RAM. The virtualization program deployed is the Oracle VirtualBox [13]. Three virtual machines were configured according to Table II.

TABLE II - CONFIGURATION OF THE VIRTUAL ENVIRONMENT

VM	Software	RAM (MB)	CPU	Disk (GB)	OS
1	Mininet	1024	1	8	Ubuntu
2	HP SDN Controller	3072	2	150	Debian
3	Prometheus Grafana CAARF-SDN MongoDB	2048	2	100	Ubuntu

The software Mininet is executing on virtual machine 1 and is responsible for managing the virtual network depicted in Figure 4. The software HP VAN SDN CONTROLLER [14] is executing on virtual machine 2 which is responsible for maintaining the SDN controller. At last, virtual machine 3 hosts the monitoring applications such as Prometheus [12] to support capturing QoE and SNMP information; CAARF-SDN to monitor and optimize traffic based on the context of the computational system; NoSQL and MongoDB [15] which are applied by CAARF-SDN as database repository, and; Grafana [16] for the visualization of results in real-time.

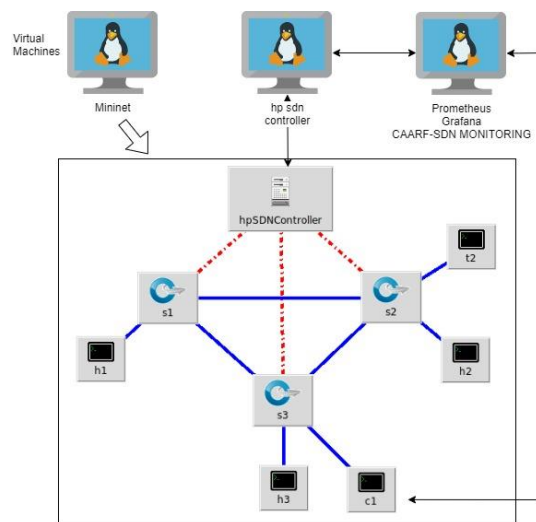


Fig. 4 Mininet visualization of the testing scenario

The virtual topology of the network depicted in Figure 4 is composed of the following elements:

- h1, h2 and h3 are the hosts and servers within this topology;
- t2 is the server that generates noise to the communication, and;
- c1 is the gateway between Mininet and the external environment, in this case the virtual machine 3.

This topology can also be visualized through the management panel of the HP VAN SDN CONTROLLER.

The Mininet network is created using a Python script. Each element of the network is created using Open vSwitch. Both Mininet and Open vSwitch were updated to support version 1.3 of OpenFlow (they originally support version 1.0). This new version impacts on how instructions are

passed to the SDN controller, considering that each provider has its own format, which reinforces the need for the CAARF-SDN's INTEGRATION module.

When the virtual machines are turned on several basic services are initialized. The CAARF-SDN's MONITORING module executing over a Node.js service works as an orchestrator of different available monitoring and analysis tasks. The MONITORING module generates a request message to the SDN controller through the INTEGRATION module each 5 seconds in order to capture the state of all the traffic variables. The MONITORING module also expects the QoE information from the presentation devices, as well as the SNMP information captured by Prometheus. At this initial moment the SDN controller's flow table is empty.

The next step is the execution of a script that configures the Mininet testing environment. Once the virtual network is created, and the contact with the SDN controller is established, the system creates the basic flow tables to provide the configured topology. At this moment, three Linux xTerm terminals are launched to initialize the traffic simulation. The Linux program sipp [17] simulates a VoIP traffic emulating the SIP protocol. Initially sipp is configured to a rate of 30 calls per seconds between client h2 and server h1.

The sipp's priority is modified within Linux. Therefore, since one of the metrics monitored is the CPU usage for each component of the network, the sipp's traffic within the virtual environment should be monitored to enable a better result evaluation.

Despite sipp generates estatistical data for each call, the estimated QoE requires data related to delay, jitter and packet loss between h1 and h2, as depicted in Figure 2. In order to acquire this information, the program mtr [18] was applied. The mtr output information were configured using JSON format in order to be sent to CAARF-SDN's MONITORING module.

In this moment, the traffic situation would be considered as normal. The QoC for each port of the components of the network, as well as the users's QoE, which would have a value, according to [2] [3], above 2 which would be the minimal acceptable to a VoIP quality. The data monitored is often stored and analyzed by the MONITORING module, in other words, the QoC is recalculated in each access to the SDN controller.

The next step of the simulation is the noise generation. For this purpose, the software iperf was deployed using server t2. The exceeding traffic will overload the system, consequently modifying the QoC and QoE. Nevertheless, the notification of optimization will only be triggered when the user's QoE is close to a critical value.

Figure 5 illustrates this situation demonstrating this correlation between CPU usage in the components of the network and variation of the user's QoE.



Fig. 5 Testing scenario for the derivation of QoC using Mininet

As previously described, as soon as an overload on the components of the network takes place (indicated on the graph of Figure 5 with label “cpu load ovs”, which is related to the CPU usage of each component of the Open vSwitch network) the QoE drops, triggering the notification process to the OPTIMIZATION module. As the noise goes down, or with the dynamic reprogramming of flow table by the DEPLOYMENT and INTEGRATION modules, the traffic status becomes regular.

When reprogramming flow tables, the optimized flow tables will have higher priority than those initially configured. Consequently, the new tables have a higher preference on the traffic forwarding within the network component.

In order to validate the strategy for obtaining QoC on the simulated environment, the CPU usage has been chosen to be monitored from the ovs-vsitchd as information of the load of the component of the network. This process when submitted to a high traffic load did not exceed 5% of total CPU usage. This was a strong limiting factor for the execution of tests in this experiment. Therefore, in order to validate the process, a multiplying value was proposed to demonstrate a CPU usage higher than the real usage.

All the data monitored and also the calculations results are stored by the MONITORING module within MongoDB. Parameters such as rx\_ratio, tx\_ratio, collision\_ratio and qos originally did not exist on the SDN controller being generated by the MONITORING module.

It is important to note that Mininet has some limitations such as [4]:

- The current Mininet based networks cannot exceed available CPU or bandwidth on the same server. This limitation has been experimented in this work since the measurement of the QoD metric CPU usage turns out to be inaccurate, and;
- Currently Mininet cannot execute applications or OpenFlow Switches non-compatible to Linux. This limitation has not been experimented in this work.

Some other generic limitations observed during the execution of this experiment due some features of Mininet were related to the observation of some specific details, as for instance when monitoring performance of a device with switching control based on dedicated ASIC compared to a device with switching control based on software on a non-dedicated hardware. This observation was motivated by the existing differences of specific features of internal components of devices which vary among producers. These differences cause an important impact on the network performance, and often are weighted when choosing these devices.

Another limitation observed is related to the fact that Mininet is based on the virtualization of systems. Therefore, the adoption of Mininet can be problematic when the virtualization overhead is a drawback when executing simulation scenarios with high computing demand. Consequently, sharing resources with the hosting server makes it difficult to experiment a real performance analysis. For instance, sharing CPU and memory between the emulated environment and the host server, would be a limiting aspect that would compromise accuracy of the KPI monitored (CPU usage).

## VI. CONCLUSIONS

This paper introduced the development of the CAARF-SDN’s MONITORING module, illustrating the derivation of QoC from the monitored QoS, QoE and QoD. Although QoC is a key aspect for optimizing traffic within this approach, this aspect is out of the scope of this paper.

The main contribution of this work is to demonstrate how different KPIs can be monitored and collected from different sources, providing an integrated view of the monitored environment aiming at improving the user’s experience. Nevertheless, the complexity of this solution is high given the need to integrate heterogeneous monitoring technologies each one specific to the monitored device.

As for future works, these experiments should be carried out in a real environment, if possible, using devices from different producers, in order to achieve a higher accuracy concerning the monitored KPIs.

## Bibliographic References

- [1] "Cisco Visual Networking Index: Forecast and Methodology, 2017–2022," Cisco, 27 02 2019. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>. [Access in 15 10 2019].
- [2] F. J. B. V. Neto, C. J. Miguel, J. A. Santos e P. N. M. Sampaio, "Context-based Dynamic Optimization of Software Defined Networks," ADVANCE 2018, 6th International Workshop on ADVANCES in ICT Infrastructures and Services, Santiago De Chile, Chile. ISBN 978-2-9561129., 11-12 Janeiro 2018.
- [3] C. J. Miguel, F. J. B. V. Neto e P. N. M. Sampaio, "Data collection in SDN networks with contextual analysis," em 14th Iberian Conference on Information Systems and Technologies (CISTI), Coimbra, Portugal. doi: 10.23919/CISTI.2019.8760999, 2019.
- [4] "Mininet," ONF, [Online]. Available: <http://mininet.org/>. [Access in 01 10 2019].
- [5] S. Taherizadeh, A. C. Jones, I. Taylor, Z. Zhao e V. Stankovski, "Monitoring self-adaptive applications within edge computing frameworks: A state-of-the-art review," Journal of Systems and Software, vol. 136, pp. 19-38, 2017. doi: 10.1016/j.jss.2017.10.033.
- [6] T. Song, Y. Kawahara e T. Asami, "Cache management algorithm of load balancer for large-scale SNMP monitoring system," IEEE Globecom Workshops (GC Wkshps), pp. 901-905, 2013.
- [7] IETF Internet Engineering Task Force, "User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)," 12 2002. [Online]. Available: <https://tools.ietf.org/html/rfc3414>. [Access in 27 09 2017].
- [8] S. R. Chowdhury, M. F. Bari e R. Ahmed, "'PayLess: A low cost network monitoring framework for Software Defined Networks,'" 2014 IEEE Network Operations and Management Symposium (NOMS), Krakow, 2014, pp. 1-9., [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6838227&isnumber=6838210>. [Access in 18 01 2018].
- [9] D. Raumer, L. Schwaighofer e G. Carle, "'MonSamp: A distributed SDN application for QoS monitoring,'" 2014 Federated Conference on Computer Science and Information Systems, Warsaw, 2014, pp. 961-968., [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6933120&isnumber=6932982>. [Access in 09 06 2018].
- [10] ITU-TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU, "P.800.1 : Mean opinion score (MOS) terminology," 07 2016. [Online]. Available: <https://www.itu.int/rec/T-REC-P.800.1>. [Access in 09 03 2018].
- [11] IETF Internet Engineering Task Force, "Cisco Systems NetFlow Services Export Version 9," 10 2004. [Online]. Available: <https://tools.ietf.org/html/rfc3954>. [Access in 01 02 2018].
- [12] The Linux Foundation, "Prometheus," [Online]. Available: <https://prometheus.io/>. [Access in 01 10 2019].
- [13] "Oracle VM VirtualBox," Oracle Inc., [Online]. Available: <https://www.virtualbox.org/>. [Access in 01 10 2019].
- [14] Hewlett Packard Enterprise, "HPE VAN SDN Controller 2.7 Administrator Guide," [Online]. Available: [http://h20628.www2.hp.com/km-ext/kmcsdirect/emr\\_na-c05028095-1.pdf](http://h20628.www2.hp.com/km-ext/kmcsdirect/emr_na-c05028095-1.pdf). [Access in 02 10 2019].
- [15] "mongoDB," mongoDB, Inc., [Online]. Available: <https://www.mongodb.com/>. [Access in 2019 10 29].
- [16] "Grafana," Grafana Labs, [Online]. Available: <https://grafana.com/>. [Access in 28 10 2019].
- [17] "sipp," [Online]. Available: <http://sipp.sourceforge.net/>. [Access in 15 10 2019].
- [18] "MTR," [Online]. Available: <https://github.com/traviscross/mtr>. [Access in 15 10 2019].

# Performance analysis of computational offloading on embedded platforms using the gRPC framework

Mateus S. Araújo<sup>1</sup>, Marcio E. F. Maia<sup>1</sup>, Paulo A. L. Rego<sup>2</sup>, and José N. de Souza<sup>2</sup>

<sup>1</sup> Campus Quixadá – Universidade Federal do Ceará  
Quixadá, Ceará, Brasil

`mateuseng_ec@alu.ufc.br`, `marcioefmaia@ufc.br`

<sup>2</sup> Departamento de Computação – Universidade Federal do Ceará  
Fortaleza, Ceará, Brasil

`pauloalr@ufc.br`, `neuman@ufc.br`

## Abstract

Embedded systems are becoming increasingly accessible to the Internet allowing the creation of new services and applications. Such systems need to communicate in a structured form in a way that uses standardized technologies for better results. Mobile systems also have a number of limited features like battery life, internal storage and processing performance. Such restrictions can be mitigated by the use of computational offloading since algorithms or applications can be executed in the cloud or other networked devices. This article is intended for the analysis of emerging technologies in cross-process communication between Linux and Android-based multiplatforms using the gRPC framework. Applications have been developed in various object-oriented programming languages for performing remote procedure calls between a single-board computer and a personal-use smartphone for processing higher order arrays and applying filters to images. Then, a series of analyzes were performed on the transferred data and the computational offloading performance of the algorithms in each platform.

## 1 Introduction

The Internet and the advance of hardware platforms is reshaping the mechanisms used to create networked embedded applications, with new opportunities to improve our lives rising on a daily basis. In that direction, forecasts made by hardware manufacturers indicate an order of 56 billions connected devices in 2020 [13]. That is a result of a reduction in the cost to create new embedded solutions, an improvement in the processing capability of these devices and a reduction in energy consumption [5]. Another dimension going under improvement is the communication using wireless and application protocols to permit embedded devices to interact using the Internet. Hence, it is now possible to create a device running a Linux distribution for embedded devices and use most of the open-source communication protocols designed for unix-based operating systems on a resource-constrained embedded device [14]. The evolution of embedded technologies and communication protocols to connect devices using the Internet is known as the Internet of Things (IoT).

The IoT is providing mechanisms for novel solutions to emerge, based on algorithms created using several research domains, such as Artificial Intelligence, Computing Vision, Machine Learning, Image and Natural Language Processing, among others. One common issue in the IoT is the low response time requirements imposed on these algorithms [10]. Response time is usually a function of the input data and the processing time, and it is affected by the processing capabilities of the devices and the communication protocols.

In IoT systems, aiming to achieve the shortest response time possible, developers usually have to decide between processing everything locally in the embedded devices, or send all the collected data to another (more capable) device to process it remotely. However, the time to send all the input data to remote devices may reduce or cancel out the benefits of processing it in a more capable device. In that direction, several work are studying the benefits and challenges of using Computation Offload to reduce the burden on embedded devices and improve the overall execution time [11].

Computational offloading is dealt with in this paper with a performance analysis on embedded platforms using the gRPC framework. Here, two processing-intensive applications were implemented, and the execution time compared against with and without offloading. Moreover, these applications were developed using 5 different programming languages, executed on three different operating systems for embedded devices and run on two different devices. The goal was to analyze their impact on the overall performance on embedded systems.

The remainder of this paper is composed of section 2 with a discussion on RPC for embedded systems. Section 3 presents the related work. Section 4 presents the experiment design to evaluate computational offloading using gRPC. Section 5 highlights the main results from the experiments carried out. Finally, section 6 presents some conclusions and future work.

## 2 RPC on Embedded systems

Remote Procedure Call (RPC) permits system modules running on different address spaces to interact. Using RPC, one module is capable of invoking procedures from another module, receiving the consequent return of the result data. RPC can be achieved by specifying procedures able to be called remotely and the individual input and output parameters in each procedure. Thus, when different parts of an application need to communicate, they do so by serializing data and sending them throughout the network. One common feature used by RPC implementations to enforce interoperability is the use of Data Description Languages (DDL), which is a language to specify data structures, to permit interaction between remote modules, regardless of the platform and language used in the procedure implementation.

A data description language commonly used by cloud applications today is Protocol Buffer. According to [7], Protocol Buffer is a neutral feature that enables data structuring for multi-level communication across the network. Protocol Buffer is an alternative to JSON (Javascript Object Notation), a widely used notation to represent data in web applications[16]. Although JSON has advantages such as good human-readability, it has several limitations as well. Data transferred using JSON is not encoded by browsers and is transferred as clear text, which presents important performance and energy consumption restrictions in embedded systems. Alternatively, Protocol Buffer is designed to have smaller memory footprint (as opposed to JSON), making it more efficient to exchange messages across cross-platform and embedded applications [7].

Using Protocol Buffers as a DDL to enforce interoperability, gRPC<sup>1</sup> is a general purpose framework to permit RPC interactions based on HTTP/2 as transport to traverse proxies and firewalls. There are a number of advantages of using the GRPC framework on various devices such as desktops and mobile devices. Among the highlighted advantages, the low-memory footprint can be cited. Applications using Protocol Buffer serialize the data in a structure called proto file, regardless of the language used. Additionally, the ease with which applications can be developed just by defining the proto file interface stands out. Once data structuring is

---

<sup>1</sup><https://grpc.io>



defined, it is transparent to the developer how communication between different platforms will take place, since GRPC and Protocol Buffer together handle data parsing between the most diverse types of object oriented languages such as Java, C ++, Python, among others.

Embedded systems are traditionally defined as resource-limited platforms with specific processing purposes. Most of these systems require light programming languages for optimal performance. In that direction, Protocol Buffer is an interesting solution, since it is optimized for low-memory footprint and reduced communication latency [15]. In addition, Protocol Buffer makes interoperability between distributed applications much simpler [7], depending only on a common file for defining the data to be transferred.

To improve the overall execution performance in terms of processing time, memory and energy consumption, developers usually have to decide between processing everything locally or to send all the collected data to another device to process it remotely. However, the time to send all code and data may reduce or cancel out the benefits of processing it in a more capable device. In summary, computation offloading is the set of techniques to permit algorithms to be offloaded to remote devices, trying to improve performance metrics in resource-constrained devices.

### 3 Related Work

The authors in [4] present performance and efficiency information for an embedded system using protocols for structured information exchange based on remote procedure calls. Such a feature, according to the paper, is an interoperable mechanism to exchange information between networked embedded platforms. They compare the use of SOAP and XML-RPC applications through data processing in embedded systems and mobile devices. For this, an engine monitoring system was developed through a Controller Area Network (CAN) in which the server was developed in C/C ++ language using XML-RPC. As a result, a Windows-based computer operates as a remote machine that receives all information through data serialization via the DS80C400 microcontroller, which in turn has CAN support and 1-wire communication by transmitting data over the network between vehicle control modules. The system proved to be efficient when RPC was used instead of SOAP for communication and information exchange between the computer and the embedded integrated chip.

RPC has proven to be an useful protocol for embedded systems. However, there is also the challenge of monitoring and controlling applications using wireless networks. The Marionette system implemented in [15] demonstrates the use of RPC to facilitate the development of applications with data being transferred between multiple devices wirelessly. It demonstrates how RPC can be used in applications needing a better distribution of their services and consequently a more efficient serialized data structure. The proposed application demonstrates the use of an RPC-based framework to permit the communication between a desktop computer and mobile devices through programming languages, such as Java, Python or Bash Script. The networked computer allows sensor data acquisition as well as sensor configuration through remote methods found on the server. In addition, the system is also capable of obtaining sensory network information through client/server applications implemented in other specific languages such as nesC for network programming using embedded systems and operating systems specific to such applications such as TinyOS.

The Any Run Computing (ARC)[6] architecture proposes a dynamic offloading model, demonstrating energy efficiency and latency reduction using resources available on cloud devices. According to the authors, computational offloading is considered advantageous compared to local execution, as internal resources such as CPU and memory usage can be used efficiently

while data is processed and organized in external services. The ARC architecture is implemented in Java with SCAMPI framework support and works both on Dalvik Android-based virtual machines and on Oracle's Hotspot virtual machines, used on desktop computers. Their performance evaluation showed how the ARC architecture was able to reduce offloading time by transferring applications that used high computational resources to the cloud. A more detailed analysis is required to understand the offloading impact in other parameters such as battery or network.

The authors on [9] demonstrate several results executing specific algorithms and techniques on mobile devices. Depending on the programming language used and the hardware architecture of the device, the offloading time may vary sharply. A Java application was developed to simulate an agent for resource access on a remote Linux machine. As more resources were requested, the slower the offloading computation time was perceived. However, if the agent was implemented in C/C++ using native libraries from the embedded device, the more efficient the overall offloading process was carried out.

The authors in [2] conducted experiments to evaluate the energy consumption of Android devices when using different communication protocols and architectural styles, such as REST, SOAP, Socket and gRPC, and executing classical sorting algorithms of different complexities and different types and input sizes. Their results show that local execution is more economic with less complex algorithms and small input data. When it comes to remote execution, REST is the most economic choice followed by Socket, they show that computation offloading can save up to 10 times as much energy when compared to local execution for some executions configurations. Surprisingly, gRPC did not achieve good results in their experiments. It is not easy to explain such a result, but we guess it might be caused by the type of application used in the experiment: sorting algorithms, in which data serialization is quite simple, decreasing the benefits gained from using gRPC.

Several studies [2, 4, 6, 9, 15] have been using remote invocation techniques to perform computation offloading between various systems to improve application's performance and to reduce device's energy consumption. However, to the best of our knowledge, few of them consider different communication protocols and none of them have evaluated the influence of different programming languages and embedded systems in this context.

## 4 Evaluating computational offloading using gRPC

Computational offloading permits application developers to migrate the execution of processing- and memory-intensive algorithms from embedded devices to more capable devices. The goal is to reduce execution time or to reduce energy consumption. Moreover, the decision to migrate it is based on the type of algorithm, network latency and size of the input parameters. In that direction, gRPC is a useful underlying technology to implement the migration of code and data.

We have implemented applications in various object-oriented programming languages that leverage the gRPC framework for performing computation offloading. Our goal is to assess the offloading performance on heterogeneous environments, using different programming languages as well as the following Linux and Android-based devices:

**BeagleBone:** a development kit based on 32-bit ARM processor, capable of performing over 3 million Dhrystone and floating-point vector arithmetic operations per second [3]. It is widely used and has support for various Linux distributions like Debian, Angstrom and Ubuntu and can also run Android for various applications. In addition, BeagleBone allows

Perf. analysis of comp. offloading on embedded platforms using gRPC Araújo, Maia, Rego and De Souza

the inclusion of image processing libraries such as OpenCV and OpenNI for object recognition in automation and robotics projects. Its main specification is a Cortex-A8 ARM AM335x processor with 1GHz clock and 512MB of DDR3 RAM. Besides, BeagleBone contains 4GB internal storage with 3D graphics accelerator and two 32-bit microcontrollers dedicated for real-time processing.

**Zenfone 5:** an Android-based smartphone running an Intel 32 bits processor, different from most of ARM-based smartphones running Qualcomm’s Snapdragon processors. It uses an Intel Atom Z2560 dual-core processor, 2GB RAM and 8GB storage [8].

Two case studies were used to measure the overall execution time with and without computational offloading. The first case study uses an application to perform multiplication of 1000x1000 matrices. The application client was implemented in Kotlin while its server version running the same algorithm was implemented in C++, Python, Java, Go, and Ruby programming languages.

In the second case study, an image processing application transfers images between a Kotlin client and Python and C++ servers in order to apply several filters using the OpenCV library. The client serializes and transfers one image to the server, which applies Cartoon, Bilateral and Gray filters [1] and sends back the three processed images.

Depending on the type of the applied filter, the final image size may vary as processing methods and parameters vary between filtering types and languages used. Figure 1 exemplifies the process of offloading and processing an image on the BeagleBone platform using the OpenCV library. The image was sent to the server, all three filters were applied and the resulting images were sent back to the client.

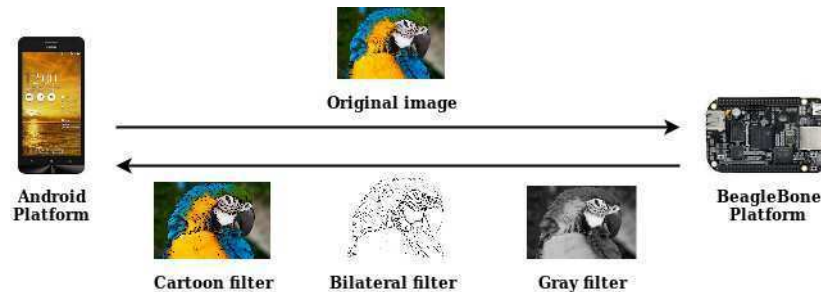


Figure 1: Overview of the image processing application workflow

The implementation of both case studies used gRPC and Protocol Buffers (Protobuf) as the underlying structure to serialize/deserialize data before sending it through the network. Using the Interface Description Language (IDL) present in Protocol Buffer, we have defined the messages to be exchanged during RPC and the Protobuf compiler generated predefined structures to each of the programming languages used. Table 1 presents the experiment design and details such as languages and operating systems used to implement the algorithms.

It is important to highlight that a Debian Linux and two Android distributions (KitKat 4.0 and Vanilla) were used to execute the applications on the BeagleBone. Android Vanilla is an OS version intended for specific use on Embedded Systems, unlike KitKat, which has features intended for smartphones [12]. For the smartphone, the application was executed on Android 5.0 (Lollipop).

Perf. analysis of comp. offloading on embedded platforms using gRPC Araújo, Maia, Rego and De Souza

<b>Factors</b>	server device operating system, server application programming language, and application.
<b>Parameters</b>	server device operating system (Debian, Android 5.0/KitKat and Vanilla), server application programming language (C++, Python, Java, Kotlin, Go, and Ruby); application (matrix multiplication and image processing).
<b>Design</b>	The experiment was executed 100 times for the matrix multiplication algorithm using 1000x1000 matrices and 50 times for the image processing algorithm using a 216 kB image.
<b>Response</b>	Offloading time for the aforementioned algorithms running on each platform using gRPC (seconds).

Table 1: Experiment Design



Figure 2: BeagleBone Black running Android OS during the performance evaluation.

## 5 Experiment results

This section presents the results of the experiments using the two case studies (matrix multiplication and image processing). Table 2 shows the mean and standard deviation of the method execution time for the matrix multiplication application (case study one). We can see that the best result was achieved using the C++ version of the application running on BeagleBone with the Debian Linux operating system. The second-best result was achieved with the Python version of the application, while the execution times are approximately the same for Java, Go and Ruby. In fact, we can see a difference of approximately 44% between the best and worst results, which shows the impact of platform choice on offloading performance. Another interesting result is that the Android Vanilla operating system presented a better performance than the Android Kitkat, which can be explained by the fact that Vanilla is an optimized version of Android for embedded devices [12].

The fact that different operating systems and languages present distinct characteristics, such as C and C++ running on top of the OS, while others run on top on a virtual machine, is another parameter to be considered, since it may present important impact in the execution performance.

As a baseline, the Local column of Table 2 shows the average execution time in seconds and the standard deviation for matrix multiplication algorithm without offloading, running on the smartphone. Here, the same pattern is present, with C++ outperforming all other languages. Additionally, we can see the benefits of offloading when we compare the execution time of the same languages running locally and remotely on a device with better processing capabilities.

Table 3 shows the mean and standard deviation of the method execution time for the image processing application (case study two). Again, as expected, the combination C++/Linux Debian outperforms the other combination, mainly because of the optimization of the Linux OS

Perf. analysis of comp. offloading on embedded platforms using gRPC Araújo, Maia, Rego and De Souza

Languages	Local (baseline)		Linux Debian		Android KitKat		Android Vanilla	
	Avg.(s)	Dev.	Avg.(s)	Dev.	Avg.(s)	Dev.	Avg.(s)	Dev.
<b>C++</b>	16.07	0.47	10.36	0.57	13.62	0.64	12.74	0.26
<b>Python</b>	18.21	0.55	12.94	1.89	16.72	0.69	14.88	0.60
<b>Java</b>	20.07	0.47	17.41	0.77	18.07	1.09	18.07	0.49
<b>Go</b>	19.97	0.57	17.39	0.78	17.50	0.65	16.67	0.25
<b>Ruby</b>	19.11	0.45	18.19	0.56	18.39	0.47	17.59	0.30

Table 2: Total execution time for the matrix multiplication application with and without offloading

for embedded platforms. The Android platforms presented statistically the same performance.

Languages	Local (baseline)		Linux Debian		Android KitKat		Android Vanilla	
	Avg.(s)	Dev.	Avg.(s)	Dev.	Avg.(s)	Dev.	Avg.(s)	Dev.
<b>Python</b>	6.89	0.58	7.24	0.97	8.15	0.6058	7.98	0.55
<b>C++</b>	6.81	0.59	5.66	0.79	5.88	0.599	5.89	0.53

Table 3: Total execution time for the image processing application with and without offloading

As a baseline, the Local column of Table 3 shows the average execution time in seconds and the standard deviation for running the image processing application on the smartphone. Comparing with the offloading times presented on Table 3, we can see that the C++ total offloading time still outperformed the local execution, even considering that the image had to be transferred from the smartphone to the Beaglebone. However, considering the Python implementation, we can see that it is worth executing the method locally because the benefits of offloading the image processing do not pay off, and the time to send the image through the network outweighs the offloading benefits.

## 6 Conclusion and Future Work

Despite using the same embedded device as a server, the choice of platform impacts overall offloading performance, causing total offloading time to vary by up to 44%. This result reveals the importance of choosing the technology stack when using embedded devices with limited resources. It is also important to highlight the advantages obtained by using gRPC as interprocess communication technology, to permit the communication between processes implemented with different programming languages. Hence, the developer can implement their components and services on the platform that suits them or perform better, which is essential in computational offloading scenarios.

The results set forth in this study indicate a wide range of possibilities and research associated with computational offloading in RPC-based embedded multiplatforms. It can be seen from the overall execution time how the programming languages and operating systems used are also relevant factors for a reduction in offloading time in most applications. Associated with this reduction, we also have the increased performance and energy savings that these analyzes can lead to overall mobile devices, as well as the use of tools that enable data transfer across the network using remote procedure calls using the open source GRPC framework.

As future work, we plan to investigate alternatives for performing offloading on embedded devices and evaluate other platforms such as Raspberry Pi and Jetson, investigate other technologies such as REST/JSON, and investigate the energy consumption of these devices.

Perf. analysis of comp. offloading on embedded platforms using gRPC Araújo, Maia, Rego and De Souza

## Acknowledgments

The authors would like to thank The Ceará State Foundation for the Support of Scientific and Technological Development (FUNCAP) for the financial support (grant number 6945087/2019).

## References

- [1] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. ” O’Reilly Media, Inc.”, 2008.
- [2] C. L. Chamas, D. Cordeiro, and M. M. Eler. Comparing REST, SOAP, Socket and gRPC in computation offloading of mobile applications: An energy cost analysis. In *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*, pages 1–6, Nov 2017.
- [3] Gerald Coley. Beaglebone black system reference manual. *Texas Instruments, Dallas*, 2013.
- [4] Suru Dissanaikie, Pierre Wijkman, and Mitra Wijkman. Utilizing XML-RPC or SOAP on an embedded system. In *24th International Conference on Distributed Computing Systems Workshops (ICDCS 2004 Workshops), 23-24 March 2004, Hachioji, Tokyo, Japan*, pages 438–440, 2004.
- [5] Gabriel B dos Santos, Fernando AM Trinta, and Paulo AL Rego. Impactos do offloading de processamento no tempo de execução e consumo energético de dispositivos m óveis. In *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, volume 36. SBC, 2018.
- [6] Alan Ferrari, Silvia Giordano, and Daniele Puccinelli. Reducing your local footprint with anyrun computing. *Computer Communications*, 81:1–11, 2016.
- [7] Gurpreet Kaur and Mohammad Muztaba Fuad. An evaluation of protocol buffer. In *Proceedings of the IEEE SoutheastCon 2010 (SoutheastCon)*, pages 459–462. IEEE, 2010.
- [8] Yoon-young Kim and Min Oh. Cover for mobile phone, May 26 2015. US Patent D730,342.
- [9] Karthik Kumar, Jibang Liu, Yung-Hsiang Lu, and Bharat Bhargava. A survey of computation offloading for mobile systems. *Mobile Networks and Applications*, 18(1):129–140, 2013.
- [10] Rainer Leupers, Miguel Angel Aguilar, Jeronimo Castrillon, and Weihua Sheng. Software compilation techniques for heterogeneous embedded multi-core systems. In *Handbook of Signal Processing Systems*, pages 1021–1062. Springer, 2019.
- [11] Yuchuan Liu, Cong Liu, Xia Zhang, Wei Gao, Liang He, and Yu Gu. A computation offloading framework for soft real-time embedded systems. In *2015 27th Euromicro Conference on Real-Time Systems*, pages 129–138. IEEE, 2015.
- [12] Anand Nayyar and Vikram Puri. A review of beaglebone smart board’s-a linux/android powered low cost development platform based on arm technology. In *2015 9th International Conference on Future Generation Communication and Networking (FGCN)*, pages 55–63. IEEE, 2015.
- [13] D Pavithra and Ranjith Balakrishnan. Iot based monitoring and control system for home automation. In *2015 global conference on communication technologies (GCCT)*, pages 169–173. IEEE, 2015.
- [14] Richard D Snyder. A cross-language remote procedure call framework. In *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, page 3822, 2017.
- [15] Kamin Whitehouse, Gilman Tolle, Jay Taneja, Cory Sharp, Sukun Kim, Jaein Jeong, Jonathan Hui, Prabal Dutta, and David Culler. Marionette: using rpc for interactive development and debugging of wireless embedded networks. In *2006 5th International Conference on Information Processing in Sensor Networks*, pages 416–423. IEEE, 2006.
- [16] Qianchuan Ye and Benjamin Delaware. A verified protocol buffer compiler. In *Proceedings of the 8th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 222–233. ACM, 2019.

# Inspiring from SDN to Efficiently Deploy IoT Applications in the Cloud/Fog/IoT ecosystem

Nada Chendeb<sup>1</sup>, Nazim Agoulmine<sup>2</sup>, and Mohammad El-Assaad<sup>1,2</sup>

<sup>1</sup> Lebanese University, Faculty of Engineering  
nchendeb@ul.edu.lb, mdalassaad@gmail.com

<sup>2</sup> University of Evry Val d'Essonne - Paris Saclay University  
nazim.agoulmine@ibisc.univ-evry.fr

## Abstract

Billion of devices are connected to the internet nowadays, more are coming in the future. Cisco assumes that 50 billion devices will be connected by 2020. It is quite difficult to manage and control the exchange of the huge amount of data generated from those connected devices. Thus the need for a new more intelligent Internet of Things (IoT) architecture for large scale networks. Based on the above needs and challenges of the IoT, Software Defined Networking (SDN) seems to be an excellent key solution. Thus, using the concepts of SDN, we aim to reduce the complexity of traditional IoT networks, this means that recent IoT networks will be more programmable, managed and controlled by software. In this work, we provide an architectural model combining SDN and IoT, and a mathematical formulation of the controller placement problem as a linear integer program optimization. At the end, we present some simulation results to show the advantages of this integration.

## 1 Introduction

Connected physical devices can communicate and send information to each other over the Internet, they can be remotely monitored, programmed and controlled. IoT applications running on the top of such infrastructure, have numerous applications in verticals such as commercial, industrial, health, etc. IoT can also be very heterogeneous in terms of functionalities and capacities. The number of deployed IoT objects is increasing significantly as a consequence the generated data from these objects is increasing significantly constituting the so called Big Data. The characteristics of diversity, dynamics of such environment associated with the generated Big Data introduce hard challenges in order to design efficient and cost-effective IoT architecture solutions. This architecture should rely on not only efficient computing infrastructure but also communication networks infrastructures that should be more dynamic, efficient and more scalable to meet up the introduced challenges (performance, cost, adaptability, etc.). We believe that SDN is a promising technology that could help IoT architecture to address these challenges. In this paper, we propose a novel architecture called Software Defined-Internet of Things (SD-IoT) architecture that aims the convergence of the two concepts. We also present a mathematical formulation to optimize the placement of SDN controllers in the IoT infrastructure. We then highlight how the proposed architecture and optimization formulation permits to improve the performances of the converged system. The rest of this paper is organized as follows: Section 2 presents an overview of IoT and SDN technologies. It also presents the current state of the art related to the integration/convergence of the SDN and the IoT. In Section 3, we propose a novel architecture that integrates IoT and SDN concepts. In section 4, we present our mathematical formulation of the controller placement problem in the form of an optimization algorithm and then some simulation results are presented. Finally a conclusion and future works are given at the end of this paper.

## 2 IoT and SDN

### 2.1 Challenges of IoT Applications Deployment

The classical approach to deploy IoT applications is to entirely deploy them in Cloud data centers. This approach makes the design of the applications simple but makes it difficult to fulfill all the non functional requirements such response time, and data rate. In this paper, we propose to leverage IoT architecture with SDN concepts to allow a better programmable instrumentation of the infrastructure based on the requirements of IoT applications at any time. This architecture is called Software Defined - Internet of Things Architecture (SD-IoT). With this approach, we envision that IoT applications could interact directly with infrastructure in order to request resources and/or deploy its own components.

### 2.2 Overview of IoT and SDN

IoT architectures have been defined in order to introduce the necessary flexibility to interconnect various devices. From the conducted state-of-art we identify several models. The basic model is a 3-layer consisting of Applications, Networks, and Perception Layers. Recent literature introduces more abstraction models, however, some other models have been proposed to add more abstractions to the IoT architecture.

Network devices are usually from different providers exhibiting proprietary and heterogeneous control functions. Control plane and data plane are both packed inside the networking devices. This packaging increases the complexity and reduces the flexibility. SDN is a solution that allows to overcome these limitation providing a software-oriented control plane (control) that is decoupled from the data (forwarding) plane. SDN introduces a three layer architecture: (a) Data plane or Device layer, (b) Control plane and (c) Applications layer. The application layer expresses the customer's needs, and communicates with the controller via northbound API. The Control plane layer is considered as the brain of the network. The controller communicates with the physical devices in the data plane via southbound APIs, e.g. OpenFlow protocol. There exist many software controllers in the market such as Ryu, Opendaylight (ODL)[3], Floodlight, NOX[4], etc.

### 2.3 State of art inter-operating SDN with IoT

The heterogeneity and the complexity of IoT devices will require a new IoT architecture with SDN capabilities to manage them and to improve the performance of the whole network. If the networks are not prepared, the flood of IoT where a lot of traffic is generated could leave the network paralysed [6].

Authors in [2] propose a framework for managing connected devices and configuring the network dynamically based on SDN. SDN is used to solve the problem of continuous changes in the network by decoupling the control plane from the data plane. The approach resembles routing table used in routers. The routing information is stored centrally. When the topology needs to be changed, the SDN controller pushes the new routing information to the nodes. Thus, it can reconfigure the system without redeploying.

In [1], authors propose an SDN-based architecture to manage the diversity of devices and networks. Their solution is based on the utilization of dockers running on the connected devices. In this architecture, IoT devices running Docker communicate together via SDN based switches and they are monitored by an SDN controller. Authors implemented their approach to validate



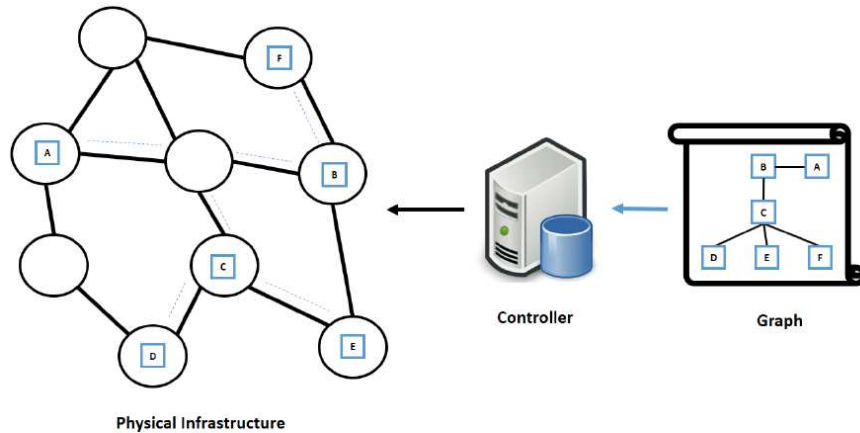


Figure 1: SDN inspired new IoT architecture

the possibility to overcome the heterogeneity of the devices as far as they are able to support containerization.

Authors in [5] propose an SDN based approach in Smart Cities. They developed a testbed highlighting how programmable SDN infrastructure can be used to significantly simplify the on-boarding and provisioning of end-to-end IoT services in a context of multi-tenant networks.

In another contribution [7], authors aim to fuse three concepts in a global IoT architecture: Mobile Edge Computing (MEC), SDN and Network Function Virtualization (NFV) to show how they can achieve better MEC employment and extend the MEC concept to provide Software-Defined Fog-enabled IoT gateways (SDF-Gateways).

Authors in [6] present a simple and general SDN-IoT architecture with NFV implementation to address the challenges of the IoT. The data layer comprising of SDN switches is not responsible of the decision making of the information being received/transmitted by/to the sensing layer rather they leave the decision making to the control layer through a southbound APIs such as OpenFlow. One of the key objectives of the proposed architecture is to replace the traditional gateways with SDN-gateways.

### 3 Our Proposed Software Defined IoT Architecture

SDN creates a high level of abstraction between the application layer and the physical layer. This centralized concept allows the controller to deal with any IoT application as a graph. The role of the controller is to guarantee the instantiating of this graph in the physical infrastructure while respecting the application requirements. Figure 1 illustrates our idea. The IoT Application specification is transformed into a graph specification of interconnected application components to submit to the controller to deploy, along with a set of requirements in terms of processing and communication resources. The controller will be in charge of the management of the deployment and execution of the application components in the physical infrastructure (the sensors and the network devices). The controller provides the required level of abstraction between the requested IoT application and the real implementation of the application components and communication links in the physical infrastructure.

We suppose that SDN concept applies also to the IoT devices themselves i.e. they are able to

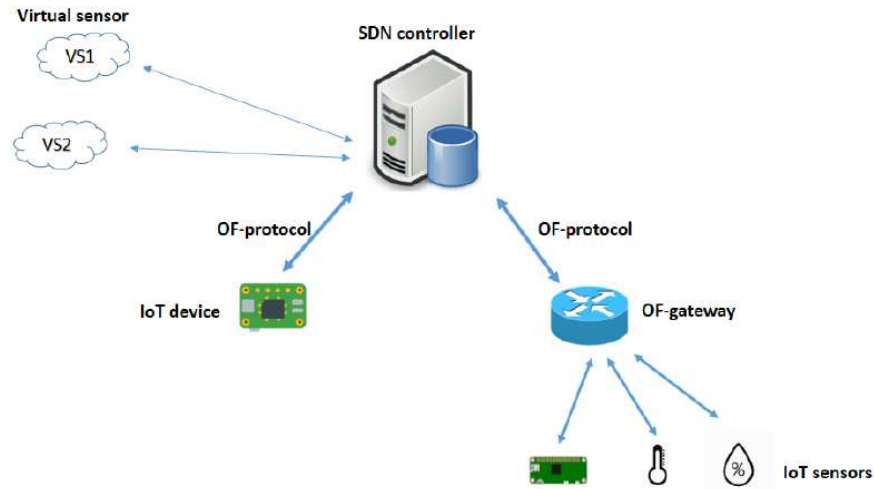


Figure 2: SD-IoT Devices connected to the controller

communicate with the controller via the OpenFlow protocol. In addition, to take into account the possibility that some IoT devices may not support the OpenFlow protocol, a gateway can help these devices to communicate with the controller by translating proprietary protocol into OpenFlow. The controller can directly update the flow table for those devices to trigger or stop the transmission of some particular data. In this way, the controller could have global control on the IoT devices and manage the generated traffic with a very coarse granularity. Figure 2 illustrates a network where IoT devices with SDN capabilities are connected directly to the controller while others are connected via an OpenFlow gateway (OF-Gateway). With this approach, the infrastructure is now decoupled from any specific IoT application. The client of the requested IoT application doesn't need anymore to worry about whether the infrastructure will be able to support the non functional requirements of its IoT application but these requirements will be taken into account by the infrastructure in a programmatic way. The physical infrastructure will be virtually shared between different IoT applications allowing for a better resource utilization.

## 4 Analytical Model of the Controller Placement Problem

In this section, we address the problem of controllers placement in the IoT infrastructure. The objective is to find the minimum number of controllers so that the response time of controllers is less or equal to a given delay bound  $d$ . The response time of the controller is an important QoS parameter. It is affected by two factors: (a) The round trip time delay between the controller and a node and (b) The sojourn time in the controller which depends on the service capacity and the load of the controller.

### 4.1 System Model

The considered system model is a set of OpenFlow switches, gateways and IoT devices deployed in a particular geographical area. The problem is to determine how many controllers should be

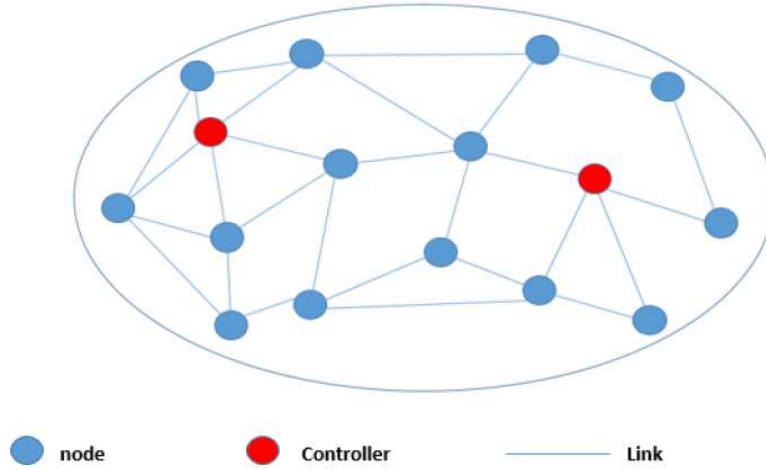


Figure 3: System Model

deployed and where so the response time constraint of the controllers is respected. We assume that a controller can be co-located with any node and share the same channel links with this node. Figure 3 illustrates our system model. The network is modeled as a graph  $G(V,E)$  where  $V$  represents the set of nodes and  $E$  represents the network links between nodes. Each link is associated with a network delay. The number of Nodes is  $N$  in the shared infrastructure. So we have  $V = v_1, v_2, \dots, v_N$  and  $v_i, 1 \leq i \leq N$ , denotes the  $i^{th}$  node.

## 4.2 Response Time and Delay Formulation

- **Network Delay Model:** The network delay denoted as  $D_{ij}^N$  is the sum of the transmission delay and the propagation delay in the network from one node to the associated controller.  $D_{ij}^N = D_{ij}^T + D_{ij}^P$  where  $D^P$  is the propagation delay and  $D^T$  is the transmission delay.
- **Controller Response Time Model:** Each controller can be modeled as an M/M/1 queuing model. Suppose packets are generated according to a Poisson process. Suppose also that all the controllers have a fixed service rate  $\mu$ . Let  $S_j$  denotes the set of nodes associated to the controller  $j$  and  $\lambda_j$  denotes the total packets arrival rate to the controller in site  $j$ , which can be calculated as follows:

$$\lambda_j = \sum_{i \in S_j} \lambda_j^i \quad (1)$$

where  $\lambda_j^i$  is the request rate of node  $i$  to the controller  $j$  [packets/s]. But arrival rate of nodes is independent of the controller site itself, and we can consider that:

$$\lambda_j = \sum_{i \in S_j} \lambda^i \quad (2)$$

According to the queuing theory, the expected mean response time  $R_j$  of controller in site  $j$  can be calculated as follows:

$$R_j = \frac{1}{\mu - \lambda_j} = \frac{1}{\mu - \sum_{i \in S_j} \lambda^i} \quad (3)$$

### 4.3 Placement Problem Formulation

The optimization problem is an NP-hard problem, which is very difficult to solve when the size of the network becomes very large. We propose instead a heuristic algorithm based on iterative approach and Dijkstra's algorithm for the least cost path problem. This algorithm aims to actively increment the number of controllers in the network by satisfying all the required constraints for all nodes in the network. The proposed algorithm is illustrated in Algorithm 1:

---

#### Algorithm 1 Controllers Placement Algorithm

---

```

1: Input  $G = (V, E, M), \delta, \mu$ 
2: Output  $X, Y$ 
3:  $R \leftarrow V$ 
4:  $C \leftarrow V$ 
5:  $D = \text{Dijkstra}(V, E)$ 
6: while  $R \neq \emptyset$  do
7:   for  $j \in C$  do
8:      $S_j \leftarrow \emptyset$ 
9:      $D_j \leftarrow 0$ 
10:     $\bar{t}_j \leftarrow 0$ 
11:    while  $\bar{t}_j \leq \delta$  do
12:       $i \leftarrow \text{NearestNode}(V, j, D, S_j)$ 
13:       $D_j \leftarrow D_j + 2D_{ij}^N$ 
14:       $\bar{t}_j \leftarrow \frac{D_j}{|S_j|+1} + \frac{1}{\mu - (\sum_{n \in S_j} \lambda^n + \lambda^i)}$ 
15:      if  $\bar{t}_j \leq \delta$  then
16:         $S_j \leftarrow S_j \cup i$ 
17:      end if
18:    end while
19:  end for
20:   $j = \text{Argmax}_j |S_j|$ 
21:   $y_j = 1$ 
22:   $C \leftarrow C - j$ 
23:  for each node  $i$  within  $S_j$  do
24:     $x_{ij} = 1$ 
25:     $R \leftarrow R - i$ 
26:  end for
27: end while
28: Return  $X, Y$ 

```

---

The network is represented by  $G = (V, E, M)$  where  $V$  is the set of nodes  $V = (v_1, v_2; \dots, v_N)$ ,  $E$  is the link's delay between the nodes,  $E = (e_{11}, e_{12}, \dots, e_{NN})$  and  $M$  represents the arrival rate packet-in requests to controller by each node.  $M = (\lambda^1, \lambda^2, \dots, \lambda^N)$ .  $\mu$  is the service rate of each controller, and  $\delta$  is the delay boundary that the average response time of each controller should not exceed.  $G$ ,  $\mu$  and  $\delta$  are the inputs of the algorithm.  $R$  represents all the nodes that are not yet associated to any controller, and  $C$  represents the candidates sites for controllers. Initially, we consider in our algorithm that all the nodes are not yet associated to any controller and  $R = V$  and all the nodes could be considered as possible candidate sites i.e.  $C = V$ .

We start our algorithm by finding the shortest path from any node to any other node in the network using Dijkstra's algorithm. Then the algorithm enters in a loop until all the nodes are

associated to at least one controller. The algorithm calculates all the possible candidate sites, and the set of nodes that can be connected to them, starting always with the node that is the nearest one to the controller. At each iteration, the algorithm computes the average response time to verify that is still less than the threshold  $\delta$ . The function  $Nearest_{Node}(V, j, D, S_j)$  selects the node  $i$  that has the minimal cost path to the controller  $j$ .  $S_j$  is the resulting associated nodes to the controller located in site  $j$ . Then, the algorithm calculates and checks the average response time  $\bar{t}_j$ . After finding the possible associated nodes to each controller, the algorithm chooses the one that has the maximum number of nodes associated to it, using the function  $Argmax_j |S_j|$ , and updates the output variable  $Y$  by setting  $y_j = 1$  to indicate that there is a controller on the site  $j$ . The algorithm removes the chosen controller from the possible candidates list  $C$ , and updates the output  $X$  to indicate the nodes that are now associated to this controller in site  $j$  by setting  $x_{ij} = 1$ .

#### 4.4 Simulation and Performance Evaluation

In order to evaluate the performance of the system, we aim to study the effect of  $\delta$ . We generated a network of  $N=120$  nodes, the links between the nodes are randomly generated following uniformly distributed pseudo random integers between  $mindelay$  and  $maxdelay$ . We fixed  $\mu$  to a high value enough to cancel the effect of the diversity in nodes' requests in the network ( $\lambda_i$ ) i.e. each controller has high performance capacity (very small service time) and is able to process all the requests in time (i.e.  $\mu \gg \lambda$ ). This is because we need to study only the effect of Network delay.

In the first scenario,  $\delta$  is set to  $mindelay$ , and therefore the number of required controllers is 120, each node shall host its own controller to satisfy the delay constraint. In scenario 3, the value of  $\delta$  is very high ( $maxdelay$ ) and therefore one controller will be sufficient to satisfy the constraints of all nodes. In scenario 2, the value of  $\delta$  is  $0.4 * maxdelay$ , and the number of needed controllers is 5. Figure 4 highlights the results of our algorithm. This figure shows that when  $\delta$  is very low, each node will host its own controller. While increasing  $\delta$ , the number of required controllers decreases and converges towards only one.

## 5 Conclusion and Future Works

In this work, we have proposed an SD-IoT architecture where SDN controllers are in charge of instantiating and controlling applications components in the infrastructure. We have proposed an algorithm to the controller placement problem in this SD-IoT architecture aiming at satisfying the delay constraints between IoT devices and the controllers. We formulated the problem, modeled the average response time of a controller taking into account the network delay and the average service time in the controller. We have performed some simulation to highlight the effect of  $\delta$  on the number of controllers to deploy. In future works, we aim to perform more simulation to study the effect of other parameters such the sampling rate of the IoT devices and the service time of the controller that depends on its architecture and price.

## Acknowledgments

This research was partially funded by the IBISC laboratory during an internship, it was also supported by the Lebanese University and CNRS Lebanon.

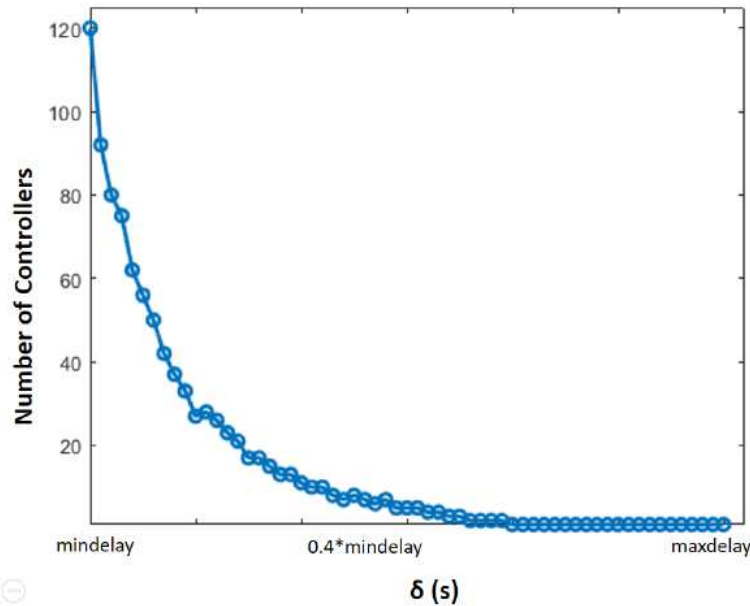


Figure 4: Number of controllers vs  $\delta$

## References

- [1] I. Bedhief, M. Kassar, and T. Aguil. Sdn-based architecture challenging the iot heterogeneity. In *2016 3rd Smart Cloud Networks Systems (SCNS)*, pages 1–3, Dec 2016.
- [2] H. Huang, J. Zhu, and L. Zhang. An sdn based management framework for iot devices. In *2014 China-Ireland International Conference on Information and Communications Technologies (CICT 2014)*, pages 175–179, June 2014.
- [3] Adrian Lara, Anisha Kolasani, and Byrav Ramamurthy. Network innovation using openflow: A survey. *Communications Surveys and Tutorials, IEEE*, 16:493–512, 03 2014.
- [4] B. A. A. Nunes, M. Mendonca, X. Nguyen, K. Obraczka, and T. Turletti. A survey of software-defined networking: Past, present, and future of programmable networks. *IEEE Communications Surveys and Tutorials*, 16(3):1617–1634, 03 2014.
- [5] L. Ogrodowczyk, B. Belter, and M. LeClerc. Iot ecosystem over programmable sdn infrastructure for smart city applications. In *2016 Fifth European Workshop on Software-Defined Networks (EWSDN)*, pages 49–51, Oct 2016.
- [6] M. Ojo, D. Adami, and S. Giordano. A sdn-iot architecture with nfv implementation. In *2016 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6, Dec 2016.
- [7] O. Salman, I. Elhajj, A. Kayssi, and A. Chehab. Edge computing enabling the internet of things. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pages 603–608, Dec 2015.

# Deploying-in-Production of a connected object-user identification approach by recognizing physical activity in a Big Data environment

Hamdi Amroun<sup>1,2</sup> and Mehdi Ammi<sup>1</sup>

<sup>1</sup>University of Paris-Saclay

<sup>2</sup>University of Paris 8, Paris, France.

Amroun.hamdi@gmail.com, Ammi@ai.univ-paris8.fr

## Abstract

In this paper, a method for identifying users of a connected object based on the recognition of the physical activities induced by the manipulation of these connected objects has been presented. Data from connected objects were first classified to recognize four types of activities: standing, sitting, lying and walking. These four activities thus classified generate four databases. Then, within each activity database, a classification model was trained to identify the different users. The model thus obtained is put into production on a Big data platform. User recognition results reached 73% accuracy in production. This study has provided a non-intrusive approach to recognizing users of a connected object in a "Big Data" environment

## 1 Introduction

The identification of users of connected objects is generally done through passwords or fingerprints as for smartphones or smartwatches [1-8]. These methods, although they are effective, present today certain limitations like the risk of forgetting the passwords or the inefficiency of the digital ones especially when the fingers of the hand present cracks [11-13].

The recognition of physical activity by using connected objects is very important in the world of research. Several works present very interesting studies, whether using a single connected object [14-19] or several connected objects [20-29]. The type of activity detected is highly variable, ranging from one activity [30] to several activities [31-40].

All research work in the field of research in physical activity recognition using connected objects stops in the training and testing stage of a classification model [40-50]. No study deals with the production of this type of algorithm.

In this paper, we propose a new approach to address the following issues: First, we propose a non-intrusive approach to recognizing the identity of the user of a connected object through the recognition of physical activity.

Secondly, the chosen model will not only be trained and tested on a small dataset but also deployed in production using a Big Data platform.

Section 2 provides an overview of the proposed approach.

## 2 Our approach to identify users of a connected object

36 people were invited to freely use a set of connected objects: a smartwatch, a smart TV and a smartphone and a connected cup for six months. The data of sensors embedded in all the connected objects have been stored in a big data platform of the Hortonworks type because the amount of data is very large that cannot be suitable in a conventional SQL SERVER type database, for instance.

The user identification process has been implemented in a previous work (Figure 1). In this work, the identification of users of a connected object has been done in two steps: Step 1: Recognize four types of activities: standing, sitting, lying and walking. Each activity type generates a database of the activity in question. So, we have four databases.

Step 2: Within each database of each activity, a classifier (Deep Neural Network) has been trained to recognize each user.

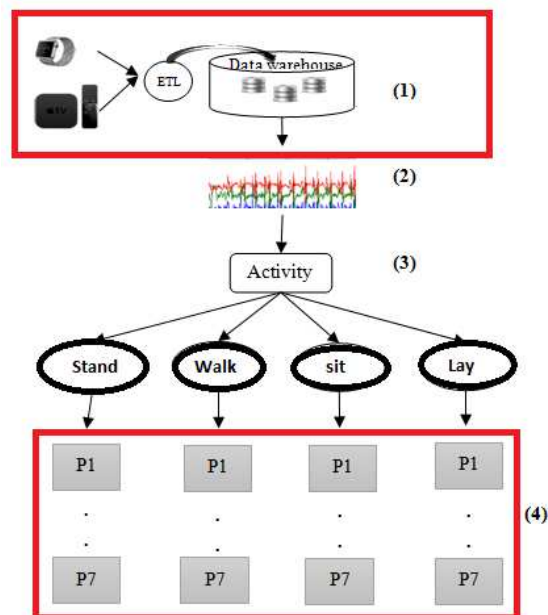


Figure 1:



Processing pipeline for the recognition of 7 users of 2 connected objects: a Smart TV and a Smart watch.

After extracting all the data from all the sensors on board the two connected objects, the data was stored in a Datawarehouse. A Deep Neural network classification algorithm was applied to classify the four activities: Stand, Walk, Sit and Lay. The four types of activities generate four sets of data. In order to detect the user of each connected object, we applied a classification of users to each set of activity data.

The algorithm trained and tested (Figure 1) is not put into production. In the practical case, we want to be able to save this algorithm in order to reuse it to recognize other types of users, it is actually the objective of putting into production this kind of machine learning algorithms.

Putting into production consists of three important notions:

- i) Make the predictions of different users available to everyone
- ii) Measure the quality of the predictions over time.
- iii) Improve prediction quality over with feedback

In order to concretize principles i) to iii), the algorithm thus trained, tested and validated, goes through four stages:

i) Deployment

Is the process of making the predictions available to everyone. For instance, the most important is to treat deployment as was the code deployment. That's has a lot of advantages : i) Flexibility: no need for complicated abstractions. ii) software deployment is very mature field. And iv) Rapid model updating with continuous deployments.

ii) Evaluation

This is exactly what we do to measure the quality of deployed models, in this case, we used the accuracy metric. For instance, the evaluation is done on line but also offline.

Offline evaluation consists of using a metric for the classification of user identities. While Online evaluation comes down to choosing a metric during the production phase. This metric is tracked during the Monitoring and Management phases.

- iii) Management : improving deployed models with feedbacks and metrics that we collect.
- iv) Monitoring : tracking model quality over time.

In order to keep maximum performance on model predictions, we need to update it. For this step, we applied an A / B testing.

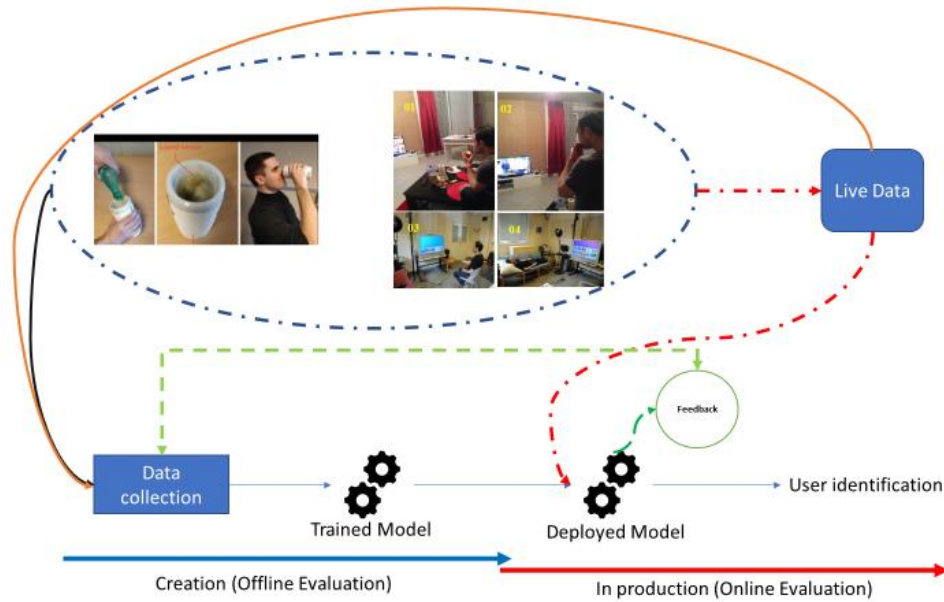


Figure 2: Explanatory diagram of the identification process of a connected object user

As shown in Figure 2, once the data were collected, an automatic learning algorithm (Deep Neural Network) was trained and tested (creation phase in Figure 2). The algorithm is then recorded in order to make predictions of users (Live Data). In order to be able to maintain good model classification performance, the model returns a feedback (metrics). The prediction results are presented in section 3.

### 3 Results and discussions

The classification results for physical activities and users are presented in Table 1.

	Smartphone, smartwatch, smart TV	Smartphone, Smartwatch, Smart Tv and Smart Cup
Debout/other	79%	78%
Sitting/other	77%	76%
Laying/Other	76%	77%
Users classification	72%	73.4%

Table 1: Classification of users and activities.

Thanks to our approach, we were able to reach a precision of more than 72% for the identification of the users of connected objects used in this study.

## 4 Conclusion

In this paper, we have proposed a methodology to predict and identify different users a connected object using physical activity recognition. This model has been put into production allowing it to be used for industrial purposes.

This work allows to see the impact of the implementation of a model of recognition of the activity and its putting into production in the real world.

## References

- [1] J. Gubbi, R. Buyya, S. Marusic, et M. Palaniswami, « Internet of Things (IoT): A vision, architectural elements, and future directions », *Future generation computer systems*, vol. 29, no 7, p. 1645–1660, 2013.
- [2] L. Atzori, A. Iera, et G. Morabito, « The internet of things: A survey », *Computer networks*, vol. 54, no 15, p. 2787–2805, 2010.
- [3] G. Kortuem, F. Kawsar, V. Sundramoorthy, et D. Fitton, « Smart objects as building blocks for the internet of things », *IEEE Internet Computing*, vol. 14, no 1, p. 44–51, 2010.
- [4] H. Sundmaeker, P. Guillemin, P. Friess, et S. Woelfflé, « Vision and challenges for realising the Internet of Things », *Cluster of European Research Projects on the Internet of Things*, European Commission, 2010.
- [5] “Keep Your Phone Safe: How to Protect Yourself From Wireless Threats.” *Consumer Reports*. Consumer Reports, June. 2013. Web.
- [6] B. Xie et Q. Wu, « Hmm-based tri-training algorithm in human activity recognition with smartphone », in *Cloud Computing and Intelligent Systems (CCIS)*, 2012 IEEE 2nd International Conference on, 2012, vol. 1, p. 109–113.
- [7] P. Sarcevic, Z. Kincses, et S. Pletl, « Comparison of different classifiers in movement recognition using WSN-based wrist-mounted sensors », in *Sensors Applications Symposium (SAS)*, 2015 IEEE, 2015, p. 1–6.
- [8] L. Fan, Z. Wang, et H. Wang, « Human activity recognition model based on Decision tree », in *Advanced Cloud and Big Data (CBD)*, 2013 International Conference on, 2013, p. 64–68.
- [9] H.-J. Kim, J. S. Lee, et J.-H. Park, « Dynamic hand gesture recognition using a CNN model with 3D receptive fields », in *Neural Networks and Signal Processing*, 2008 International Conference on, 2008, p. 14–19.
- [10] L. Zhang, X. Wu, et D. Luo, « Recognizing Human Activities from Raw Accelerometer Data Using Deep Neural Networks », in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 2015, p. 865–870.
- [11] P. Casale, O. Pujol, et P. Radeva, « Human activity recognition from accelerometer data using a wearable device », in *Iberian Conference on Pattern Recognition and Image Analysis*, 2011, p. 289–296.
- [12] T. Zebin, P. J. Scully, et K. B. Ozanyan, « Human activity recognition with inertial sensors using a deep learning approach », in *SENSORS*, 2016 IEEE, 2016, p. 1–3.
- [13] T. Datta et K. Manousakis, « Using SVM for user profiling for autonomous smartphone authentication », in *MIT Undergraduate Research Technology Conference (URTC)*, IEEE, 2015, p. 1–5.
- [14] Kayacik H. G., Just M., Baillie L., Aspinall D., Micallef N., “Data Driven Authentication: On the Effectiveness of User Behaviour Modelling with Mobile Device Sensors”, *Proceedings of the 3rd Mobile Security Technologies Workshop*, Held as part of the IEEE S&P Symposium, (MoST-2014), May 2014.
- [15] W. H. Lee and R. B. Lee, “Multi-sensor authentication to improve smartphone security,” in *Proceedings of the 1st International Conference on Information Systems Security and Privacy*, pp. 270–280, February 2015.
- [16] J.-S. Wu, W.-C. Lin, C.-T. Lin, et T.-E. Wei, « Smartphone continuous authentication based on keystroke and gesture profiling », in *Security Technology (ICCST)*, 2015 International Camahan Conference on, 2015, p. 191–197.
- [17] S. F. Weizhi Meng, Duncan S. Wong and J. Zhou. Surveying the development of biometric user authentication on mobile phones. In *Communications Surveys & Tutorials*, IEEE, vol. 17, no. 3, pages 1268–1293, 2014.
- [18] M. Antal and L. szlo Zsolt Szabó. An evaluation of one-class and two-class classification algorithms for keystroke dynamics authentication on mobile devices. <http://www.ms.sapientia.ro/~manyi/research/43.pdf>, 2015.
- [19] L. Cai and H. Chen. Touchlogger: Inferring keystrokes on touch screen from smartphone motion. In *Proceedings of the 6th USENIX conference on Hot topics in security*, 2011.
- [20] A. D. L. Daniel Buschek and F. Alt. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2015.
- [21] E. M. I. M. Mario Frank, Ralf Biedert and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. In *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pages 136–148, 2013.

- [22] J. Z. Ben Draffin and J. Zhang. Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction. In 5th International Conference, MobiCASE 2013, 2013.
- [23] Y. Z. Hui Xu and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In Symposium On Usable Privacy and Security (SOUPS 2014), 2014.
- [24] X. Z. Lingjun Li and G. Xue. Unobservable re-authentication for smartphones. In NDSS, The Internet Society, 2013.
- [25] F. G. Da Silva et E. Galeazzo, « Accelerometer based intelligent system for human movement recognition », in Advances in Sensors and Interfaces (IWASI), 2013 5th IEEE International Workshop on, 2013, p. 20–24.
- [26] Z.-Y. He et L.-W. Jin, « Activity recognition from acceleration data using AR model representation and SVM », in Machine Learning and Cybernetics, 2008 International Conference on, 2008, vol. 4, p. 2245–2250.
- [27] Z. He et L. Jin, « Activity recognition from acceleration data based on discrete cosine transform and SVM », in Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on, 2009, p. 5041–5044.
- [28] S. Pirttikangas, K. Fujinami, et T. Nakajima, « Feature selection and activity recognition from wearable sensors », in International Symposium on Ubiquitous Computing Systems, 2006, p. 516–527.
- [29] T. Zebin, P. J. Scully, et K. B. Ozanyan, « Human activity recognition with inertial sensors using a deep learning approach », in SENSORS, 2016 IEEE, 2016, p. 1–3.
- [30] S. Ha et S. Choi, « Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors », in Neural Networks (IJCNN), 2016 International Joint Conference on, 2016, p. 381–388.
- [31] Y. Chen et Y. Xue, « A deep learning approach to human activity recognition based on single accelerometer », in Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, 2015, p. 1488–1492.
- [32] M. Zeng et al., « Convolutional neural networks for human activity recognition using mobile sensors », in Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on, 2014, p. 197–205.
- [33] Z. He et L. Jin, « Activity recognition from acceleration data based on discrete cosine transform and SVM », in Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on, 2009, p. 5041–5044.
- [34] S. M. Kia, E. Olivetti, et P. Avesani, « Discrete cosine transform for MEG signal decoding », in Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on, 2013, p. 132–135.
- [35] D. Tran et A. Sorokin, « Human activity recognition with metric learning », in European conference on computer vision, 2008, p. 548–561.
- [36] F. Eyben, M. Wöllmer, et B. Schuller, « Opensmile: the munich versatile and fast open-source audio feature extractor », in Proceedings of the 18th ACM international conference on Multimedia, 2010, p. 1459–1462.
- [37] L. Mo, F. Li, Y. Zhu, et A. Huang, « Human physical activity recognition based on computer vision with deep learning model », in Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International, 2016, p. 1–6.
- [38] H. Yalçın, « Human activity recognition using deep belief networks », in Signal Processing and Communication Application Conference (SIU), 2016 24th, 2016, p. 1649–1652.
- [39] X. Yin et Q. Chen, « Deep metric learning autoencoder for nonlinear temporal alignment of human motion », in Robotics and Automation (ICRA), 2016 IEEE International Conference on, 2016, p. 2160–2166.
- [40] K. Nakadai, T. Mizumoto, et K. Nakamura, « Robot-Audition-based Human-Machine Interface for a Car », in Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on, 2015, p. 6129–6136.
- [41] S. Jia, T. Lansdall-Welfare, et N. Cristianini, « Gender Classification by Deep Learning on Millions of Weakly Labelled Images ».
- [42] T. Liu, M. Li, S. Zhou, et X. Du, « Sentiment classification via l2-norm deep belief network », in Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, p. 2489–2492.
- [43] G. Hinton et al., « Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups », IEEE Signal Processing Magazine, vol. 29, no 6, p. 82–97, 2012.
- [44] G. E. Hinton et R. R. Salakhutdinov, « Reducing the dimensionality of data with neural networks », Science, vol. 313, no 5786, p. 504–507, 2006.
- [45] V. Nair et G. E. Hinton, « Rectified linear units improve restricted boltzmann machines », in Proceedings of the 27th international conference on machine learning (ICML-10), 2010, p. 807–814.
- [46] X. Cui et V. Goel, « Maximum likelihood nonlinear transformations based on deep neural networks », IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no 11, p. 2023–2031, 2016.
- [47] J. Ray, B. Thompson, et W. Shen, « Comparing a high and low-level deep neural network implementation for automatic speech recognition », in Proceedings of the 1st First Workshop for High Performance Technical Computing in Dynamic Languages, 2014, p. 41–46.
- [48] L. Bao et S. S. Intille, « Activity recognition from user-annotated acceleration data », in International Conference on Pervasive Computing, 2004, p. 1–17.
- [49] S. E. El-Khamy, H. A. Elsayed, et M. M. Rizk, « C45. Classification of OFDM signals using higher order statistics and clustering techniques », in Radio Science Conference (NRSC), 2012 29th National, 2012, p. 541–549.
- [50] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, et P. Havinga, « Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey », in Architecture of computing systems (ARCS), 2010 23rd international conference on, 2010, p. 1–10.

# Machine Learning Supporting Brazilian Public Health Care policies

Raimundo Valter, Silas Santiago,  
Ronaldo Ramos, Mauro Oliveira  
Computer Science Dept., IFCE  
Fortaleza, Brazil

Email: {valter.costa, silas.santiago, ronaldo, mauro}@ifce.edu.br

Luis Odorico M. Andrade,  
Ivana Cristina de H. C. Barreto  
Public Health Dept., FIOCRUZ  
Fortaleza, Brazil

Email: {odoricomonteiro, ivana.barreto}@fiocruz.br

**Abstract**—Health data monitoring is a crucial activity to reduce maternal, neonatal and infant mortality rates by supporting public health policies decisions. Available data in Brazilian health databases point that it is possible to predict death risk in the early stages of gestation and infant development. In this research, we consider the information availability still in the gestational period to propose different death risk prediction models for this public of interest. We also detail the data mining process to apply machine learning-based techniques in death risk classification for maternal, neonatal and infant patients. We present an experiment pipeline to estimate average performance and evaluated machine learning models with different features combinations. Additionally, we show a web service which provides multiple predictive models by information availability. Results show Random Forest obtaining better performance when compared to the other machine learning methods.

**Index Terms**—Brazilian health data, data mining, information availability.

## I. INTRODUCTION

Historically, data analysis was always a guide in the decision-making process. Modern computing techniques and the vast amount of information available in Brazil by public transparency points to new opportunities. When it comes to linked data, though, there are many challenges. Despite the difficulties, researches in this area continue to demonstrate that it is possible to associate data and extract solutions immediately.

The World Health Organization (WHO) [1], [2] reports many of the maternal and infant occurred deaths, are due to gestation or parturition complications and can usually be avoided by performing simple self care actions. For this purpose, population health parameters monitoring is a key activity to reduce maternal (gestation and puerperium), neonatal and infant mortality rates.

When considering this context, collecting some data during each gestation stage period allows generating relevant information to identify death risk for mothers and babies. The available web tool Intelligent Governance in Health Systems (GISSA) is a system that supports governance in public health care. GISSA consists of a set of components which allow data collection, integration and visualization to support the decision-making process. This work was applied in the context of GISSA's data and provided a new tool for this system.

Researches [3]–[5] has demonstrated the correlation between data collected by Brazilian Government and the death of babies and mothers using machine learning techniques. Although information is collected at the time of the child's birth, this data can be inferred earlier in gestational development. This suggests that risk of death can be assessed months before birth, allowing family health professionals make simple interventions like teach self care actions, or even identify and refer the case to specialized medical care. Figure 1 outlines three periods of interest in case of mothers (pregnancy up to puerperium) and babies (neonatal and infant).

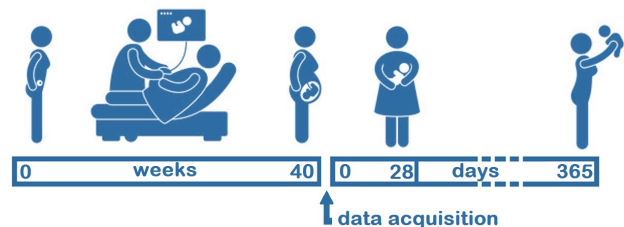


Fig. 1. Periods: gestational (0 to 40 weeks), neonatal (0 to 28 days), and infant (0 to 365 days).

Using this data collected at different gestational fuses allows the proposal of different pattern recognition models to the death risk identification for maternal, neonatal and infant patients. We could apply these predictive models to different stages of gestation and baby development. This individualized or joined information may be used to support public health strategies.

This paper details the data mining process in GISSA datasets to evaluate the application of supervised machine learning methods for neonatal, infant and maternal mortality for death risk prediction. The main contribution of this work includes a proposal of a web service which provides multiple predictive models ordered by information availability. A Proof of Concept (PoC) is also presented to demonstrate its use by GISSA portal in the family health care program running by federal government.

The remainder of this paper is organized as follow. Section II overviews related works about infant and pregnancy mortality risk prediction with machine learning. Section III details

data mining process used to build and evaluate a couple of machine learning-based death risk prediction models. Section IV presents and discusses performance results. Section V ends with some conclusions and future works.

## II. RELATED WORK

In [3], it is applied Fuzzy logic to death prediction of children group in the neonatal period. In this study, the authors identify a set of characteristics of interest: newborn birth weight, gestational age at parturition, Apgar score, and previous report of stillbirth. These features showed to be enough for the fuzzy inference. From the 24 rules identified by specialists, with it is possible to predict neonatal death with an accuracy of 90.0%. This study points for the model applicability from the child's birth, since 3/4 of these attributes are measured only in birth, been not possible using them to predict mortality conditions in early gestational stages.

In [6], authors present and evaluate the Intelligent Health Analysis System (LAIS), to support decision-making in preventive actions involving pregnant mothers and newborns. This system uses data mining techniques to generate death risk alerts using probability-based methods for training and evaluation of predictive models. The authors applied data from the Mortality Information System (SIM) and Live Birth Information System (SINASC) databases available on the DATASUS portal. Results showed that the probabilistic algorithm Naive Bayes performed better when compared to other machine learning techniques. The obtained accuracy and Area Under the Receiver Operating Characteristic Curve (AUROCC) were 98.2% and 92.1%, respectively.

Another research, performed in [4], presents a practical and straightforward approach to identify whether the infant mortality coefficient of a given city will be above or below the Brazilian national mean rate. By the use of regression trees model, it is sufficient to observe the total of prenatal medical appointments and mother's educational level hitting 65.0% of the cases.

## III. METHODOLOGY

Since the project workflow focuses on the data mining process from data collection to the deployment, it is applied Cross Industry Standard Process for Data Mining (CRISP-DM) [7] as the predominant methodology.

### A. Data Mining Steps in Gissa

In order to better represent the methodology applied, we simplified CRISP-DM in four main steps. Figure 2 summarizes macro activities that guide data mining based approach in this paper.

### B. Business Understanding

This step was about the business domain understanding of GISSA and included the comprehension of the entities, relationships, and fields in the databases. As an artefact, we produced a data dictionary which describes tables of SINASC and SIM databases.

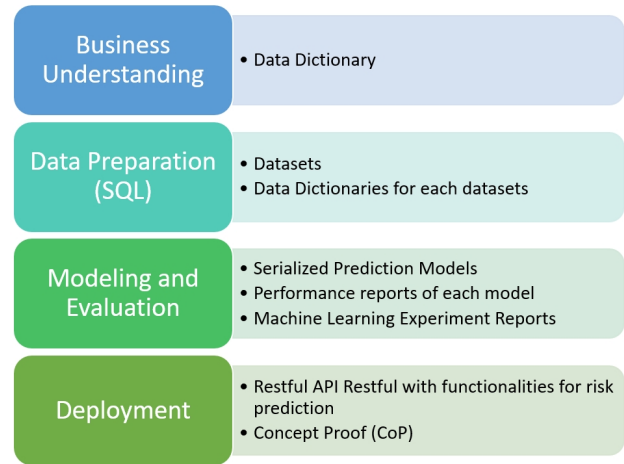


Fig. 2. Data mining steps in GISSA.

### C. Data Preparation

Raw data in GISSA's scenario is from relational databases, which enables applications to efficiently store and query with Structure Query Language (SQL) large amount of data (about 1.5M samples) [8]. The dataset and its description were built based on SIM and SINASC tables through the data preparation phase summarized in a couple of steps:

- 1) **Select Data:** a selection of columns and rows of interest in SIM and SINASC tables;
- 2) **Integrate Data:** tables union is done defining data classification as well (samples that incur in death or not). Some fields appear with missed values due to lack of information coming from different tables;
- 3) **Clean Data:** filling default missed values and replacement of inconsistent values in the table resulted by integration step;
- 4) **Construct Data:** some features are extracted based on each problem definition;
- 5) **Format Data:** Filled registers (without ignored information) are selected randomly and recorded in Comma-Separated Values (CSV) format.

Figure 3 represents the data preparation process, which involves applying SQL scripts resulting in each dataset.

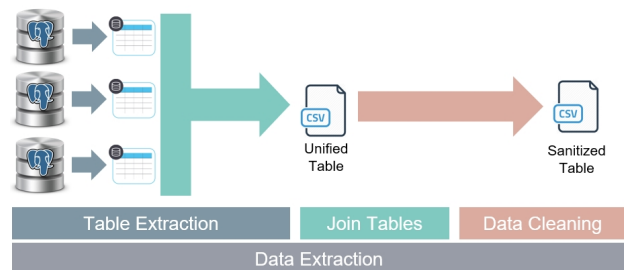


Fig. 3. Overview for data preparation process.

After the data preparation process, datasets classification

TABLE I  
DATASETS COMPOSITION

Dataset	dead	alive	total
maternal dataset	508	2531	3039
neonatal dataset	657	682	1339
infant dataset	911	952	1863

remains in two groups (dead or alive) for binary classification as in Table I. Women in pregnancy or puerperium stages compose the maternal dataset. The neonatal dataset is composed of newborns children, from 0 up to 28 days of life. The infant dataset is composed of children from 0 up to 365 days of life.

After this phase, each dataset is standardized with Standard-Scaler class, available at Scikit-Learn library [9]. This operation results in zero mean and scaling data to unit variance, considering each feature separately in all samples selected. Each value is scaled by the expression:  $x_{scaled} = (x - \mu) / \sigma$ , where  $\mu$  and  $\sigma$  represent, respectively, the mean and standard deviation for a given feature in dataset.

Was performed Exploratory Data Analysis (EDA) to verify the generated datasets. EDA allows verifying data variables quality by graphics visualization and statistical measurements. All these actions aim to prevent biased and overfitting models.

#### D. Modeling

We perform some tasks in order to model and assess applicability: (1) data loading and preprocessing; (2) exploratory data analysis; (3) hyperparameters optimization; (4) cross-validation executions. After these steps, models are ranking by AUROC and accuracy. Figure 4 shows the sequence of steps to guide modelling and deployment in API restful used by GISSA portal. This pipeline is applied to the three datasets considering each group of features selected.

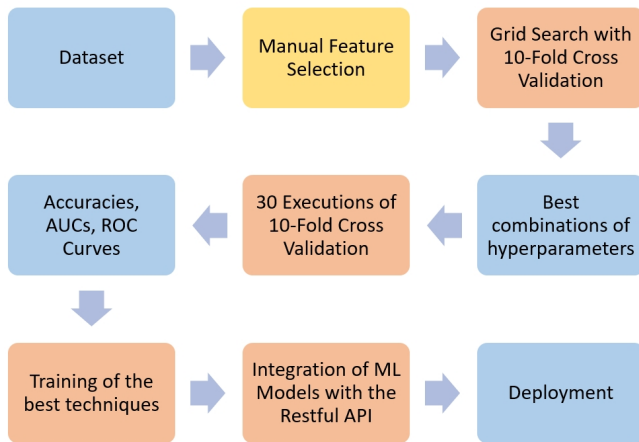


Fig. 4. Experiment overview for model selection and evaluation.

For each dataset, we listed the attributes in order of information availability. This feature arrangement allows the definition and evaluation of multiple predictive models, depending on the available information. Besides that, this strategy also

enables the prediction in different periods of interest, once we proposed multiple models with increasing feature numbers.

The features in maternal dataset are ( $F_1$ ) Birthplace, ( $F_2$ ) Education level (mother), ( $F_3$ ) Child's race, ( $F_4$ ) Child's gender, ( $F_5$ ) Number of healthy parturition, ( $F_6$ ) Gestational age (in weeks), ( $F_7$ ) month starting prenatal, ( $F_8$ ) Child positioning for parturition, ( $F_9$ ) Parturition type, ( $F_{10}$ ) Assisted parturition, ( $F_{11}$ ) Induced parturition, ( $F_{12}$ ) Cesarean occurrence before parturition, ( $F_{13}$ ) Birth indicative, ( $F_{14}$ ) Robson classification [10] for pregnancy, ( $F_{15}$ ) Apgar 5 minutes for child at birth, ( $F_{16}$ ) Age of child at death, ( $F_{17}$ ) Death occurred in relation to parturition and ( $F_{18}$ ) Death indicative of child.

The features in neonatal and infant datasets are ( $F_1$ ) Age of father at birth, ( $F_2$ ) Age of mother at birth, ( $F_3$ ) Level of education (mother), ( $F_4$ ) Marital status of mother, ( $F_5$ ) Number of prenatal consultations, ( $F_6$ ) Start month of prenatal consultations, ( $F_7$ ) Start week of prenatal consultations, ( $F_8$ ) mother's Brazilian Code of Occupation (BCO), ( $F_9$ ) Number of previous pregnancies, ( $F_{10}$ ) Number of stillbirths, ( $F_{11}$ ) Number of live births, ( $F_{12}$ ) Number of cesarean parturition, ( $F_{13}$ ) Number of healthy births, ( $F_{14}$ ) race of mother, ( $F_{15}$ ) gender of child, ( $F_{16}$ ) Pregnancy type, ( $F_{17}$ ) Birth occurrence place, ( $F_{18}$ ) Robson classification, ( $F_{19}$ ) Assisted parturition code, ( $F_{20}$ ) Cesarean occurrence before parturition begins, ( $F_{21}$ ) Cesarean status before parturition begins, ( $F_{22}$ ) Birthplace, ( $F_{23}$ ) Apgar 1 minute index for child at birth, ( $F_{24}$ ) Apgar 5 minutes index for child at birth, ( $F_{25}$ ) Weight of child at birth, ( $F_{26}$ ) Race of child and ( $F_{27}$ ) Malformed occurrence status.

#### E. Evaluation

The performance of supervised classifiers Naive Bayes [11] (NB), Decision Tree (DT) [12] and Random Forest (RF) [13] was measured and evaluated for the binary classification task.

The application of the Grid Search optimization strategy combined with the K-Fold Cross Validation (CV) technique makes it possible to obtain different model performance estimates for each hyperparameters combination. From these results, we can choose the most appropriated model (with the lowest CV error). Besides, the use of CV maximizes the confidence of the values of the selected hyperparameters, ensuring a better generalization (reducing overfitting).

The hyperparameters adjustment in supervised algorithms We performed Decision Tree and Random Forest parameter optimization by the following. For Random Forest, we considered the parameters `n_estimators`, `criterion` and `max_depth` of the RandomForest Classifier class available in the Scikit-learn library. Table II shows the parameters and tested values for this classifier.

For the Decision Tree, we considered `criterion` and `splitter` parameters to find the best-optimized combination. The Table III presents the tested values.

Since Gaussian Naive Bayes computes *a priori* and *a posteriori* probabilities from datasets, there are no parameters to be tuned.

TABLE II  
EVALUATED PARAMETERS FOR RANDOM FOREST

Parameters	Description	Tested Values
n_estimators	Number of trees in the forest	10, 50, 100
max_depth	Maximum depth of the tree	5, 10, 15, 20
criterion	Function to measure the quality of a split	"gini", "entropy"

TABLE III  
EVALUATED PARAMETERS FOR DECISION TREE

Parameters	Description	Tested Values
criterion	Function to measure the quality of a split	"gini", "entropy"
splitter	strategy used to choose the split at each node	"best", "random"

To obtain the best combination of parameters, the Grid Search technique combined to K-Fold Cross Validation was performed with  $k = 10$ . The optimal values for Random Forest were *criterion* = "gini", *max\_depth* = 10 and *n\_estimators* = 100. For the Decision Tree technique were found *criterion* = "gini", and *splitter* = "best".

1) *Cross-Validation Experiment*: For each supervised estimator, the experiment pipeline was run 30 times, for the estimation of a confidence interval and average performance benchmarks. Algorithm 1 details the experiment process. For a given dataset  $D$ , a group of supervised machine learning techniques  $T$  and features set  $A$ , the experiment firstly generates a set of features combinations  $C$  (TOP 01, TOP 02, ..., TOP  $M$ ) and evaluate each technique for a different features subset by cross-validation.

#### F. Deployment

It was serialized the selected predictive models for each addressed classification scenario (maternal, neonatal and infant) and made available in a restful API. This software modularization allows simple integration with any system, either web or mobile. It was generated a total of 27 predictive models for neonatal mortality, 27 models for infant mortality and 18 models for maternal mortality.

For each scenario, the model represents a classifier which receives a features vector of a given size. For example, considering neonatal mortality risk, model TOP04 is trained with the Gaussian Naive Bayes algorithm and receives as input a vector with four attributes. In this context, GISSA web system works as the PoC, consuming services provided by a restful API, in order to demonstrate the proposed models. Figure 5 illustrates the architecture.

Still considering TOP04 ( $F_1$  father's age at birth in years,  $F_2$  mother's age at birth in years,  $F_3$  mother's level of education - 4 means married -,  $F_4$  mother's marital status - 2 means 8 to 11 years of study completed) model, GISSA application must inform the scenario and a feature vector as a POST request according to the format.

#### Algorithm 1 Experiment pseudocode

```

D ← GetDataset()
A ← {a1, a2, ..., am}
T ← {t1, t2, ..., tk}
C ← FeaturesCombination(A)
foreach t ∈ T do
    thyperparameters = GridSearch(t, D, A)
end foreach
foreach round ∈ 30 rounds do
    foreach c ∈ C do
        foreach t ∈ T do
            S ← Subset(D, c)
            S ← FeatureStandardization(S)
            ACCcv, AUROCCcv ← CrossVal(S, folds =
10)
        end foreach
    end foreach
end foreach
foreach c ∈ C do
    foreach t ∈ T do
        results[c][t] = ComputeMetrics()
    end foreach
end foreach
S ← BestCombinations(results)

```

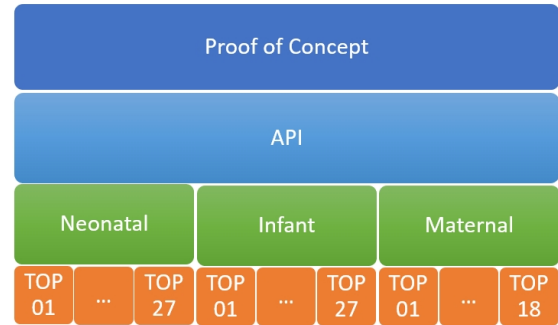


Fig. 5. PoC and API architecture.

```

POST http://<server>:5001/predict
{
  "data": "[21.0, 19.0, 4.0, 2.0]",
  "model": "MMInfantil",
}

```

The API executes prediction for the selected model, returning the class value (0 for alive or 1 for death prediction) and the death probability as a POST response, in this case: 'class': 1, 'prob': 0.79.

#### IV. RESULTS

It is performed a series of experiments to choose the best predictive model algorithms considering performance results. The average values for accuracy and AUROCC are presented, according to the experiment procedures detailed in section III.



All models and its results are presented in Tables IV, V and VI.

It is presented the Receiver Operating Characteristic (ROC) curve, where the  $Y$  axis represents the sensitivity and the  $X$  axis represents  $1 - \text{specificity}$ . Sensitivity refers to the likelihood of a test be positive, given that the individual has died. Specificity refers to the likelihood of the test be negative, once the individual is alive [14]. For the Figures 6, 7 and 8 blue dots describes the mean ROC graph that represents all 30 experiments randomly initialized of the considered model, the gray shadow (when observable) in background is a composition of all results separately. This shaded area shows that the ROC curve profile does not differ from the mean case for the set of selected predictor variables regardless of data separation between training and test groups.

The ratio between the accounting of correctly classified examples and the total of evaluated examples. This metric is accepted for evaluation and describes the accuracy of classification entirely [15]. AUROCC represents the overall performance of an estimator, once this metric considers all computed values of sensitivity and specificity. The more the estimator ability to discriminate against individuals with and without risk of death, the more the curve will approximate to the upper left corner and AUROCC will approximate to 1 [14].

The predominance of the Random Forest algorithm as the best technique for the evaluated datasets is notorious. For the neonatal and infant mortality datasets, the Random Forest algorithm obtained better performance for 24 of the 27 predictive models evaluated. For the maternal mortality dataset, this algorithm presented better accuracy and AUROCC for all models.

For the neonatal dataset, the combination that scored highest AUROCC and accuracy was TOP26, with 0.8876 and 93.90%, respectively. The area under the ROC curve for this combination is shown in Figure 6. This combination concerns with a Random Forest model with 26 predictive attributes.

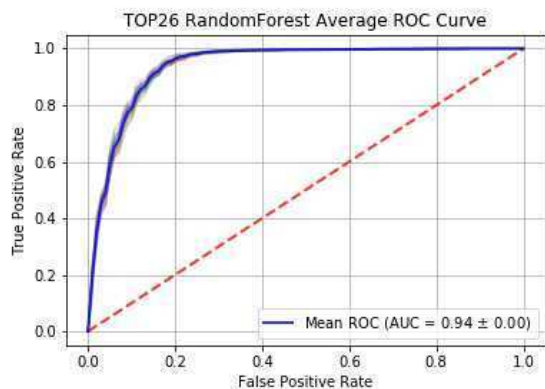


Fig. 6. ROC curve for neonatal death risk.

For the infant dataset, the best combination was also the TOP26, with 0.9909 and 99.73%, respectively. This similarity

TABLE IV  
NEONATAL MORTALITY EXPERIMENTS

Features Set	Classifier	Mean AUROCC	Mean ACC
TOP01	GaussianNB	0.5180	54.07%
TOP02	GaussianNB	0.5407	55.21%
TOP03	GaussianNB	0.5554	57.83%
⋮	⋮	⋮	⋮
TOP17	RandomForest	0.7454	82.02%
TOP18	RandomForest	0.7459	81.86%
TOP19	RandomForest	0.7448	81.87%
TOP20	RandomForest	0.7504	82.73%
TOP21	RandomForest	0.7517	82.73%
TOP22	RandomForest	0.7511	82.79%
TOP23	RandomForest	0.8155	88.27%
TOP24	RandomForest	0.8261	89.75%
TOP25	RandomForest	0.8394	90.82%
<b>TOP26</b>	<b>RandomForest</b>	<b>0.8876</b>	<b>93.90%</b>
TOP27	RandomForest	0.8872	93.94%

TABLE V  
INFANT MORTALITY EXPERIMENTS

Features Set	Classifier	Mean AUROCC	Mean ACC
TOP01	GaussianNB	0.5277	54.34%
TOP02	GaussianNB	0.5495	56.15%
TOP03	GaussianNB	0.5760	60.70%
⋮	⋮	⋮	⋮
TOP17	RandomForest	0.7512	82.69%
TOP18	RandomForest	0.7533	82.81%
TOP19	RandomForest	0.7518	82.70%
TOP20	RandomForest	0.7903	86.32%
TOP21	RandomForest	0.7893	86.25%
TOP22	RandomForest	0.7893	86.20%
TOP23	RandomForest	0.8452	91.28%
TOP24	RandomForest	0.8521	92.00%
TOP25	RandomForest	0.8744	93.15%
<b>TOP26</b>	<b>RandomForest</b>	<b>0.9909</b>	<b>99.73%</b>
TOP27	RandomForest	0.9906	99.82%

is associated with the composition of the neonatal and infant datasets, given that both differ in the way how class attribute is determined for each instance. The ROC curve for this combination is shown in Figure 7.

Lastly, for the maternal dataset, the TOP15 combination was the one that obtained the highest value for accuracy and AUROCC, with 0.9163 and 97.50%, respectively. The ROC curve for this combination is shown in Figure 8.

## V. CONCLUSION

This paper evaluated three problem scenarios for death prediction to support decision-making in health management. From the data mining process in GISSA portal, it was possible to build and evaluate a set of machine learning models trained with neonatal, infant and maternal data with different feature combinations.

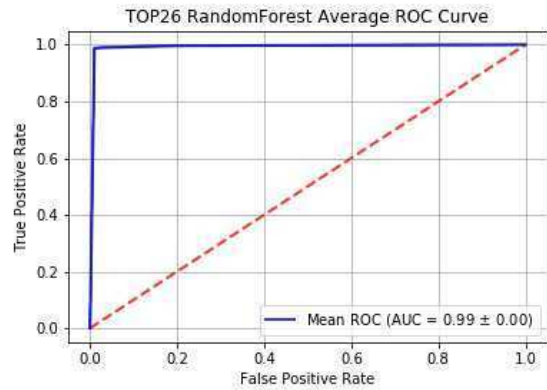


Fig. 7. ROC curve for infant death risk.

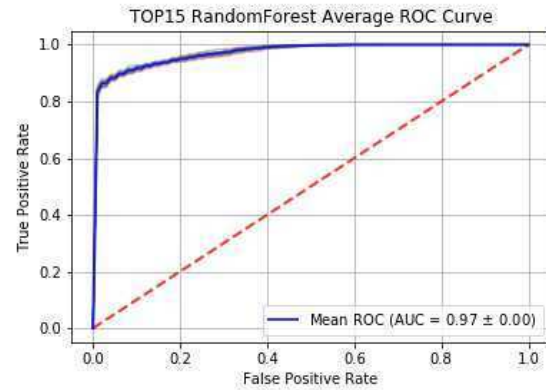


Fig. 8. ROC curve for maternal death risk.

TABLE VI  
MATERNAL MORTALITY EXPERIMENTS

Features Set	Classifier	Mean AUROCC	Mean ACC
TOP01	RandomForest	0.6120	61.20%
TOP02	RandomForest	0.6913	71.64%
TOP03	RandomForest	0.8277	92.81%
⋮	⋮	⋮	⋮
TOP10	RandomForest	0.8840	95.49%
TOP11	RandomForest	0.8892	95.77%
TOP12	RandomForest	0.8946	96.14%
TOP13	RandomForest	0.8957	96.26%
TOP14	RandomForest	0.9060	97.11%
<b>TOP15</b>	<b>RandomForest</b>	<b>0.9163</b>	<b>97.50%</b>
TOP16	RandomForest	0.9147	97.39%
TOP17	RandomForest	0.9133	97.38%
TOP18	RandomForest	0.9143	97.41%

In this approach, we consider the information availability as a guideline to generate, evaluate and select the predictive models. Each model represents the best algorithm for a feature combination. The Random forest estimator is the predominant method in feature combinations for the three scenarios, which indicates its generalization capability for health data. To demonstrate the utility of our approach, we built a microservice to serve all these models for each scenario. A PoC was also implemented to demonstrate the use of the restful API in GISSA portal.

Future works include expanding the restful API system adding new predict models (services). Intends to evaluate other supervised methods for described scenarios or even apply semi-supervised approaches to deal with labelled and unlabeled datasets.

#### ACKNOWLEDGEMENT

Our sincere acknowledgements to the Financier of Studies and Projects (FINEP) for funding this research. We also thank the AVICENA™ team and Instituto Atlântico™ for provide access to data and support from IT technicians and health care specialists.

#### REFERENCES

- [1] W. H. Organization. (2018) Newborns: reducing mortality. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality>
- [2] ——. (2018) Maternal mortality. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>
- [3] L. F. C. Nascimento, P. M. S. R. Rizol, and L. B. Abiuzi, "Establishing the risk of neonatal mortality using a fuzzy predictive model," *Cadernos de Saúde Pública*, vol. 25, pp. 2043 – 2052, 09 2009.
- [4] A. D. P. Chiavegatto Filho, "Uso de big data em saúde no brasil: perspectivas para um futuro próximo," *Epidemiologia e Serviços de Saúde*, vol. 24, pp. 325 – 332, 06 2015.
- [5] C. L. da Silva, "Lais, uma solução baseada em classificadores para geração de alertas em sistema de saúde," Ph.D. dissertation, Universidade Estadual do Ceará, 2017.
- [6] R. Ramos, C. Silva, M. W. L. Moreira, J. J. P. C. Rodrigues, M. Oliveira, and O. Monteiro, "Using predictive classifiers to prevent infant mortality in the brazilian northeast," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Oct 2017, pp. 1–6.
- [7] R. Wirth and J. Hipp, "Crisp-dm: Towards a standard process model for data mining," in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Citeseer, 2000, pp. 29–39.
- [8] J. Grus, *Data science from scratch: first principles with python*. " O'Reilly Media, Inc.", 2015.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [10] J. P. Vogel, A. P. Betrán, N. Vindevoghel, J. P. Souza, M. R. Torloni, J. Zhang, Ö. Tunçalp, R. Mori, N. Morisaki, E. Ortiz-Panozo *et al.*, "Use of the robson classification to assess caesarean section trends in 21 countries: a secondary analysis of two who multicountry surveys," *The Lancet Global health*, vol. 3, no. 5, pp. e260–e270, 2015.
- [11] I. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub, 2005.
- [12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.
- [13] Y. Liu and H. Wu, "Water bloom warning model based on random forest," in *Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2017 International Conference on*. IEEE, 2017, pp. 45–48.
- [14] B. Lopes, I. C. d. O. Ramos, G. Ribeiro, R. Correa, B. d. F. Valbon, A. C. d. Luz, M. Salomão, J. M. Lyra, and R. Ambrósio Junior, "Bioestatísticas: conceitos fundamentais e aplicações práticas," *Rev Bras Oftalmol*, vol. 73, no. 1, pp. 16–22, 2014.
- [15] A. Dainotti, A. Pescape, and K. C. Claffy, "Issues and future directions in traffic classification," *IEEE network*, vol. 26, no. 1, pp. 35–40, 2012.

# AuFa - Automatic Detection and Classification of Fake News Using Neural networks

Vinícius N. Barbosa, Carina T. de Oliveira, Reinaldo B. Braga

Federal Institute of Education, Science and Technology of Ceará (IFCE)  
viniciusnb14@gmail.com, (carina,reinaldo)@lar.ifce.edu.br

## Abstract

The increased proliferation of fake news on social networks has a significant impact on the information received by the society. The malicious use of information can compromise the democracy, as well as manipulate the opinions of people exposed to such news. These impacts have boosted new search directions in an attempt to classify and identify this news. Therefore, we propose AuFa, a solution to automatically detect and classify fake news through neural networks algorithms. Preprocessing techniques used are in the collected database to analyze the patterns of news, as well as to reduce noise and eliminate null information. The results obtained showed that the supervised neural network algorithm (MLPClassifier), obtained satisfactory performance to be used in the proposed solution.

KEYWORDS: Fake News. Neural networks. Natural Language Processing.

## 1 Introduction

Society has been looking for quick and practical ways to communicate and perform daily tasks. At the same time, there is an evident growth in the number of Internet users exchanging information worldwide. A survey performed [4] shows that in 2016, there were 81.4 million mobile phone internet users in Brazil, which accounted for nearly 40 percent of the Brazilian population. By 2021, these figures forecast are to increase to 112.7 million and 51.8 percent respectively.

In this way, access to information has become easier by the fact that any smartphone, tablet or notebook makes news available, either through trusted websites or through social networks. Social networks stand out in this rapid sharing of data, but much of the information disseminated in these environments is not true. For example, a Reuters Institute report [12] shows that 66% of respondents in Brazil use social media as a news source. While social networks have increased the ease of disseminating information, their popularity has potentiated the problem of fake news, accelerating the speed and scope at which such news are disseminated.

People daily are bombarded by information from a variety of subjects and sources, leading to one of today's biggest problems, the so-called Fake News. According to Google Trends<sup>1</sup>, which is a Google tool where the interest of a particular topic is exposed over time, showed that the term Fake News was widely used in the surveys conducted by Brazilians in 2018 and that it continues to grow in 2019. In addition, Fake News had already gained prominence in the international press in 2016, when the U.S. presidential election took place. Technology companies did an analysis and discovered various Internet content that contained Fake News, most of which denigrated a US presidential candidate of that election [1]. Another example of the problems that the spread of this fake news cause, according to [3], It is the fact that happened in 2014, where false news ended with the death of a woman in Brazil. This came after rumors circulated on the Internet accusing her of kidnapping children to perform black magic rituals.

---

<sup>1</sup><https://trends.google.com.br/trends/?geo=BR>

According to [7], young people tend to consume less of the older media, such as newspapers, radio and TV, because they believe that, besides being uninteresting, they are repetitive. Thus, they use social networks as the main source of information, resulting in young people more focused on thinking and believing in issues arising from manipulated information. News posted on social networks and messaging apps have a low level of reliability [10], mainly caused by the lack of a Fake News editorial board and filtering or nonexistence. Still, Fake News can reach an audience that “...wants to believe them, consumes them even if they suspect them, because they want to see their ideology and prejudices confirmed ...” [5]. A recent report by the Jumpshot Tech Blog showed that Facebook referrals represented 50% of total fake news site traffic and 20% of total trusted site traffic<sup>2</sup>. According to [13], Facebook and Google are trying to solve the Fake News problem as they receive strong criticism for not using a tool that could prevent the spread of fake news.

The challenges of fake news not only received huge public attention but also attracted growing attention from the academic community. One of the main goals of researchers is to detect these types of news in an attempt to minimize negative results. Thus, several works in the academic and commercial areas are looking for automatic mechanisms to recognize Fake News among a set of real news on the Internet and in the social media.

In this context, this paper presents AuFa, a solution to automatic detect and classify Fake News using neural networks. The purpose of this solution is to recognize the existing patterns between real and false news and thus classify the texts into *TRUE* or *FAKENEWS*, returning to their respective odds between 0% and 100%. The news used for the training and testing of this application was extracted from the *kaggle* repository<sup>3</sup>, which consisting of 20.800 news related to various subjects. In addition, it is intended to increase this number of news by searching other databases available on the web.

This proposal can influence the impact of false news in the world and serve as a model for other professionals in the field. The mechanism of detection of this news is not limited to the linguistic characteristics found in the texts. The results can assist either in developing new tools that help combat fake news or in any other area involving machine learning.

## 2 Related Works

Marumo [8] has proposed a deep learning for fake news classification by text summarization. A database of true and false news is used, and the news patterns are analyzed using the preprocessing techniques: Summary and Word2Vec. The purpose of this paper was to present a model capable of classifying texts between fake news and real news, using Deep Learning and Traditional Machine Learning. He used the algorithms Long Short-Term Memory, Random Forest and Support Vector Machine. From the obtained results, the experiments with Long Short-Term Memory returned the highest accuracy, with the value of 79.3% of correct answers.

With regard to the topic of fake news and the application of Machine Learning, the authors of Monteiro [9] proposed the development of a system for classifying fake news with news texts in Portuguese. This work used machine learning methods to discover, classify and store fake news texts, for later application. Extract Transform Load ETL step are software tools whose function is to extract data from various systems and apply to a Data warehouse. For this, a dataset was created and the Logistic Regression, Naive Bayes and Support Vector Machine (SVM) methods were evaluated. After performing the tests, the Logistic Regression, Naive Bayes and SVM algorithms obtained accuracy of 90.33%, 89.27% and 90.52% hits respectively.

---

<sup>2</sup><https://www.jumpshot.com/data-facebooks-fake-news-problem/>

<sup>3</sup><https://www.kaggle.com/c/fake-news/data>

AuFa

Vinícius Nunes, Reinaldo Braga

The objective of Leal [6] work was to study and implement fake news classification and identification techniques for the 2018 presidential election. Machine learning, data mining and textual algorithms were used in the implementation. The dataset was built using data collected on the social network Twitter, in order to obtain messages and news regarding presidential election candidates. After training and testing the algorithms, the results reached an accuracy of 90% of correct answers.

The work [11] presents an application web called FakeCheck that demonstrates the results obtained in the project Automatic Detection of False News for Portuguese, funded by *CNPq's* Institutional Program for Scientific Initiation Scholarships (*PIBIC*) and by *CAPES*. The project aims to study methods for the automatic detection of false news using Natural Language Processing (NLP) and Machine Learning (ML). When receiving text, the system applies methods to extract linguistic attributes from that text and uses them in a machine learning model that classifies the news as true or false. Attributes extracted from the text are applied to a Support Vector Machine classifier, which automatically infers the news class (true or false). In the tests performed, in a controlled test environment, the system obtained about 89% accuracy (overall accuracy).

The work of Castelo [2] deals with a study to detect fake news recorded on the web. They propose a new detection approach that uses topic-independent features has been applied. To the work, a new database was created from the selection of Politifact, BuzzFeed and OpenSources.co sites such as unreliable news sites and 242 most visited news sites, the top 500 Alexa news sites as trusted sites. The algorithms used were Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forest (RF). All morphological, psychological and word count resources were extracted from the database and applied to cross validation with the metric precision of performance to increment the algorithm, after the tests they obtained a maximum accuracy of 86%. They found that from this approach, a rating was more accurate with cross-domain news, outperforming content-based approaches.

The differential of AuFa in relation to the presented works: (I) its database fed is daily with new news, which allows the algorithm to be more effective in detecting false news; (II) the detection of this news is not limited to the linguistic characteristics found in the texts. Classification occurs by analyzing the news content, making it necessary to use the Web Scraping feature. These resources together provide support for the purpose of this work.

### 3 Proposal

The objective of this work is to develop an intelligent neural network based solution to detect false news, proposing a model according to Figure 1, divided into 3 phases: Classification, Web and Database. The idea is to provide the user with the possibility to know if a news story is real or false information.

The classification layer according to the model proposed in Figure 1, the user enters a news item in the system, this news goes through a preprocessing, where text is cleared, removing special characters that do not add useful information, as well as double spaces, line breaks, and StopWords such as: a, to, from, one, and others. Then all the text is turned lowercase so that all news follows a pattern.

With the preprocessed data, vectoring is performed where the data is converted to numeric values so that the algorithm can interpret this data and then return the class *TRUE* or *FAKENEWS* of the entered text. If the probability of classification is less than 90%, a web search is performed, entering the second layer of the application.

The second layer is responsible for performing a web search on the respective topic of the

AuFa

Vinicius Nunes, Reinaldo Braga

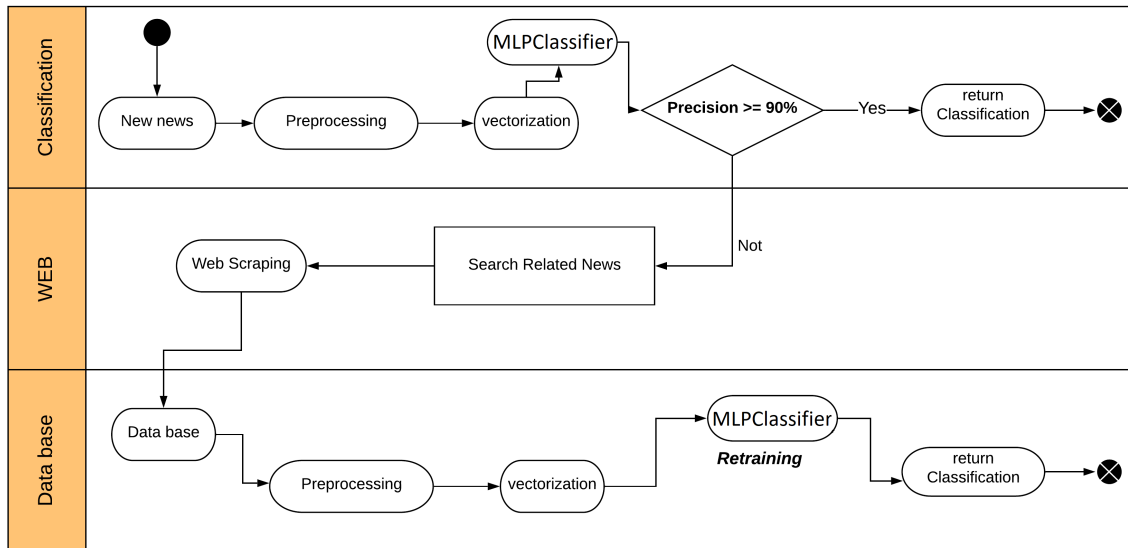


Figure 1: AuFa Activity Diagram.

news informed by the user at the beginning of the application. The search that will be carried out closely resembles plagiarism detectors, which scan the text and look for similar results on the web. Likewise, this search will go through the sites considered reliable using a similarity grouping technique, which is widely used in search engine algorithms. With the grouped texts, we will use the Web Scraping technique that is able to extract data from a website. This extracted data will be stored in a database thus entering the third layer.

The database layer will be responsible for storing the news returned by the Web layer; this data will initially be stored in a text file for easy manipulation. With the saved data, the preprocessing and vectoring procedures of the new stored texts are applied. After that, the MLPClassifier algorithm will be retrained and thus will return the proper rating *TRUE* or *FAKENEWS* of the initial news reported by the user.

### 3.1 Dataset

The database chosen to train and test the algorithm was the dataset "Fake News - Build a system to identify unreliable news", this dataset was obtained in the kaggle repository<sup>4</sup>, developed by UTK Machine Learning Club and all data is in English. This dataset was divided into id, title, author, text, and label. Id is a unique ID for each of the news, the title is the title of the news, the author is the author responsible for the news, and the text is the news itself and, finally the label that is the class that marks potentially reliable or unreliable news.

This dataset has 20,800 news on various subjects, but 39 of this news has null fields and, therefore, are discarded. Of the remaining 20761, 10374 is reliable news and 10387 are unreliable news, as shown in Table 1. The news is already sorted with their respective classes: 0 to reliable and 1 for unreliable. This large amount of data contributes greatly to the evaluation process of an algorithm, and the fact that the news is balanced helps prevent problems in the process of training algorithms in the Area of AI (Artificial Intelligence). For these reasons, this dataset

<sup>4</sup><https://www.kaggle.com/c/fake-news>

AuFa

Vinícius Nunes, Reinaldo Braga

was chosen for use in this proposal.

Dataset	
<b>Real News</b>	10374
<b>Fake News</b>	10387
<b>Null</b>	39
<b>Total News</b>	20800

Table 1: Dataset news.

After some testing it was possible to extract statistics from the samples contained in the dataset, as shown in Table 2. The data show that true news presents a greater number of words and sentences compared to Fake news.

	<b>FakeNews</b>	<b>True</b>
<b>Average words per news</b>	1950	2608
<b>Average sentences per news</b>	14.10	21.07

Table 2: Sample Information.

Additional information on how data was collected, from which sources and how these data were obtained, are not provided in the repository available for download of the dataset.

## 3.2 Experiments and Tests

To perform the experiments, a database available on the kaggle site is used, as explained in the section 3.1. After data collection, preprocessing is initiated using the Pandas and NLTK (Natural Language Toolkit) libraries of the Python programming language, in order to eliminate null fields, clean up data that was “dirty” (special characters that do not contribute to the algorithm learning), remove StopWords and standardize the text to lowercase.

The measurement metric used is made up of four classifications, as shown below:

- True Positive (TP): All data from the positive class that has been sorted correctly.
- True negative (TN): All negative class data that has been sorted correctly.
- False Positive (FP): All data that was from the negative class and was classified as positive.
- False Negative (FN): All data that was from the positive class and was classified as negative.

To evaluate the results obtained, the accuracy, precision, recall and F-measurement (F1) metrics are used. Their formulas are in the following equations:

$$accuracy = (TP + FN)/Total$$

$$precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$F1 = 2 * (precision * Recall)/(precision + Recall)$$

AuFa

Vinicius Nunes, Reinaldo Braga

Accuracy is a general assessment of hits, making a relationship to the total instances of the dataset. Precision is an assessment between hits in relation to the sum of hits and misses. The recall is about how often you rate the classifier correctly, that is, the number of correct results divided by the number of results that should have been submitted. The F-measurement (F1) is the combination of precision and recall in a uniformly weighted measurement, which is the harmonic mean of the two, that is, square of the geometric mean divided by the arithmetic mean.

With the clean and preprocessed dataset, the testing phase of the algorithms begins, where NaiveBayes, SGDClassifier, SVM (Support Vector Machines) and MLPClassifier (Multilayer Perceptron) are tested. Among the tested algorithms, the one that has presented the most satisfactory accuracy, precision, recall and F-measurement (F1) was the MLPClassifier. The values obtained are presented in detail in the section 4, corresponding to the results.

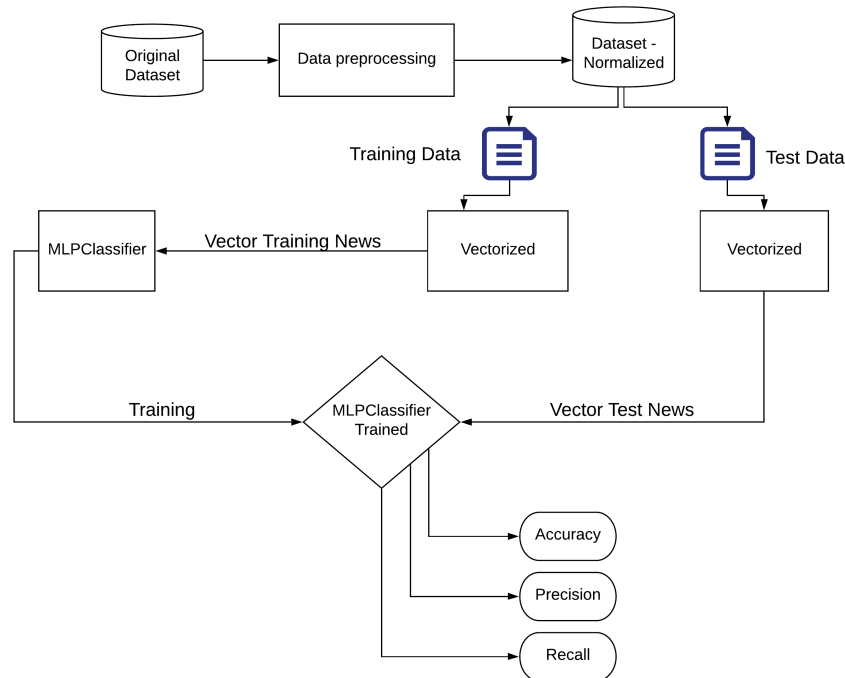


Figure 2: Flow for algorithm evaluation.

The normalized baseline data were divided into 75% for training and 25% for testing, as shown in Figure 2, as it is a widely used form in the literature. However, this division will depend greatly on the amount of data in the dataset and the way it is evaluated. Since the classifier “only understands numbers”, you must convert raw data, which is in text format, to a numeric format. This must happen before passing the data to the classifier. In addition, it should be noted that some words in the training phase would be more frequent, such as prepositions and articles. These words tend to recur in all documents and do not usually carry very significant information for classification. With that in mind, the TF-IDF (term-inverse frequency of documents) measure is used to limit the importance of these words that repeat throughout documents so that they do not cause more influence than necessary.



Then the vectorized training data is submitted to the MLPClassifier algorithm, this training is inserted in the supervised machine learning context, where each data sample used has a label informing which classification it fits. Thus, the general idea is to make the network learn the standards for each class type, so when new news is provided to the network, the algorithm will be able to establish to which class the news belongs. After training, the model is subjected to test data so that its accuracy, precision, recall and F-measurement (F1) can be analyzed.

## 4 Results

Initially a comparison was made between the Naive Bayes, SGDClassifier, SVM and MLPClassifier algorithms so that the algorithm that best suited the data could be chosen, their specifications and documentation can be found in Scikit-Learn<sup>5</sup>. For testing, the Fake News dataset obtained from kaggle<sup>6</sup> was divided into 75% for training and 25% for testing, which correspond respectively to 15.570 news for training and 5.191 test news. With the divided dataset, the training news was applied to each algorithm, which after being trained generated their models that were submitted to the test data to be evaluated. After that, we used the confusion matrix, which is a table that shows the classification frequencies for each class of the model, which helps in the algorithm evaluation process. With the confusion matrix, it is possible to take the data returned from the algorithms and thus calculate their respective: accuracy, precision, recall and F-measurement (F1). The results obtained from each algorithm are found in Table 3.

Algorithms	Accuracy	Precision	Recall	F-Measure (F1)
Naive Bayes	75,68%	77,84%	73,16%	75,42%
SGDClassifier	92,10%	91,55%	92,53%	92,03%
SVM (Support Vector Machines)	93,70%	95,88%	94,44%	95,15%
<b>MLPClassifier</b>	<b>96,44%</b>	<b>97,62%</b>	<b>95,36%</b>	<b>96,47%</b>

Table 3: Algorithm Metrics.

The Naive Bayes algorithm, although a good text classification algorithm, was the least successful in the classification, because Naive Bayes does not take into account the correlation between the factors present in the text.

The SGDClassifier and Support Vector Machines (SVM) algorithms have obtained similar results, since SGDClassifier is a linear classifier that uses the Stochastic Descending Gradient as a training method, which resembles the SVM estimator.

MLPClassifier as seen in Table 3 obtained the best results in the testing process, due to its universal approach capability and its flexibility to form quality solutions for a wide range of attributes of the same algorithm of learning. Seen as artificial neural networks, they have been successfully applied to the most diverse problems.

The Web layer shown in Figure 1, which is responsible for the related news search process, is still in the search process to develop the best way to search the news on the web. Thus, the expected results for this layer are related to the search for news by similarity, being able to return the most similar news and thus add knowledge to the algorithm, improving the classification process.

<sup>5</sup><https://scikit-learn.org/stable/index.html>

<sup>6</sup><https://www.kaggle.com/c/fake-news>

## 5 Conclusion

The need to combat Fake News and its consequences in society, influencing people's thinking and decision-making, were the initial cause for the development of this work so that it could detect and combat these types of news. Several studies are being conducted and numerous technologies are being introduced to deal with the fake news. However, readers need to look for reliable sources and distinguish between real and false news.

In this study, the proposed solution was to use the Multi-layer Perceptron (MLP) neural network algorithm that obtained satisfactory results as seen in the results section, to create a model that was able to learn the patterns of preprocessed texts and for detect false news.

As future work is expected to improve the accuracy, precision of the MLPClassifier algorithm, start the process of implementing layer 2 of the model proposed in Figure 1 to search and group related news for dynamism. In learning from the algorithm, increase the database with more fake and true news texts and finally apply the idea of AuFa in some software so that the end user can use it.

## References

- [1] Rafael Batista. A divulgação de notícias falsas, conhecidas como fake news, pode interferir negativamente em vários setores da sociedade, como política, saúde e segurança. <https://mundoeducacao.bol.uol.com.br/curiosidades/fake-news.htm>, 2018. Accessed: 06/07/2019.
- [2] Sonia Castelo, Thais Almeida, Anas Elghafari, Aécio Santos, Kien Pham, Eduardo Nakamura, and Juliana Freire. A topic-agnostic approach for identifying fake news pages. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 975–980. ACM, 2019.
- [3] Rosanne D'Agostino. Três anos depois, linchamento de fabiane após boato na web pode ajudar a endurecer lei. <https://g1.globo.com/e-ou-nao-e/noticia/tres-anos-depois-linchamento-de-fabiane-apos-boato-na-web-pode-ajudar-a-endurecer-lei.ghtml>, 2017. Accessed: 06/07/2019.
- [4] Statista Research Department. Internet usage in brazil - statistics facts. <http://www.digitalnewsreport.org/survey/2018/brazil-2018/>, 2017. Accessed: 08/10/2019.
- [5] Xosé Hermida. Fake news tornam o jornalismo de qualidade mais necessário do que nunca, diz diretor do el país. <https://bit.ly/2K2zLqK>, 2018. Accessed: 06/07/2019.
- [6] Ithalo Henrique de Sousa Leal. O uso de aprendizagem de máquina para identificação e classificação de fake news no twitter referentes a eleição presidencial de 2018, 2018. Monografia (Bacharelado em Ciência da Computação), Faculdade Doctum de Caratinga.
- [7] Regina Marchi. With facebook, blogs, and fake news, teens reject journalistic “objectivity”. *Journal of Communication Inquiry*, 36(3):246–262, 2012.
- [8] Fabiano Shiiti Marumo. Deep Learning para Classificação de fake news por sumarização de texto., 2018. Monografia (Bacharelado em Ciência da Computação), Universidade Estadual de Londrina.
- [9] Roger Oliveira Monteiro, Rodrigo Ramos Nogueira, and Greisse Moser. Desenvolvimento de um sistema para a classificação de fakenews com textos de notícias em língua portuguesa. 2019.
- [10] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David Levy, and Rasmus Kleis Nielsen. Reuters institute digital news report. 2017.
- [11] Thiago A. S. Pardo Rafael A. Monteiro, Roney L. de Sales. Detecção automática de notícias falsas para o português. 2018. <https://nilc-fakenews.herokuapp.com/about>. Acessado em 11/12/2019.
- [12] Digital News Report. Urban brazil. <http://www.digitalnewsreport.org/survey/2018/brazil-2018/>, 2018. Accessed: 11/10/2019.
- [13] Nick Wingfield, Mike Isaac, and Katie Benner. Google and facebook take aim at fake news sites. *The New York Times*, 11:12, 2016.

# Integrating Blockchain with IoT for a Secure Healthcare Digital System

Nada Chendeb<sup>1</sup>, Nour Khaled<sup>1</sup>, and Nazim Agoulmine<sup>2</sup>

<sup>1</sup> Lebanese University, Faculty of Engineering  
nchendeb@ul.edu.lb, nouna.nour97@gmail.com

<sup>2</sup> University of Evry Val d'Essonne - Paris Saclay University  
nazim.agoulmine@ibisc.univ-evry.fr

## Abstract

In the past few years, the number of wireless connected devices has increased to a number that could reach billions in the next few years. While cloud computing is being seen as the solution to process this data, security challenges could not be addressed solely with this technology. Blockchain is the technology that underpins Bitcoin, it introduces manners to provide a fully autonomous secure system, by using smart contracts. Multi-layer BC is a very powerful solution to overcome many IoT challenges. This paper illustrates how Blockchain works, what are the IoT challenges, and how it can be integrated with Blockchain. We proposed in this work a multi-layer IoT/blockchain based architecture customized and designed to be used in the medical field. With this information interact many parties including the doctors, health service providers, insurance companies and pharmacies. The ultimate goal being to solve the problem of scalability and performance.

## 1 Introduction

Data generated from IoT devices is increasing dramatically. One of the most interesting applications of IoT is e-health or intelligent medical care. Medical data generated by IoT devices is critical and sensitive to any unauthorized access. This data should be protected carefully because it is directly related to patient's life. Security concerns are more accentuated in the medical field and need a special attention especially when this field embraces IoT.

In parallel to the advancement in the e-health domain, a new technology that was conceived first to secure financial transactions of the famous cryptocurrency bitcoin, was and still developing to find its applications in many domains including the medical field. Blockchain technology is a peer-to-peer technology that provides a global consensus and assures that no one can alter or change previously validated transactions. Blockchain is a very good solution for security but it still suffer from some problems especially when used by IoT devices. In this case, arise multiple problems such as scalability, complexity and architectural based problems. In this work, we aim to build an intelligent medical system in which all partners interact in an IoT/Cloud environment with protocols for communication, management and sharing of private data using blockchain technology. We aim mainly to solve the problems related to the integration between IoT and blockchain.

The remaining of this paper is organized as follow: section 2 contains some preliminaries to understand the topic and section 3 contains a discussion of the related work. In section 4 we present our design and solution for a new architecture integrating blockchain and IoT and in section 5 we discuss this architecture and explain how it works in different medical data exchange scenarios. Finally, section 6 concludes the paper and gives some perspectives.

## 2 Preliminaries and Definitions

### 2.1 What is Blockchain?

Blockchain is an information technology that allows transactions to be verified by a group of unreliable actors. It provides a distributed, immutable, transparent, secure and auditable ledger [1]. This is mainly a distributed database of all transactions or digital events that have been executed and shared among participating parties. Once entered in the blockchain, information can never be modified or erased. The blockchain contains a certain and verifiable record of every single transaction ever made [2]. In a blockchain network, whenever a new transaction (record) is created, a new block is automatically generated stating the date and the time (known as a “timestamp”) when the record was entered in the block. Each new block is automatically linked to its previous block, all the way to the originating block, using the previous block’s “Hash”. Every time a new block is created, it is broadcasted in real-time to all connected computers that participate in the blockchain network. These computers are known as “nodes”. While creating a new block, the node uses its own Private Encryption (Crypto) Key and the Public Crypto Key of the receiving node that is also a node in the network. Without its Private Key, no node can create a new block [3]. When we talk about information security, blockchain can be majorly divided into two parts: hashing and digital signatures [4].

### 2.2 Smart Contracts

A smart contract is a computer program that directly controls the transfer of digital information or assets between parties under certain conditions[5]. Blockchain is ideal for storing smart contracts because of the technology’s security and immutability. Smart contract data is encrypted on a shared ledger, making it impossible to lose the information stored in the blocks.

### 2.3 Multi-layer Blockchain

Multi-layer blockchain can combine the benefits of blockchain and IoT and gives a solution for all the problems that cannot be solved by using one technology alone. By combining these two technologies, multi-layer blockchain gives a next-generation platform for IoT devices that are based on the blockchain technology. This is a multi-layered architecture, the main advantage of this architecture is to solve the problems faced by current blockchains mainly the lack of scalability. The IoT devices could be the nodes in the private blockchains, some of them are also part of the next layer public blockchain. We will base our proposed architecture on this main idea.

### 2.4 Mining

Mining is the concept of validating a block, it varies between different types of blockchain. What’s important to know is the fact that a lot of computational power is needed for becoming a ”Miner”. Miners are usually rewarded, and mining might affect the whole system’s performance.

## 3 Review of Literature

We start our discussion by the work done in [1] where we find three types of integration:

**IoT- IoT:** This approach could be the fastest one in terms of latency, and security since it can work offline. This approach is also applicable in scenarios involving only IoT devices.

**IoT- Blockchain:** In this approach, all the interactions go through blockchain, enabling an immutable record of interactions. Nevertheless, recording all the interactions in blockchain would involve an increase in bandwidth, which is one of the challenges.

**Hybrid approach:** Where only part of the interactions and data take place in the blockchain and the rest are directly shared between the IoT devices. In this approach fog/cloud computing [7] could come into play to complement the limitations of blockchain and IoT.

The last approach, may be a good candidate for our solution, the challenge posed by this approach is to optimize the split between the interactions that occur in real-time and the ones that go through the blockchain.

If we want to move to the architectures and models, authors in [8] speak about the Cloud Storage, Overlay network, etc. So, instead of saving the IoT data over blockchain, we use cloud storage servers to save the patient data. The cloud storage groups user's data in identical blocks associated with a unique block number. These clouds are connected to overlay networks, once the data stored in a block, the cloud server sends the hash of the data blocks to the overlay network. According to [9] and [10], the design consists of three core tiers that are smart home, cloud storage, and overlay network. Smart devices are located inside the smart home tier and are centrally managed by a miner. Smart homes constitute an overlay network along with Service Providers (SP), cloud storage, and users' smartphones. All transaction to or from the smart home are stored in a local private BC. Smart home miner is a device that centrally processes incoming and outgoing transactions to and from the smart home.

Furthermore, the architecture is more clarified in [11], where there is also three layers: device, fog, and cloud. At the edge of the network, the device layer is used to monitor the various public infrastructure environments and sends the filtered data that is consumed locally to the fog layer and uses the request services.

Most importantly, the authors of [12] make a full architecture that consists of three tiers: Tier 1 contains the constrained and unconstrained nodes (devices - sensors) and patient (gateway - aggregator), Tier 2 groups N Authorities (hospitals - labs - medical Insurance organization - medical research institute ...) and Tier 3 for the EHRs cloud providers (cloud storage servers for EHRs records).

Based on this study, and in order to meet the objective of our research, we will adopt the hybrid approach for IoT - blockchain integration with cloud storage of patients' data. A multi-layer blockchain is also a good candidate to alleviate the scalability problem. We think that not all IoT generated data in real time should be stored in the public blockchain, so a private blockchain for home related sensory data of the patient is a good solution, periodical report generated from the gateway could be stored in the public blockchain of a higher layer. A three-tier architecture is a good architecture for the system we will propose in the next section. This system will be mainly based on the solution proposed in [13] within the same research project.

## 4 The Proposed Three-Tier Blockchain Architecture

### 4.1 Baseline Solution/Design

Recently, members of the research project we are working on already proposed a basic architecture to secure medical data using blockchain [13], [14]. They proposed a communication protocol between nodes based on publish/subscribe model which is a model used specifically by IoT devices. They also propose an access control and management scheme based on the use of smart contracts, they define multiple smart contracts for publishers of IoT data, subscribers to

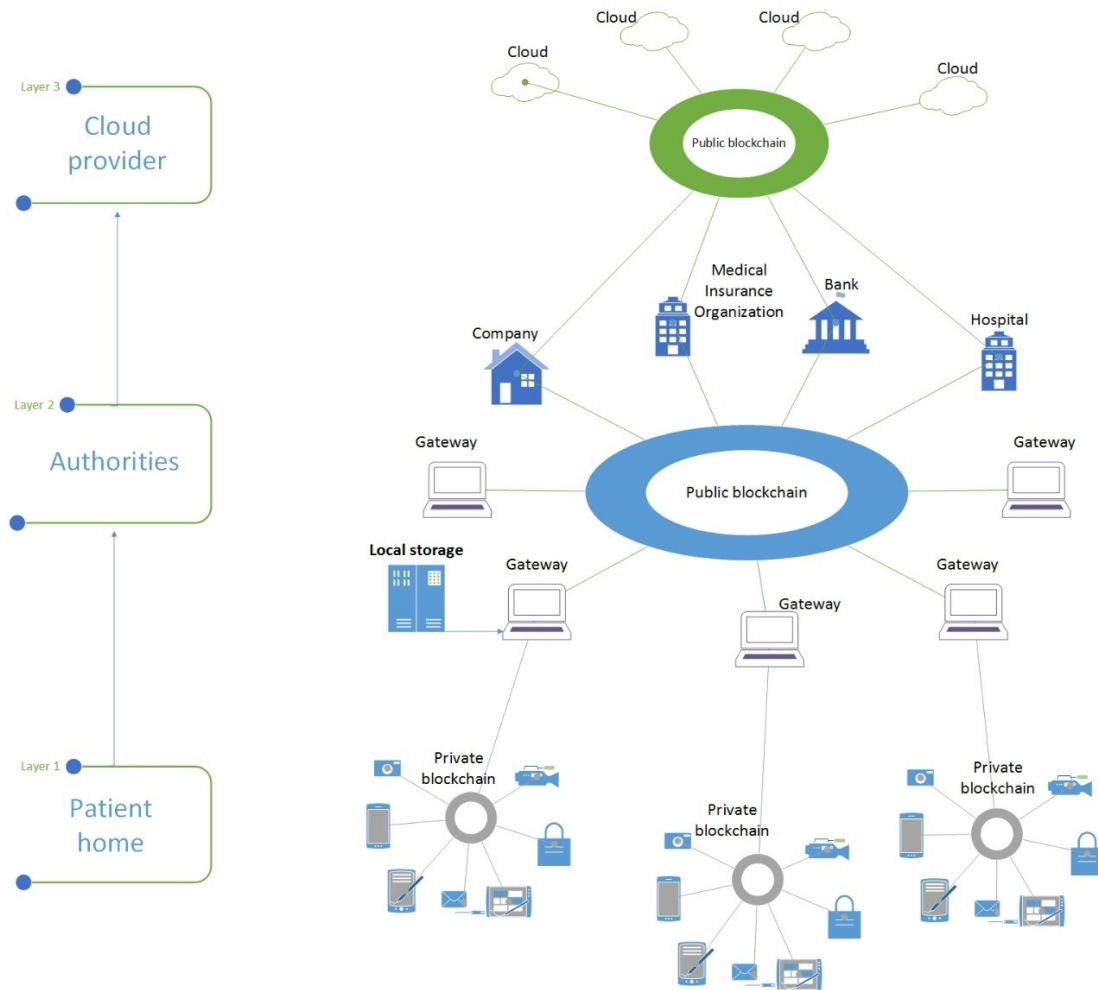


Figure 1: Integration of Blockchain with IoT, the three-layer architecture

that data and access permission done by the data owners. These smart contracts when executed define who can access to any generated data from any IoT device. Because IoT devices have capacity limitations, an off-chain storage was adopted to store the private data. Therefore, data is stored in distributed storage and hash of the data location is stored in the blockchain. The proposed solution is a one-layer blockchain solution, which suffers from scalability limitations. In addition, IoT devices are data generators only and could not be considered as nodes in the blockchain making the integration of blockchain in the IoT very poor. We think that it is necessary to leverage this solution toward a better integration and better performance.

## 4.2 System Architecture

The system architecture is composed of three layers as shown in figure 1

**Layer 1:** This is the layer related to the “patient”; it includes all the IoT nodes gathering

information from a specific patient; medical information or any other type of information describing his environment. A main node in this blockchain is a powerful computer that will act as a gateway to higher layer blockchain. Note that each patient will have its proper blockchain.

**Layer 2:** This is the core layer or the “authorities” layer; it contains representative nodes for all interacting parties in the medical field who have interest in the data related to patients such as hospitals, medical centers, labs, etc. The gateways in the first layer are also members of this blockchain.

**Layer 3:** This is the higher layer or the “cloud provider” layer. In fact, IoT devices are in general completed by processing and storage capacities in the cloud. A blockchain at the cloud level is necessary to control the interaction between cloud providers in order to grant access for patient’s data to each others wherever it is located.

To understand how the system will work, we recall that we are using the publisher/subscriber’s model described in [13], smart contracts for access management and off-chain database for storage. In the core blockchain (layer 2), and based on the description of the access management in [13] we consider that:

○ **Gateways:** are the publishers that generate all the data related to a certain patient (medical data). Publishers specify who can access and who has the permission to read/write/modify its data in the cloud (using smart contracts).

○ **Authorities:** are the subscribers that are able to access the data of the publishers in the cloud. They also have the right to write and modify data according to the access rules specified by the publishers.

In the lowest layer, we propose to use private blockchain. In this platform, there is only one miner, which is the gateway in our case. Moreover, all IoT devices are generating data so the gateway pre-processes this data and generates records to higher layers. Hence, the gateway replaces all the IoT devices and acts like a publisher of the data generated from these devices.

### 4.3 Workflow in Different Scenarios

#### 4.3.1 Scenario 1: The gateway collects IoT data and generates a new record

In this scenario, the entire network will be active. In the private BC (BlockChain), the sensors and their gateway are the nodes, so the miner is the gateway because it is the most powerful node in its private blockchain. Every device has to authenticate with the network before starting to send data by using public and private keys. These two keys are specific to each device. The gateway saves all the keys in its local storage to easily recognize any device that authenticates with it.

With every private blockchain, there might be a local storage. Note that only filtered, processed and abnormal data that reflects critical situations should be stored in the cloud. After completing the registration, the device starts creating a new block. This block, once validated by the gateway will be added to the patient’s private BC (Figure 2). All the data collected is saved in the off-chain database (at the gateway local storage). The gateway processes the data and creates periodically medical records. However, the gateway has to be registered in the remaining two layers’ blockchains: layer 2 to communicate with different types of authorities, and layer 3 to save some periodic records and the emergency data. All information in the next steps only concern the periodic/emergency records.

In the layer 2 blockchain, we have mainly the interactions between the patients and different types of authorities; we may also have interactions between the authorities themselves. Here we have the Proof of Work (POW) [15], and the Proof of Stake (PoS) to validate any transaction

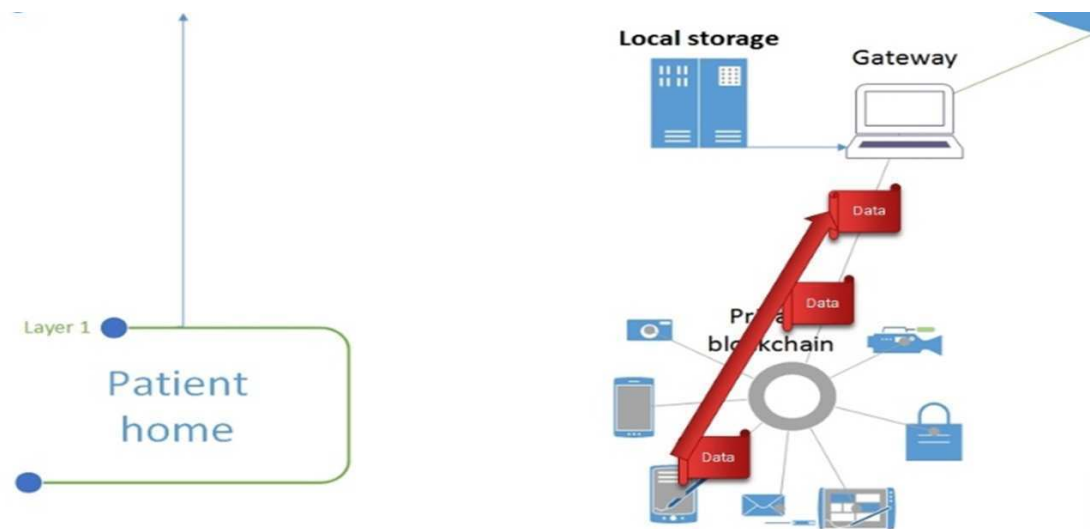


Figure 2: The gateway collects data from a device

based on the previous ones. This means that there should be many miners to mine the blocks and validate them. After the registration, the gateway creates a new block in the public BC (Ethereum is a good choice), to tell the concerned authorities (healthcare specialists who care of this patient) that there is newly created data (Figure 3). The gateway saves the record in the cloud layer, meanwhile the cloud produces (creates) a block to note that new data is created.

#### 4.3.2 Scenario 2: Gateway/Authority Want to Access Patient's Medical Record

In this scenario, the records containing patient's data have already been stored in the patient's cloud provider. The gateway/authorities that have permission need to access a record that has already been stored. The main work is concentrated in layer 3 when all cloud providers are linked, using public blockchain. The control of access and permission is done by using the cloud contract. The patient has a direct access to the data in the cloud by using its account; it is simply the data owner. When it stores the record, the cloud gives back the ID of this record. Using this ID, the patient can access this record. The cloud creates then a transaction to note that the patient has accessed the record of that ID on this date. The authority sends its ID to the appropriate cloud and waits for the ACK. The ACK in this case is the result of finding that the authority's ID is one of the allowed IDs of the list in the cloud's contract.

#### 4.3.3 Scenario 3: Patient Visits and Interact with an Authority

A new block is added to the layer 2 BC when a patient visits an authority. Consequently, the same block will be added to all authorities and a related block will be added to the cloud's provider blockchain. When the patient finishes its visit, the authority adds a new block to the public BC that includes the ID of the authority, the ID of the patient and some information about the data stored in the off-chain storage of this authority (this data might be medical or



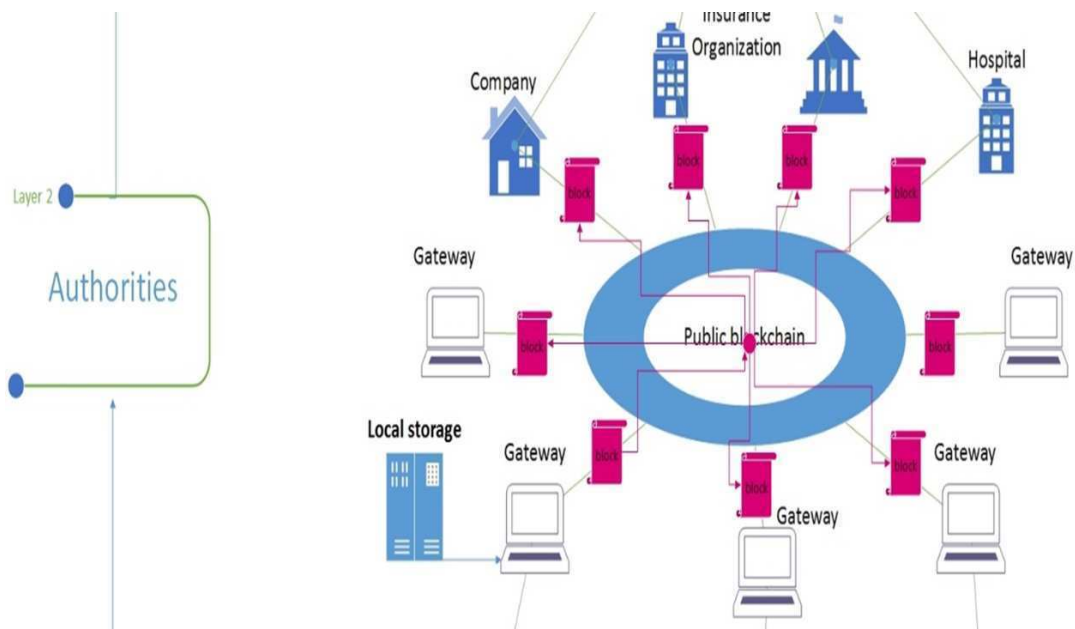


Figure 3: Gateway adds new block to the ledger of BC in layer 2

administrative). The visited authority creates in the cloud's blockchain a block to note that the patient of this ID has visited the appropriate authority. It even notes the place where the data had been stored.

## 5 Conclusion and Future Works

In this work, we proposed a new distributed blockchain cloud architecture model to meet the design principles required to efficiently manage the raw data streams produced by numerous IoT devices. We were able to verify the possibility of using blockchain technology with IoT and vertical applications, by proposing, implementing and testing a multi-layer structure. The proposed architecture was designed to support high availability, real-time data delivery, high scalability, security, resiliency, and low latency.

To complete this work, multiple performance tests regarding mining time, difficulty variation, and blockchain size should be achieved which will help further enhancements in this domain. Finally, what is important for the future of this technology, and this proposal, in particular, is to implement this architecture in a real system, and continue to evaluate the performances.

## Acknowledgments

This research was supported by The Lebanese University and CNRS Lebanon.

## References

- [1] Ana Reyna, Cristian Martin, Jaime Chen, Enrique Soler and Manuel Díaz, "On blockchain and

- its integration with IoT. Challenges and opportunities”, *Future Generation Computer Systems*, vol. 88, pp. 173-190, 2018
- [2] Michael Crosby, Nachiappan, Pradhan Pattanayak, Sanjeev Verma and Vignesh Kalyanaraman, ”Blockchain Technology”, *Sutardja Center for Entrepreneurship and Technology*, 2017
- [3] Gemalto, ”Blockchain Security: 3 Ways to Secure Your Blockchain”, 04 December 2018. [Online]. Available: <https://blog.gemalto.com/security/2018/12/04/blockchain-security-3-ways-to-secure-your-blockchain/>
- [4] Preeti Seth, ”Smart Contracts: The Blockchain Technology That Will Replace Lawyers”, SYSTweak, 13 06 2018.[Online]. Available: <https://blogs.systweak.com/an-insight-into-hashing-digital-signature-in-blockchain/>
- [5] Ameer Rosic, ”Smart Contracts: The Blockchain Technology That Will Replace Lawyers”, 2016 . [Online]. Available: <https://blockgeeks.com/guides/smart-contracts/>
- [6] Ameer Rosic, ”What is Ethereum? [The Most Comprehensive Guide Ever!]”, Blockgeeks, 2016 . [Online]. Available: <https://blockgeeks.com/guides/ethereum/>
- [7] Hany F. Atlam, Robert J. Walters and Gary B. Wills, ”Fog Computing and the Internet of Things: A Review”, *sensors*, 15 January 2019
- [8] Ashutosh Dhar Dwivedi, Gautam Srivastava, Shalini Dhar and Rajani Singh, ”A Decentralized Privacy-Preserving Healthcare Blockchain for IoT”, *Big Data Cogn*, vol. 10, 2018
- [9] Ali Dorri, Raja Jurdak, Salil S. Kanhere and Praveen Gauravaram, ”Blockchain for IoT Security and Privacy: The Case Study of a Smart Home”, vol. 161, March 2017
- [10] Ali Dorri, Salil S. Kanhere and Raja Jurdak, ”Blockchain in Internet of Things: Challenges and Solutions”
- [11] P. K. Sharma, M.-Y. Chen and J. H. Park, ”A Software Defined Fog Node Based Distributed Blockchain Cloud Architecture for IoT”, *Special Section on intelligent systems for the internet of things*, vol. 10.1109, 2017
- [12] Shaimaa Badr, Ibrahim Gomaa and Emad Abd-Elrahman, ”Multi-tier Blockchain Framework for IoT-EHRs Systems”, *ScienceDirect*, vol. 141, p. 159–166, 2018
- [13] Nabil Rifi, Nazim Agoulmine, Nada Chendeb Taher and Elie Rachkidi, ”Towards using blockchain technology for IoT data access protection”, *IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*, 2017
- [14] Nabil Rifi, Nazim Agoulmine, Nada Chendeb Taher and Elie Rachkidi, ”Blockchain Technology: Is It a Good Candidate for Securing IoT Sensitive Medical Data?”, *Wireless Communications and Mobile Computing journal*, 2018
- [15] Georgios Konstantopoulos, ”Understanding Blockchain Fundamentals, Part 2: Proof of Work and Proof of Stake”, loom, 8 Dec 2017. [Online]. Available: <https://medium.com/loom-network/understanding-blockchain-fundamentals-part-2-proof-of-work-proof-of-stake-b6ae907c7edb>

# Analysis of Interoperability in Public Health Systems\*

Leonardo Nascimento<sup>1†</sup>, Renato Freitas<sup>1‡</sup>, Cesar Olavo<sup>1§</sup>, Mauro Oliveira<sup>1¶</sup>, Ivana Holanda<sup>2||</sup>, and Odorico Monteiro<sup>2\*\*</sup>

<sup>1</sup> Federal Institute of Ceara (IFCE), Fortaleza, Ceará, Brazil  
leonardobrendoti@gmail.com, renato.freitas@ppgcc.ifce.edu.br  
cesar.olavo2011@gmail.com, amauroboliveira@gmail.com

<sup>2</sup> Department of Public Health - FIOCRUZ, Fortaleza, Ceará, Brazil  
ivana\_barreto@ufc.br, odorico@ufc.br

## Abstract

Health information systems (HIS) are constantly evolving in both complexity and data volume. Nevertheless, not every HIS share the same data standard or are otherwise able to communicate with one another. The need for communication among these systems has resulted in a worldwide effort to develop interoperability standards for healthcare systems. Brazil also has adopted measures to foster interoperability in public health services, notably with the issue of the ordinance 2.073 of August 31, 2011 by the Ministry of Health, which regulates the adoption of international standards. This article presents an analysis of the different types of interoperability in public health systems, using GISSA, an intelligent system to support decision making in maternal and child health, to showcase them. A prototype was implemented that addresses the problem of interoperability from a structural viewpoint, by aggregating new services to the GISSA legacy version and also from a semantic viewpoint, by enabling the coexistence, in one system, of the two main electronic health record standards, i.e, FHIR and OpenEHR.

## 1 Introduction

In the health sector, medical entities generate a dense volume of information on a daily basis in diverse formats. One of the priorities highlighted by these entities is directly linked to the integration of these ISs with their partners, ensuring the exchange of information. This integration of ISs is a fundamental feature of the concept of systems interoperability, object of this work. The complexity of medical activity requires interoperable health information systems (HIS), a worldwide trend. Countries such as Canada, France, the United Kingdom, the United States, Australia, New Zealand, among others, are taking measures to enable interoperability in public services.

This paper analyzes the types of interoperability in public health systems. In addition to the interoperability stratification, an architecture that solves interoperability problems in the aggregation of new services in legacy health systems is presented. As proof of concept for the architecture presented the framework GISSA (Intelligent Governance in Health System) was

---

\*Other people who contributed to this document include Maria Voronkov and Graham Gough.

Leonardo, Renato, Cesar, Ivana, Odorico and Mauro

used, an intelligent decision-making system in maternal and child health. The types of interoperability addressed in GISSA permeate the semantic levels, Health Information Standards, and aggregation of new services.

This paper's organization is presented hereafter. Section two presents related works that use HIS and interoperability concepts. Section three describes theoretical Foundation about interoperability and GISSA. Section four addresses the main contribution of this article: The health system interoperability classification and identification. Finally, section five presents the conclusion of this paper.

## 2 Related Works

In [3], a research is described whose purpose is to deepen the identified issues that are related to the problem of semantic interoperability on the patient information record in the openEHR standard. The results express that the use of specific tools may be relevant in the search for semantic interoperability between systems and that the informality in the definition of the terms that will be used in information system vocabularies can prevent a clear understanding of the desired meaning. The architecture proposed by [1] enabled the creation of a software called SmartBeat, whose purpose was the development and evaluation of an intelligent system for the management of heart failure in senior people. The results ensure interoperability between the SmartBeat system and the e-Prescription (software service for electronic prescription). In [4], an architecture for developing a SOA-based EHR system is described, taking into account interoperability between legacy systems. The results indicate that in the literature there is a deficiency precisely in the definition of an interoperable architecture for specific systems, and a deployment architecture was defined.

## 3 Theoretical Foundation

### 3.1 Interoperability

In Computer Science, interoperability is the ability of a system to share, communicate, and exchange applications and information with other systems that have disparate structures and data[8]. For [2], interoperability is defined as the quality capable of making one system or process use the information and / or functions of another system or process by adhering to common standards. In [6] 's view, interoperability is understood as a continuous process centered on ensuring that systems or processes exchange information.

### 3.2 GISSA

The GISSA project is a framework that emerged from the LARIISA platform [7]. It is an intelligent governance system to support decision-making in health environments, focusing on the Ministry of Health's Cegonha Network project, whose goal is to preserve the health of both mother and child, especially in the early years [5], being supported by Studies and Projects Funding. The GISSA framework is made up of a series of components that make it possible to collect, integrate and view information relevant to decision-making procedures [9]. Figure 1 shows the GISSA architecture.

Leonardo, Renato, Cesar, Ivana, Odorico and Mauro

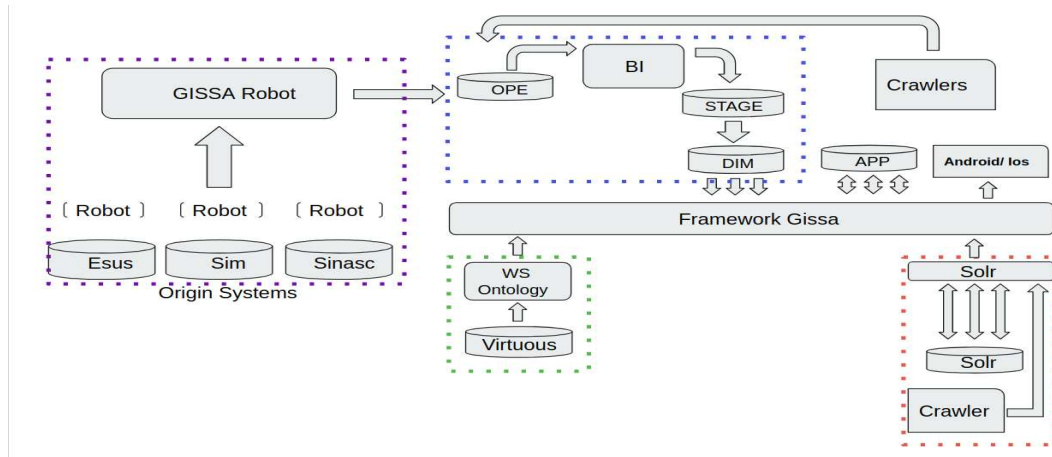


Figure 1: Gissa Architecture

## 4 GISSA interoperability

As stated in the Introduction, this paper analyzes the types of interoperability in public health systems, stratifying these types into independent models, highlighting semantic aspects, standardized Electronic Health Records (RES) and aggregating new services in legacy health systems, **GISSA semantic interoperability**, **GISSA RES interoperability** and **GISSA Interoperability in aggregating new services**.

### 4.1 GISSA semantic interoperability

The ontological view of GISSA it is a portal where ontologies and data are interconnected to face application development challenges, where there is a need to semantically integrate heterogeneous data sources. As it stands, the portal publishes two SUS databases available on the GISSA platform.

### 4.2 GISSA RES interoperability

Is being implemented a gateway and a Hub capable of integrating the multiple existing systems in the public and private health sectors allows. despite the different data formats using micro-services technology is being implemented for interoperability between systems with different standards, such as FHIR and OpenEHR. Structurally, this architecture comprises the Richardson Maturity Model, reaching the four implementation levels of an API RESTful, HATEOAS.

### 4.3 GISSA Interoperability in aggregating new services

Interoperability in GISSA happens through an abstraction layer that integrates types of data and services, enabling communication between the systems involved. The identification of interoperability is based on the application of the concepts presented in the Interoperability (technical, semantic, organizational, political and human, inter-community, legal and international). Figure 2 shows the interoperable architecture for aggregating new services.

Leonardo, Renato, Cesar, Ivana, Odorico and Mauro

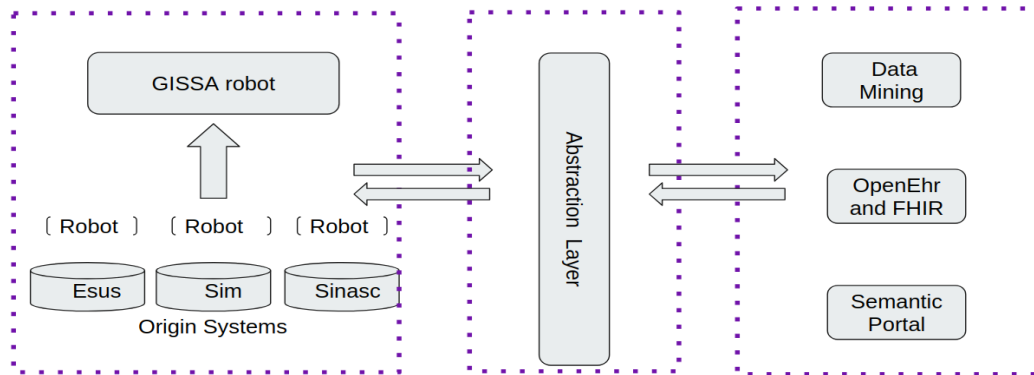


Figure 2: Interoperable Architecture for aggregating new services

## 5 Conclusion

Adding new services in legacy environments is not always an easy task, especially if the implemented system is not well documented. Although the GISSA framework was used as proof of concept of the proposed interoperability in the aggregation of new services, the study and analysis done in this paper can be used in other similar environments. In addition, the proposed architecture allows dealing with different levels of information, offering a layer of abstraction in which new data can be integrated. In the GISSA, interoperability was identified at the service aggregation level, health information representation standard (FHIR and OpenHER) and ontology.

## References

- [1] Margarida Amorim de Beir. Arquiteturas para interoperabilidade de sistemas de informação na área da saúde: caso de demonstração-prescrição eletrônica de medicamentos, 2018.
- [2] Ernani Marques dos Santos. Implementing interoperability standards for electronic government: An exploratory case study of the e-ping brazilian framework, 2008.
- [3] Fernanda Farinelli and Maurício Barcellos Almeida. Interoperabilidade semântica em sistemas de informação de saúde por meio de ontologias formais e informais: um estudo da norma openehr.
- [4] Josimar de Souza Lima et al. Uma arquitetura de software para implementação de um ehr utilizando soa considerando a interoperabilidade entre sistemas legados, 2016.
- [5] Gabriel Lopes, Vânia Vidal, and Mauro Oliveira. A framework for creation of linked data mashups: A case study on healthcare, 2016.
- [6] Paul Miller. Interoperability: What is it and why should i want it?, 2000.
- [7] Mauro Oliveira, Carlos Hairon, Odorico Andrade, Regis Moura, Claude Sicotte, JL Denis, Stenio Fernandes, Jerome Gensel, Jose Bringel, and Herve Martin. A context-aware framework for health care governance decision-making systems: A model based on the brazilian digital tv, 2010.
- [8] Amit P Sheth. Changing focus on interoperability in information systems: from system, syntax, structure to semantics, 1999.
- [9] Cristiano Silva, Joyce Quintino, Oton C Braga, Ronaldo Ramos, Odorico Monteiro, and Mauro Oliveira. Laís, um analisador baseado em classificadores para a geração de alertas inteligentes em saúde, 2017.

# SECURITY ISSUES OF HEALTHCARE IOT DEVICES

Sadry Fievet and Karima Boudaoud

Université Côte d'Azur, CNRS, I3S, France  
karima.boudaoud@univ-cotedazur.fr, sadry@orange.fr

## Abstract

The recent adoption of connected objects in the health sector raises the question of security of personal data that are collected and transmitted by these objects. Actually, cybercriminals show an increasing interest in stealing and reselling personal data thanks to the illegal revenue they can benefit from. To evaluate the security measures implemented by these objects, we have conducted penetration tests (pentests) on healthcare IoT devices targeting the mobile application designed for the connected objects and the network communications between the connected object and the smartphone of health professionals. The goal of this work in progress paper is to present the preliminary results of the pentests related specifically to network attacks.

## 1 Introduction

The tremendous expansion of the number of connected objects is spreading to more and more sectors today. The medical sector is no exception, the adoption of connected objects is growing. Although the benefit they bring to doctors, nurses and healthcare staff is the primary motivation for their success, we believe that the protection of the personal data they contain is the first characteristic that should be taken into account when they are purchased. A computer attack on these objects could lead to a disclosure of data of patients and hospital staff, the encryption of these same data for financial purposes (ransomware), or their alterations / modifications in a criminal purpose.

International regulations for the manufacturing, marketing and use of health appliances mean that manufacturers must comply with a certain number of standards. The results of the pentests presented in this work in progress paper concerns healthcare IoT devices that are compliant with the standards requirements of the medical sector [1].

This paper presents the results of pentests carried out on a medical IoT device that measures the body composition by bioelectrical impedance. In order to establish a protocol of experimentation, we started from the following postulate:

Computer security is defined by adding protections on various levels, namely *confidentiality*, *integrity* and *availability* of processed data. It is therefore enough for only one of these protections to be breakable in order to question the notion of computer security. We will present two types of network attacks that undermine the confidentiality and integrity of personal data. All of these attacks belong, as we will see in the next section, to a new type of cyberattacks specific to the medical sector named MEDJACK [2].

The rest of the paper is organized as follows. Section 2 gives an overview about the MEDJACK attack; Section 3 describes the pentests results regarding network attacks targeting Wi-Fi and BLE protocols; and, finally, Section 4 concludes this paper and gives an overview about future works.

## 2 MEDJAK Cyberattack

In 2015, a new kind of cyberattack, called MEDJACK, has been identified by the company TrapX [2]. MEDJAK targets healthcare IoT devices to compromise the information system of a hospital, clinic, laboratory or any other infrastructure that deals with health data. Actually, in this kind of infrastructures, healthcare IoT devices are considered as the weak points of the defence system and the most exploited gateways in the attacks targeting the information

system of the medical sector. Cybercriminals target specifically healthcare IoT devices because of the restricted processing power of this kind of devices. The methodology of a MEDJACK attack follows the following steps:

1. **Step 1:** The cybercriminal tries to gather as much information as possible on the targeted healthcare IoT devices in order to adapt his attack parameters.
2. **Step 2:** The cybercriminal attacks the targeted healthcare IoT device to gain access to the hospital network through the device.
3. **Step 3:** Being inside the hospital network, the cybercriminal starts a new recognition and identification phase to find the potential machines to attack.
4. **Step 4:** After finding a location where sensitive data are stored, such as personal health related or financial data, the cybercriminal extracts this and send them to an external server before deleting all the intrusion traces.

In this work in progress paper we will focus specifically on Step 2 that aims to compromise the healthcare IoT device.

### 3 Pentests Results

We now present our pentests results carried out on a specific IoT device that measures the body composition by bioelectrical impedance and use both WI-FI and Bluetooth (BLE version) network protocols. Due to the widespread use of these two protocols, our pentests are representative for a wide class of healthcare and medical IoT devices.

For the pentests, we used: 1) specific Wi-Fi card to inject network packets [3], 2) an Android smartphone to simulate the victim and 3) two machines, running on Debian OS, to attack the IoT device and the Smartphone.

To conduct the pentests, we have followed the usual methodology. We have first scanned the network protocols to identify the vulnerabilities and to collect some useful information (such as the security level, if communications were encrypted, etc...). Then, depending on the results, either we exploited the vulnerabilities or we analyzed more deeply the security implementation and in the context of this work we have done a code analysis to identify other weak points (i.e. No class defined for SSL certificate check) to select the right attack.

#### 3.1 Network Attacks on Wi-Fi Protocol

To identify quickly the potential vulnerabilities of the Wi-Fi protocol, we used the vulnerability scanner Drozer [4]. The first results indicated that there was no vulnerability that can be exploited with network injection attacks. In addition, we have seen that the HTTPS protocol is systematically used, which means that the network should be well protected. Therefore, we analyzed the mobile application code, particularly the Java class managing the authentication, where we have identified a recurrent problem related to the implementation of HTTPS. We have noticed that there were no classes related to the creation or verification of SSL certificates [5]. Finally, we isolated the function managing

```
protected void doInBackground(Void... arg0) {
    HttpClient client = new DefaultHttpClient();
    HttpPost httppost = new HttpPost(ConfigurationApi.SERVICERLOGIN);
    try {
        List<NameValuePair> namevalpair = new ArrayList();
        namevalpair.add(new BasicNameValuePair(NotificationCompat.CATEGORY_EMAIL, LoginActivity.this.usernamefromuser));
        namevalpair.add(new BasicNameValuePair("password", LoginActivity.this.passfromuser));
        httppost.setEntity(new UrlEncodedFormEntity(namevalpair, CharEncoding.UTF_8));
        HttpResponse httpResponse = client.execute(httppost);
        LoginActivity.this.responsecode = httpResponse.getStatusLine().getStatusCode();
        LoginActivity.this.responseStr = EntityUtils.toString(httpResponse.getEntity());
    }
}
```

**Figure 1.** Function managing mobile application authentication to the server

the sending of the connection credentials to the server. The third code line, see Figure 1, of this function confirms our intuition regarding a bad implementation of the HTTPS protocol. `HttpsURLConnection` object should have been used instead of `HttpClient` and `HttpPost` objects [6]. We have then decided to launch the MITM (Man in the Middle) attack [7] to intercept and modify data.

##### 3.1.2 The MITM attack

The goal of the MITM attack that we have launched was to intercept the data exchanged between the smartphone and the server (where are stored the medical data collected from the IoT health device) when the victim connects to a web site in order to access to the medical data and after being authenticated. Theoretically, this kind of attacks has no chance of success against HTTPS. Actually, even if a cybercriminal gains a MITM position [7], she/he will intercept encrypted data, which will be useless. In practice, since 2009 [8], cybercriminal could bypass HTTPS protections and access to unencrypted data that can then be intercepted, modified or replayed.





## 5 Acknowledgement

We thank Ibrahim ABDELKADER, Dhiaeddine GUEZGUEZ and Sarah LOUKHMI who have also participated in the pentests.

## 6 References

- [1] ISO13485:2016 <https://www.iso.org/standard/59752.html>
- [2] MEDJACK, Anil Chacko, Thaier Hayajneh: Security and Privacy Issues with IoT in Healthcare. EAI Endorsed Trans. Pervasive Health Technol. 4(14): e2 (2018) <https://eudl.eu/pdf/10.4108/eai.13-7-2018.155079>
- [3] Wi-Fi card [https://www.alfa.com.tw/products\\_detail/8.htm](https://www.alfa.com.tw/products_detail/8.htm)
- [4] Drozer Android vulnerability scanner <https://labs.f-secure.com/tools/drozer/>
- [5] Android SSL certificate class <https://developer.android.com/reference/android/net/http/SslCertificate>
- [6] HttpURLConnection class <https://developer.android.com/reference/java/net/ssl/HttpsURLConnection>
- [7] MITM (Man in the Middle attack) [https://en.wikipedia.org/wiki/Man-in-the-middle\\_attack](https://en.wikipedia.org/wiki/Man-in-the-middle_attack)
- [8] SSLStrip (Moxie Marlinspike) <https://moxie.org/software/sslstrip/>
- [9] Rogue access point [https://en.wikipedia.org/wiki/Rogue\\_access\\_point](https://en.wikipedia.org/wiki/Rogue_access_point)
- [10] Bettercap <https://www.bettercap.org/>
- [11] GATT <https://www.bluetooth.com/specifications/gatt/>
- [12] Hcitol & Gatttool <https://github.com/pcborenstein/bluezDoc/wiki/hcitol-and-gatttool-example>
- [13] Gattacker <https://github.com/securing/gattacker>

# Heimdall: An Authorization Framework Based on Blockchain for Sensitive Data Access

Bruno L. A. Batista<sup>1</sup>, José Neuman de Souza<sup>2</sup>, and Joaquim Celestino Júnior<sup>3</sup>

<sup>1</sup> Fortaleza University,  
Fortaleza, Ceará, Brazil  
bruno.lopes@unifor.br

<sup>2</sup> Ceará Federal University,  
Fortaleza, Ceará, Brazil  
neuman@ufc.br

<sup>3</sup> Ceará State University,  
Fortaleza, Ceará, Brazil  
joaquim.celestino@uece.br

## Abstract

The Internet of Things (IoT) is transforming the way that we interconnect the devices, collect data and make computation over these data. But to achieve the widespread adoption of IoT it's necessary to improve the security of the IoT. In this paper, we propose Heimdall, a distributed smart contract-based framework that provides access control for sensitive or personal data collected by IoT devices that follow the prerogatives of GDPR and LGPD laws.

## 1 Introduction

The Internet of Things (IoT) is a new communication paradigm that is transforming the modern wireless communication [4]. A whitepaper published by IHS Technology shows the growth of IoT devices based on market estimative since 2015 until 2025, as we can see in Figure 1. There are considerable market for IoT-based services businesses. Healthcare is the one of IoT impact areas that project to form a biggest impact [1]. The annual economic impact caused by IoT in 2025 is estimated by range of US\$2.7 trillion to US\$6.2 trillion [11].

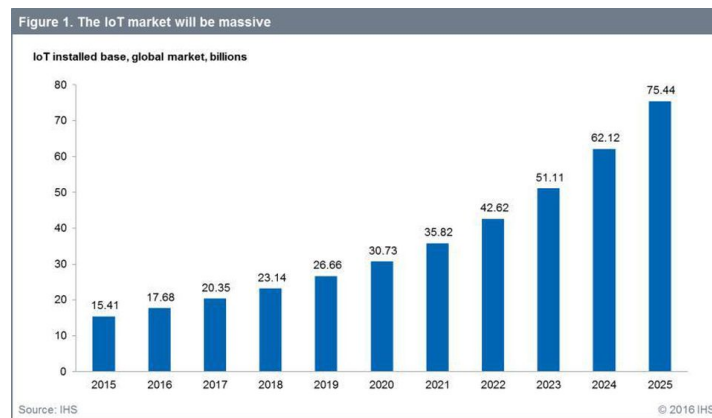


Figure 1: Global market of IoT devices [10]

However many researchers are point to many related security problems, as such as, data confidentiality, privacy and trust. These components are critical to large-scale adoption of IoT in the most diverse segments of contemporary society [12]. In this scenario, the Blockchain it being used by many researchers together with IoT to provide security for IoT devices. A paper published in Journal of Medical System [2] did a systematic review of all Blockchain and eHealth related articles since 2010. The authors get a total of 84 related publications, where they did a process of disposal and reading and they get a subset of 18 publications with relevant contributions, where 33,33% of this publications deals specifically with privacy and security in eHealth.

In 2016 the European Union approved the General Data Protection Regulation (GDPR), a law that guarantee the protection of data and identity of European Union citizens [8]. Also in Brazil was approved in 2018 the General Data Protection Law (Lei Geral de Proteção de Dados, LGPD) that is a version of GDPR adapted to Brazil reality [6].

Nowadays, any medium or huge IoT project has dozens of millions of IoT devices acting as sensor or actuator, collecting sensitive data from people or processing this data for the most varied purposes. In the GDPR or, in the Brazilian counterpart, the LGPD, all citizens have total control of your personal data and the power to to authorize, revoke and update the access of your personal. Thence all IoT projects must be in compliance with the legal impositions of countries, without the project will have negative implications and will not be viable.

So we raise the following question: Is it possible to create a framework of access control that allows the users to define whom, when and what may be accessed from sensitive and personal data collected by IoT devices? We assume it is possible to create this framework using Blockchain to achieve the expected goal it providing to the user a way to define access control to your data in a fine-grained way.

The main objective of this work is project a framework of access control of sensitive and personal data based on smart contract. The users must be able to: Authorize or revoke the data access without depends on a trusted third party; To define what subset of your sensitive or personal data may be accessed or not based on access rules; To audit whom accessed your sensitive or personal data and be able to see when this occurred.

To achieve this objective is necessary develop a smart contract that rules the access control, create the protocol that defines the possible operations on the access control framework and the involved actions, define the syntax and semantics of access rules based on JSON or YAML, design a distributed mechanism to verify the access rules and create a multidimensional index of sensitive and personal data to facilitate queries on stored data.

## 2 Related Work

Therefore many researches has been proposed some solutions for authentication and authorization mechanism for IoT-based solutions using Blockchain [7, 13, 14, 9]. However, some solution is based on assumptions that cannot be generalized in a broader context such as assume the all sensor that collect data belongs to the owner of data [7], some solutions don't care about encryption of stored data [14]. Notwithstanding all solution work with Blockchain [7, 13, 14, 9], all of them has at least one centralized software component in your architecture, allowing a malicious user to compromise the entire system.

So the purpose of this paper is to create a smart contract-based access control solution that will operate fully distributed on a Blockchain. All communication between the parties involved must be encrypted and users' sensitive and personal data will be stored in a distributed file system, increasing the resilience and scalability of the solution.

### 3 Solution Architecture

We will show the architecture of the proposed solution for smart contract-based access control that we are calling Heimdall (Nordic god that has the mission to guard the access to Bifrost bridge that interconnect the sky with the earth) and comment on the main components. Figure 2 shows the architecture of the proposed solution.

The proposed solution will be tested in the eHealth scenario, at first. In the scenario we will have some patients with Wireless Body Area Network (WBAN) of sensor collecting biomedical data from them. All sensor of WBAN will be connect to a gateway that belongs to an IoT network and the one of objectives of the gateway is to encrypt the collected data and send to a docker container.

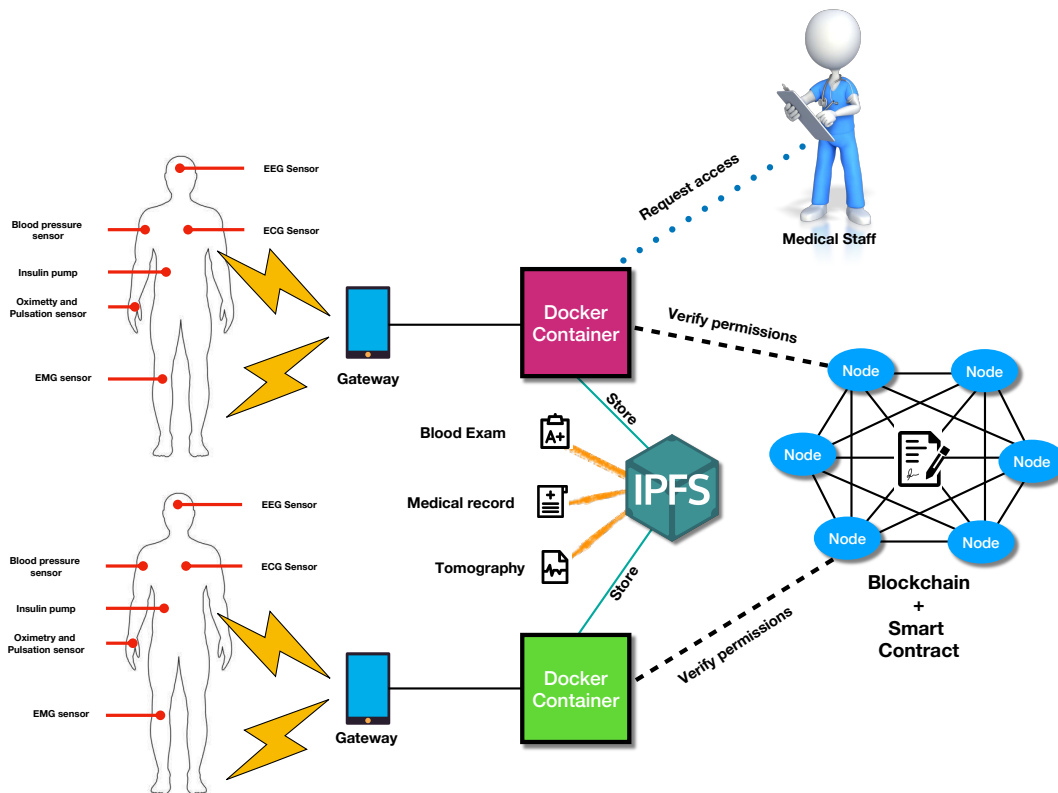


Figure 2: The Heimdall Architecture [10]

The docker container will act like a agent that manages all the data collected by the sensor of WBAN or from other resources (e.g. digitized tomography or blood exam). All data will be encrypted and stored in a distributed file system by the container to increase the storage scalability of the solution. IPFS will be used as a distributed file system by providing a block storage model addressed to high throughput content with hyperlinks addressed to content [5].

Another feature of docker container is to receive the request to access sensitive or personal data (e.g. request from medical staff) and analyze with the user who made a request has the authorization to access the request data. All authorization rules will be stored in the Blockchain

using a smart contract for that.

So the user can authorize, revoke or update them access to your sensitive or personal data executing the actions exposed by the smart contract since the user provides to smart contract his credentials. If a malicious user attacks the docker container, at worst, only the sensitive or personal data of the related user will be compromised. To compromise the whole system the Blockchain must be compromised and we use a permissive Blockchain implementation called Hyperledger Fabric [3] that provides some security features that increase the confidence and the trust of Blockchain.

## References

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys Tutorials*, 17(4):2347–2376, Fourthquarter 2015.
- [2] Susel Alonso, Jon Basañez, Miguel Lopez-Coronado, and Isabel De la Torre Díez. Proposing new blockchain challenges in ehealth. *Journal of Medical Systems*, 43, 01 2019.
- [3] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the Thirteenth EuroSys Conference*, page 30. ACM, 2018.
- [4] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.
- [5] Juan Benet. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561*, 2014.
- [6] C. Brook. What is the LGPD? Brazil’s version of the GDPR, 2018.
- [7] Dinh C. Nguyen, Pubudu Pathirana, Ming Ding, and Aruna Seneviratne. Blockchain for secure ehrs sharing of mobile cloud based e-health systems. *IEEE Access*, PP:1–1, 05 2019.
- [8] GDPR. General data protection regulation, 2016.
- [9] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–5, Oct 2017.
- [10] Sam Lucero. Iot platforms: enabling the internet of things, 2016.
- [11] J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs. Disruptive technologies: Advances that will transform life, business, and the global economy, 2013.
- [12] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and Imrich Chlamtac. Internet of things: Vision, applications and research challenges. *Ad hoc networks*, 10(7):1497–1516, 09 2012.
- [13] Vidhya Ramani, Tanesh Kumar, An Braeken, Madhusanka Liyanage, and Mika Ylianttila. Secure and efficient data accessibility in blockchain based healthcare systems. In *IEEE Global Communications Conference (GLOBECOM) 2018, At Abu Dhabi, UAE, 12 2018*.
- [14] Qi Xia, Emmanuel Sifah, Kwame Omono Asamoah, Jianbin Gao, Xiaojiang Du, and Mohsen Guizani. Medshare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access*, PP:1–1, 07 2017.

# Kidney Failure Detection Using Machine Learning Techniques

Kílvia L. de A. Almeida<sup>1</sup>, Lucília Lessa<sup>1</sup>, Anny B. S. Peixoto<sup>1</sup>, Rafael L. Gomes<sup>1</sup>, and Joaquim Celestino Jr<sup>1</sup>

State University of Ceará (UECE), Fortaleza, Brazil  
[kilvia.leticia, rafaellgom, celestino]@larces.uece.br  
lucilia.lessa@uece.br  
beatriz.peixoto@aluno.uece.br

## Abstract

Renal insufficiency is the loss of kidney function, which is responsible for the filtration of residues, salts and liquids present in the blood. Being considered a silent disease, as the symptoms are often only detected in advanced stages of the disease. Loss of kidney function can be measured by the Glomerular Filtration Rate, which represents the volume of fluid that is filtered into the Bowman capsule, located in the glomerulus, per unit of time. It is an important indicator for detection, evaluation and treatment of kidney failure. According to a census conducted in 2018 by the Brazilian Society of Nephrology (SBN), the number of chronic dialysis patients in Brazil from 2002 to 2017 increased by 159.4%, while Acute Kidney failure has been showing a mortality rate around 50%. Early detection is a goal to be pursued by those dealing with public health. In this work, a system was developed to detect kidney failure early and thus increase the chances of treatment for these patients. For this, several Machine Learning techniques were used, where through these techniques were calculated their respective accuracy. The data were using the MIMIC-II database and it was shown that techniques such as decision trees and random forests provided good results and could be important life saving strategies.

## 1 Introduction

Kidney failure is the loss of kidney function, which has a range of functions, such as: filtering salts, impurities and blood fluids; regulate the level of water in the body and the level of potassium, sodium, phosphorus and calcium in the blood; and eliminate medicines and toxins from the body and release hormones in the blood. Its functioning can be measured by calculating the Glomerular Filtration Rate(GFR), which uses empirical mathematical equations based on serum creatinine dosage. GFR is important in the clearance of a substance that is freely filtered by the glomeruli and does not undergo tubular resorption or secretion [5], so it is commonly used as the standard measure of renal function and is an important indicator for detection, evaluation and treatment of both chronic kidney failure(CKD) and acute kidney failure(AKF).

In its normal state, the kidney filters the blood and eliminates the end products of metabolism, preserving other solutes. However, in most renal diseases GFR decreases over time as a result of a decrease in the total number of nephrons or a decrease in GFR per nephron, which may be reduced even before symptoms appear, and is related to disease severity level [8].

Renal insufficiency is a silent disease where symptoms are only detected in advanced stages of the disease through various tests, such as blood tests, and can be classified as CKD or AKF. In the case of CKD there is a gradual loss of kidney function, resulting in an irreversible stage of the disease. While AKF is the acute reduction of renal function due to injury, which may

---

Work supported by FUNCAP, CAPES and CNPq.

occur in hours or days. However, there may be a possibility of reversal of AKF if diagnosed and treated early. According to a census conducted in 2018 by the Brazilian Society of Nephrology (SBN) [12] the number of chronic dialysis patients in Brazil from 2002 to 2017 has increased by 159.4%, while AKF has a higher rate of mortality around 50% [13].

This work proposes a system aiming to identify patients with kidney disease using machine learning techniques. To obtain the desired results, we used the MIMIC-II database, which has data from patients who went through the intensive care unit of Beth Israel Deaconess medical center between 2001 and 2012 [7]. In order to compare the results obtained will be taken into consideration three existing studies that aim to diagnose patients with chronic kidney failure. This paper is organized as follows. Section 2 presents the background related renal insufficiency and machine learning techniques. Section 3 describes some related works, while Section 4 details the proposed system. Section 5 shows the experimental results and discussion. Finally, Section 6 concludes the paper and presents some direction for future works.

## 2 Theoretical Foundation

This section presents a brief study on Kidney Failure and some Machine Learning techniques: Decision Trees (DT), Random Forest (RF), and Support Vector Machine (SVM).

### 2.1 Kidney Failure

Creatinine is used in the calculation of GFR because its filtration is performed by the glomerulus and it is not reabsorbed, secreted or metabolized by nephrons, the basic functional unit of the kidney responsible for producing urine. Thus, the amount of creatinine excreted through the urine is directly related to the amount of creatinine filtered by the glomerulus, but it should be noted that creatinine is not completely filtered by the kidneys, only 15% to 20% of it, as well as the others filtrate substances [5].

Measurement of creatinine clearance, the result of all the processes by which a substance suffers in nephrons, acts as a marker of GFR because it is a freely filtered and unabsorbed compound. Therefore, practically all creatinine content in the final urine came from glomerular filtration [5] and is commonly used as the standard measure of renal function, an important indicator for detecting kidney failure.

Kidney failure, considered a silent disease, is due to decreased renal function, which may be reversed or progressively decline. It can be classified as CKD or AKF. In CKD there is a slow, progressive and irreversible loss of kidney function, so the diagnosis can be made by the relationship between GFR and the level of renal function. The importance of GFR in CKD goes beyond detecting it, and is also one of the parameters to determine the need for indication of renal replacement therapy (RRT) which consists of intervening either through dialysis, kidney transplantation, among others, to offer a quality patient's life and try to prevent the patient from dying. It should be borne in mind that the sooner a patient begins medical follow-up, before starting RRT, the less likely they are to die. Patients referred 3 months earlier had 13% deaths, while those referred less than a month earlier had 29% deaths [3]. In addition, this early referral reduces the therapeutic cost per patient and decreases the hospitalization rate per year.

CKD can cause various problems in the body such as: anemia due to decreased red blood cell production; increased uric acid, causing what is called gout; more acidic blood due to decreased excretion of body acids, among others. CKD may be directly related to other diseases such as diabetes and hypertension, as these diseases cause damage to the small blood vessels of the



kidney. It is extremely important to detect CKD in its early stages, as the disease may progress to a stage of terminal kidney failure and if the patient does not receive the necessary treatment, he or she may die within a few months. However, even when undergoing treatment, their life expectancy decreases compared to people of the same age who do not have terminal kidney failure, besides the fact that 40% of these patients are diabetic.

AKF is the acute reduction of renal function due to injury, which may occur within hours or days. The most common causes of AKF occur when the kidneys are deprived of oxygen and may be due to surgery or side effects of a drug. In addition to oxygen deprivation, some physical problems such as kidney stones and autoimmune diseases can lead to AKF. Treatment for AKF is more varied depending on what is causing kidney damage.

## 2.2 Machine Learning

Machine learning consists of a set of methods that can automatically identify patterns in data and use them to predict future data or perform other types of decision making in a scenario of uncertainty. Machine learning has wide applicability in health sciences field, and these techniques have become successful when applied in this area. In this paper we use Decision Tree, Random Forest and Support Vector Machine.

Decision trees (DT) consist of a tree-like structure in which each node represents a check of a characteristic, with each branch representing one of the possible outcomes of the check, and each leaf node representing a classification. This technique is based on statistical models that use supervised training for data classification and prediction, where they apply the split-to-conquer strategy. To construct this model, entropy or gini can be used. Entropy is the calculation of information gain based on a measure used in information theory. It is used to measure the degree of purity or impurity of a data set. In order to identify whether entropy is high, i.e. with a high level of disorder, one must analyze whether the entropy calculation result is closer to 0 or 1, where 1 means a highly impure set and 0 a group fully pure. The gini index, on the other hand, consists of a statistical measure of distribution, being used to measure the degree of heterogeneity of the data so that when the index equals 0 we have a pure node.

Random Forest (RF) is a supervised machine learning technique, which consists in randomly creating a combination of decision trees using a method called bagging most often. Sorting is done by voting, where all trees that make up the forest give a result and the class that gets the most votes will be used as a result for that entry.

The support vector machine (SVM) is yet another supervised machine learning technique based on the Statistical Learning Theory, which allows to classify a particular set of data points, a kernel function, by mapping them to a multidimensional feature space. SVM is widely used in the field of bioinformatics due to its ability to provide high precision, handle large data and its flexibility with respect to modeling various data types. Some of the key features of SVMs that make it attractive to use include good generalization its ability is defined as the measure of efficiency in classifying data that does not belong to the set used in your training; large robustness the Overfitting is not present in SVMs, making them robust when confronted with large objects such as images; convexity of the objective function convexity is present, since the quadratic function, objective function to be optimized, has only one global optimal; theoretical basis is well established within Mathematics and Statistics.

### 3 Related Work

The section discusses some works considered relevant for this proposition and that use the same techniques used in our work for detection of kidney failure.

Levey et al. [9] has developed a more accurate formula for determining GFR in adults. This formula is called CKD-EPI, Formula 1, where  $Scr$  means creatinine,  $k$  and  $\alpha$  are variables that have different values according to gender: for women it is 0.7 and -0.329 and for men 0.9 and -0.411, respectively. The relationship between estimated GFR and gender occurs through creatinine level. The CKD-EPI equation has predefined values according to gender and ethnicity.

$$CKD = 141 * \min(Scr/k, 1)^{\alpha} * \max(Scr/k, 1)^{-1,209} * 0.993^{Age} * 1,018[if\ female] * 1,159[if\ black] \quad (1)$$

Another equation named MDRD, represented by Formula 2, has a fixed ratio of 0.74 between GFR and gender, regardless of the level of patient's creatinine. Thus, this difference can be observed through the formulas 1 and 2, where in MDRD if the patient is female, the equation is multiplied by 0.742. In CKD-EPI, although there is a gender-dependent fixed value, there is a relationship between creatinine and the patient's gender.

$$MDRD = 175 * Scr^{-1,154} * Age^{-0.203} * 0.742[if\ female] * 1.212[if\ black] \quad (2)$$

Levey et al. demonstrated that even when the formulas showed good accuracy, CKD-EPI was able to achieve higher accuracy than MDRD. This is due to some restrictions existing in the MDRD formula, such as reduced accuracy with increased GFR and in different ethnicities. Therefore, in this work, we chose to use the CKD-EPI.

Ahmad et al. [1] applied the SVM technique to detect chronic kidney failure. The work was divided into 5 stages, namely: data collection, data preparation, data grouping and classification. Data collection was performed from the UCI Machine Learning database which was created by Dr. P. Soundarapandian. M.D. This database has only 400 patients and 25 available parameters. In the data preparation phase among the parameters available in the database, only five were selected to be used in the classification, being chosen through statistical methods and interviews. These parameters were: blood pressure, creatinine, compressed cell volume, hypertension factor and anemia factor, but Ahmad et al. did not use some extremely important biomarkers to detect kidney problems that were present in the database, such as urea. In the data grouping phase, the data selected in the previous step were divided into two groups, a test group with 30% of the total data and a training group with 70% of the total data. Finally, in the classification phase was used the language R which has a package with the implementation already ready for SVM. Before the data was passed to the SVM construct it was necessary to convert it to a data frame that could be read by the method.

Al-Hyari et al. [2] uses Decision Tree to diagnose Chronic Kidney Failure. In this approach the data were collected in collaboration with Prince Hamza Hospital in Amman. The database has a record 102 patients. Fifteen parameters were used here: age, weight, gender, blood pressure, hemoglobin, glucose, urea, creatinine, sodium, calcium, potassium, total protein serum, albumin and phosphorus. From these parameters the training and classification by the DT is performed. If the DT classifies that the patient has CKD, the GFR is calculated to inform at what stage of the disease the patient is. For this calculation, Al-Hyari et al. chose to use the Cockcroft-Gault formula. However studies have already shown that the accuracy of this formula is lower than the MDRD and CKD-EPI formula.

Salekin et al.[11] uses Random Forest to detect CKD using the same database as Ahmad et al.[1]. The authors made a selection of attributes considering the most relevant to the model. The formulas Cockcroft-Gault, 1 and 2 are used in order to compare with their results. In our work we use the CKD-EPI formula as an attribute to aid detection.

## 4 Proposal

This work proposes a system capable of detecting kidney failure through the analysis of six biomarkers directly related to renal function, together with GFR, body mass, age, sex and ethnicity. It uses various supervised machine learning techniques, such as: decision tree(DT), support vector machine(SVM) and random forest(RF). These techniques were chosen due to their wide applicability and the capacity to provide high precision in bioinformatics [10][4].

In order to select the parameters, a study was conducted through biomarkers that are directly related to renal function and other physical characteristics that could also aid in the diagnosis. From this study, some attributes were selected with the help of a renal specialist, thus choosing 7 attributes that are most affected by the loss of renal functions and 4 related to physical characteristics that may affect the determination of the disease, so in total, we selected 11 attributes. This attributes are GFR consists of the rate of glomerular filtration, i.e the rate of volume of blood filtered by the kidneys. This rate is extremely important in determining whether the kidney is functioning properly, and is extremely simple to calculate. There are several equations that perform the calculation of GFR, however, for this work will be used the CKD-EPI; Age in addition to being directly related to the calculation of GFR, when the human body is aging some substances present in the body tend to decrease; Sex and Ethnicity are used exclusively in the calculation of GFR due to the difference between the amount of components, such as creatinine, in a man's body in relation to a woman's body, just as this difference is also shown in relation to ethnicities; Body Mass people with a high BMI need a higher blood flow to irrigate the whole body, this creates an overload for the kidneys that need to work harder to filter all the blood; Creatinine when the kidneys lose their ability to filter blood, this substance increases in value and is easily clinically evaluated by blood testing; pH Blood and Urea when the kidneys lose their ability to filter blood, urea values rise and the blood tends to become more acidic; Phosphorus when the kidney begins to slow down, it tries to adapt by maintaining phosphorus levels. However, as insufficiency progresses, this adaptation no longer has an effect and phosphorus levels increase and the kidneys can no longer produce the required amount of vitamin D; Potassium the kidneys are responsible for eliminating about 90% of the potassium present in the body, when the kidney does not function properly the potassium rate rises. A fall or rise in potassium levels in the body can lead to cardiac arrhythmia and sudden death; Glucose when the human body goes through a long period of time with the high glucose rate, this can lead to kidney damage known as diabetic nephropathy. Thus, the measurement of blood glucose is associated with the relationship between kidney failure and diabetes, a disease that affects 40% of patients with terminal kidney failure.

The purpose of the algorithm is to assist health professionals in the early diagnosis of chronic or acute kidney failure, as this is a disease in which symptoms manifest late. Thus, a binary classification of has KF or not. Differently from the existing studies, we sought to classify KF regardless of type, i.e, AKF will also be taken into account as both need early diagnosis.

The first phase of the algorithm consists of grouping data from each patient's urea, creatinine, glucose, blood pH, phosphorus and potassium tests according to the time of collection along with age and body mass. For each set the GFR is then calculated from the creatinine, gender, age and ethnicity values. Sex and ethnicity are used exclusively to obtain the value of

Accuracy					
DT	RF	Linear	RBF	Sigmoid	Polynomial
87%	80%	71%	79%	71%	77%

Table 1: Cross Validation Accuracy

glomerular filtration. The use of GFR is directly linked to renal function. Decreasing this rate indicates some type of insufficiency and together with the selected biomarkers, it is possible to more accurately detect if the patient has kidney failure.

After cluster formation and GFR calculation, this data set will serve as input to machine learning algorithms. This data set consists of records of several healthy and kidney failure patients that will be divided into two groups, one for testing and one for training. After training phase each machine is saved to a file so that sorting can be performed from them.

Through the generated files begins the classification phase, which consists of loading the information of the machines to perform the classification of the desired set. To classify a patient, the machine must receive the 8 attributes, that is the algorithm can receive just the 8 parameters, with GFR already calculated, or all the parameters without the GFR.

## 5 Experiments and Results

In order to validate the algorithm, we decided to use the MIMIC-II database updating the patients diagnosis in renal insufficiency (chronic or acute kidney failure) and healthy.

The following strategies were adopted for training: in the case of DT classification, the gini criterion was used and the best split was always chosen; with the RF classifier, the gini criterion was also used and a forest with ten decision trees was stipulated; Finally, the SVM used linear, polynomial, sigmoid and rbf functions.

For the algorithm evaluation phase, the cross validation technique, the confusion matrix and the classification report were used. Cross-validation uses K-Fold, which is a technique for evaluating machine learning models. The confusion matrix helps in the evaluation of the algorithm, indicating if it has a good ability to predict the desired class, which in the case of this work consists of KF, using 4 metrics: True Positive (TP) when the class to be predicted was correctly classified, in this case kidney failure; False Positive (FP) when the class to be predicted is incorrectly classified, the patient is classified as healthy; True Negative (TN) when the non-predictive class has been correctly classified, in this case a healthy patient; and False Negative (FN) when the non-predictive class has been misclassified, the healthy patient is classified as kidney failure.

Finally the rating report which has 4 metrics also important like Precision provides the percentage of hits when classifying the class to be predicted; Recall number of examples correctly classified as belonging to a class divided by the total samples belonging to that class, regardless of the given classification. When using binary classification the recall of the desired class is also known as sensitivity and for unwanted class specificity; F1-Score fundamental when you have a disproportionate data set, which combines recall with accuracy to indicate overall model quality; and Accuracy indicates how well the machine is able to correctly classify both the desired class and the unwanted class.

For this work a K-Fold with 5 groups was used and the results obtained for this technique and the others approached are presented in Table 1, 2 and 3 .

From the results obtained with all these techniques, can be observe that the algorithm has

	RF		DT		SVM Linear	
	Normal	Kidney failure	Normal	Kidney failure	Normal	Kidney failure
Normal	356	11	362	5	314	53
Kidney failure	133	226	91	268	155	204
	SVM Sig		SVM RBF		SVM Poly	
	Normal	Kidney failure	Normal	Kidney failure	Normal	Kidney failure
Normal	284	83	364	21	341	26
Kidney failure	124	235	128	231	138	221

Table 2: Confusion Matrices

	RF		DT		SVM Linear	
	Normal	Kidney failure	Normal	Kidney failure	Normal	Kidney failure
Precision	0.73	0.95	0.80	0.98	0.67	0.79
Recall	0.97	0.63	0.99	0.75	0.86	0.57
F1-Score	0.83	0.76	0.88	0.85	0.75	0.66
Accuracy	0.80		0.87		0.71	
	SVM Sig		SVM RBF		SVM Poly	
	Normal	Kidney failure	Normal	Kidney failure	Normal	Kidney failure
Precision	0.70	0.74	0.73	0.92	0.71	0.89
Recall	0.77	0.65	0.94	0.64	0.93	0.62
F1-Score	0.73	0.69	0.82	0.76	0.81	0.73
Accuracy	0.71		0.79		0.77	

Table 3: Classification Report

a good ability to classify both renal and normal patients. Due to the sensitivity and specificity provided by the recall, Table 3, we have that the algorithm is easier to predict healthy patients, this happens due to the difficulty in detecting the AKF. When the insufficiency affects only one kidney, the other tends to compensate, thereby maintaining laboratory measurements at normal levels. In addition, when creatinine levels are elevated, it becomes more difficult to assess whether the patient has AKF or CKD, mainly because this work does not include urinary tract obstruction and non-steroidal anti-inflammatory drugs that are directly related to the AKF [6]. This difficulty in identification may have been one of the factors that led the authors [11] [2] [1] to disregard the detection of AKF.

According the results, the six supervised learning machines used were able to achieve the table 3. DT stands out because it has the lowest values of PF and FN.

## 6 Conclusion and Future Works

Through the proposed algorithm it was possible to identify the difficulty to diagnose kidney failure, especially when taking into account acute kidney failure. It is necessary to develop more works and studies in this area in order to improve the identification of AKF, in this way assisting the doctors to identify this illness early and providing to the patient the possibility of recovering the renal function.

As future work, we pretend to use neural networks and deep learning as a classifiers and evaluating their performance; turn the classification into a multi-class classification, ie, the

patient is classified as renal insufficiency acute or chronic, and healthy; inform the patient the stage of their disease; and associate CKD with diseases other than diabetes, like sepsis.

## References

- [1] Mubarak Ahmad, Vitri Tundjungsari, and Dini Widiyanti et al. Diagnostic decision support system of chronic kidney disease using support vector machine. In *2017 Second International Conference on Informatics and Computing (ICIC)*, pages 1–4. IEEE, 2017. <https://ieeexplore.ieee.org/abstract/document/8280576/>.
- [2] Abeer Y Al-Hyari, Ahmad M Al-Tae, and Majid A Al-Tae. Clinical decision support system for diagnosis and management of chronic renal failure. In *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pages 1–6. IEEE, 2013. <https://ieeexplore.ieee.org/abstract/document/6716440>.
- [3] Pasqual Barretti. Indicações, escolha do método e preparo do paciente para a terapia renal substitutiva (trs), na doença renal crônica (drc). *J Bras Nefrol*, 26(3, supl. 1):47–49, 2004. <http://bjn.org.br/export-pdf/1196/v26n3s1a15.pdf>.
- [4] Dongsheng Che, Qi Liu, and Khaled Rasheed et al. Decision tree and ensemble learning algorithms with their applications in bioinformatics. In *Software tools and algorithms for biological systems*, pages 191–199. Springer, 2011. [https://link.springer.com/chapter/10.1007/978-1-4419-7046-6\\_19](https://link.springer.com/chapter/10.1007/978-1-4419-7046-6_19).
- [5] Tereza Neuma de Souza Brito, Arthur Renan de Araújo Oliveira, and Adrielly Karingy Chaves da Silva. Taxa de filtração glomerular estimada em adultos: características e limitações das equações utilizadas. *RBAC*, 48(1):7–12, 2016. [http://www.rbac.org.br/wp-content/uploads/2016/05/ARTIGO-1\\_VOL-48\\_1\\_-2016-ref-370.pdf](http://www.rbac.org.br/wp-content/uploads/2016/05/ARTIGO-1_VOL-48_1_-2016-ref-370.pdf).
- [6] Eunjeong Kang, Minsu Park, and Peong Gang Park et al. Acute kidney injury predicts all-cause mortality in patients with cancer. *Cancer medicine*, 2019. <https://onlinelibrary.wiley.com/doi/full/10.1002/cam4.2140>.
- [7] Joon Lee, Daniel J Scott, and Mauricio Villarroel et al. Open-access mimic-ii database for intensive care research. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 8315–8318. IEEE, 2011. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6339457/>.
- [8] Andrew S Levey, Josef Coresh, and Kline Bolton et al. K/doi clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *American Journal of Kidney Diseases*, 39(2 SUPPL. 1), 2002. <https://experts.umn.edu/en/publications/kdoi-clinical-practice-guidelines-for-chronic-kidney-disease-eva>.
- [9] Andrew S Levey, Lesley A Stevens, and Christopher H Schmid et al. A new equation to estimate glomerular filtration rate. *Annals of internal medicine*, 150(9):604–612, 2009. [http://annals.org/acp/content\\_public/journal/aim/20181/0000605-200905050-00006.pdf](http://annals.org/acp/content_public/journal/aim/20181/0000605-200905050-00006.pdf).
- [10] Ana Carolina Lorena, André Carlos Ponce de Leon Carvalho, et al. Introdução às máquinas de vetores suporte (support vector machines). 2003. [https://www.seer.ufrgs.br/rita/article/view/rita\\_v14\\_n2\\_p43-67/3543](https://www.seer.ufrgs.br/rita/article/view/rita_v14_n2_p43-67/3543).
- [11] Asif Salekin and John Stankovic. Detection of chronic kidney disease and selecting important predictive attributes. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 262–270. IEEE, 2016. <https://ieeexplore.ieee.org/abstract/document/7776352/>.
- [12] Fernando Saldanha Thomé, Ricardo Cintra Sesso, and Antonio Alberto Lopes et al. Brazilian chronic dialysis survey 2017. *Brazilian Journal of Nephrology*, (AHEAD), 2019. [http://www.scielo.br/scielo.php?pid=S0101-28002019005013101&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0101-28002019005013101&script=sci_arttext).
- [13] L Yu, BF dos Santos, and EA Burdmann et al. Diretrizes da amb e sociedade brasileira de nefrologia para insuficiência renal aguda [internet]. *São Paulo: Sociedade Brasileira de Nefrologia (SBN)*, 2007. <http://www.bjn.org.br/details/1203/pt-BR/insuficiencia-renal-aguda>.

# Educational robotics at K-12 schools in the southeast of Mexico.

Cinthia Gonzalez-Segura<sup>1</sup>, Michel Garcia-Garcia<sup>1</sup>, Jorge Rios-Martinez<sup>1\*</sup>,  
Sergio Gonzalez-Segura<sup>1</sup> and Luis Basto-Diaz<sup>1</sup>

<sup>1</sup>Universidad Autonoma de Yucatan  
{gsegura, michel.garcia, jorge.rios, sgsegura,  
luis.basto}@correo.uady.mx

## Abstract

This paper presents a descriptive study about the use of robots in the teaching of mathematics for elementary children and the use of different models of educational robotic kits which were programmed to address issues such as the Cartesian plane, fractions, polygons and areas and arithmetic operations. During four semesters, fourteen elementary schools located in the eastern part of the State of Yucatan, were visited to perform activities with robots. More than two thousand children attending the six grades of elementary education and their respective teachers have participated. It was observed that the use of robots as new pedagogical tool greatly contributes to motivate students during the teaching-learning process of mathematic subjects.

## 1 Introduction

The advances in science and technology continuously re-shape our daily activities. However, a large part of society remains detached from these changes, thereby excluding individuals in the use of tools that would allow them a full development in today's technological world. Robotics is an area of great potential for the development of modern society. Robots have been used for various purposes, in education, for example, their use promises to encourage the students' creative development in scenarios where robots are applied as tools, tutors, or even pairs (Alimisis, 2013), (Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013). Several countries, conscious of educational robotics potential, have started strategies supported by robotics to enhance education in basic levels. For example, in South Korea a program was developed to produce robotic assistants that support the teaching of English in preschool (Yun et. al., 2011). Other authors discuss the role of new technologies to achieve sustainable development in developing communities (Dias, Mills-Tetty, & Nanayakkara, 2005). The interest that robotics arouse in both educators and students should be exploited to make it a tool to support the quality

---

\* Corresponding author

and innovation in classrooms. This paper describes the project “A day of science and technology in your school”, which consists in visiting elementary schools with a group of university students and professors, exposing ways to use technology for educational purposes to participants. The region where the project was performed belongs to a low-income region in Mexico which has minimum technological infrastructure, most part of the children reached by the project had not seen a physical robot in their life. The following section depicts the theoretical framework on which the performed activities are based. Next, the planning carried out previously to the events held in schools is described. Then the experience and information obtained during the events are presented, and finally, the results, conclusions and future work are discussed.

## 2 Educational Robotics

Many efforts have been made to improve the quality of education, one of them is to include technological innovation in education, specifically robotics. According to (Stager, 2010) it is possible to include robotics in education as an independent subject; as an aid in the teaching of concepts in science, technology, engineering and mathematics; as thematic units to model machines and systems; as a specific medium to solve a problem related to a formal topic of the Curriculum and finally, as a material that learners can use to express their ideas freely. Critical factors for the success of educational activities with robotics are the following: a good suggestion on how to start the activity, appropriate materials, enough time, and a non-coercive, collaborative and non-competitive atmosphere. Last suggestion seems to contrast with the many robotic contests which seek to spread the use of robots by rewarding the better performance on programming robots for specific tasks. In a classroom, the rewards can result in increased stress and posterior frustration. Educational robots are a subset of educational technology employed to facilitate learning and improve the educational performance of students. Robots add social interaction and perform roles not only as a tool but also as a guardian or as a pair (Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013). Robots should not be considered as a replacement of teacher but as assistants letting the teacher to focus on high level cognitive tasks of teaching process. It has been reported that the main motivation to include robotics in education comes from the positive role of robots in educational activities, the development of creative thinking and the improvement on problem-solving skills, however, the lack of adequate teacher training is an obstacle (Karim, Lemaignan, & Mondada, 2015). Closer to the objectives of the present article is the work by (Pinto Salamanca, Barrera Lombana, & Pérez Holguín, 2010), involving preschools and elementary schools. That project employed robots to assist in the areas of math, science, read-write and computer programming, taking advantage of the natural motivation of children towards robots. There were activities such as the layout of plane geometric figures using robots. The teachers from the involved schools expressed a lack of knowledge in recent technological didactic tools and at the same time showed interest in knowing them. Finally, this project concluded that education can be improved if solid foundations for change in teaching methods are proposed and new models that can contribute to changing modern society are created. In Colombia, (Jimenez Jojoa, Bravo, & Bacca Cortes, 2010) developed a course for children and teenagers based on mobile robotics with the aim of arousing the scientific interest and practicing the problem-based learning. At the end, the participants expressed their interest in studying a career related to engineering. Kathia Pitti et al described a project (Pitti Patiño, y otros, 2012) from the Republic of Panama, involving students and teachers from six middle schools (grades 7, 8 and 9; ages 13-15) in which Robotics was employed as a tool to support the teaching-learning process, focusing on subjects such as mathematics, physics and computer science. During the project there were activities with mazes, missions, competitions with challenges, among others. The authors assert that robotics facilitates and encourages the teaching and learning of sciences and technologies becoming a useful tool to understand complex and abstract concepts. The authors observed



that students from schools located in rural areas got a better performance than those located in urban areas. Also, as a tool to implement constructivism and constructionism in the project TERECOP robotics was aimed at the education of future teachers.

Creativity and logical thinking, keys in STEM practice, were approached by TIC TAC project (Muntaner-Perich, y otros, 2013) which was carried out with children of 5th grade of elementary education, in some of the most disadvantaged areas of India. Workshops and activities were designed in which technology was applied in areas such as math, physics and computer science. Teachers reinforced the importance of teamwork, learning through practice and the trial-and-error method. Besides serving as support in the teaching of math and science-related topics, a robot can also be used to learn a second language. In South Korea, authors of (Yun, y otros, 2011) have experimented with a teleoperated robot that helps in English education at preschool. In Taiwan, (Chang, Lee, Chao, Wang, & Chen, 2010) studied the role of robots as instructional tools in the teaching of a second language in elementary level students and concluded that the use of robots improves the interactivity and motivation of both students and pupils. According to (Barranco Candanedo, 2012) some challenges must be faced. The first of these is the fear of teachers to use educational technology new for them, especially in those that are found in regions away from major cities, who were taught in traditional way and prefer to teach in turn in traditional way. Another major challenge is the lack of infrastructure and robotic equipment as result of elevated costs. In the other side, open source philosophy and the decrease in the price of electronics can help to alleviate the latter. Finally, there is the challenge to generate own robots designed and built in the region which could be adapted more easily to the specific educational needs. In Mexico, there are private schools that include some type of robotics courses in their curricula and promote the participation of students at international competitions. The amendments made to the Law of Science and Technology in Mexico proposed to promote scientific and technological vocations since the first educational cycles and in general they put education as a pillar of sustainable development. In recent years the government has sought to bring science and technology close to children and young people with the incorporation of activities related to the National Week of Science and Technology. In the State of Yucatan, public elementary schools have begun to participate and they do so by putting their best effort, however, the impact is limited since there is no sufficient scientific and technological materials, either because schools are unaware of or do not have the resources to acquire them. The apathy towards mathematics (Míguez, 2004), exists among students from long ago because math and science are seen as boring and difficult to learn which is reflected in the results of national assessments such as ENLACE (Secretaría de Educación Pública, 2013). Most students do not recognize science as a tool to understand their world and the consequences are serious because it means that professional expectances are limited and the number of potential scientists, mathematicians and engineers (STEM) is lower. It lacks both fostering the use of technology or appropriate digital content and studies to assess the impact of technological devices in Mexican schools.

### 3 Project “A day of science and technology in your school”

The project aimed to promote and encourage the interest in science and technology among the students of elementary education in the East of the State of Yucatán, Mexico. It was based on the spreading of academic topics related to the educational plans for elementary schools, and the installation of six stands denominated: Electronics, Robotics, Animations, Videos, Mathematical Challenges and Origami. In charge of the stands were professors and undergraduate students of computer science.


The overall objective of the stand was to show a set of programmed robots with routines for the explanation of the Cartesian plane, the addition and subtraction of fractions, areas and perimeters, among others. It was used the interactive exhibition, a didactic technique focused on the student, consisting of the oral presentation of a topic to achieve objectives related to the learning of theoretical

knowledge or information of various types. The exhibitions were made by university students and their professors, who previously designed a slide presentation. Interactive presentations consisted of a set of activities for each school grade. In all cases, the session begins with the interactive exhibition in which basic concepts about motors, sensors, gears and axes are explained, among others. Then, the academic subject is presented by means of examples and finally the evaluation activities are carried out. The academic topics, activities and materials are shown in Table 1.

Grade	Topic	Activity	Material
1-6	Introduction	Each robot is presented and explained.	Lego NXT Robots with different designs and routines.
1	Counting series (2 by 2 and 3 by 3)	The exhibitor asks children to count with different increments and at every step the robot advances controlled by a child.	Tribot NXT programmed to be controlled remotely via Bluetooth.
2	Number line	Children control the robots making them to move on the number line, adding and subtracting steps.	Robot with wheels (tank) programmed to be controlled remotely via Bluetooth.
3	Geometric shapes	Children put the robots at the asked coordinates making a triangle, then a rectangle. They were requested to find areas of both figures.	Cartesian plane drawn on the floor and four different robots, at least one of them remotely controlled. One robot is required for each point indicated with coordinates.
4	Decimal numbering system	Children are asked to guide the robot towards the sheet printed with the number pronounced. The child must control the robot to guide it towards the right answer. Some questions are done, such as how many units, tens, etc. are in the printed quantities.	Several robots, at least one remotely controlled. Four numbers of 6 or more digits are put on the floor, in a different order.
5	Fractions	Children are chosen to guide and locate the robot at the coordinates indicated on the Cartesian plane, with fractional quantities. The children must perform adding and subtraction operations to move the robot to the correct position.	Cartesian plane drawn on the floor and a robot with a Wireless control routine.
6	Coordinates and scales	Children are chosen to locate the robot at the coordinates indicated on the Cartesian plane. An equivalence is established between units and km, and the children are questioned about the distances between pairs of objects.	Cartesian plane drawn on the floor and several robots with Wireless control routines.
1-6	Conclusion	Direct interaction with robots, motors and sensors.	Lego NXT robots with different mechanical designs and routines.

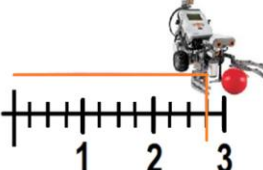
**Table 1. Activities at the Stand of Robotics**

Fig. 1 shows examples of the activities designed for fifth and sixth grades. The topics presented in each grade were selected after reviewing the contents of the textbooks that are handled in the country's public schools. The main difference in the methodology implemented with respect to traditional teaching is the use of robots to move in a three-dimensional space, instead of just working on the notebook or the board. In addition, the active participation of children is sought during all activities, through interaction with robots.




**5° Grade**  
**Question 2**

If the number line represents the path of the Tribot robot, what fraction of kilometers did it advance?



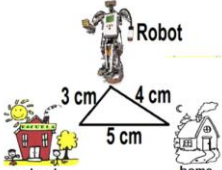
**Answers**

a)  $\frac{3}{4}$  Km  
b)  $\frac{5}{4}$  Km  
c)  $1\frac{1}{4}$  Km



**6° Grade**  
**Question 1**

¿ What is the distance between the robot and the school, if the triangle represents a 1 cm scale map: 2 km?



**Answers**

a) 5 Km  
b) 6 Km  
c) 12 Km

Figure 1. Activities for fifth and sixth grades

Fourteen elementary schools were visited over four semesters. About 100 teachers and 2,300 children participated in total. During the presentations two students from the bachelor's degree in Computer Science were helping, who were also in charge of building and programming the robots, as well as to explain and answer questions from children and their teachers. It was decided to work with a maximum of 6 groups and one classroom per session in order to optimize efforts, see Fig. 2.



Figure 2. An example of a session with children

## 4 Analyzing results

A survey was applied at the beginning and another at the end. These surveys included questions related to the topics presented in the stand of Robotics and were applied the day prior to the visits. A sample of students was selected randomly to answer the initial multiple-choice surveys, guided by their

own teachers. At the end of the event, the children of the initial sample were provided with another survey with similar questions but changing some data and possible responses. The final survey was conducted by the presenters of the event, each stand applied it to the group at the end of their presentation. While the children answered their final surveys, the teachers answered an opinion survey about the event, in which they were also asked to suggest topics that they considered relevant.

Once digitized and analyzed the collected information, it was carried out a quantitative analysis with the surveys applied to children and a qualitative analysis based on the opinions of teachers. The children offered effusive and spontaneous signs of gratitude, asking that the event occurs again in their school.

The sample of students surveyed was 300 students, 10 children per group, randomly selected in each of the 5 schools visited during the first stage of the project.

To analyze whether the results obtained in the initial and final tests regardless of grade differ significantly, the T-Student statistic was used for paired samples. The verification of the assumption of normality to the difference of the paired data was carried out with the Shapiro Wilk and Kolmogorov Smirnov test statistics, using a significance level of 5%. To compare whether there is a difference for each academic degree, Z statistics were used for large samples to compare the percentages of the initial and final tests. Eighty percent of the surveyed teachers rated the event as excellent, nine percent as good and eleven percent did not answer the question. Among the positive reviews that the visited teachers expressed about the stand of Robotics, it was mentioned that it arouses the interest of children to learn more about the technology, particularly about robots, their construction and programming, which will be useful in their future. It was also mentioned that the event allows children to have a close encounter with the technology, which may not have at their disposal daily. They thought it was useful the explanation about how the movements of robots are programmed, especially to let them know that mathematical and logical concepts are required to program the desired routines. They expressed that it was very helpful to provide them with new perspectives to develop fun tech toys without resorting to violence.

Regarding the content, they liked the funny way in which technology is applied to teach math, that the different types and characteristics of the robots can be applied also to teach natural sciences, that the logical reasoning of children is stimulated and that the concept of science is expanded to include also robotics. It was also said that it is beneficial to include physical routines as part of the activities of the stand, when referring to the imitation that children made of the movements of a humanoid robot.

Similarly, some comments reflected certain areas of opportunity observed, such as: insufficient time (30 minutes per stand), the large number of children avoided that everyone could participate, and they went out of control sometimes. It was also reported that some explanations were given too fast.

Finally, in terms of attitude and interest observed in children during the event, eighty percent of teachers reported seeing their own students highly participatory and motivated, sixteen percent saw them interested and four percent did not respond to the question. The indifferent, distracted and very distracted options were not selected.

University collaborators and professors expressed that: it is a good experience to publicize and teach technologies related to their area of study; prior knowledge about the covered topics was of great help for the development of activities and to support the exhibitors; it is very nice to provide useful tools to motivate children and teachers to new technologies; the project is of great interest to the collaborators and even influences their personal goals. Besides, the teachers of the visited schools consider important that the university involves more with the surrounding population through projects such as this, in addition to providing them with new techniques applicable to their classroom.

On the other hand, to analyze the academic impact of the event, to the random sample of participating children, an initial diagnostic survey was applied at the beginning of the event and a final survey after the event. Table 1 shows the results obtained as well as the difference between both surveys for each academic grade in elementary education. The first column corresponds to the academic level, the second and third column shows the percentage of correct answers for all the participants that were

obtained before and after the event, respectively. The last column shows the differences between these two.

LEVEL	INITIAL	FINAL	DIFFERENCE
1°	40%	59%	11%
2°	60%	39%	-21%
3°	59%	53%	-6%
4°	52%	50%	-2%
5°	11%	8%	-3%
6°	58%	30%	-28%
<b>AVERAGE</b>	58%	46%	12%

**Table 2. Academic performance of the participating children**

The values in Table 2 were obtained as follows: after grading the exercises performed by the students of each grade in both instruments, the success ratio was obtained by dividing the number of correct answers by the number of participants of each grade and each school. Subsequently, the averages by school grade in both instruments were obtained and the difference between both values was obtained, by each grade, as well as the final average of both instruments. As can be seen, in the general average field a decrease of 1% was obtained. However, when performing a T-student statistical test for related samples, this decrease proved to be non-significant with a p-value of 0.4062 and a 95% confidence interval. This indicates that, the difference in the average performance of the students in the initial and final tests is not significant. Besides, in all grades it was noted a decrease in the percentage of correct answers of the earlier to the latter, which could be interpreted as a negative effect in the event. However, it is important to mention that even when it was emphasized to the teachers who applied the initial surveys to their own students that the surveys were not to evaluate them but the project, there were many corrected answers in the initial surveys, which did not happen with the final surveys.

## 5 Conclusions and future work

Educational robotics can be a useful tool to teach STEM topics to basic level students as it raises curiosity of children and can be programmed to function in diverse activities and for different education levels. The project "A day of science and technology in your school" got a participation of more than two thousand children attending the six grades of the elementary education in Mexico, and their respective teachers, who number about one hundred. The effort was fruitful because the spread of this new technology between children of basic education had a positive impact on their perception of math and sciences application. We think that a significant learning can be reached because the students can link abstract concepts with concrete objects and activities while they interact with robots. The feedback collected from students and teachers about the robotics session was very encouraging however the quantitative analysis must be improved, first by extending the sessions and its duration and second by focusing on one specific grade and topic. The present project was not designed to train teachers with robotics activities, but we conclude that they must be included on the definition of activities to match the subject they aim to reach with their students according the national plan of education. It was confirmed as many authors declare in the literature that robots are attractive for children to work with

mathematics, but it is necessary to design a complete pedagogical initiative in order to keep its usefulness when initial enthusiasm decreases. The design of activities in the present project was inspired by the ideas of Constructionist learning theory by Sigmund Papert, but a more formal development of hands-on exercises must be deployed in order to test and compare learning results.

## References

- Alimisis, D. (2013). Educational robotics: Open questions and new challenges. *Themes in Science & Technology Education*, 63-71.
- Barranco Candanedo, A. A. (2012). La robótica educativa, un nuevo reto para la educación panameña. *Teoría de la educación, educación y cultura en la sociedad de la información.*, 9-17.
- Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y., & Chen, G.-D. (2010). Exploring the Possibility of Using Humanoid Robots as Instructional Tools for Teaching a Second Language in Primary School. *Educational Technology & Society*, 13-24.
- Dias, M. B., Mills-Tettey, G. A., & Nanayakkara, T. (2005). Robotics, Education, and Sustainable Development. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, (pp. 4248 - 4253). Barcelona, Spain.
- Jimenez Jojoa, E. M., Bravo, E. C., & Bacca Cortes, E. B. (2010). Tool for Experimenting With Concepts of Mobile Robotics as Applied to Children's Education. *Education, IEEE Transactions on*, 88-95.
- Karim, M., Lemaignan, S., & Mondada, F. (2015). A review: Can robots reshape K-12 STEM education? *Proceedings of the 2015 IEEE International Workshop on Advanced Robotics and its Social impacts*. Lyon, France.
- Miguez, M. Á. (2004). El rechazo hacia las matemáticas. Una primera aproximación. *Acta Latinoamericana de Matemática Educativa* (pp. 292-298). México: Comité Latinoamericano de Matemática Educativa.
- Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Technology for education and learning*, 1-7.
- Muntaner-Perich, E., Niell Colom, M., Peracaula Bosch, M., Estebanelli Minguell, M., de la Rosa Esteva, J., & Freixenet Bosch, J. (2013). El Proyecto TIC TAC. Tecnología, Educación y Desarrollo Humano. *VI CONGRESO UNIVERSIDAD Y COOPERACIÓN AL DESARROLLO*. Valencia, España.
- Pinto Salamanca, M. L., Barrera Lombana, N., & Pérez Holguín, W. J. (2010, Julio). Uso de la robótica educativa como herramienta en los procesos de enseñanza. *Ingeniería, investigación y Desarrollo*, 10, 15-23.
- Pittí Patiño, K., Muñoz Arracera, L. E., Moreno, I., Serracín Pittí, J. R., Quintero, J., & Quiel, J. (2012). La robótica educativa, una herramienta para la enseñanza-aprendizaje de las ciencias y las tecnologías. *Teoría de la Educación: Educación y Cultura en la Sociedad de la Información*, 74-90.
- Secretaría de Educación Pública. (2013). *ENLACE*. Retrieved from <http://www.enlace.sep.gob.mx/ba/>
- Stager, G. S. (2010). A Constructionist Approach to Teaching with Robotics. *Constructionism*. Paris, France.
- Yun, S. a.-T., Yun, S., Shin, J., Kim, D., Kim, C., Kim, M., & Choi, M.-T. (2011). Engkey: Tele-education Robot. In *Social Robotics* (pp. 142-152). Springer Berlin Heidelberg.

# A proposal methodological to Assure Quality using Data Profiling Techniques

César Guerra-García<sup>1</sup>, Héctor G. Pérez González<sup>1</sup>, Reyes Juárez-Ramírez<sup>2</sup>,  
Victor Menéndez-Domínguez<sup>3</sup> and J. Talburt<sup>4</sup>

<sup>1</sup>Autonomous University of San Luis Potosí, México

<sup>2</sup>Autonomous University of Baja California, México

<sup>3</sup>Autonomous University of Yucatán, México

<sup>4</sup>University of Arkansas at Little Rock, U.S.A

{cesar.guerra, hectorgerardo}@uaslp.mx, reyesjua@uabc.edu.mx,  
mdoming@uady.mx, jrtalburt@ualr.edu

## Abstract

For any organization it is necessary to satisfy their business objectives besides using data to implement organizational processes, for that reason, it is indispensable to have knowledge of how these data satisfy the preset quality requirements. These requirements could be expressed through data quality (DQ) dimensions. In some scenarios, models and methodologies of DQ assessment require of mechanisms to control the level of quality of data. Thus, proposing a methodology with a qualitative diagnosis of DQ dimensions and using data profiling techniques to measure some of these dimension, will have a major impact on the processes of appropriate use of data.

## 1 Introduction

Nowadays, both public and private organizations understand the value of data. Well managed information has an incalculable value for the organizations. During years, the data management has acquired a growing importance in companies, because data constitute one of the main assets of them, and without data, it is almost impossible that corporations can align with their organizational strategy. Although significant works of research have been done on the notion of DQ, their main focus is on the objective quality attributes of data (Eshraghian and Harwood 2015). According to standard (ISO-25012 2008), the concept of DQ is defined as "*the degree to which the characteristics of the data are suggested conditions and needs when used under specific conditions*". With this definition in mind, before considering any operation or process, it is very important to assess the suitability degree of the use of data involved in a task, according to the context in which the data are. In this sense, the technique of data profiling is one of that could help to diagnose the DQ in specific contexts, which is the "*data analysis systems to understand its content, structure, quality and dependencies*" (Olson 2003). Indeed, data profiling and monitoring the defects of data are useful activities for assessing DQ in specific contexts. Although, nowadays there exists some methodologies and tools to carry out the

data profiling processes (described in section 2.2), in our context, it is possible to find some specific needs such as: assessment of some dimensions of DQ using techniques and tools of data profiling, definition of roles and responsibilities for the DQ control, organization of the process through the use of artifacts, besides of reporting to the organization about the DQ diagnosis, depending on user types to involve roles that interacts with these data. Until now, there are not any proposals, which include these mentioned points. Thus, with the aim of establishing all the previous points in our proposal, the specific process of Project Planning established in the standard ISO 12207:2008 was used as a reference, this process determines “*the scope of project and technical activities, identifies process outputs, project activities, deliverables, establishes schedules for project task conduct, and required methods and techniques to accomplish project activities*”.

## 2 Related Areas

### 2.1 Methodologies for Data Quality Assessment

Different DQ dimensions have been defined by several authors (from different point of view) and even they have been defined in the standard (ISO-25012 2008). This proposal is based on the DQ dimensions introduced in this standard. To make a comparative study of existing methodologies for DQ assessment, we consider some aspects, including dimensions used, cost and types of data and information systems involved. Some of these methodologies are *AIMQ* (Lee, Strong et al. 2002), *CIHI* (Long and Seko 2005), *AMEQ* (Su and Jin 2007) and *IQM* (Eppler and Muenzenmayer 2002). *AIMQ methodology*: it is the only methodology of information quality based on benchmarking. It classifies the DQ dimensions according to interest of users. *CIHI methodology*: it focuses on the control of DQ of the data set stored in the Canadian Institute of Health Information, specifically in the monitoring of the size, heterogeneity and quality of the stored data. The DQ evaluation is based on a four-level hierarchical model. *AMEQ methodology*: it provides a rigorous basis for Product Information Quality assessment and improvement in compliance with organizational goals. The methodology is specific for the evaluation of DQ in manufacturing companies. *IQM methodology*: it conceived the provision of a quality framework adapted to Web data. Among its entries, besides of the quality criteria, it has tools and techniques used to measure the DQ. After studying the characteristics of these audit methodologies, we consider that they are very useful, depending on its features and goals. However, according with some aspects like: the focus on the business processes of organization, definition of roles, artifacts for documenting the process and the inclusion of data profiling techniques for the DQ evaluation; we conclude that, except AMEQ that utilizes the organizational processes in its process modeling, the rest of the methodologies do not consider it. They do not use roles neither include data profiling techniques.

### 2.2 Data Profiling Models, Techniques and Tools

Nowadays, some data profiling methods and techniques (described later on) also contribute to the necessary assessment for the DQ control in any kind of IS, where the fundamental approach is performed on data collections. The DQ dimensions more widely used to assess DQ are correctness, completeness and accuracy. One of the models available is (Olson 2003), which consists of various inputs of data and metadata, and as outputs, corrected metadata and information related with data. Oracle Corporation developed a system for profile data, it is oriented to the thorough investigation and close monitoring of its quality (Corporation 2010). With a tool named Oracle Data Profiling, the user has the possibility to discover and infer rules based on data and monitor their quality. Microsoft offers a tool named Data Quality Services (Corporation 2012), with techniques and mechanism of data profiling, such as: candidates keys profiling, column profiling, data profiling using patterns, etc. The Embarcadero Company is noted for the CA ERwin Data Profiler tool, the user combines the



analysis and data modeling in a practical way. In its own model highlights 4 key activities: analysis column, integration with data models, the discovery of keys and Analysis of attributes (Enterprise 2009). Other tool is Informatica PowerCenter (Corporation 2017), an enterprise platform that offers access, research, data profiling and data integration from any data source.

### 3 Methodology for Analysis and Evaluation of DQ based on Data Profiling Techniques

The methodology guides to the establishment of the DQ control, based on the diagnosis of DQ dimensions and processes related with data profiling of relational databases, where activities and tools are involved, as a guide for its implementation. It to unlike the others audit methodologies is based on business processes, besides it defines roles for a better organization in the execution of its phases and activities. It also proposes well-defined artifacts that can document the implementation of the methodology. The methodology is represented in 3 key phases: *Analysis*, *Evaluation* and *Transition*.

**1. ANALYSIS.** At this phase, the current status of a particular organizational process is studied, this implies to consider the types of users, data types, DBA, etc., for preparing the infrastructure for the application of data profiling techniques and the survey of diagnostic of DQ dimensions. In next iterations new organizational processes should be diagnosed (Table 1). This phase encompasses the next two activities. *1.1 Diagnosis*, it must be executed to get a first assessment of current status of the selected organizational processes. For doing so, it is necessary to consider the databases used by the selected organizational processes, user types and existing roles, the types and formats of data handled by the organization. *1.2 Election of Requirements*, the DQ dimensions selected will be involved throughout the execution cycle of the methodology for each one of selected organizational processes.

Input products	Description of the organizational process.
Output products	Identified information, selected dimensions for the DQ evaluation.
Activities	1.1. Diagnosis.                      1.2 Election of Requirements.
Methods, techniques and tools	Expert judgment, brainstorming, artifact for the Diagnosis of organizational process (for sake of space, it could be seen in the author's web page).
Roles	Business analyst, Data Quality analyst.

**Table 1:** Phase of Analysis.

**2. EVALUATION.** An evaluation of the DQ level of a relational database should be performed. This implies the use of some techniques like structure profiling, relational profiling, data rules profiling and the implementation of surveys for the diagnosis of DQ dimensions (Table 2).

Input products	Result of the diagnosis of organizational process, data source, metadata source.
Output products	Data profiled, metadata profiled, result of survey for diagnostic of DQ dimensions.
Activities	2.1. Structure profiling.           2.2. Relational profiling. 2.3. Data rule profiling.         2.4. Conductions of Survey for diagnostic of DQ dimensions.
Methods, techniques and tools	Profiling of table structures, and its functional dependences, data rules profiling, data profiling tools, questionnaire of the survey for the diagnostic of DQ dimensions (for sake of space, it could be seen in the author's web page).
Roles	Business analyst, DQ analyst, database administrator, database designer.

**Table 2:** Phase of Evaluation.

In this phase, the team should execute these activities: *2.1 Structure profiling*. It consists of investigate each one of the columns and rows of tables in the source systems, applying a set of techniques to calculate statistical information and their metadata. *2.2 Relational profiling*. The main aim of this activity is to determine possible relationships and functional dependencies between tables or business objects, and discovering primary and foreign keys. With this activity is possible to evaluate the degree of consistency. *2.3 Data rules profiling*. Activity aimed to the researching,

discovering, verification and validation of data rules. It helps to specify the degree of conformity, which determines whether the data has attributes that adhere to standards or conventions. *2.4 Conduction of a Survey for the diagnostic of DQ dimensions.* The survey is a system for collecting information to describe, compare and explain knowledge and behavior. In this process a qualitative diagnostic of DQ is done.

**3. TRANSITION.** In this phase the organizational process analyzed is monitored, continuing to the analysis. It should be reported the status of the DQ, to all roles and members involved in the organizational business process. It implements activities related to the process of monitoring and alerting of DQ. Table 3 shows their characteristics. *3.1 Monitoring and control.* The aim is to notify, and alert events related with the detection of a poor DQ in any of the selected business processes. People in charge should ensure the beginning for repeating the phase of analysis in a new process.

Input products	Result of the survey for the diagnostic of DQ dimensions.
Output products	Artifacts, notifications and alerts.
Activities	3.1 Monitoring and control.
Methods, techniques and tools	Notification, and alerts of Data Quality.
Roles	DQ Analyst.

**Table 3:** Phase of Transition.

## 4 Conclusions

This main contribution of this paper is a methodology for the analysis, control and evaluation of DQ, through data profiling techniques and the application of surveys for the diagnostic of DQ dimensions, to different types of users. We empirically obtained the DQ dimensions used in the methodology, the types of users to which the questionnaire should be applied, the roles and responsibilities defined, and the output products of the analysis phase. For the success in execution of the methodology in a real scenario, we think that it is necessary to consider the systemic method as a combined and integrated system of all phases and activities. As future work, we will continue with the analysis of the rest of processes defined in standard 12207:2008, identifying artifacts and processes. Besides, we are going to implement the methodology in a real environment into an organization.

## References

- Corporation, I. (2017). "Informatica Power Center." 2019, from <https://http://www.informatica.com/mx/products/data-integration/powercenter.html>.
- Corporation, M. (2012). "Data Quality Services." from <https://docs.microsoft.com/en-us/sql/data-quality-services/introduction-to-data-quality-services?view=sql-server-2017>.
- Corporation, O. (2010). "Oracle data quality for data integrator and oracle data profiling 11G.", 2019, from <http://www.oracle.com/technetwork/middleware/data-integration/oracledq-datasheet-128404.pdf>.
- Enterprise, S. (2009). "CA Edwin Data Profiler." 2019, from <http://www.dbta.com/DBTA-Downloads/Software/CA-ERwin-Data-Profiler-1581.aspx>.
- Eppler, M. and P. Muenzenmayer (2002). *Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology*. Proceeding of the Seventh International Conference on Information Quality.
- Eshraghian, F. and S. A. Harwood (2015). Information product: How information consumers' perception of 'fitness for use' can be affected. International Conference on Information Quality, ICIQ 2015.
- ISO-25012 (2008). ISO/IEC 25012: Software Engineering-Software product Quality Requirements and Evaluation (SQuaRE)-Data Quality Model.
- Lee, Y. W., D. M. Strong, B. K. Kahn and R. Y. Wang (2002). "AIMQ: a methodology for information quality assessment." *Information & Management* **40**(2): 133-146.
- Long, J. and C. Seko (2005). A cyclic-hierarchical method for database data-quality evaluation and improvement. *Advances in Management Information Systems-Information Quality Monograph (AMISIQ)*.
- Olson, J. E. (2003). *Data Quality: The Accuracy Dimension*. San Francisco, CA, USA, Morgan Kaufmann Publishers.
- Su, Y. and Z. Jin (2007). A Methodology for Information Quality Assessment in the Designing and Manufacturing Process of Mechanical Products. *Information Quality Management: Theory and Applications*. Hershey, PA, USA: 190-220.

# Document Database for Scientific Production

Jared D.T. Guerrero-Sosa, Víctor Hugo Menéndez-Domínguez, María-Enriqueta Castellanos-Bolaños and Francisco Moo-Mena

Universidad Autónoma de Yucatán, Mérida, México.

jared.guerrero,mdoming,enriqueta.c,mmena@correo.uady.mx

## Abstract

The aim of this paper is to present the use of a document database for the storage of information related to scientific production, which can be retrieved through various digital repositories. In the introduction it is briefly presented what digital repositories of scientific production consist of. Then the methodology and its implementation are presented, which are based on the discovery of knowledge through data sets. Finally, the results, conclusions and future work are presented.

**Keywords:** Document database, MongoDB, scientific production, Scopus.

## 1 Introduction

Scientific research has become more accessible thanks to digital repositories, which are responsible for storing, preserving and being a tool for disseminating content (Ip, Morrison & Currie, 2001). There are several repositories of indexed scientific production, which is one that can be located in a database that collects citations, including through periodic evaluations. They use statistical quantitative indicators to measure scientific productivity and with them determine the quality of the publications. Internationally, the main ones are Scopus and Web of Science. There are studies that compare characteristics, strengths and weaknesses of the mentioned repositories and Google Scholar (Falagas, Pitsouni, Malietzis & Pappas, 2007) (Bakkalbasi, Bauer, Glover & Wang, 2006) (Jacso, 2005). However, the first two have greater credibility for institutions to consider the quality and relevance of publications, such is the case of the National Council of Science and Technology (CONACYT) and the National System of Researchers (SNI) in Mexico.

Repositories such as Scopus make their web services available for the recovery of information about authors, institutions and products of scientific relevance (Elsevier, 2019). On the other hand, there are other open access repositories, such as the national repository of Mexico, which allows anyone to harvest the metadata of the stored production through the OAI-PMH protocol (CONACYT, 2017). However, if you want to recover all the production associated with an institution, a geographical area or over a certain area of study, the volume of data is usually very large, and for that NoSQL databases are well-adapted thanks to their ability to store and process large amounts of data effectively (Han, Haihong, Le & Du, 2011). One type of NoSQL database is the document one. This

work aims to present a proposal for the recovery and storage of the production associated with researchers, which is found in Scopus and open access repositories that use the OAI-PMH protocol, by using a document and non-SQL database.

## 2 Methodology and implementation

Document databases store data in structures called documents by using BSON or JSON formats. One of their advantages is the flexibility in the design of the documents, since they do not follow a specific scheme and not all documents must have the same descriptive fields. Taking into account the following aspects: 1) each type of scientific publication does not have the same descriptive data (for example, a journal article has ISSN and a book chapter may have more than one ISSN (physical and electronic)), 2) the production stored in Scopus consists of more than 40 thousand different sources, which is a large amount of data even selecting that associated with an institution, geographical area or type of publication, 3) the API of Scopus can return queries in JSON format, MongoDB has been chosen to use this database for greater flexibility, scalability and performance of data in non-established formats.

The methodology used in this research was a variant of the discovery of knowledge in databases (Figure 1). Next, the actions executed in each of the phases for the scientific production of a Mexican public university are described. However, it is intended that these phases be replicated to obtain the production of a specific geographical area, by type of production or to evaluate institutions belonging to an evaluation instance for scientific production.

The Autonomous University of Yucatán (UADY) is the most important educational institution in south-eastern Mexico. It has 15 faculties distributed in five campuses and a research center focused on two areas of study, social science and biomedical science.

As of June 1st, 2018, at UADY there were 824 full-time Professors. From this data, a document database was designed in MongoDB.

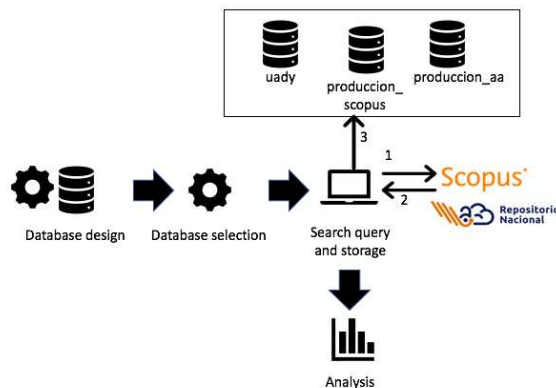
*Database design.* Three databases were created: one for the information of the researchers to obtain their production (named **uady** in Figure 1); the second one called **produccion\_scopus** database contains all the results of the query to this repository through its API; and **produccion\_aa** database for production stored in open access repositories.

*Selection.* It consists of learning the domain of knowledge, considering priority knowledge that is relevant and the goals of the application. Due to the content of the high impact publications and its web service, Scopus was chosen, in addition to the repositories that make up the Mexican national repository of CONACYT, which allow metadata harvesting through the OAI-PMH protocol.

*Recovery.* To recover the production associated with a researcher, it is necessary to know his Scopus code. However, it can have more than one key, so his identifiers were first located. Subsequently, a new consultation was made with Scopus requesting the production associated with the researcher considering as an additional parameter the institution to which he belongs.

*Storage.* The elementary data of each researcher is recovered, and for each publication, descriptive data (title, authors, keywords, identifiers, among others).

*Analysis.* At this stage, the results stored in the documents are observed, in order to visualize the recovered information in the corresponding context. It is possible to analyse this through statistics, data mining, and other techniques.



**Figure 1:** Phases of the methodology.

### 3 Results

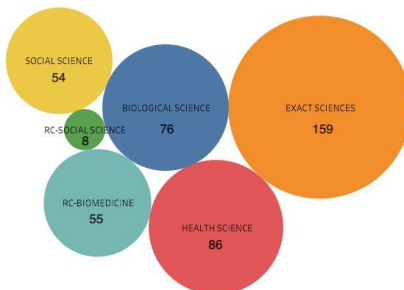
The use of various collections in the proposed databases allows a more efficient distinction to be made for each object stored. Of the 824 full-time Professors, various types of scientific publications were found in Scopus and in the open access repositories belonging to the national repository, which are presented in Table 1.

Type of publication	Total
Papers	2441
Chapters	66
Conferences	218
Books	12
Others	121

**Table 1:** UADY production stored in the document database.

From these data, it has been possible to carry out various studies, such as: an index system to measure productivity and scientific relevance from various digital repositories (Guerrero Sosa, Menéndez Domínguez & Castellanos Bolaños, 2018), the representation of the scientific collaboration using graph theory (Guerrero-Sosa, Menendez-Domínguez, Castellanos-Bolaños & Curi-Quintal, 2019) and the use of an ontological model for the representation of scientific production (Guerrero-Sosa, Menéndez-Domínguez, Castellanos- Bolaños & Gómez-Montalvo, 2019).

On the other hand, in Figure 2 it is presented by campus how many professors have scientific production stored in the document database.



**Figure 2:** Professors production at UADY by campus.

## 4 Conclusion

This paper presented the use of digital repositories of scientific production as a source of data for the evaluation of the productivity of researchers. The use of a document database for the storage and management of information was proposed, considering the advantages associated with big data. Subsequently, the methodology was presented based on the discovery of knowledge through data sets. Finally, the results were presented, which are a first approach to relevant information on the state of current research in a Mexican university. It should be mentioned that based on this proposal, it is possible to perform more complex tasks and with a greater analysis of the results, so what is presented is the basis for the production of various studies related to scientific production.

## References

- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(7). <https://doi.org/10.1186/1742-5581-3-7>
- CONACYT. (2017). Repositorio Nacional. Retrieved October 25, 2017, from <https://www.repositorionacionalcti.mx/>
- Elsevier. (2019). Scopus Search API. Retrieved April 7, 2019, from <https://dev.elsevier.com/documentation/ScopusSearchAPI.wadl>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2007). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*. <https://doi.org/10.1096/fj.07-94921sf>
- Guerrero-Sosa, J. D. T., Menendez-Domínguez, V., Castellanos-Bolaños, M.-E., & Curi-Quintal, L. F. (2019). Use of graph theory for the representation of scientific collaboration. In N. T. Nguyen (Ed.), *11th International Conference on Computational Collective Intelligence*. Hendeaye: Springer.
- Guerrero-Sosa, J. D. T., Menéndez-Domínguez, V., Castellanos-Bolaños, M. E., & Gómez-Montalvo, J. (2019). Use of an Ontological Model to Assess the Relevance of Scientific Production. *IEEE Latin America Transactions*, 17(9), 1424–1431.
- Guerrero Sosa, J. D. T., Menéndez Domínguez, V. H., & Castellanos Bolaños, M. E. (2018). Sistema de índices para valorar la calidad de la producción académica y la investigación, a partir de repositorios digitales y metadatos. In M. E. Prieto-Méndez, S. J. Pech-Campos, & A. Francesa-Alfaro (Eds.), *X Conferencia Conjunta Internacional sobre Tecnologías y Aprendizaje* (pp. 45–52). Cartago: CIATA.org-UCLM.
- Han, J., Haihong, E., Le, G., & Du, J. (2011). Survey on NoSQL database. In *Proceedings - 2011 6th International Conference on Pervasive Computing and Applications, ICPCA 2011*. <https://doi.org/10.1109/ICPCA.2011.6106531>
- Ip, A., Morrison, I., & Currie, M. (2001). What is a learning object, technically? In W. Fowler & J. Hasebrook (Eds.), *Proceedings of WebNet 2001- World Conference on the WWW and Internet* (pp. 580–586). Association for the Advancement of Computing in Education (AACE). Retrieved from <https://files.eric.ed.gov/fulltext/ED466945.pdf>
- Jasco, P. (2005). As we may search - Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537–1547.

# Gesture Recognition in Sign Languages: Methods and Approaches

Allan Ojeda-Pat and Francisco Moo-Mena

Universidad Autónoma de Yucatán, Mérida, Mexico  
allan.ojedaa@gmail.com, mmmena@correo.uady.mx

## Abstract

In the world of computing, computer vision is a highly studied field due to the great advantages it provides. For example, the recognition of objects by means of a photograph or using a camera in real time allows to extract important information that may be useful to the user. Using sign language is an effective way in which people with hearing troubles can communicate with any other person, either by emergency or by social inclusion. A problem arises due to the gap and the lack of interaction between deaf and hearing people, and the lack of effective systems capable of helping to minimize this gap. The main goal of this work is to describe the methods and work done for sign language recognition using convolutional neural networks.

## 1 Introduction

Unlike computers, humans can perform sorting and location activities intuitively. The goal of computer vision is to be able to give similar capabilities to intelligent systems in order to replicate the skills of the human visual system [1, 2]. Given the advances in technology and computer vision methods, they have been used today in a wide variety of applications. Such as human interaction computer, robotics, and different real-world problems that can be improved with a focus on computational vision [1, 3]. One approach to solving a problem is sign language. Using sign language is an effective way in which people with hearing problems can communicate with any other person, either by emergency or by social inclusion. A problem arises due to the gap and the lack of interaction between deaf and hearing people, and the lack of effective systems capable of helping to minimize this gap [4, 5, 6]. The object of study of this work is to describe the methods and recent work in the task of sign language recognition focused mainly on the use of convolutional neural networks.

### 1.1 Sign language

Sign language is considered an effective way of communication between people with deafness and people who listen. In general, sign language is composed of two types of movements, manual and non-manual signs. The manual signs consist of hand and finger movement. Non-manual signs are made up of facial expression movements, such as lip, eyebrow, head and other similar movements. Non-manual signals complement the meaning of a hand signal [7]. Sign language recognition is considered a challenge in the area of computer vision [8].

### 1.2 Object and gesture recognition

A simple scheme used for object recognition is based on two primary components: a descriptor and a classifier. A descriptor is responsible for extracting features and generating candidates to consider in an image or video. A classifier is responsible for classifying candidates or descriptors, based on their shape, color, or by means of any other defined parameter. Finally, a window generation module is responsible for extracting the connected regions that form candidates and representing a decision or response. [9]. See figure 1.

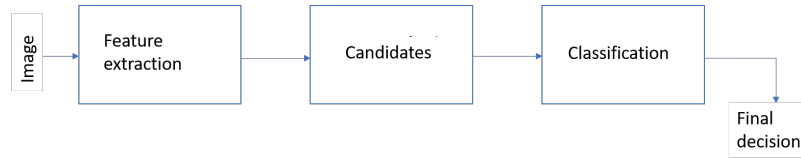


Figure 1: Basic recognition scheme.

Gesture recognition includes object recognition, gesture tracking and classification. Gesture recognition includes a video [10, 11] as input. For gesture recognition two phases can be defined. The first phase is known as a low-level part that is responsible for receiving the video input provided. In this phase, activities such as background subtraction and object segmentation are implemented in order to improve the image and focus on important objects. Later in this stage the detection of objects within the video or frame is applied, for instance a person's hands [12]. The second phase is known as a high-level part that is responsible for tracking the objects and an analysis of the activity that is carried out (classification) [12]. To perform the classification of candidates there are algorithms with different methodologies to achieve the purpose, a few will be described later. Gesture recognition can be seen as the problem of sign language recognition due to the similar approach and methodologies used [13].

### 1.3 Descriptors and classifiers

The first step is to extract features of an image or sequence of images from a video. This leads to one or more feature vectors also known as descriptors. This information will help the computer to perform the classification of gestures, serving as a new representation of the image [14, 15]. A computer vision system typically feeds machine learning algorithms for subsequent classification [1]. Some more popular descriptors are Histogram of Oriented Gradients (HOG), Haar-like features, Transformation of invariant characteristic of scale (SIFT) and Speeded-Up Robust Features (SURF) [1, 15].

A classifier will use these features obtained by some method to discriminate between the possible candidates and thus generate a final classification [14]. The most common algorithms belong to machine learning classifiers with a supervised method. They learn a mapping of inputs to outputs, given a tagged set of input and output [1]. Other ones are based on assembly methods, which are responsible for combining the predictions of many base classifiers to improve the generalization of a single classifier. Some of the most common machine learning methods are: Support Vector Machine (SVM) [1] and AdaBoost [16].

## 2 Convolutional Neural Networks (CNN)

Convolutional neural networks are a kind of deep learning models. CNNs replace the two primary steps of feature extraction and candidate classification, through a single neural network that trains end-to-end with RAW pixel values as input and a final classification as output [17]. CNN is a powerful tool in solving tasks based on computer vision for object recognition and classification. CNN avoids the complex feature extraction process. One of the complications in its development is the large amount of data needed to train a CNN. With a long training database, complex problems can be solved with a minimum margin of error [1, 18]. Due to the large number of images that are processed, powerful GPUs are required to train a CNN [19].

### 2.1 Static and Dynamic Sign Language Recognition

A work to recognize 26 letters of the alphabet and 10 numerals of the American language is proposed in [20]; they trained a CNN with still images and reported an accuracy value of 96%. Next we describe the works for dynamic movements through video. In [21], spatial attention based on a 3D convolutional



Table 1: Summary of works.

Author	Year	Used Method	Database	Training Data	Accuracy	Movement Type	Portability
Masood et al. [20]	2018	CNN	American Sign Language	RGB Images	96%	Static	No
Huang [21]	2018	Spatial attention + CNN	Chinese Sign Language	RGB Video + Depth	95.3%	Dynamic	No
Masood et al. [22]	2018	CNN + Recurrent Networks	Argentinean Sign Language	RGB Video	95.2%	Dynamic	No
Kishore et al. [18]	2018	CNN	Indian Sign Language	RGB Video	92.88%	Dynamic	Yes
Yuangcheng et al. [23]	2018	CNN + Recurrent Networks	American Sign Language	RGB Video + Depth	69.2%	Dynamic	No
Pigou et al. [14]	2015	CNN	Italian Sign Language	RGB Video + Depth + Joints	91.7%	Dynamic	No

neural network is proposed for the task of Chinese sign language recognition. For this, the CNN was fed with features of an RGB video source and a depth data camera. After the extraction they use different techniques to classify a vocabulary of 500 words. The proposal gets an accuracy value of 95.3%.

In [22] a real-time sign language recognizer is proposed using sequence videos. The CNN was trained with the Argentine language database for 46 signs. They reported an accuracy value of 95.2%. In [18] an approach is made to classify sign language through video on a mobile device. They recognize the Indian sign language for 200 different signs using a CNN by capturing continuous 2D video in selfie mode as input. They reported an accuracy value of 92.88%. In [23] a system was developed for recognizing the American sign language using continuous RGB videos for 27 words. They reported an accuracy value of 69.2%. For the Italian language, a 20 sign recognition system was developed in [14], using a Kinect camera as input data. The authors claim to generalize the classification for any user doing the movements, reporting an accuracy value of 91.7%. The table 1 shows a summary of the works for sign language recognition using a CNN.

### 3 Conclusion

A general introduction to the issue of gesture recognition and its focus on the resolution of sign language recognition was presented. Convolutional neural networks are currently a powerful computational tool for the detection, description and classification of gestures, objects and signs. Due to the particularity of machine learning, CNN are effective in getting features and for the subsequent classification of movements. To ensure that a recognition system has a minimum margin of error, a large amount of training data must be considered. Diverse literature was found for sign language recognition for different languages. Each author defines their own necessary approach for classification and that decision defines the success rate together with the available dataset. For the Mexican language recognition, no related work was found using a CNN.

### 4 Acknowledgments

The first author appreciates the financial support received from CONACYT for the realization of this work as part of his Master's studies.

### References

- [1] S. Khan, H. Rahmani, S. Shah, M. Bennamoun, G. Medioni, and S. Dickinson, *A Guide to Convolutional Neural Networks for Computer Vision*. Morgan and Claypool, 2018.
- [2] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976 – 990, 2010.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*. Texts in Computer Science, Springer London, 2010.
- [4] F. Camacho and J. Labrandero, "Redes neuronales convolucionales aplicadas a la traducción del lenguaje verbal español al lenguaje de señas boliviano," *Revista Ciencia, Tecnología e Innovación*, vol. 12, pp. 755 – 762, 06 2016.

- [5] X. Chai, G. Li, X. Chen, M. Zhou, G. Wu, and H. Li, "Visualcomm: a tool to support communication between deaf and hearing persons with the kinect," *Proceedings of the 15th*, pp. 3–4, 2013.
- [6] M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to arabic sign language recognition," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 551–557, 2014.
- [7] P. Kumar, P. Roy, and D. Dogra, "Independent bayesian classifier combination based sign language recognition using facial expression," *Information Sciences*, vol. 428, pp. 30–48, 2018.
- [8] G. Rao, K. Syamala, P. Kishore, and A. Sastry, "Deep convolutional neural networks for sign language recognition," in *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, pp. 194–197, Jan 2018.
- [9] Coursera, "Detección de objetos," November 2018. Retrieved from <https://www.coursera.org/learn/deteccion-objetos>.
- [10] A. Asadi-Aghbolaghi, A. Clapes, M. Bellantonio, H. Escalante, and V. Ponce-López, "Deep learning for action and gesture recognition in image sequences: a survey," *Springer Verlag*, pp. 539–578.
- [11] D. Leite, J. Duarte, L. Neves, de Oliveira J., and A. Giraldi, "Hand gesture recognition from depth and infrared kinect data for cave applications interaction," *Multimedia Tools and Applications*, vol. 76, pp. 20423–20455, Oct 2017.
- [12] M. Cristani, R. Raghavendra, A. Del-Bue, and V. Murino, "Human behavior analysis in video surveillance: A social signal processing perspective," *Neurocomputing*, vol. 100, pp. 86 – 97, 2013. Special issue: Behaviours in video.
- [13] C. Sun, T. Zhang, B.-k. Bao, C. Xu, and T. Mei, "Discriminative Exemplar Coding for Sign Language Recognition with Kinect," vol. 43, no. 5, pp. 1418–1428, 2013.
- [14] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *Lecture Notes in Computer Science*, pp. 572–578, Springer, 2015.
- [15] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing*. Boca Raton: CRC Press., 2018.
- [16] R. Schapire, "Explaining adaboost," *Schölkopf B., Luo Z., Vovk V. (eds) Empirical Inference.*, pp. 37 – 52, 2013.
- [17] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. Li, "Large-scale video classification with convolutional neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.
- [18] P. Kishore, G. Anantha-Rao, K. Eepuri, T. Maddala, and D. Anil-Kumar, "Selfie sign language recognition with convolutional neural networks," *International Journal of Intelligent Systems and Applications*, vol. 10, pp. 63–71, 10 2018.
- [19] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid cnn-hmm for continuous sign language recognition," 09 2016.
- [20] S. Masood, H. Thuwal, and A. Srivastava, "American sign language character recognition using convolution neural network," pp. 403–412, 2018.
- [21] J. Huang, W. Zhou, H. Li, and W. Li, "Attention based 3d-cnns for large-vocabulary sign language recognition.," *IEEE Transactions on Circuits and Systems for Video Technology*. PP., 2018.
- [22] S. Masood, A. Srivastava, H. Thuwal, and M. Ahmad, "Real-time sign language gesture (word) recognition from video sequences using cnn and rnn," pp. 623–632, 01 2018.
- [23] Y. Ye, Y. Tian, M. Huenerfauth, and J. Liu, "Recognizing american sign language gestures from within continuous videos," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2064 – 2073, 2018.

