Universidad de Málaga

Escuela Técnica Superior de Ingeniería de Telecomunicación

Programa de Doctorado: Ingeniería de Telecomunicación

TESIS DOCTORAL

Methods for Self-Healing based on traces and unsupervised learning in Self-Organizing Networks

Autora:

Ana Gómez Andrades

Directora:

Raquel Barco Moreno

2016

AUTOR: Ana Gómez Andrades

http://orcid.org/0000-0001-6063-6738

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

UNIVERSIDAD DE MÁLAGA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

Reunido el tribunal examinador en el día de la fecha, constituido por:

Presidente: Dr. D. ______________________________

Secretario: Dr. D. ______________________________

Vocal: Dr. D. ______________________________

para juzgar la Tesis Doctoral titulada *Methods for Self-Healing based on traces and unsupervised learning in Self-Organizing Networks* realizada por D^{a}. Ana Gómez Andrades y dirigida por la Dra. D^{a}.Raquel Barco Moreno, acordó por

______________________________ otorgar la calificación de

______________________________ y para que conste,

se extiende firmada por los componentes del tribunal la presente diligencia.

Málaga a ______ de __________ del ______

El Presidente::

El Secretario:

El Vocal:

Fdo.: ______________

Fdo.: ______________

Fdo.: ______________

Acknowledgements

Many years have passed since I came to this University to study Technical Telecommunications Engineer in Sound and Image. Such was my passion for telecommunications that I decided to pursue learning throughout my career. And, here I am finishing my PhD. Looking back, I can wholeheartedly state that the thesis has been a life changing experience, very stressful at times, but worth every minute since it has been the cornerstone of my career progression. Still, none of that would have been possible without the support of those around me. I would therefore like to thank all of those helping me to pursue my thesis and for making this a memorable experience.

Firstly, I would like to express my heartiest gratitude to Raquel Barco, my supervisor, for giving me the opportunity to participate in this project and work in this team, for her trust and continued support in my work (especially in the most critical moments), for guiding me in the right direction, for assisting me in writing papers, reviews and this thesis, no matter what time of the day.

I would like to express my special appreciation and thanks to Inmaculada Serrano, for the confidence and enthusiasm she has always shown for my reports, for her encouragement and insightful comments during our productive meetings, for incenting me to widen my research from business perspectives and focus on the real needs of Ericsson. I especially appreciate her commitment to support this research in any possible way.

My sincere thanks also goes to Pablo Muñoz for his continuous support of my PhD study and related research, for his motivation and immense knowledge, for the stimulating discussions and for all the fun we have had in the last years. His guidance helped me in all the time of research. It is impossible for us to forget how many hours we spent dealing with the configuration of the LTE simulator. Nevertheless, it is rewarding to know that we succeeded.

I want to thank everyone for the confidence that they have one more time placed in me by allowing me the chance to work with RCA team. Thanks all concerned for giving me this opportunity and thanks to Adolfo Fernández, Jaime Rodríguez and Rubén Morales for making me feel valued and part of the team from the first day. More especially, I would like to express my warm thanks to Adolfo for actively contributing to make this possible, for being truly inspirational, for generously assisting me so effectively in every possible way and for making me feel very comfortable in a completely new environment inside of a new culture. I especially enjoyed our fun and friendly conversations on the way to work and, naturally, those afternoons working at Starbucks. Thanks also to all the people at GCD RAN US Ericsson for their kind welcome and support during my staying in Plano (USA). Thanks to Helen Zeng for accepting to supervise my international stay. Special mention to Paolo, for receiving me with a smile on his face every day and for making the office a nice place to work. I would also like to extend my thanks to the AT&T CU, especially to Michal Morawski and David Burkholder for recognizing my work done and contributions in the R&D activities and also for economically supporting my stay in US. This stay would not have been a positive and enriching experience in human terms without the hospitality of Gokcen. I have no words to express my gratitude to Gokcen and for what she has done to make me feel at home every day. She has taught me many things at a personal level: a new way of life, a new form of seeing the world and a new way to interpret the needs presented. I will not forget our Swing dance parties. Specially, I would like to thank Gokcen and Micah for taking care of me, for being my friends from the first time, for helping me in my daily tasks and for all the fun we have had. I also wish to thank the people I had the great opportunity to meet in Dallas, with whom I have spent many happy moments.

More than being proud of working for an extraordinary project, I am proud of working with extraordinary colleagues. Thus, I would like to thank Isa, Ale, Juan, Emil, David, Lydia and Sergio for making the lab a pleasant place to work. Special mention to Isa for her support and help in using our "beloved" LTE simulator, which has given us many headaches. Thanks to Isa, Ale, Juan, Emil and David for actively participating in our technical discussions, sharing our concerns and providing good advices. Thanks to all my colleagues for the good time spent together in Glasgow, Valencia and Santiago de Compostela. Thanks to all my colleagues in Ericsson for creating such a nice working environment, specially to Patricia D. and Patricia C. for also contributing to this thesis.

Last but not least, I would like to strongly thank to all the people who have been with me all these years. A great thanks to my parents and my sister for the motivation and encouragement they have provided me through my entire life. I would also like to express my most sincere gratitude to Rafa for his love and patience, for being always by my side, for helping me whenever necessary and, most important of all, for his great sense of humor that makes me laugh even in the most difficult moments.

Contents

Abstract

With the advent of Long-Term Evolution (LTE) networks and the spread of a highly varied range of services, mobile operators are increasingly aware of the need to strengthen their maintenance and operational tasks in order to ensure a quality and positive user experience. Furthermore, the co-existence of multiple Radio Access Technologies (RAT), the increase in the traffic demand and the need to provide a great variety of services are steering the cellular network toward a new scenario where management tasks are becoming increasingly complex. As a result, mobile operators are focusing their efforts to deal with the maintenance of their networks without increasing either operational expenditures (OPEX) or capital expenditures (CAPEX). In this context, it is becoming necessary to effectively automate the management tasks through the concept of the Self-Organizing Networks (SON).

In particular, SON functions cover three different areas: Self-Configuration, Self-Optimization and Self-Healing. Self-Configuration automates the deployment of new network elements and their parameter configuration. Self-Optimization is in charge of modifying the configuration of the parameters in order to enhance user experience. Finally, Self-Healing aims reduce the impact that failures and services degradation have on the end-user. To that end, Self-Healing (SH) systems monitor the network elements through several alarms, measurements and indicators in order to detect outage and degraded cells, then, diagnose the cause of their problem and, finally, execute the compensation or recovery actions.

Even though mobile networks are become more prone to failures due to their huge increase in complexity, the automation of the troubleshooting tasks through the SH functionality has not been fully realized. Traditionally, both the research and the development of SON networks have been related to Self-Configuration and Self-Optimization. This has been mainly due to the challenges that need to be faced when SH systems are studied and implemented. This is especially relevant in the case of fault diagnosis. However, mobile operators are paying increasingly more attention to self-healing systems, which entails creating options to face those challenges that allow the development of SH functions.

On the one hand, currently, the diagnosis continues to be manually done since it requires considerable hard-earned experience in order to be able to effectively identify the fault cause. In particular, troubleshooting experts thoroughly analyze the performance of the degraded network elements by means of measurements and indicators in order to identify the cause of the detected

anomalies and symptoms. Therefore, automating the diagnosis tasks means knowing what specific performance indicators have to be analyzed and how to map the identified symptoms with the associate fault cause. This knowledge is acquired over time and it is characterized by being operator-specific based on their policies and network features. Furthermore, troubleshooting experts typically solve the failures in a network without either documenting the troubleshooting process or recording the analyzed indicators along with the label of the identified fault cause. In addition, because there is no specific regulation on documentation, the few documented faults are neither properly defined nor described in a standard way (e.g. the same fault cause may be appointed with different labels), making it even more difficult to automate the extraction of the expert knowledge. As a result, this a lack of documentation and lack of historical reported faults makes automation of diagnosis process more challenging.

On the other hand, when the exact root cause cannot be remotely identified through the statistical information gathered at cell level, drive test are scheduled for further information. These drive tests aim to monitor mobile network performance by using vehicles to personally measure the radio interface quality along a predefined route. In particular, the troubleshooting experts use specialized test equipment in order to manually collect user-level measurements. Consequently, drive test entail a hefty expense for mobile operators, since it involves considerable investment in time and costly resources (such as personal, vehicles and complex test equipment). In this context, the Third Generation Partnership Project (3GPP) has standardized the automatic collection of field measurements (e.g. signaling messages, radio measurements and location information) through the mobile traces features and its extended functionality, the Minimization of Drive Tests (MDT). In particular, those features allow to automatically monitor the network performance in detail, reaching areas that cannot be covered by drive testing (e.g. indoor or private zones). Thus, mobile traces are regarded as an important enabler for SON since they avoid operators to rely on those expensive drive tests while, at the same time, provide greater details than the traditional cell-level indicators. As a result, enhancing the SH functionalities through the mobile traces increases the potential cost savings and the granularity of the analysis.

Hence, in this thesis, several solutions are proposed to overcome the limitations that prevent the development of SH with special emphasis on the diagnosis phase. To that end, the lack of historical labeled databases has been addressed in two main ways. First, unsupervised techniques have been used to automatically design diagnosis system from real data without requiring either documentation or historical reports about fault cases. Second, a group of significant faults have been modeled and implemented in a dynamic system level simulator in order to generate an artificial labeled database, which is extremely important in evaluating and comparing the proposed solutions with the state-of-the-art algorithm. Then, the diagnosis of those faults that cannot be identified through the statistical performance indicators gathered at cell level is automated by the analysis of the mobile traces avoiding the costly drive test. In particular, in this thesis, the mobile traces have been used to automatically identify the cause of each unexpected user disconnection, to geo-localize RF problems that affect the cell performance and to identify the impact of a fault depending on the availability of legacy systems (e.g. Third Generation, 3G). Finally, the proposed techniques have been validated using real and simulated LTE data by analyzing its performance and comparing it with reference mechanisms.

Resumen

Con la llegada de las redes LTE y la proliferación de una gran variedad de servicios, los operadores de telefonía móvil son cada vez más conscientes de la necesidad de reforzar las tareas de operación y mantenimiento con el fin de garantizar una experiencia de usuario de calidad. Además, la coexistencia de múltiples tecnologías de acceso radio (*Radio Access Technologies*, RAT), el aumento de la demanda de tráfico y la necesidad de proporcionan una gran variedad de servicios están dirigiendo las redes móviles hacia un escenario en el que las tareas de mantenimiento son cada vez más complejas. Debido a ello, los operadores de telefonía móvil están concentrando sus esfuerzos para abordar la gestión de la red sin aumentar los gastos operativos ni los gastos de capital. En este contexto, se hace necesario automatizar de manera eficiente las tareas de gestión a través de las redes auto-organizativas (*Self-Organizing Networks*, SON).

En concreto, las funciones SON cubren tres áreas: auto-configuración, auto-optimización y auto-curación. La primera de ellas, auto-configuración, automatiza el despliegue de los nuevos elementos de red así como la configuración de sus parámetros. La segunda, auto-optimización, está encargada de modificar la configuración de los parámetros con el fin de mejorar la experiencia del usuario. Finalmente, el sistema de auto-curación tiene como objetivo reducir el impacto que los fallos y la degradación de los servicios tienen en el usuario final. Para ello, un sistema de auto-curación monitorea los elementos de la red a través de alarmas, medidas e indicadores con el fin de detectar cortes del servicio o celdas degradadas, posteriormente, diagnosticar la causa de dichos fallos y, finalmente, ejecutar las acciones correctoras o compensadoras.

A pesar de que las redes móviles son cada vez más proclives a los fallos, debido al enorme incremento de su complejidad, la automatización de las tareas de resolución de problemas a través de la funcionalidad de auto-curación no se ha realizado completamente. Tradicionalmente, tanto la investigación como el desarrollo de las redes SON ha estado relacionada con las funcionalidades de auto-configuración y auto-optimización. Esto se ha debido principalmente a los desafíos que hay que afrontar cuando se quiere estudiar e implementar un sistema de auto-curación. Esto es especialmente relevante en el caso de la diagnosis de fallos. Sin embargo, los operadores de telefonía móviles están prestando cada vez más atención a los sistemas de auto-curación, lo que implica proponer opciones para hacer frente a dichos desafíos que permitan el desarrollo de las funcionalidades de auto-curación.

Por un lado, actualmente, la diagnosis sigue siendo una tarea realizada manualmente dado que requiere una considerable experiencia, obtenida con una dura labor, con el fin de ser capaz

de identificar la causa del fallo. En concreto, los expertos en resolución de problemas analizan exhaustivamente el rendimiento de los elementos de red degradados con el fin de identificar la causa de las anomalías detectadas. Por lo tanto, automatizar las tareas de diagnosis implica conocer los indicadores que tienen que ser analizados y cómo mapear los síntomas identificados con las causas de los fallos. Este conocimiento es adquirido a lo largo del tiempo y se caracteriza por ser específico para cada operador en base a sus estrategias y las características de la red. Además, los expertos típicamente resuelven los problemas sin documentar el proceso ni guardar los indicadores analizados junto con la etiqueta de la causa del fallo. Debido a que no hay una regulación específica sobre cómo documentarlo, los pocos fallos documentados no están bien definidos ni descritos de manera estándar, lo que da lugar a que un mismo fallo sea nombrado con distinta etiqueta, dificultando aún más el proceso de automatizar la extracción del conocimiento experto. Como resultado, esta falta de documentación y de un histórico de casos etiquetados hace que la automatización del proceso de diagnosis sea aún más difícil.

Por otro lado, cuando la causa exacta no puede ser identificada remotamente por medio de los estadísticos recogidos a nivel de celda, los drive test son planificados para obtener información más detallada. Mediante los drive test se toman medidas radio de la red móvil a lo largo de rutas previamente definidas. En concreto, los expertos utilizan equipos de prueba especializados para recoger medidas manualmente. Como consecuencia, los drive test suponen un alto coste para los operadores, dado que conllevan una inversión considerable en tiempo y recursos (tales como personal, vehículos y equipos de prueba). En este contexto, el *Third Generation Partnership Project* (3GPP) ha estandarizado la recogida automática de medidas de campo (por ejemplo, mensajes de señalización, medidas radio e información de localización) a través de la funcionalidad de las trazas móviles y su funcionalidad extendida *Minimization of Drive Tests* (MDT). En concreto, las trazas móviles están consideraras como un factor importante para mejorar las funcionalidades SON dado que evitan a los operadores tener que depender de los caros drive test mientras que al mismo tiempo proporciona muchos más detalles que los tradicionales indicadores a nivel de celda. Como resultado, mejorar las funcionalidades de auto-curación a través de las trazas móviles aumenta el ahorro en los costes y mejora la granularidad del análisis.

Por tanto, en esta tesis, distintas soluciones son propuestas para abordar los desafíos que evitan el desarrollo de los sistemas de auto-curación con especial énfasis en la fase de diagnosis. Para ello, la falta de casos etiquetados ha sido abordada de dos maneras distintas. En el primero se han utilizado técnicas no supervisadas para diseñar automáticamente sistemas de diagnosis a partir de datos reales sin necesitar ni datos etiquetados ni documentación sobre el comportamiento de los fallos. En el segundo se ha modelado e implementado un grupo de fallos significativos en un simulador dinámico con el fin de generar un conjunto de datos etiquetados que es sumamente importante para evaluar y comparar las soluciones propuestas con algoritmos disponibles en el estado del arte. Posteriormente, la diagnosis de los fallos que no pueden ser analizados mediantes estadísticos a nivel de celda ha sido automatizada mediante el análisis de las trazas móviles para evitar los costosos drive test. En concreto, en esta tesis, las trazas móviles se han utilizado para identificar automáticamente la causa de cada desconexión de usuario indeseada, para geo-localizar problemas radio que afectan al rendimiento global de la celda y para identificar el impacto que provoca un fallo dependiendo de si hay sistemas preexistentes (tales como la tercera generación móvil, 3G). Finalmente, se han validado las técnicas propuestas utilizando tanto datos reales como simulados analizando su rendimiento y comparándolo con técnicas de referencia.

Acronyms

2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
ANR	Automatic Neighbor Relation function
ARQ	Automatic Repeat Request
ATOL	Average active time on LTE
BLER	Block Error Rate
BMU	Best Matching Unit
BN	Bayesian Network
CAPEX	Capital Expenditures
CCO	Capacity and Coverage Optimization
CH	Coverage Hole
CM	Configuration Management Parameters
CMC	Connection Mobility Control
COC	Cell Outage Compensation
CTR	Cell Traces

DER	Diagnosis Error Rate
DL	Downlink
eNB	evolved Node B
EPC	Evolved Packet Core
EPS	Evolved Packet System
ER	Error Rate
ERA	Ericsson RAN Analyser
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDD	Frequency-division Duplexing
FPR	False Positive Rate
FNR	False Negative Rate
GCD	Global Competence Domain
GPS	Global Positioning System
GSM	Global System for Mobile communications
GPRS	General Packet Radio Service
HARQ	Hybrid Automatic Repeat Request
HO	Handover
IP	Internet Protocol
IMS	IP-Multimedia Subsystem
KPI	Key Performance Indicators
LTE	Long-Term Evolution
M2M	Machine to Machine
MAC	Medium Access Control

MDT	Minimization of Drive Test
MME	Mobility Management Entity
NAS	Non-Access Stratum
NGMN	Next Generation Mobile Networks
OAM	Operation, Administration and Maintenance
ODG	OSS Data Gateway
OFDMA	Orthogonal Frequency Division Multiplex Access
OPEX	Operational Expenditures
OSS	Operations Support System
PAPR	Peak-to-Average Power Ratio
PBD	Percentile-Based Discretization
PCI	Physical Cell Identity
PDCP	Packet Data Convergence Control
PDF	Probability Density Functions
PF	Performance Functions
PI	Performance Indicators
PM	Performance Management Parameters
PRB	Physical Resource Blocks
P-GW	Packet Data Network Gateway
QAM	Quadrature Amplitude Modulation
QoS	Quality-of-Service
QPSK	Quadrature Phase Shift Keying
RAC	Radio Admission Control

RACH	Random Access Channel
RAN	Radio Access Networks
RAT	Ratio Access Technology
RB	Resource Block
RBC	Radio Bearer Control
RCA	Root Cause Analysis
RE	Resource Element
RF	Radio Frequency
RLC	Radio Link Control
RLF	Radio Link Failure
RRM	Radio Resource Management
RSRP	Reference Signal Receive Power
RSRQ	Reference Signal Receive Quality
RSSI	Received Signal Strength Indicator
SC-FDMA	Single Carrier Frequency Division Multiple Access
S-GW	Serving Gateway
SH	Self-Healing
SINR	Signal-to-Interference-plus-Noise Ratio
SOM	Self-Organizing Maps
SON	Self-Organizing Networks
TA	Timing Advance
TDD	Time-division Duplexing
TER	Total Error Rate

TPS	Trace Processing Server
TTT	Time-To-Trigger
UE	User Equipment
UL	Uplink
uRAT	underlying Ratio Access Technology
WCDMA	Wideband Code Division Multiple Access

Chapter 1

Introduction

In this chapter, the aim of this thesis is described, presenting the motivation, the objectives and the structure of this document.

1.1 Motivation

Traditionally, mobile networks have been an effective way of communicating information anywhere at any time. Furthermore, in recent years, the high development of mobile technologies along with the proliferation of a huge variety of applications has transformed the mobile phone into a essential device for the new lifestyle. One thing is certain, mobile phones have changed our day-to-day life, modifying not only the way we communicate but also the way we view the world, and even the way we act. Consequently, cellular networks have extremely developed in the last years. In particular, the number of mobile subscriptions is growing across all regions, exceeding the population in many countries. This is largely due to the emergence of new subscribers in developing regions (as phones become more affordable) and the increased number of devices per individual in mature markets. For example, the number of mobile subscriptions worldwide exceeds 7 billion by end 2015, representing a global penetration of 97% [1]. As a result, the large increase in the number of connected devices, mobile applications and data consumption per device has caused a significant growth of the mobile data traffic.

In this context, where consumers expect high-quality user experience and continual network development but with steady prices, mobile operators seek to continually evolve their networks. These include higher data speed, development of new functionalities and services, capacity improvements to handle new applications and better spectrum efficiency. As a result, mobile networks are currently evolving from either Second Generation (2G) or Third Generation (3G) towards the Fourth Generation of mobile networks known as Long-Term Evolution (LTE) [2]. In this complex scenario, mobile operators have to efficiently address the maintenance and operational tasks in order to meet the consumer demand, ensuring a high quality level and enhancing the user's experience. In addition, mobile operators aim to ensure minimum Capital

Expenditures (CAPEX) and minimum Operational Expenditures (OPEX). Namely, mobile operators have to achieve a balance between maximizing profits with lower prices and delivering high-quality services for the customer.

The key solution for reaching this objective involves the automation of the management and operational tasks. And this is exactly what the development of Self-Organizing Networks (SON) [3, 4] aims to cover. This new paradigm arrives with the main purpose of automating the manual tasks by incorporating intelligence into the network, thus achieving substantial savings. Therefore, this issue has raised great interest among operators and the research community. It was in 2008 when the Next Generation Mobile Networks (NGMN) Alliance established the principles and use cases of SON [5, 6]. Later, the 3rd Generation Partnership Project (3GPP) further developed the requirements of SON, including them in its standards. In particular, 3GPP grouped the main SON functionalities into three categories [3, 7, 8]: Self-Configuration [9], Self-Optimization [9] and Self-Healing (SH) [10].

- Self-Configuration: it is responsible for automating the setup of the newly deployed network elements, auto-configuring their radio parameters (such as power, transmission frequency...).
- Self-Optimization: it is in charge of dynamically reconfiguring network parameters based on the changing conditions of the environment (e.g. traffic, the presence of new neighboring cells...) in order to enhance the quality of the services.
- Self-Healing: it aims to automatically recover the network elements from outage or service degradations that worsen user performance. To that end, this system (see Fig. 1.1) automatically collects the alarms, indicators and measurements of the network in order to detect faulty cells, diagnose the root cause of their problem and, based on that, provides the recovery and/or compensation action. There are situation in which the operator can decide to compensate or even recover a detected fault without diagnosed it at all. For example, it is quite common to blindly recover most of the simple faults via a cell reset.

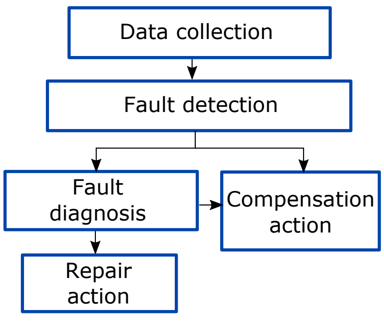

Figure 1.1: Phases of a Self-Healing system [7]

Due to the considerable interest raised by SON functionalities, in the last years, several research projects and consortiums have been focused in the area of SON, such as CELTIC Gandalf [11], FP7 E3 [12], FP7 SOCRATES [13], SELF-NET project [14], UniverSelf [15], SEMAFOUR [16] and COMMUNE [17]. Nevertheless, not all SON functionalities have been addressed to the same extent. Especially, Self-Healing functionality has been the least researched due to the significant limitations that must be faced, even though it plays an increasingly important role in the automation of the network management. Furthermore, the research of each phase of a SH system (see Fig. 1.1) presents different hurdles, reason that has led to a great research gap among those phases. In particular, the detection of failures is one of the most studied [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29] focusing on performance visualization and on modeling the normal behavior of the system in order to detect abnormal cells. However, regarding the automation of the diagnosis phase also called Root Cause Analysis (RCA) in the RAN of cellular networks, only few references can be found such as [30, 31, 32, 33, 34].

It is a common practice on the part of the mobile operators to solve the problems without properly documenting and storing the indicators and their values that led them to identify the fault cause. Note that, when automating a manual process, it is essential to have records about real situations and personal field experience in order to properly design this type of systems. Therefore, the lack of documented knowledge and the lack of historical databases of fault cases are the main reasons why the diagnosis phase of a self-healing system has not been as developed as the detection phase. In particular, the scarce number of historic reported faults greatly complicates the research, development and evaluation of diagnosis systems. Furthermore, within this scenario, all the diagnosis systems available in literature [30, 31, 32, 33, 34] are supervised systems (i.e. systems that requires labeled cases), which is not in line with the real limitations.

In addition, the advent of the technological developments and the big data paradigm [35] presents advanced techniques to further develop SON systems in general, and SH in particular. This, together with the standardization of mobile traces features [36] and the Minimization of Drive Tests (MDT) [37, 38, 39, 40], provides the opportunity to analyze large amount of data related to signaling messages, radio measurements and location information. As a result, current technological development enables the research of new diagnosis systems based on mobile traces, in contrast with what has been done so far. In particular, all previous studies related to SH systems [18, 19, 20, 41, 21, 22, 23, 24, 25, 42, 43, 30, 31, 32, 33, 34] were conducted by mean of cell level statistics stored by the Operation, Administration and Maintenance system (OAM) which causes loss of information about user performance and the geo-location of faults.

Consequently, due to the lack of studies related to SH system in general and diagnosis systems in particular in the RAN of cellular networks, this thesis is devoted to overcome the barriers that have prevented its research and development, with a particular focus on overcoming the lack of historical databases of fault cases and improving the diagnosis by means of mobile traces.

1.2 Challenges and objectives

The main goal of this thesis is to overcome the limitations that hinder significant advances in the area of Self-Healing systems. Fig. 1.2 summarizes the scenario, indicating in each step the particular challenge that must be faced.

Self-Healing systems (Fig. 1.1) aim to automate troubleshooting tasks through automatic systems capable of analyzing cell-level indicators, measurements and alarms in order to automatically detect problematic cells, diagnose the root cause of the problem and, finally, propose the recovery and/or the compensation actions [7]. These systems may be manually designed by troubleshooting experts or automatically built by using machine learning algorithms. Since the first approach requires the involvement of specialists with expertise and time to build complex models, the second one has been welcomed as the most feasible solution [31, 32, 33, 34], as it automates the design process while reducing the intervention of the experts. These systems need historical databases of fault cases, composed of the values of the KPIs at the problematic hour along with the identified fault cause (i.e. the label). However, troubleshooting experts do not tend to save a label identifying the fault cause together with the values of the KPI, resulting in the absence of labeled cases (Fig. 1.2-1), which is the main obstacle that experts have to deal with when they need to design a Self-Healing system. Furthermore, in the existing references on techniques for self-healing in cellular networks [12, 13, 17, 18, 22, 25, 42, 43, 30, 31, 32, 33, 34, 44, 45], the studies are validated using proprietary simulators and different types of causes, making results not comparable or repeatable. These challenges are faced in this thesis by following two different objectives. The first objective (Obj.1) is to artificially generate large sets of labeled cases that model the behavior of real networks under not only normal situation but also in the presence of faults. Having historical databases will allow to design self-healing systems, to validate and compare different techniques in order to choose the most appropriate one. The second objective (Obj.2) is to design unsupervised diagnosis techniques. Unsupervised techniques are those not requiring the fault cause to be trained. Since in live networks the values of KPIs are regularly recorded, those unsupervised diagnosis systems can be directly used in real networks without the need of labeled cases.

For the detection and diagnosis phases of self-healing systems, having a set of labeled cases would ease the design and validation of new techniques. Conversely, for assessing compensation and fault recovery functions, some configuration changes should be done in the network and their impact on the performance should be measured. However, proving those solutions under realistic conditions is limited because the operators are logically reluctant to introduce unpredictable changes in their networks (e.g. power, antenna tilts). This restraints the design of new systems and also impairs their evaluation under realistic conditions (Fig. 1.2-2). Because of that, research on compensation and recovery has been limited to simulations of cell outage [12, 13, 17, 44, 45]. As a result, the whole Self-Healing system has never been evaluated together, but only different functionalities (e.g. diagnosis, compensation, etc.) in isolation. To address this challenge, this thesis has the objective to provide a methodology to design and evaluate any self-healing functionality (Obj.3).

Within the scope of troubleshooting process, the diagnosis phase is the most time consuming task for the experts, since they spend a great deal of time and effort identifying the reason of the problems. Thus, several approaches can be found in the literature on automatic diagnosis in mobile networks (such as [30, 31, 32, 33, 34]). However, the operators are unwilling to apply these diagnosis techniques in their real networks because they require a deep understanding of the related theory (e.g. Bayesian Networks), which, in addition, do not follow their way of working and reasoning. Hence, this is established as another challenge to be confronted (Fig. 1.2-3). In relation to this challenge, the objective (Obj.4) is to design a diagnosis system that reasons in a similar way to cellular networks troubleshooting experts.

Furthermore, in this context of diagnosis systems, the state-of-the-art references are designed to automatically identify faults whose behavior is previously known, since they are built based either on the expert knowledge or on labeled cases. As a result, another challenge is that experts have to manually analyze huge amounts of data and indicators in order to find new faults (Fig. 1.2-5). Thus, another objective of this thesis (Obj.6) is to design a diagnosis system that is able to identify non-common faults, whose typical behavior is not known by diagnosis experts.

Once a self-healing system has been designed and implemented, its parameters should be reconfigured as the mobile network evolves. At this point, the next challenge that must be addressed in this thesis refers to the manual configuration of Self-Healing systems (Fig. 1.2-4). Thus, this process should be as autonomous as possible in order to reduce the intervention of the experts (Obj. 5), since their time is a scarce resource.

Furthermore, state-of-the-art Self-Healing systems have been focused on problem solving by means of cell-level information. That is, these systems analyze cell-level indicators in order to fix the problems of the cells. The main drawback of that approach is that those indicators represent aggregated and statistical values which do not provide user-level details and geo-located information (Fig. 1.2-7). Therefore, this kind of systems is not capable of neither analyzing the user performance nor localizing the problems. Thus, operators have to resort to the traditional drive tests and then analyze the actual user-level performance and the cause of the unintended service disconnections. This has the drawback of requiring highly specialized measurement equipment and also being a time-consuming task, since experts have to manually analyze the user data and manually collect the user-level measurements (Fig. 1.2-6). This challenge leads to the objectives of identifying problems at user level (Obj.7) and of locating critical problematic areas within cells (Obj.8).

Finally, the last challenge is that these Self-Healing systems are based solely on LTE information to diagnose and solve the problems of the network (Fig. 1.2-8). Therefore, it prevents identifying the impact that either the faults or the recovery actions have on other Radio Access Technology (RATs), such as 2G or 3G. The objective (Obj. 9) is to assess the impact that a fault has in an LTE network depending on the existence or not of underlying RATs in a multi-RAT scenario.

In summary, the specific objectives to address the previous challenges are the following:

- (Obj. 1) To provide a representative set of fault cases. In order to cover the lack of labeled validation datasets, the first objective of this thesis is to model a reduced and specific set of KPIs along with the most common fault causes in LTE networks. In this way, a reliable dataset, that is, a set of labeled data related to different fault categories, will be available for further research. The aim is to offer a useful database to the research community facilitating the evaluation and comparison of different techniques in order to choose the most appropriate one for each use case.
- (Obj. 2) To propose an unsupervised diagnosis system for cellular network, i.e. a automatic diagnosis system that does not required labeled cases. Up to now, existing state-of-the-art diagnosis systems for cellular networks are based on supervised techniques.
- (Obj. 3) To propose a methodology for the design and validation of SH system. Namely, in this thesis a complete procedure to design and evaluate any Self-Healing functionality under the same scenario should be proposed, allowing repetition of the experiments by third parties and comparison with other techniques.

- (Obj. 4) To propose a diagnosis system that follows experts' way of reasoning. In particular, the proposed system should be easy to interpret with a clear design based on previously defined KPI-cause relations. The aim is to model the tasks of the experts taking into account the effects that faults cause over the KPIs. These KPI-cause relations may be manually defined by experts or automatically determined through automatic systems (either supervised or unsupervised techniques).
- (Obj. 5) To propose an automatic method to configure SH systems. Once these systems are designed, many parameters need to be configured at deployment time and during its life time as the network evolves. Therefore, in order to release experts from these responsibilities, in this thesis an automatic system to configure SH parameters should be

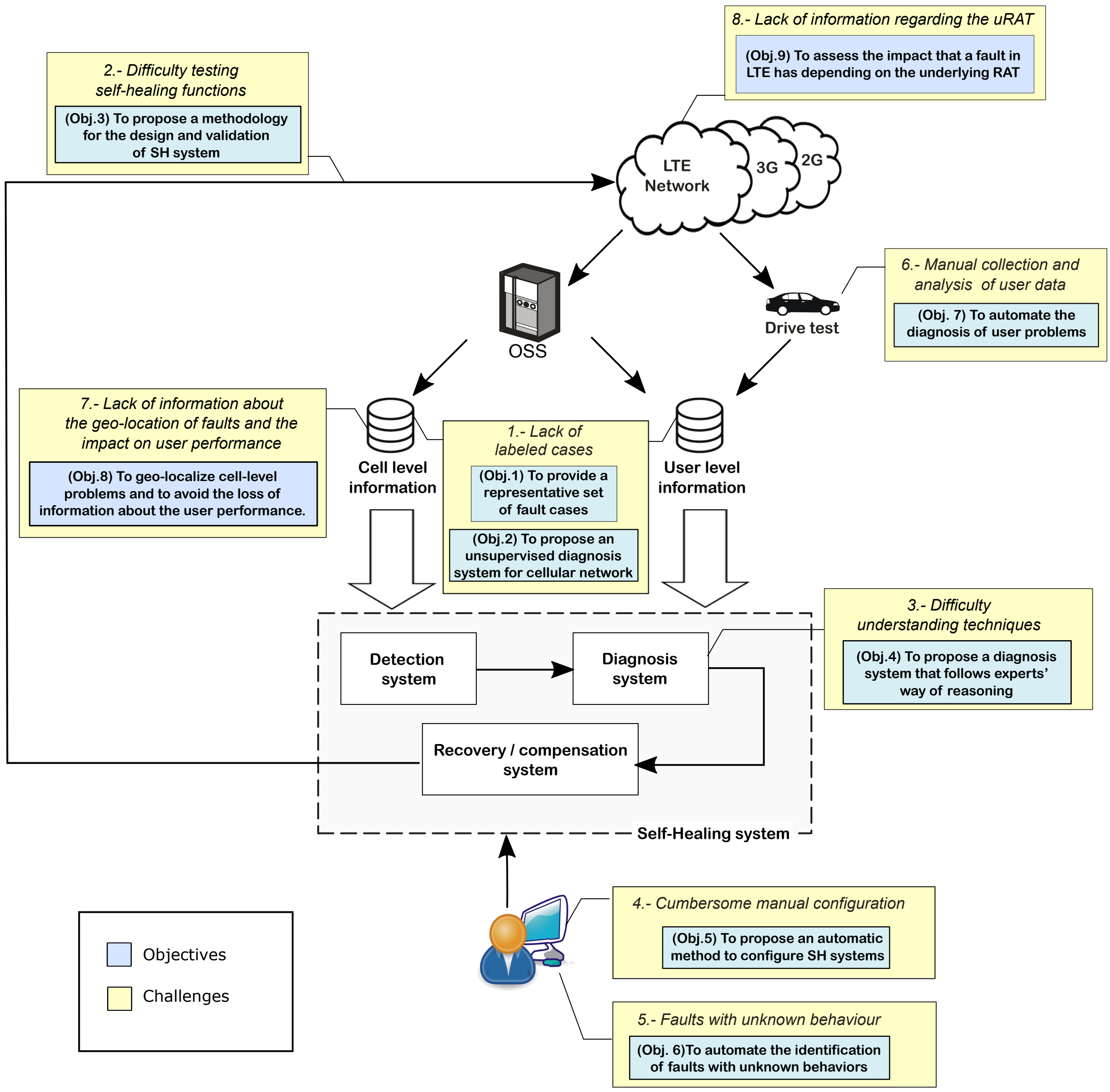

Figure 1.2: Challenges and objectives faced in this PhD.

designed.

- (Obj. 6) To automate the identification of faults with unknown behaviors. As stated earlier in this section, one of the most difficult and time-consuming tasks for the expert is the identification of new faults since it requires analysis of a vast amount of data. Taking into account this objective, together with Obj.2, in this thesis, a diagnosis system capable of identifying unknown patterns among the whole real dataset and without any prior information about the existence of faults should be proposed.
- (Obj. 7) To automate the diagnosis of user problems. Another objective defined in this thesis seeks to design a framework to identify the reason for each unintended service disconnection. The idea is to automatically discern between problematic and non-problematic connections, identifying the particular connection's stage where the issue happens. Toward this objective, the proposed method should be based on the information of the mobile traces or the Minimization of Drive Test (MDT), which have been standardized by the 3GPP in order to automate the collection of user-level information avoiding the traditional drive test.
- (Obj. 8) To geo-localize cell-level problems and to avoid the loss of information about the user performance. To that end, a procedure for the automatic assessment of problems of a cell as part of self-healing systems using mobile traces should be developed. The idea is to use the unsupervised method defined in this thesis to find user-level problems by means of mobile traces and then to propose a method to aggregate those problems at cell level, providing their geo-localization over the map.
- (Obj. 9) To assess the impact that a fault in LTE has depending on the underlying RAT. The objective is to analyze how the information reported in the mobile traces of other underlying RATs can be taken into account from the point of view of LTE Self-Healing systems. The goal is to propose an automatic method that classifies a specific problem based on its impact both on LTE and on the co-existing uRAT, since depending on that the required remedial action will be different.

1.3 Document structure

This report has been organized in seven chapters grouped into three blocks (see Fig. 1.3) for an easier understanding. The first block is devoted to presenting the background and the knowledge required to understand the rest of the report. Chapter 1 introduces this thesis, detailing the motivation of the research and the objective that must be addressed. Chapter 2 provides a review of the theoretical basis of the techniques used. In particular, it is focused on Self-Healing systems in LTE networks. Therefore, in the first section, the LTE technology is presented, describing its main characteristics. Then, an introduction to the SON paradigm is provided, paying particular attention to self-healing functionality.

Since this thesis is presented as a set of publications ("*tesis por compendio de publicaciones*"), the second block consists of the research papers, grouping them into different chapters according to their topic. In order to guide the reader through the report, the second block begins with Chapter 3, in which an insight into each paper is presented detailing the relations among them.

Therefore, Chapter 3 constitutes a comprehensive summary of the main published papers and outlines the research methodology, indicating the technologies and tools employed. The structure of this second block is directly related to the objectives presented in Section 1.2. Subsequently, each paper illustrates the specific problem to be solved, then, it details the proposed solution to the problem, and, finally, evaluates the main results of the associate study. It is worth noting that, each paper presents its specific state-of-the-art explained in detail and specially focused on its topic.

Chapter 4 (Obj.1, Obj. 3, Obj. 4, Obj. 5) includes the two papers that contribute to define a methodology for the design, validation and configuration of SH system. In particular, the first paper presents a procedure to design and evaluate any Self-Healing function based on a dynamic system-level LTE simulator. To that end, this paper defines a model of the KPIs and the most common faults that can be used in a simulator to artificially generate labeled cases. In addition, a diagnosis system based on fuzzy logic is also proposed. Then, the second paper of this chapter provides an unsupervised method to automatically configure SH systems. Then, this methodology and the provided labeled dataset are then used in the rest of the thesis to evaluate the proposed systems.

Chapter 5 (Obj. 2, Obj. 6) presents an automatic diagnosis system based on unsupervised techniques.

Chapter 6 (Obj. 7, Obj. 8, Obj. 9) compiles three papers that are based on the analysis of the information provided by mobile traces. Therefore, in this chapter, the theoretical basis of mobile traces and the characteristics of the user-level measurements are presented. More specifically, the first paper presents an automatic system to analyze mobile traces in order to diagnose user problems. Then, the second paper proposes a technique to automate the identification and geo-localization of cell problems by means of mobile traces. To that end, this paper combines the previously explained user-level diagnosis with the unsupervised system proposed in Chapter 5 to identify new user-level behaviors. The third paper proposes a method to merge the information gathered from LTE and its co-existent uRAT in order to be able to discern different types of impacts of a specific fault.

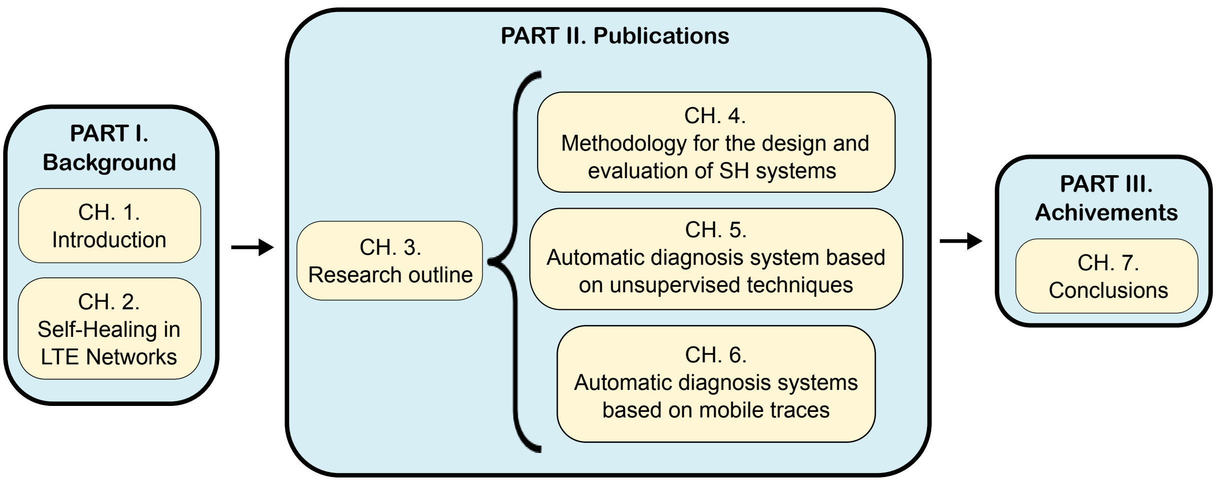

Figure 1.3: Organization of the thesis report.

Finally, the third block consists of Chapter 7, which presents an overview of the results and summarizes the main conclusions of the research and future lines of research.

This thesis also includes a brief summary in Spanish as appendix A.

Chapter 2

Self-Healing in LTE networks

This chapter is devoted to a general presentation of those aspects of LTE network required to follow the rest of this thesis. Thus, the first part summarizes the characteristics of LTE Radio Access Networks (RANs) and presents the basic concepts of SON. Subsequently, the different functionalities of SON are presented in detail, paying special attention to self-healing systems since they play an important role in the content of this thesis. The second part of the chapter describes the state of the art regarding self-healing systems in mobile communication networks. In particular, in this chapter a general overview is presented, while the state of the art associated to each particular contribution of the thesis is elaborated in detail in its corresponding paper.

2.1 Radio Access Technologies

LTE network has been designed to substantially improve the user experience by reducing latency times and increasing packet data rate, while, at the same time, ensuring an efficient use of spectrum. This technology is an all-IP packet-based network, that is, all offered services (either voice or data) are provided over a packet switching architecture, this increases the robustness and efficiency of the network since it improves the integration of the multimedia services.

2.1.1 Network architecture

LTE system, also known as Evolved Packet System (EPS), consists of the Evolved Packet Core (EPC) and the air interface named Evolved Universal Terrestrial Radio Access Network (E-UTRAN), see Fig. 2.1.

The core network is responsible for controlling the network access, managing the user mobility and providing interconnection with other systems. Its main elements are:

- Mobility Management Entity (MME): is the key control-node responsible for
  - The control plane functions for mobility between LTE and 2G/3G access networks.
  - NAS signalling security

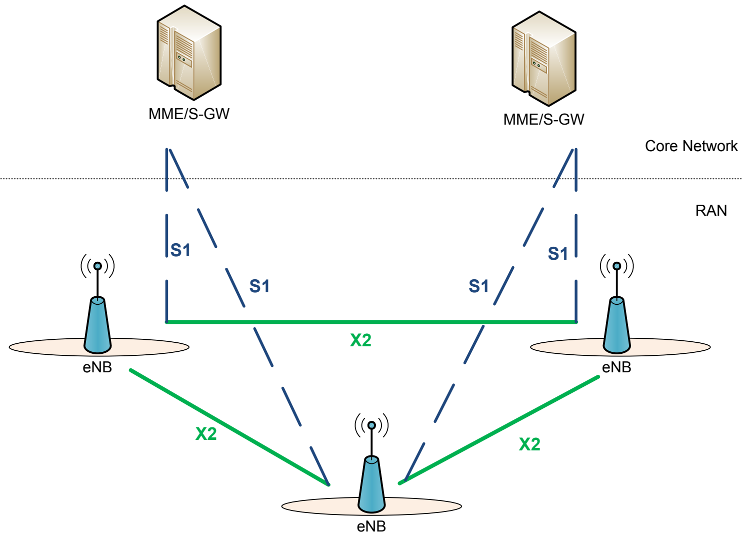

Figure 2.1: LTE network architecture

- Access stratum security control.
- Idle state mobility handling.
- Bearer management functions including dedicated bearer establishment;.

- Serving Gateway (S-GW): is the termination point of the packet data network interface towards E-UTRAN routing the user data packets. SGW functions include:
  - Transport level packet marking in the uplink and the downlink.
  - Handling user data functions such as routing and forwarding packets.
  - Gathering accounting information per user and per bearer.
- Packet Data Network Gateway (P-GW): is responsible for acting as a gateway -between 3GPP and non-3GPP technologies for mobility functions. The most important functions of the P-GW are the following:
  - User Equipment (UE) IP address allocation.
  - Policy enforcement features
  - Per-user based packet filtering for quality-of-service (QoS) differentiation

These network elements are connected with the evolved Node B (eNB) by means of the S1 interface. In turn, the eNBs are interconnected through the X2 interface. The main functions carried out by the eNB are:

- Radio resource management.
- Header compression and user plane ciphering.
- UL and DL bearer level admission control.
- Selection of MME at UE attachment.
- Routing of user plane data towards S-GW.
- Measurement configuration for mobility and scheduling functions.
- User data transmission from and towards the UE through the air interface

2.1.2 Physical layer

The physical layer of the air interface covers several functions such as cell search, power control, multiplexing and channel coding, modulation and physical radio resources allocation. Furthermore, the physical layer of LTE manages DL and UL transmission between the UE and the eNB. In particular, the physical layer supports two multiple access schemes: Orthogonal Frequency Division Multiplex Access (OFDMA) in DL and Single Carrier Frequency Division Multiple Access (SC-FDMA) in uplink, which will be discussed in detail later in this chapter. Additionally both OFDMA and SC-FDMA are supported by using Frequency-division Duplexing (FDD) and Time-division Duplexing (TDD), respectively.

OFDMA is a multiplexing scheme that split the available bandwidth into several narrow subcarriers. In OFDMA, a specific number of subcarriers are allocated to each individual user for a predetermined amount of time. OFDMA signal is characterized by being multi-carrier with one data symbol per subcarrier. This allows the parallel transmission of data symbols in subcarriers and therefore the simultaneous low data rate transmission from several users. Within the OFDM signal it is possible to choose between three types of modulation for the LTE signal:

- QPSK: 2 bits per symbol.
- 16QAM: 4 bits per symbol.
- 64QAM: 6 bits per symbol.

However, in the uplink direction, SC-FDMA is used as the access technique in order to mitigate the negative aspects of OFDMA regarding the high peak-to-average power ratio (PAPR) which is inacceptable for the mobile since they significantly increase the power consumption in the UE terminals. SC-FDMA utilizes a single-carrier transmitting signal, transmitting the data symbols in series over one wideband signal, with higher rate and more bandwidth than OFDMA.

In LTE, a resource element (RE) is the smallest unit and corresponds to one symbol (OFDM or SC-FDMA) in the time domain, and one subcarrier in the frequency domain. The transmission of the data, the REs are organized into physical resource blocks (PRBs), forming the smallest unit that can be scheduled. One PRB occupies 180 kHz in frequency since it is composed of 12 subcarriers with a 15 kHz spacing) and it has a duration of 0.5ms. In each PRB, the user data is transmitted along with the control data, the synchronization signaling and reference symbols (or pilot symbols) to allow for channel estimation.

2.1.3 Link layer

The link layer is further subdivided into three sublayers:

- Medium Access Control (MAC): is in charge of controlling access to the radio channel by means of the dynamic resources allocation and its main functionalities are:
  - Mapping between logical channels and transport channels.
  - Scheduling information reporting.
  - Error correction through HARQ.
  - Priority handling between logical channels of one UE.
- Radio Link Control (RLC): is responsible for concatenation, segmentation and reassembly

of those data packets send between the eNB and the UE. It ensures reliable data transfer since it provides error correction through automatic repeat request (ARQ).

- Packet Data Convergence Control (PDCP) The main functions of this sublayer are the transference of the IP packets associated to both user plane and control plane data. It is therefore responsible for header compression and decompression of IP data, as well as, ciphering and deciphering of user plane data and control plane data.

2.1.4 Network layer

The protocols of the network layer are specific of the control plane and they are over the same physical and link layer as the user plane.

- Radio Resource Control (RRC): is in charge of establishing the connection between the eNB and the UE through which the air interface is controlled. Among these mechanisms, it is worth stressing the management, establishment, configuration, maintenance and release of point to point Radio Bearers.
- Protocolo Non-Access Stratum (NAS): its messages are transferred between the MME and the UE encapsulated in the RRC message, so they are transparent to the air interface. This protocol is responsible for controlling the connection with the transport network carrying out functions such as: support the mobility of the UE and the session management procedures to establish and maintain IP connectivity between the UE and a PDN GW.

2.1.5 Radio Resource Management (RRM)

One of the main goals of LTE networks is to provide high variety of services to a high number of users, whilst also ensuring Quality of Service (QoS) requirements. In order to guarantee this QoS, the fast-changing conditions of the mobile system (due to the variation of the propagation conditions, the traffic variations and the mobility of the UE) have to be faced. As a result, it is necessary to dynamically and efficiently manage the radio resources. Particularly significant among these functionalities are:

- Radio Admission Control (RAC): its main task is to admit or reject the establishment requests for new radio bearers. The goal of RAC is to ensure high radio resource utilization by accepting radio bearer requests if radio resources are available without degrading the in-progress sessions. As a result, a new bearer is established if there are enough resources available to ensure proper QoS for the new radio bearer request and for the in-progress sessions. Otherwise, the radio bearer requests are rejected since they cannot be accommodated.
- Radio Bearer Control (RBC): involves the establishment, maintenance and release of radio bearers along with the configuration of radio resources associated with them. Furthermore, RBC is in charge of the maintenance of the radio bearers of in-progress sessions at the change of the radio resource situation due to mobility procedures.
- Connection Mobility Control (CMC): the aim of CMC is to guarantee that the services are offered even when the UEs move around the geographical area. Namely, CMC oversees the management of radio resources related to idle or connected mode mobility. Cell selection

and reselection procedures: take place when the UE is in idle mode, so it does not have any established connection with a MME. Cell selection functionality allows the user to camp on a cell. Then, through the cell reselection, the UE monitors also the neighboring cells in order to determine if there is a cell with better conditions to camp. Handover (HO) procedure: takes place when the UE is in connected mode and thus it can send and receive data. To properly perform both HOs and cell reselections, the UE takes and reports measurements of the radio environment to the network in order to make the correct mobility decisions. In particular, the physical layer measurements considered in the mobility decisions are:

- Reference Signal Receive Power (RSRP) [46]: is defined as the linear average over the power contributions of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth.
- Reference Signal Receive Quality (RSRQ) [46]: is defined as the ratio $N \times RSRP/(E-UTRAcarrierRSSI$, where N is the number of RBs of the E-UTRA carrier RSSI measurement bandwidth. The measurements in the numerator and denominator shall be made over the same set of resource blocks. The E-UTRA carrier RSSI comprises the linear average of the total received power observed only in OFDM symbols containing reference symbols, in the measurement bandwidth, over the N resource blocks by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise etc.

2.2 Self-Organizing Networks

The spectacular growth that the mobile networks have experienced in the last years has led to an increase in their complexity. As a result, mobile operators have focused their attention on automating the traditional mobile tasks in order to save CAPEX and OPEX by introducing SON. In this context, the SON concept has been established as a major necessity for mobile network management. In particular, the main objectives of SON are:

- To reduce time and cost of installation processes.
- To reduce OPEX by decreasing human effort in connection with monitoring, optimizing, diagnosing, and healing of the network.
- To reduce CAPEX by efficiently using the network elements and spectrum.
- To improve user experience and service quality.
- To improve network performance by optimizing network efficiency and service quality.

These objectives are achieved by different SON algorithms, each of which is focused on the automation of a particular mobile task. All these algorithms follow the same phases: collection of network measurements; data analysis in order to identify the status of the network; decision-making process in order to determine the required action; and, finally, the execution of the selected actions over the network. Based on the source of the data, the information analyzed by the SON algorithms can be classified into the following types [7]:

- Configuration management parameters (CM): they determine the configuration of each network element.
- Performance management parameters (PM): they are counters that provide information

about the status of the network elements and the events that have taken place during the measurement period. Examples of counters are the number of dropped calls, the number of successful handovers, etc.

- Alarms: they are messages that are triggered when there is a specific failure in a network element.
- Key Performance Indicators (KPIs): they are calculated by combining several counter or measurements, in order to obtain a meaningful performance measure. An example of KPI is the handover success rate, defined as the ratio between the number of successful handovers and the total number of handovers (considering both the successful and the failed handovers).
- Drive tests: they are field measurements taken by the troubleshooting experts using specialized equipment in order to monitor the performance of the mobile network in certain areas. These measurements are localized by Global Positioning System (GPS) and they provide user-level information such as the actual coverage and interference that the UE perceived.
- Mobile traces [36] and Minimization of Drive Test (MDT) [37, 38, 39, 40]: this is also user-level information taken from specific UEs, the difference with the drive test is that this measurements are automatically collected by the network element (e.g. eNB).
- Context information: related to the environment, such as weather, geographical conditions, type of area (e.g. rural area), UE distribution (e.g. uniform), etc.

Furthermore, depending on the network element in which the SON algorithms are executed, the architecture of SON functions can be classified in: centralized; distributed; and hybrid architectures. In case of centralized SON architecture, the algorithms are run at the network management level, while in a distributed SON solution, the network nodes are in charge of executing the SON algorithm, exchanging their related messages directly with each other. Finally, the hybrid SON architecture is characterized by having part of the SON algorithm in the network management level and part in the network elements.

According to the scope of application of the SON functions, they can be categorized into three main groups [3, 7] according to the 3GPP guidelines (Fig. 2.2): Self-Configuration [9], Self-Optimization [9] and Self-Healing [10].

2.2.1 Self-Configuration

Self-Configuration is in charge of the installation and setting up of a new eNB prior to its operation, as well as, the execution of the periodical updates of network elements. The main tasks of Self-Configuration systems are (Fig.2.2):

- Hardware installation and initial configuration of the eNB, requiring minimum human intervention.
- Network Authentication: once the eNB has been physically installed, the new eNB identify the associated Network Element Manager in order to perform the authentication process.
- Software installation, transport and radio parameters setting up and network integration: the new eNB downloads the required parameter settings and software packages in order to

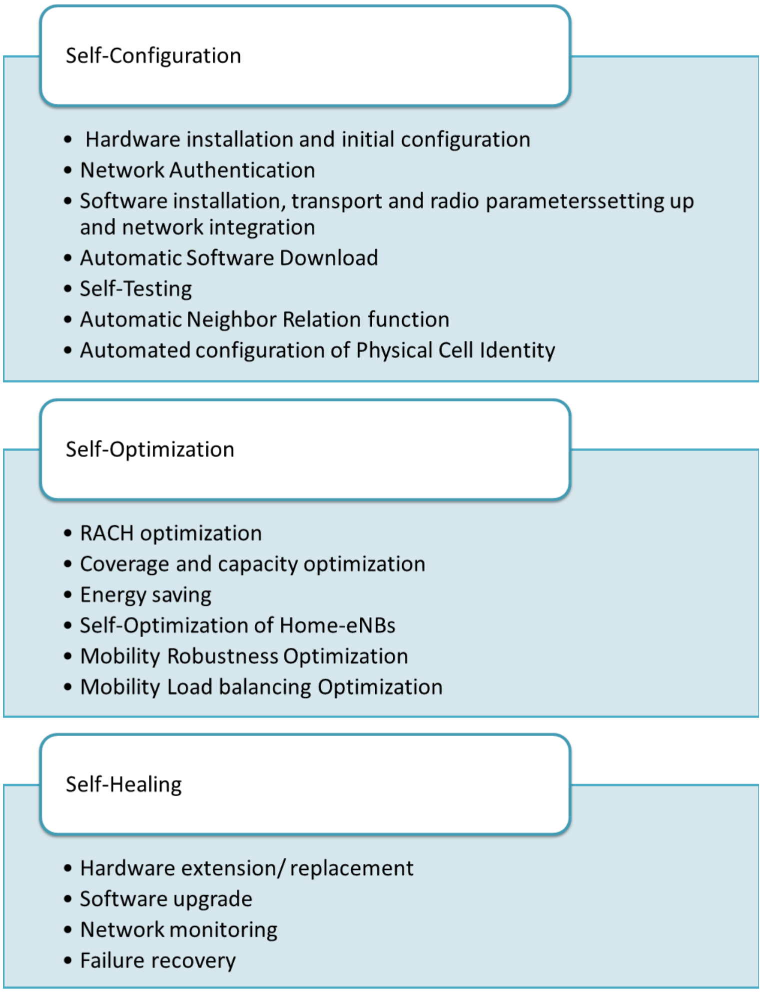

Figure 2.2: Classification of the SON use cases [2, 9, 6].

configure transport and radio parameters and update the neighbor network elements.

- Automatic Software Download: through this functionality, the eNB automatically downloads the new software packages and firmware. This process can be triggered by either the eNB or the management server.
- Self-Testing: this functionality checks the hardware and software of the eNB in order to guarantee its proper operation.
- Automatic Neighbor Relation function (ANR): it is in charge of managing the neighbor cell relations through the Neighbor Relation Table. In particular, this function adds new neighbors to the table, or removes them if they are outdated.
- Automated configuration of Physical Cell Identity (PCI): PCI is a cell identifier that determines the cell identity group. There are a limited number of possible values, so it is crucial to ensure that neighboring cells do not present the same PCI in order to avoid collision or confusion.

2.2.2 Self-Optimization

Once the eNB has been installed, Self-Optimization functionality automatically and continually configures the parameters of the systems over time in order to ensure that the network elements operate to its best level of efficiency according to the traffic profile and the network environment (i.e. topology, propagation, interference). Among the use cases of Self-Optimizing systems [2, 9, 4], the following functions can be found:

- Radio Access Channel (RACH) optimization: this function aims to continually configure the random access procedure in order to optimize the random access parameter setting and thus maximize the end-user experience. This process is very important to reduce the number of accesses failures and improve the performance of the call setup and handover procedures.
- Coverage and capacity optimization: the goal of this use case is to adjust the antenna parameters (such as antenna tilts) or the transmitter power levels in order to maximize the serving area at the same time the capacity is optimized and the inter-cell interference levels are minimized.
- Energy saving: energy consumption is directly related to operator's OPEX. Thus, its reduction involves decreasing power consumptions in the network elements. In particular, this can be achieved by temporarily shutting down unused capacity when not needed, reducing active carries for off-peak times, putting some eNB in sleep mode, etc.
- Mobility load balancing optimization: this automatic function adapts the mobility parameters in order to share traffic and share the load of congested cells with their neighbors with available resources to provide effective service for users while maintaining overall capacity.
- Mobility robustness optimization: this solution automatically detects and correct those errors and suboptimal settings in the mobility configuration, that may lead to a degradation of user performance

2.2.3 Self-Healing

Self-Healing is a continuous process by which the network elements are monitored in order to reduce the negative impact of a problem on the user performance and thus to improve the user experience. To that end, according to [6], Self-Healing covers the following use cases (Fig. 2.2):

- Hardware extension /replacement: the goal is to replace the faulty hardware of a network element that causes service outage and extend the hardware to increase the capacity, requiring minimal operator intervention.
- Software upgrade: this activity is overlapped with a Self-Configuration activity and it involves the automatic software download and automatic upgrade, minimizing software failures.
- Network monitoring: it comprises the collection and analysis of the network data (such as alarms, KPIs, user measurements and configuration parameters) with the objective of detecting faulty cells and diagnosing the fault causes (i.e. root cause). A typical fault case monitored by SH system is the cell outage.
- Failure recovery: based on the identification of a fault, this function determines the ade-

quate recovery actions to solve or compensate the problem.

Among all SON use cases, the network monitoring is the one addressed in this thesis, focusing in particular in the diagnosis phase. Its goal is to automate the troubleshooting work which still remains manual, since it requires a high degree of technical knowledge that can only be gained through many years of experience. Furthermore, they usually are operator-specific and rarely documented. During this process, troubleshooting experts often focus on the worst offender cells, that is, the top ten cells that either present the worst performance or present a great change in their performance during the observation period. Then, for those selected cell, the experts analyze all collected data in order to find the symptoms that may reveal the cause of the degradation. According to the symptoms and based on their experience and knowledge, they identify the most possible root cause, thus allowing the execution of the corrective actions. For example, a typical fault case is the cell outage, which is usually compensated by its neighboring cells while the on-site repair actions are carried out.

The automation of these time-consuming and expensive tasks by means of a SH system provides clear benefits for mobile operators:

- It reduces the time spent finding the fault cause and resolving the problem.
- It enhances the performance of the network since the degradation time is significantly minimized.
- It reduces cost since the number of troubleshooting experts is considerably reduced.

For this purpose, the combination of both network monitoring and failure recovery use cases of a SH system comprises the following functions 2.3:

- Data collection: it comprises the collection of all the required network data, such as alarms, counters, KPIs, MDT, etc.
- Fault detection: it aims to identify those cells that present performance degradations. To

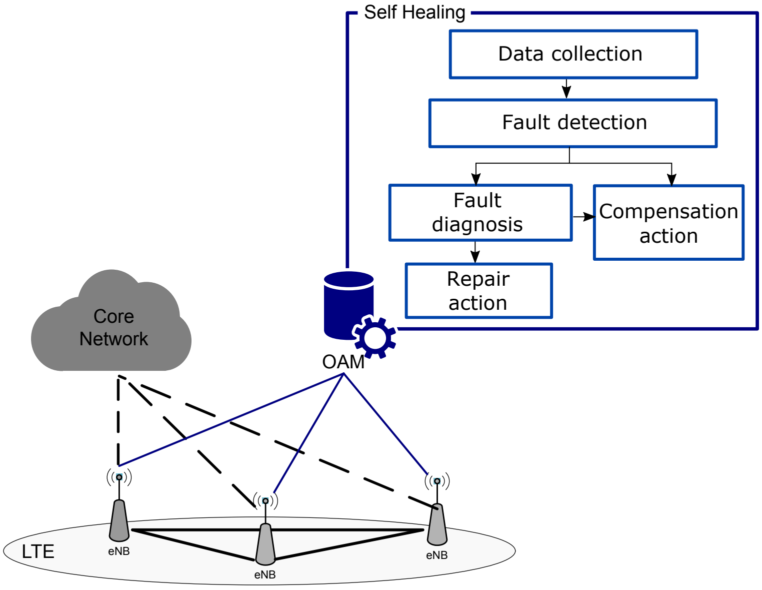

Figure 2.3: Phases of a Self Healing system in LTE network.

that end, it is first necessary to define the most significant performance indicators and model their normal and expected behavior (e.g. by defining a threshold). Then, the detection system automatically identifies the degraded cells by comparing the selected indicators with their normal profile.

- Fault diagnosis: this function is triggered by the detection system when a deteriorated cell has been found. In particular, it analyzes the rest of alarms, indicators and measurements not considered in the detection phase in order to identifying the root cause of the problem. In a mobile network, there is a wide range of problems that can cause performance degradation and they can be categorized in: Hardware faults, Software faults, configuration faults, environmental changes, etc. [4].
- Repair action: based on the identification of a fault, this function determines the adequate actions to solve the problem. These repair actions may be executed remotely (e.g. by modifying a parameter or rebooting a system) or locally (by replacing an equipment).
- Compensation action: in contrast with the previous function, the compensation system aims to improve the user experience and minimize the degradation of the cell performance. The aim is to reduce the effects of a fault by configuring both the faulty cell and its neighbors while the fault is solved by the repair system. A typical example of this functionality is the cell outage compensation.

Chapter 3

Research outline

The first part of this chapter presents a concise description of the research papers of this thesis and the connection among them. In particular, these papers are summarized, justifying their thematic unit and indicating their objectives and contributions. The second part of the chapter is devoted to the research methodology. Therefore, in that section, the techniques, equipment and tools used in this thesis are presented.

3.1 Description of the publications

Each of the papers composing this thesis addresses at least one of the previously presented challenges. Fig. 3.1 summarizes the scenario and the relation of the papers to the challenges and objectives, indicating how they are organized in this report. In particular, this figure presents each paper as a unique block, specifying for each one: the achieved objectives, the used data sources; and, the interaction among the papers.

1 Methodology for the Design and Evaluation of Self-Healing LTE Networks [I] (Chapter 4.1).

In order to accomplish all objectives of this thesis, the starting point is to address all those challenges related to the design and evaluation of Self-Healing systems. This part is essential to meet the rest of the objectives since it provides the basis to devise the systems presented in this thesis and assess their performance, as well as to compare them to other state-of-the-art techniques. Therefore, the first paper of this thesis defines an approach to design and evaluate any SH functionality. In particular, it proposes to do a feasibility study of the systems in a controlled environment before being developed in a real network. To that end, this paper also presents a Fault Model that describes at implementation level typical faults of an LTE network related to the main categories (coverage, mobility and interference). Furthermore, it also details how the most relevant KPIs for the identification of those faults should be calculated in any LTE simulator. As a result, researchers can apply the proposed Fault Model to their LTE RAN simulators in order to artificially test their

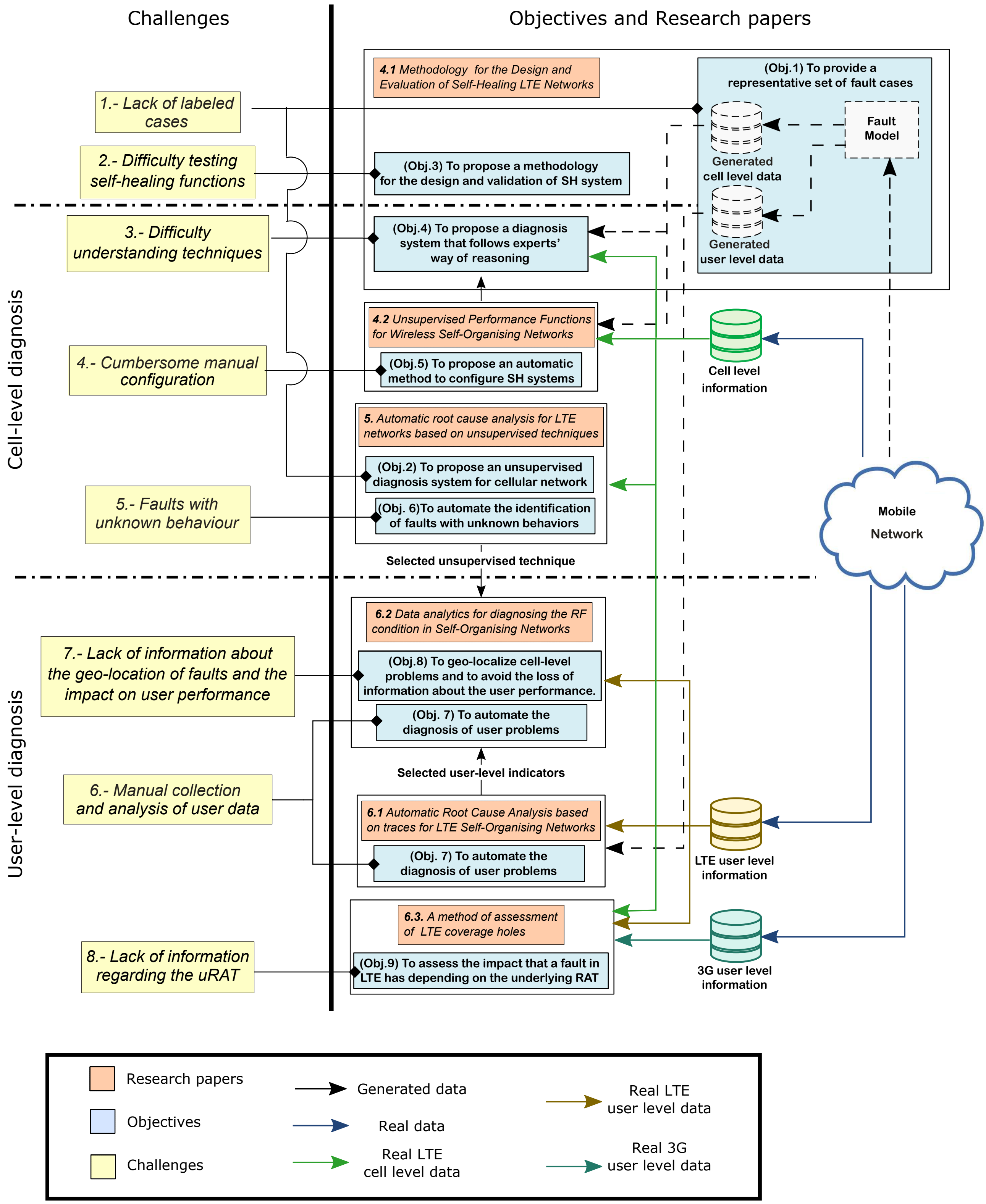

Figure 3.1: Challenges, Objectives and Research papers.

self-healing systems. Then, as a second important contribution of this paper, a diagnosis system that reasons in a similar way to cellular network troubleshooting experts have been devised, by following the methodology proposed in the first part of the paper. This system identifies the cause of the problems by applying a set of fuzzy rules to the KPIs. The main advantage of using fuzzy logic is that troubleshooting experts can easily understand and interpret the results and the behavior of the system. In addition, evaluation results show that the defined fuzzy system provides fault identification with high success rate. This system and the Fault Model have been tested over a dynamic system-level LTE simulator, emulating the behavior of the network in the presence of faults and thus generating an artificial dataset composed of labeled cases. In particular, in this paper, two datasets have been obtained, one using a hexagonal scenario and one using an irregular scenario which is based on a real LTE network. Those datasets are shared online [47] in order to cover the lack of labeled cases and allow research community to replicate and compare the experiments. Furthermore, some examples of how the diagnosis system works in a live network have been discussed.

2 Unsupervised Performance Functions for Wireless Self-Organizing Networks [II] (Chapter 4.2).

Once the objective to develop and assess Self-Healing functionalities have been accomplished, the next step is to avoid the cumbersome manual configuration of those systems. Therefore, the second paper in this thesis is devoted to automating the calculation of the configuration parameters. It is worth noting that those parameters depend on the behavior and characteristics of the indicators, since SH systems analyze the KPIs in order to detect problems and identify their cause. As a result, in this paper, an unsupervised approach to model the behavior of the KPIs (i.e. their profile under normal and abnormal conditions) is presented. More specifically, the indicators are modeled by means of a pair of functions (named performance functions) that are easy to understand by troubleshooting experts. To that end, the performance functions are automatically designed using the statistical behavior of real data through a low computational complexity algorithm. This facilitates adaptation of the performance functions as the characteristics of the network change over time, without any significant cost. The reliability and effectiveness of the proposed method has been assessed comparing the results with reference approaches.

3 Automatic root cause analysis for LTE networks based on unsupervised techniques [III] (Chapter 5).

At this point, the next objective is to design a diagnosis system that can work without labeled training cases. In addition, the aim is to automatically identify new patterns of faults whose behaviors are neither known nor documented. Therefore, the next research work aims to design a diagnosis system with the most appropriate unsupervised methods. During this study, different unsupervised techniques have been examined, studying in detail the performance and characteristics of both unsupervised Bayesian Networks [48] and Self-Organizing Maps (SOM) [49]. The research work on unsupervised Bayesian Network has been presented in an international conference [XV] while the diagnosis system based on SOM is detailed in this paper [III]. The main reason of selecting Self-Organizing Maps (SOM) was that this technique does not require the discretization of the inputs (unlike the unsupervised Bayesian Network) which leads to a loss of information that is crucial for a successful diagnosis. As a result, the core of the diagnosis system presented in this paper is based on a SOM. In particular, this method is used to automatically model the behavioral patterns of the data. Then, the Ward's Hierarchical method [50] is used to group those patterns into clusters based on their similarity. Furthermore, in order to avoid the user intervention, an automatic approach to automatically identifying the total number of clusters that combines the Davies-Bouldin index [51] and the Kolmogorov-Smirnov test [52] is presented. Then, the clusters are labeled based on their statistical behavior associating each one to a specific fault. Moreover, with the aim of increasing the accuracy of the system, a novel adjustment process based on the so-called Silhouette index [53] is presented to refine the diagnosis solution. Finally, the effectiveness of the developed diagnosis system is tested with the Fault Model defined in Chapter 4.1, analyzing its complexity and comparing its performance with reference mechanisms. In addition, this system has been put into practice in a real LTE network, demonstrating its usability and effectiveness.

4 Automatic Root Cause Analysis based on traces for LTE Self-Organizing Networks [IV] (Chapter 6.1).

Automating the troubleshooting tasks by means of just cell-level information does not allow an evaluation of the problems that directly affect each user's performance. Therefore, so far the collection and analysis of user-level measurements remain manual, through traditional drive tests. In this paper, this challenge is addressed, proposing a framework to automate the diagnosis of each unintended service disconnection through the analysis of the mobile traces in order to determine the cause of the connection's release. The proposed method is characterized by a top-down model that identifies the nature of the release (normal, access failure, abnormal release ...) and defines the cause-indicator relations required to determine the RF conditions of the user at that moment. Furthermore, in this study, the proposed RF classification, implemented with a rule-based system, has been evaluated in a simulated LTE network and then applied in a real LTE network illustrating its benefits and potentials.

5 Data analytics for diagnosing the RF condition in Self-Organizing Networks [V] (Chapter 6.2).

Assessing the RF conditions in each area of the mobile network is a task traditionally done by drive testing. Therefore, in this paper, a new methodology to automate this process by means of mobile traces is proposed. To that end, each user is used as a sampler of the air interface, automatically collecting its information through the mobile traces instead of using the traditional equipped vehicle. Their RF condition is analyzed using the user-level indicators presented in Chapter 6.1 along with the SOM-based system proposed in Chapter 5 which allows the identification of unknown RF conditions. Then, the RF conditions of the users of a given cell are aggregated in order to determine the overall performance of that cell. Furthermore, the aggregation method takes into account the geo-location of the users to quantify the severity of the RF problems based on their concentration. This allows the detection of the problems even if they only affect a small area within a cell. In addition, this paper presents a new way of visualizing the RF condition at each point of the network in one single map, thereby avoiding the individual analysis of one chart for each indicator.

The proposed system has been assessed in two different real LTE networks, comparing the results with a state-of-the-art diagnosis system using OAM KPIs. The results show that this system achieves a more efficient way of identifying RF issues, reaching areas that cannot be covered by drive testing.

6 A method of assessment of LTE coverage holes [VI] (Chapter 6.3).

The aim of this paper is to address the lack of information about the underlying RAT when analyzing the symptoms of the problems in a LTE network. This lack of information is particularly meaningful in the case of coverage holes, since they are characterized by causing different impact on the users depending on whether or not an uRAT is available to maintain the service released by the LTE network. Therefore, in order to address such challenge, this paper proposes a method capable of analyzing the performance of the users both in LTE and in the underlying RAT at the same time. In particular, this approach detects cells with coverage holes, diagnosing their type and severity. To that end, it performs an Inter-technology follow-up of those users that leave LTE technology to continue their services in the underlying RAT and then quantifies the effect of the coverage hole by means of a new inter-technology indicator estimated from the mobile traces of both RATs. The main advantage of classifying the coverage hole based on their impact is that the expert can design their particular remedial action, making them case-specific. The proposed system has been validated using data from a live LTE network and its co-located 3G network, showing its effectiveness in detecting coverage holes and diagnosing their type.

3.2 Research Methodology

The studies carried out in this thesis were conducted throughout a structured research methodology, characterized by an engineering-oriented approach. In particular, this research methodology is divided into the following steps (Fig. 3.2):

- Background review: The work started with a survey of the state-of the art related to self-healing in cellular networks. In this way, the main challenges that needed to be confronted to facilitate progress in the area of self-healing systems were identified. This literature review was particularly focused on the diagnosis functionality and the existing automatic techniques. This research phase also included the study of the most typical faults of LTE networks along with the identification of the most appropriate performance indicators for their diagnosis. Since typical faults, KPIs and their relation is not documented in scientific literature, this study had to be carried out in close cooperation with operators and vendors. Furthermore, over the thesis duration, in order to address the nine challenges, several techniques from the fields of statistics, data analytics and artificial intelligence, which had not been previously applied to the problem under study, had to be studied.
- Problem formulation: The next step involved the formulation of the problem for each identified challenge, defining what issues had to be investigated and looking for the most adequate techniques to solve them.
- Data preparation: this phase is further divided into two subtasks depending on whether the data is obtained from the real network or generated by a simulator.

- Real data preparation: This subtask consists of collecting the raw data from the network elements of either LTE or 3G networks. In particular, the data collected for this thesis is comprised of the configuration parameters (CM) of the network element, the performance measurements (PM) that monitor the network performance and the mobile traces (i.e. LTE traces and 3G traces) that contains the signaling messages (i.e. events) exchanged between the network elements and the radio measurements that reflect user performance. Then, the raw data has been processed by means of the official Ericsson tools: Trace Processing Server (TPS) has been used to decode the mobile traces in order to obtain the user-level data, while OSS Data Gateway (ODG) and Ericsson RAN Analyser (ERA) have been used to calculate the performance indicators of the cell from CM and PM. Then, the processed data have been stored in databases using MySQL as database management system.
- Simulated data generation: in order to facilitate the analysis and evaluation of the systems, a suitable controlled environment is needed. To that end, in this thesis, the mobile network has been modeled and implemented over a dynamic-system level simulator [54] developed in MATLAB. During the start-up phase, the main configuration parameters are initialized, the scenario along with the users is created and the power measurements (i.e. propagation losses, interferences...) are calculated. The scenario models a macro-cellular environment, where the number of cells and their position are configurable in the simulator. After that, in each simulation step, the radio resources are managed and shared according to the request of the users as they moved along the scenario. At the end of each simulation loop (i.e. a group of several simulation steps), the performance indicators, the statistics and the user-level measurements are estimated. This simulator had to be upgraded during this thesis to include the main faults and related KPIs, as well as the proposed algorithms.

- System Design: during this phase different techniques to overcome each challenge were analyzed and the best solution was devised and implemented.
- Evaluation and conclusions:
  - Validation with simulated data: the designed systems were assessed, evaluating their performance with simulated data and comparing their diagnosis error rate with baseline systems.
  - Demonstration in real networks: all the proposed systems were also tested in live networks, showing their effectiveness in the diagnosis of problems. The automatic diagnoses were compared with those obtained by troubleshooting experts through the manual analysis of the data with the ERA tool.
- Dissemination: Finally, the most relevant results of the research have been published in high impact journals and in both international and national conferences.

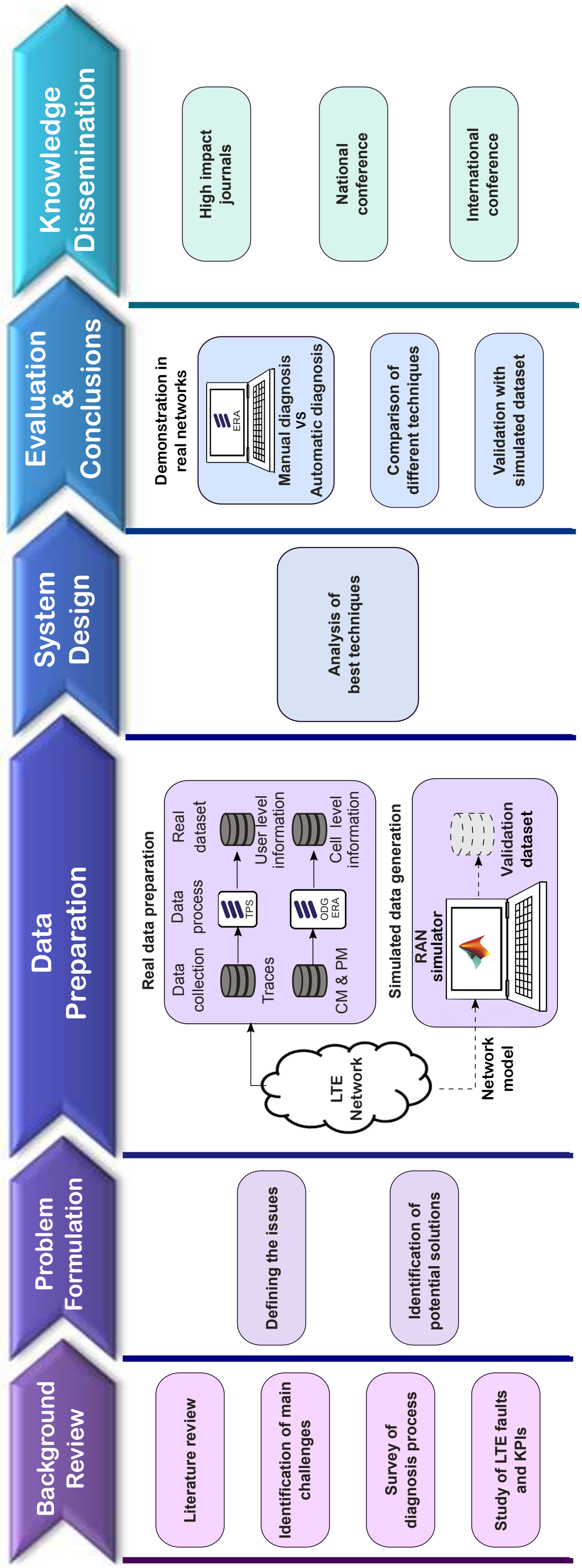

Figure 3.2: Research Methodology.

Chapter 4

Methodology for the design and evaluation of SH systems

Paper I:

A. Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano and R. Barco, "Methodology for the Design and Evaluation of Self-Healing LTE Networks" *IEEE Transactions on Vehicular Technology*, DOI 10.1109/TVT.2015.2477945, Published Online 2015

Abstract: Self-Healing networks aim to detect cells with service degradation, identify the fault cause of their problem and execute compensation and repair actions. The development of this type of automatic system presents several challenges to be confronted. The first one is the scarce number of historic reported faults that greatly complicates the evaluation of novel self-healing techniques. For this reason, in this paper, a system model to simulate faults in LTE networks along with their most significant Key Performance Indicators (KPIs) is proposed. Secondly, the expert knowledge required to build a Self-Healing system is usually not documented. Therefore, in this paper, a methodology to extract this information from a collection of reported cases is proposed. Finally, following the proposed methodology, an automatic fuzzy logic-based system for fault identification in LTE networks is designed. Evaluation results show that the fuzzy system provides fault identification with high success rate.

Paper II:

A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano,"Unsupervised Performance Functions for Wireless Self-Organizing Networks" *Wireless Personal Communications*, DOI: 10.1007/s11277-016-3435-1, Online 2016

Abstract: Traditionally, in cellular networks, troubleshooting experts have manually analyzed Key Performance Indicators (KPI), so that they could identify the cause of problems and fix them. With the emergence of Self-Organizing Networks (SON), Self-Healing systems are designed to automate those troubleshooting tasks. With that aim, the behavior of the KPIs (i.e. their profile under normal and abnormal conditions) needs to be modeled. Since the behavior of the KPIs is network-dependent and it changes as the network evolves, their profile should be automatically defined and readjusted depending on the characteristics of the network. Therefore, in this letter, an automatic process to model the KPIs based on the real data taken from the network is proposed. In particular, this method is characterized by designing a pair of functions (named performance functions) from the statistical behavior of real data without requiring any information about the existence of faults (i.e. unsupervised learning). Results have shown the reliability and effectiveness of the proposed method in comparison to reference approaches.

Chapter 5

Automatic diagnosis system based on unsupervised techniques

Paper III:

A. Gómez-Andrades, P. Muñoz, I. Serrano and R. Barco, "Automatic root cause analysis for LTE networks based on unsupervised techniques" *IEEE Transactions on Vehicular Technology*, vol. 65 , pp. 2369 - 2386, DOI: 10.1109/TVT.2015.2431742, 2016

Abstract: The increase in the size and complexity of current cellular networks is complicating their operation andmaintenance tasks. While the end-to-end user experience in terms of throughput and latency has been significantly improved, cellular networks have also become more prone to failures. In this context, mobile operators start to concentrate their efforts on creating self-healing networks, i.e., those networks capable of troubleshooting in an automatic way, making the network more reliable and reducing costs. In this paper, an automatic diagnosis system based on unsupervised techniques for Long-Term Evolution (LTE) networks is proposed. In particular, this system is built through an iterative process, using self-organizingmaps (SOMs) and Ward's hierarchical method, to guarantee the quality of the solution. Furthermore, to obtain a number of relevant clusters and label them properly from a technical point of view, an approach based on the analysis of the statistical behavior of each cluster is proposed. Moreover, with the aim of increasing the accuracy of the system, a novel adjustment process is presented. It intends to refine the diagnosis solution provided by the traditional SOM according to the so-called silhouette index and the most similar cause on the basis of the minimum Xth percentile of all distances. The effectiveness of the developed diagnosis system is validated using real and simulated LTE data by analyzing its performance and comparing it with reference mechanisms.

Chapter 6

Automatic diagnosis systems based on mobile traces

Paper IV:

A. Gómez-Andrades, R. Barco, I. Serrano, P. Delgado, P. Caro-Oliver and P. Muñoz, "Automatic Root Cause Analysis based on traces for LTE Self-Organizing Networks" *IEEE Wireless Communication*, In Press 2016

Abstract: Within the functionality of self-healing in self-organizing networks, automatic root cause analysis is typically focused on identifying problems at the cell level, based on the statistics gathered by the operation, administration, and maintenance system. Therefore, mobile operators lose visibility of the problems that directly affect users' performance. Conversely, this article presents a complete strategy to systematically identify, on the basis of information at the user level (by means of mobile traces), why user disconnection occurs (i.e., the cause of the connection's release). The proposed automatic root cause analysis is characterized by a top-down model and provides a comprehensive classification of faults when they are caused by radio-related problems. First, it classifies the connections according to the type of release, and subsequently, it determines the specific fault cause based on the event information for those connections abnormally released. The proposed method for identification of the radio-related cause has been applied in both an LTE simulator and a real LTE network, illustrating the usefulness of the approach.

Paper V:

A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano, "Data analytics for diagnosing the RF condition in Self-Organizing Networks" submitted to *IEEE Transactions on Mobile Computing*, (Minor revision) 2016

Abstract: The current trend in the management of mobile communication networks is to increase the level of automation in order to enhance network performance while reducing Operational Expenditure (OPEX). In this context, the 3rd Generation Partnership Project (3GPP) has presented different solutions. On the one hand, Self-Organizing Networks (SON) include self-healing capabilities, which allow operators to automate their troubleshooting tasks in order to identify and solve the problems of the network. On the other hand, the use of mobile traces or Minimization of Drive Tests (MDT) are proposed to automate the collection of user's measurements and signalling messages. This paper proposes to combine both solutions, SON and traces, with the purpose of quickly detecting and solving issues related to the radio interface. That is, the user information gathered by the cell traces function is used to perform an automatic diagnosis of the RF condition of each cell. In addition, the proposed approach allows to precisely locate RF problems based on the assessment of the RF condition. Mobile traces constitute large sets of data, whose analysis requires the application of big-data analytics techniques. The proposed system has been evaluated in two different live LTE networks, demonstrating its validity and utility.

Paper VI:

A. Gómez-Andrades, R. Barco and I. Serrano, "A method of assessment of LTE coverage holes", submitted to EURASIP Journal *on Wireless Communications and Networking*, 2016

Abstract: Detecting coverage holes in mobile networks is still an important problem that needs to be addressed, mainly in Long-Term Evolution (LTE) networks, which have been recently deployed. In particular, those areas suffering from lack of coverage present a high concentration of dropped connections. However, in current multi-RAT network architecture the coexistence of different Radio Access Technologies (RAT) saves lots of connections by redirecting them to an underlying RAT through an inter-RAT handover procedure. As a result, the impact of LTE coverage holes is totally different in multi-RAT networks than in single RATs. Thus, this paper presents a method and a new metric to classify different kinds of coverage holes, which can require different remedial actions. The proposed solution has been validated in a live LTE network, showing its effectiveness in analyzing areas with coverage problems.

Chapter 7

Conclusions

This final chapter summarizes the research carried out in this thesis, presenting the major findings of the conducted studies. That is, the main contributions of each paper are highlighted. In addition, some future lines of work in the covered topics are suggested.

7.1 Contributions

This thesis report has dealt with different use cases in the field of Self-Healing networks. To that end, a thorough survey to identify the main challenges that prevent the development of this kind of systems has been carried out. As a result, this thesis targeted nine objectives, which tried to overcome the identified challenges. Thus, following the overall structure of this report, the contributions are separately presented according to the initially defined objectives.

- **To propose a methodology for the design and validation of SH systems (Obj.3).**
  - At a first stage, an entire framework to design and evaluate any self-healing functionality has been proposed. This method encompasses three tasks: 1) the analysis of the KPI behavior in order to statistically model their features and to determine the cause-KPI relations; 2) the feasibility study of the system to compare different automatic techniques over a controlled environment; and, 3) the implementation of the system in a real network using its particular data.
  - In addition, the most typical faults in cellular networks and the most common KPIs have been modeled, explaining their characteristics and indicating how they can be simulated. In particular, this Fault Model describes a short but specific set of faults related to the main categories in cellular networks (i.e. mobility, coverage and interference). In addition, the most appropriate KPIs to analyze those faults are modeled. It is also worth mentioning that it is not a closed model; rather it admits more faults and KPIs. As a result, by implementing this Fault Model, all the self-healing functionalities can be evaluated together under the same scenario, including both cell-level or user-level SH systems.

  - A highlight of the proposed Fault Model is the possibility to be implemented in any RAN simulator. Therefore, this thesis also describes a practical way of incorporating the Fault Model in any simulator.
  - Following the proposed practical implementation, this Fault Model has been implemented in a dynamic system-level simulator. This allows the generation of faults over some cells of the scenario monitoring their impact on the affected cell and its neighbors through the proposed KPIs. First, the Fault Model was put into practice using a regular and hexagonal LTE scenario to evaluate the results in a control environment. Then, an irregular scenario was configured using the parameters and characteristics of a real LTE network of an urban area. This experiment reveals that the statistical behavior of the faults is similar in both scenarios.

- **To provide a representative set of fault cases (Obj.1).**
  - As indicated throughout the report, the lack of labeled cases is one of the main challenges that need to be confronted in the design of SH systems. Thus, in this thesis, particular emphasis has been placed on providing a reliable dataset. As a result, the Fault Model implemented in a dynamic system-level simulator using a regular and simple LTE scenario has been used to generate a significant and useful database. Thanks to that, an exhaustive dataset has been collected to artificially prove self-healing functionalities. In addition, it has been made available online in [47] to facilitate the study, replication and comparison of future experiments.
  - Using the proposed Fault Model implemented in this simulator, user-level information has also been gathered. As a result, an artificial dataset with the user-level information for the modeled faults has been obtained. This has allowed the analysis of the diagnosis system at user-level.
- **To propose a diagnosis system that follows experts' way of reasoning (Obj.4).**
  - In order to overcome this challenge, a diagnosis system based on fuzzy rules has been proposed in this thesis along with a robust scheme to build it. The main benefit of this technique is that it follows the reasoning and human thinking. This facilitates its interpretation and understanding.
  - The proposed system has been assessed in the simulator using both the regular and the irregular scenarios. Thus, the performance of the system has been studied, demonstrating that this system, besides being simple, provides a high success rate and even better that the traditional rule-based diagnosis. Furthermore, it has also been particularized for some cells, proving that the proposed diagnosis system can be configured either for the whole network or on a cell basis. Finally, the studies carried out in this thesis have gone one step further and the proposed system has been successfully applied in a live network.
- **To propose an automatic method to configure SH systems (Obj.5).**
  - Even though self-healing systems are automatically designed, they normally require some configuration parameters. In particular, some parameters should be configured in relation to the characteristics of the KPIs, e.g. when a KPI is considered normal or abnormal. Therefore, another contribution of this thesis is the automation of this con-

figuration process. Namely, a new method to model the features of the KPIs through a pair of performance function has been proposed. This method automatically designs the performance functions using real data from the network where the diagnosis system is going to be deployed, in an unsupervised manner and thus preventing expert intervention. Furthermore, its low computational cost and easiness of being implemented make this method an ideal process to use online for the configuration and reconfiguration of the Self-Healing systems as the network evolves.

  - The proposed approach has been evaluated using the fuzzy-based diagnosis system proposed in this thesis and the automatic diagnosis systems available in the literature. The results have demonstrated that the utilization of this performance functions improves the accuracy of the diagnosis compared to the previous approaches.

- **To propose an unsupervised diagnosis system (Obj.2). To automate the identification of faults with unknown behavior (Obj.6).**

  The manual analysis of performance data to find new faults is one of the most time consuming tasks for troubleshooting experts. As a result, this thesis contributes to overcome this challenge by proposing a new system based on unsupervised techniques. Therefore, the proposed approach automatically identifies faults whose behavior may be previously unknown without requiring any prior information. In particular, in this context, several contributions can be emphasized:

  - When analyzing a particular dataset, the first unknown is the total number of clusters, i.e. the number of fault causes. Thus, this thesis proposes an unsupervised method that combines the Kolmogorov-Smirnov test with the Davies-Bouldin index in order to automatically determine the number of clusters that are statistically different in the dataset.
  - In order to be trained with unlabeled cases and to automatically identifying unknown patterns, this thesis proposes the use of Self-Organizing Maps to model the found patterns through its neurons and the use of Ward's Hierarchical method to group the neurons of SOM into as many clusters as detected based on their similarity. As a result, a system capable of working with unlabeled cases and of diagnosing a group of previously unknown patterns has been designed.
  - Once a group of patterns have been identified, a procedure to statistically analyze their behavior is proposed. Thus, the experts only have to study those behavioral patterns instead of the whole dataset. By performing this analysis, the expert can determine the relevance of each pattern and assign it a meaningful name. This ensures the validation of the clustering by the experts by determining the relevance of each pattern. Consequently, this knowledge is incorporated into the system, guaranteeing its autonomy during the exploitation stage.
  - Another important contribution of this thesis is the inclusion of an adjustment process to ensure accuracy in the final diagnosis. In particular, this adjustment phase is focused on improving the diagnosis of those faulty cells whose diagnosis is not clear because its behavior is just between two or more patterns. Therefore, the proposed approach analyses the worst-cases and determines which fault is the most similar one among the possible options based on the Silhouette index.

  - The whole diagnosis system has been evaluated using the simulated dataset generated in this thesis report. As a result of the assessment, it can be asserted that another valuable part of the approach is its high accuracy, which is better than the reference mechanisms.
  - The complexity of the iterative procedures used in the design stage is also discussed, showing that the unsupervised clustering is the most complex phase although it is only executed over a few iterations. Furthermore, the results with the simulated dataset also prove the low run time of the whole system.
  - This thesis report also demonstrates the operation of the whole system in a real network. First, the construction of the system using real data is explained step-by-step. As a result, the identified clusters are analyzed in-depth providing its main characteristics. Then, the designed system has been used to automatically diagnose problematic cells over time, showing how the diagnosis evolves over time as the characteristics or the configuration of the cell change.

- **To automate the diagnosis of user problems (Obj.7).**

  In order to accomplish this objective, in this thesis report a complete strategy to automate the collection and analysis of the user data has been proposed, resulting in several meaningful contributions:

  - First, this approach proposes to capitalize on the mobile traces and the Minimization of Drive Test (MDT) for automating the collection of the user-level information. In particular, a detailed study has been performed, determining the most important events and measurements to analyze the user's disconnection.
  - Second, a top-down model to automatically identify the reason of the user disconnection is proposed, distinguishing between problematic and non-problematic disconnections. Furthermore, in this last case, the method classifies the abnormal disconnections in access failure or dropped connection depending on the particular connection's stage where the issue lies. Thanks to this new approach plenty of details about the analyzed connection can be provided.
  - Third, for those RLF or connections that dropped due to bad RF conditions, a complete technique to automate the diagnosis of the specific RF causes is also presented since this information is not reported in the mobile events. In this study, the necessary knowledge to build an automatic diagnosis system of RF causes has been provided. Thus, a survey of the available methods to automate the process has been presented along with the required indicators and the set of RF causes that can be identified. In particular, those RF causes have been analyzed in detail, determining how they can be identified through the proposed set of user's indicators. As a result, using this method, the cause of the release can be classified as coverage hole, cell edge, lack of dominant, mobility problems or interference. It is important to highlight that a big benefit of the proposed strategy is that only requires a few standard indicators to determine the RF problem.
  - Finally, the root cause analysis to determine the radio-related cause has been evaluated through a rule-based system both in an LTE simulator and in live LTE network.

The results reveal the benefits of identifying and geo-locating the radio causes that adversely affect each connection. In addition, they also prove the coherence of the RF classification.

- **To geo-localize cell-level problems and to avoid the loss of information about the user performance. (Obj.8).**

  The diagnosis of a problematic cell on the basis of the traditional OAM KPIs does not allow to identify the geo-location of the faults and to know the impact on the users. This problem is particularly relevant when attempting to assess the RF condition of a cell. Thus, to face this challenge, this thesis proposes the use of mobile traces to diagnose cell-level problems. To that end, a complete approach to automate the identification of the RF conditions of a cell without losing the location information has been proposed. More deeply, the main contributions of this thesis in this context are the following:

  - An unsupervised approach to automatically analyze the user's performance from the information reported in the mobile traces has been presented. In particular, this method proposes the use of unsupervised systems to identify unknown RF conditions patterns. Furthermore, among the available unsupervised systems, SOM has been used to do the clustering because: it allows the visualization of the results based on the input variables; it allows the identification of unknown patterns without requiring the number of clusters in advance; and it can work with missing data.
  - A novel procedure to assess the RF problems of a cell based on the RF conditions diagnosed for each of its users is proposed. Therefore, this method proposes the aggregation of the user-level diagnosis in order to determine whether or not the cell has RF problems, and in that case, indicate which one is the most relevant.
  - The diagnosis of RF problems that only affect localized small areas has also been addressed in this thesis. In particular, in this scenario, the number of affected users is relatively low compared to the total number of users in the cell, but they have the characteristic of being concentrated in the problematic area. As a result, the proposed method identifies RF issues based on the level of concentration that the impacted users present. This allows the detection of densely concentrated problems, thus differentiating them from those problematic users that are sparse and isolated.
  - Another contribution is the RF condition maps that show in a single chart the RF condition at each point of a network. A highlight of this map is its readability since it directly shows the RF condition in each area preventing experts to individually analyze the five typical maps (one for each user measurement: $RSRP_{serving}$, $RSRQ_{serving}$, $RSRP_{strongest}$, NumCells, $TA_{Relative}$) in order to manually combine that information and identify the whole RF condition. As a result, this map facilitates the RF checking process reducing the time and effort required for troubleshooting.
  - The whole system has been assessed in two live networks proving its value in checking the RF conditions of a cell and demonstrating its validity and effectiveness with real data. In addition, the results have been compared to those obtained with a state-of-the-art diagnosis system, concluding that the proposed systems not only provides lower total error rate but also allows to localize RF problems and diagnose those that are concentrated in a small area.

- **To assess the impact that a fault in LTE has depending on the underlying RAT (Obj.9).**

  Considering only the information of LTE networks, it is not possible to discern the impact that a specific fault has on the whole mobile system. Therefore, the contributions made in this thesis to tackle this problem are the followings:

  - A method to combine data gathered from LTE with data from the co-existing uRAT has been proposed. In particular, the mobile traces are used to analyze the performance of the users when they leave LTE technology to continue their session in the co-existing uRAT. Through this method, an inter-technology track of the user is performed providing its inter-technology event flow.
  - Another key contribution is an inter-technology metric that estimates the active time that the users are on LTE, so it is calculated at user-level based on the proposed inter-technology event flow and then aggregated at cell-level to determine the overall impact.
  - The proposed metric has been used to design the detection and diagnosis phases of a self-healing system. The main benefit is its ability to identify coverage holes and classify them depending on their impact on LTE and the co-existing uRAT. This classification is very useful and necessary, since the compensation or recovery action has to be specific for each situation.
  - This system has been assessed in a live mobile network where LTE coexists with 3G. The obtained results show the success of the system in detecting and diagnosis coverage holes.

7.2 Future work

Possible lines of research that might continue the work in this thesis are the following:

- The diagnosis system automatically designed through unsupervised techniques can be translated into a set of fuzzy rules. As a result, the unknown patterns automatically identified can be included in the fuzzy system proposed in this thesis, making it easier to interpret.
- Within Self-Healing systems, the compensation and the recovery are also important phases that should be further studied. Traditionally, the research of those systems has been mainly focused on outage problems. Therefore, the proposed methodology and its Fault Model can be used to investigate both compensation and recovery systems to solve more variety of problems. Furthermore, the proposed Fault Model can be expanded with more kind of faults and KPIs.
- The diagnosis systems proposed in this thesis identify the behavior of the cells through the analysis of performance indicators. However, there are disturbance variables that influence the behavior of the cell, causing a deterioration of the performance indicators which is not due to a problem. For example, the weather typically alters the behavior of the cells so it should be taken into account in order to improve the automatic diagnosis. As a result, the proposed diagnosis system can be further enhanced by including more information apart

from the typical performance indicator (such as weather, geographical features, etc.).

- The common basis of the topics addressed in this thesis is the automation of every tasks involved in troubleshooting process. However, the choice of the most appropriate KPIs has been manually done based on expert knowledge and the features of the problems to be diagnosed. Therefore, the automation of the KPIs selection among all existing indicators is a meaningful line of investigation.
- With the technological development and the emergence of machine to machine (M2M) paradigm, the design of Self-Healing systems in this heterogeneous environment becomes a necessity. Therefore, as a future work, this study can be extended to adapt the proposed methodology and the techniques to design diagnosis systems for M2M scenarios.

7.3 Publications and projects

The following subsections present the publications related to this thesis.

7.3.1 Journals

Publication arising from this thesis

	Publication arising from this thesis	IF	Journal Rank
I	A. Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano and R. Barco, "Methodology for the Design and Evaluation of Self-Healing LTE Networks" *IEEE Transactions on Vehicular Technology*, Online 2015	2.243	Q1 (14/82) Telecommunications
II	A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano, "Unsupervised Performance Functions for Wireless Self-Organizing Networks" *Wireless Personal Communications*, Accepted 2016	0.701	Q4 (63/82) Telecommunications
III	A. Gómez-Andrades, P. Muñoz, I. Serrano and R. Barco, "Automatic root cause analysis for LTE networks based on unsupervised techniques" *IEEE Transactions on Vehicular Technology*, vol. 65 , pp. 2369 - 2386, 2016	2.243	Q1 (14/82) Telecommunications
IV	A. Gómez-Andrades, R. Barco, I. Serrano, P. Delgado, P. Caro-Oliver and P. Muñoz, "Automatic Root Cause Analysis based on traces for LTE Self-Organizing Networks" *IEEE Wireless Communication*, Accepted 2015	4.148	Q1 (3/82) Telecommunications
V	A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano "Data analytics for diagnosing the RF condition in Self-Organizing Networks" *IEEE Transactions on Mobile Computing*, Under review (Minor revision) 2016	2.456	Q1 (11/82) Telecommunications
VI	A. Gómez-Andrades, R. Barco and I. Serrano "A method of assessment of LTE coverage holes", EURASIP Journal *on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic

Publication related to this thesis

	Publication related to this thesis	IF	Journal Rank
VII	E. J. Khatib, R. Barco, A. Gómez-Andrades, P. Muñoz and I. Serrano, "Data mining for fuzzy diagnosis systems in LTE networks" *Expert Systems with Applications*, 42 (21), 7549-7559, 2015	2.981	Q1 (19/130) Computer science, artificial intelligence
VIII	E. J. Khatib, R. Barco, A. Gómez-Andrades and I. Serrano, "Diagnosis based on genetic fuzzy algorithms for LTE Self-Healing" *IEEE Transactions on Vehicular Technology*, vol. 65 (3), pp. 1639 - 1651, 2016	2.243	Q1 (14/82) Telecommunications
IX	P. Muñoz, R. Barco, I. Serrano and A. Gómez-Andrades, "Correlation-Based Time-Series Analysis for Cell Degradation Detection in SON" *IEEE Communications Letters*, vol. 20 (2) , pp. 396 - 399, 2016	1.291	Q2 (34/82) Telecommunications
X	E. J. Khatib, R. Barco, A. Gómez-Andrades and I. Serrano "Modelling LTE solved troubleshooting cases" *Journal of Network and Systems Management*, Under review 2016	1.078	Q3 (77/143) Computer science, information systems
XI	P. Muñoz, R. Barco, I. de-la-Bandera, E. J. Khatib, A. Gómez-Andrades and I. Serrano, "Root Cause Analysis based on Temporal Analysis of Metrics toward Self-Organizing 5G Networks" *IEEE Transactions on Vehicular Technology*, Under review 2016	2.243	Q1 (14/82) Telecommunications
XII	P. Muñoz, R. Barco, E. Cruz, A. Gómez-Andrades, E. J. Khatib and N. Faour, "A method for identifying faulty cells using a classification tree-based UE diagnosis in LTE", EURASIP Journal *on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic
XIII	I. de-la-Bandera, R. Barco, P. Muñoz and A. Gómez-Andrades, "Fault Compensation Algorithm based on Handover Margins in LTE Networks", *EURASIP Journal on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic

	Patent related to this thesis
XIV	P. Muñoz, R. Barco, I. Serrano and A. Gómez-Andrades, "First network node, method therein, computer program and computer-readable medium comprising the computer program for determining whether a performance of a cell is degraded or not". , 2016

7.3.2 Conferences and Workshops

Conferences arising from this thesis

	Conferences arising from this thesis
XV	L. Flores-Martos, A. Gómez-Andrades, R. Barco and I. Serrano, “Unsupervised system for diagnosis in LTE networks using Bayesian networks” *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1-5, Glasgow 2015
XVI	A. Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano and R. Barco “Simulador de fallos en una red LTE para sistemas de diagnosis”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014
XVII	L. Flores-Martos, A. Gómez-Andrades and R. Barco, “Diagnosis no supervisada basada en Redes Bayesianas para redes LTE”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014

Conferences related to this thesis

	Conferences related to this thesis
XVIII	I. de-la-Bandera, R. Barco, A. Gómez-Andrades, P. Muñoz and I. Serrano, “Compensación de Celdas Degradadas en Redes LTE”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014
XIX	E. J. Khatib, A. Gómez-Andrades and R. Barco, “Diagnosis en LTE Self-Optimizing Networks basada en algoritmos genéticos”, *XXIII SImposium nacional de la Unión Científica Internacional de Radio, Santiago de Compostela*, 2013

7.3.3 Related projects

This thesis was partially funded by the following projects:

- Optimi-Ericsson, ref.59288, Junta de Andalucía (Agencia IDEA, Consejería de Ciencia, Innovacion y Empresa) y ERFD.
- Proyecto de Investigación de Excelencia P12-TIC-2905, Junta de Andalucía.

7.3.4 Research stays

In addition, this thesis involved a three month stay in Plano, Texas (USA), collaborating with GCD RAN Ericsson and AT&T CU in automatic root cause analysis.

Appendix A

Summary (Spanish)

A.1 Introducción

A.1.1 Antecedentes y justificación

Tradicionalmente, las redes móviles han permitido comunicar información de manera eficiente a cualquier parte. En los últimos años, el gran desarrollo de las tecnologías móviles junto con la proliferación de una gran variedad de aplicaciones ha convertido el teléfono móvil en un dispositivo esencial para el nuevo estilo de vida. Una cosa es cierta, los teléfonos móviles han cambiado la vida diaria, modificando no solo la manera en la que nos comunicamos sino también la forma en la que vemos el mundo e incluso la forma en la que actuamos. Por consiguiente, las redes móviles han experimentado un gran desarrollo en los últimos años. En concreto, el número de subscripciones móviles está creciendo en todas las regiones, excediendo la población en algunos países. Esto se debe principalmente al surgimiento de nuevas subscripciones en los países en desarrollo, puesto que los móviles se han vuelto más asequibles, y al aumento en el número de dispositivos por individuos en el caso de las redes maduras. Por ejemplo, el número de subscripciones móviles superó los 7 mil millones a finales de 2015, representando una penetración global del 97% [1]. Como resultado de ello, el gran incremento en el número de dispositivos, aplicaciones móviles y consumo de datos ha causado un crecimiento significativo en el tráfico de los datos móviles.

En este contexto, en el que los consumidores esperan una alta experiencia de usuario y un desarrollo continuo de la red pero con precios estables, los operadores móviles buscan constantemente la forma de avanzar y evolucionar sus redes. Lo cual incluye el aumento de la velocidad de transmisión, el desarrollo de nuevas funcionalidades y servicios, la mejora de la capacidad de manejar nuevas aplicaciones. Como resultado, las redes móviles están evolucionando hacia la cuarta generación conocida como *Long-Term Evolution* (LTE) [2]. En este complejo escenario, los operadores de telefonía móvil tienen que tratar eficientemente las tareas de gestión y operación con el fin de cumplir con la demanda del consumidor, garantizando alta calidad y mejorando la experiencia del usuario. Además, los operadores tienen como objetivo minimizar los gastos de capital (CAPEX) y los gastos operativos (OPEX).

La solución clave para conseguir estos objetivos implica automatizar las tareas de gestión y operación. Y esto es exactamente lo que las redes auto-organizativas (*Self-Organizing Networks*, SON) [3, 4] pretenden abordar. Este nuevo paradigma llega con el principal objetivo de automatizar las tareas manuales mediante la incorporación de inteligencia en las redes móviles. Por tanto, este tema ha despertado un gran interés entre los operadores y la comunidad científica. Fue en 2008 cuando la alianza *Next Generation Mobile Networks* (NGMN) establece los principios y casos de uso de SON [5,6]. Más tarde, el organismo de estandarización 3rd *Generation Partnership Project* (3GPP) sigue desarrollando los requisites de SON, incluyéndolos en sus estándares. En concreto, 3GPP agrupa las principales funcionalidades SON en tres categorías [3,7,8]: auto-configuración [9], auto-optimización [9] y auto-curación (*Self-Healing*, SH) [10]. La auto-configuración es la encargada de automatizar la instalación y puesta en funcionamiento de los nuevos elementos desplegados en la red. La auto-optimizatión, por el contrario, reconfigura dinámicamente los parámetros de la red según van cambiando las condiciones del entorno con el fin de mejorar la calidad del servicio ofrecido. Por último, la auto-curación pretende recuperar automáticamente los elementos de la red de cortes o degradaciones del servicio que empeoran el rendimiento del usuario. Para ello, este sistema (see Fig. 1.1) automáticamente recoge datos de la red (tales como alarmas, indicadores o medidas) con el fin de detectar celdas con problemas, diagnosticar la raíz del problema y, en base a ello, proporcionar las acciones mediante las que recuperar o compensar el fallo detectado.

Debido al considerable interés que han suscitado las funcionalidades SON, en los últimos años, varios proyectos de investigación se han enfocado en el área de SON, tales como CELTIC Gandalf [11], FP7 E3 [12], FP7 SOCRATES [13], SELF-NET project [14], UniverSelf [15], SEMAFOUR [16] and COMMUNE [17]. Sin embargo, no todas las funcionalidades SON se han abordado con la misma profundidad. Especialmente, la auto-curación ha sido la menos investigada debido a las importantes limitaciones que hay que afrontar; a pesar de tratarse de una función fundamental para automatizar la gestión de las redes. Además, la investigación de cada una de las fases de un sistema de auto-curación (Fig. 1.1) presenta diferentes obstáculos, razón por la cual se encuentran grandes diferencias en el esfuerzo dedicado a la investigación de cada una de ellas. En concreto, la detección de fallos es una de las fases más estudiadas [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29], las cuales se centran en visualizar y modelar el comportamiento normal del sistema con el fin de detectar celdas con un comportamiento anómalo. Sin embargo, para la automatización de la fase de diagnosis en el área de la red de acceso, sólo se encentra un grupo reducido de referencias tales como [30, 31, 32, 33, 34].

Es una práctica común, por parte de los operadores de telefonía móvil, resolver los problemas de la red sin documentar ni almacenar adecuadamente los indicadores y sus valores que permitieron a los expertos identificar la causa real del fallo. A la hora de automatizar un proceso manual es esencial tener registrados las situaciones reales y la experiencia del experto con el fin de poder diseñar adecuadamente los sistemas de diagnosis. Por lo tanto, la falta de documentación junto con la falta de casos reales etiquetados son la principal causa por la que los sistemas de diagnosis no han sido tan estudiados como los sistemas de detección. Además, dentro de este contexto, todos los sistemas de diagnosis disponibles en la literatura [30, 31, 32, 33, 34] utilizan técnicas supervisadas, es decir, técnicas que requieren casos etiquetados, lo cual no es compatible con las limitaciones reales de este tipo de sistemas. La llegada de los avances tecnológicos y el paradigma del big data [35] presenta técnicas avanzadas para desarrollar en profundidad los sistemas SON en general, y los sistemas de auto-curación en particular. Esto junto con la estandarización de las funcionalidades de las trazas móviles [36] y la *Minimization of Drive Tests* (MDT) [37, 38, 39, 40], ofrecen la oportunidad de analizar gran cantidad de datos relacionados con los mensajes de señalización intercambiados por los equipos de la red, medidas radio e información de localización. Como resultado de ello, los avances tecnológicos posibilitan la investigación de nuevos sistemas de diagnosis basados en trazas móviles, en contraste con lo que se ha estado realizando hasta ahora. En concreto, todos los estudios previos relacionados con la auto-curación [18, 19, 20, 41, 21, 22, 23, 24, 25, 42, 43, 30, 31, 32, 33, 34] se han desarrollado en base al uso de medidas estadísticas a nivel de celda proporcionadas por el sistema de Operación, Administración y Gestión (*Operation, Administration and Maintenance system*, OAM) lo cual conlleva una pérdida de información sobre el rendimiento del usuario y la geo-localización de los fallos. Por consiguiente, debido a la falta de estudios relacionados con los sistemas de auto-curación en general y con los sistemas de diagnosis en particular, esta tesis está dedicada a superar las barreras que han evitado su investigación y desarrollo, con especial atención a superar la falta de casos etiquetados y mejorar la diagnosis mediante el uso de las trazas móviles.

A.1.2 Objetivos

El principal objetivo de esta tesis es superar las limitaciones que impiden avances significativos en el área de los sistemas de auto-curación. Fig. 1.2 resume dicho escenario, indicando en cada punto el desafío concreto al que hay que hacer frente y el objetivo asociado que se pretende abordar en esta tesis. En concreto, los objetivos fijados son:

- (Obj. 1) Proporcionar un conjunto representativos de casos de fallo: Con el fin de hacer frente a la falta de casos etiquetados necesarios para la validación, el primer objetivo de esta tesis es modelar un conjunto reducido y significantes de indicadores de rendimiento (*Key Performance Indicators*, KPI) junto con las causas de fallo más comunes en las redes LTE. De esta forma, la base de datos generada estará disponible para futuras investigaciones. El objetivo es ofrecer una base de datos útil a la comunidad investigadora facilitando la evaluación y comparación de distintas técnicas con el fin de elegir la más apropiada en cada caso de uso.
- (Obj. 2) Proponer un Sistema de diagnosis no supervisado para redes móviles, es decir, un sistema de diagnosis automático que no requiera casos etiquetados. Hasta ahora, los sistemas de diagnosis existentes en el estado del arte de las redes móviles se han basado en técnicas supervisadas.
- (Obj. 3) Proponer una metodología para el diseño y evaluación de sistemas de auto-curación. Es decir, en esta tesis, se debe proponer un procedimiento completo para diseñar y evaluar sobre el mismo escenario cualquier funcionalidad de un sistema de auto-curación, permitiendo así la repetición de experimentos por terceros y la comparación con otras técnicas.
- (Obj. 4) Proponer un sistema de diagnosis que siga el razonamiento experto. En concreto,

el sistema propuesto debe ser fácil de interpretar con un diseño claro basado en la relación entre los KPI y las causas de los fallos. El objetivo es modelar las tareas de los expertos teniendo en cuenta los efectos que los fallos causan sobre los KPIs. Estas relaciones entre causa y KPIs pueden ser manualmente definidas por los expertos o automáticamente determinadas mediante sistemas automáticos (ya sea usando técnicas supervisados o no supervisadas).

- (Obj. 5) Proponer un sistema automático para configurar sistemas de auto-curación. Una vez que los sistemas son diseñados, sus parámetros tienen que ser configurados en el momento de su instalación y durante su ciclo de vida a medida que la red evoluciona. Por ello, con el fin de liberar a los expertos de esta responsabilidad, en esta tesis se debe diseñar un sistema automático para configurar los sistemas de auto-configuración.
- (Obj. 6) Automatizar la identificación de fallos cuyo comportamiento es desconocido. Como se ha mencionado anteriormente, una de las tareas más difíciles y que mayor tiempo consumen es la identificación de fallos nuevo, dado que requiere el análisis de una gran cantidad de datos. Teniendo en cuenta este objetivo, junto con el Obj.2, en esta tesis se debe proponer un sistema de diagnosis capaz de identificar patrones desconocidos entre toda la base de datos reales y sin tener ninguna información previa sobre la existencia o no de fallos.
- (Obj. 7) Automatizar la diagnosis de problemas de usuario. Otro objetivo definido en esta tesis persigue diseñar un esquema para identificar la razón de cada desconexión del servicio inesperada. La idea es discernir automáticamente entre conexiones problemáticas y no problemáticas, identificando la fase de la conexión concreta en la que se produce el fallo. Para alcanzar este objetivo, el método propuesto se debe basar en la información proporcionada por las trazas móviles o los *Minimization of Drive Test* (MDT), que han sido estandarizados por el 3GPP con el fin de automatizar la recopilación de información a nivel de usuario evitando así los tradicionales drive test.
- (Obj. 8) Geo-localizar problemas de celda evitando la pérdida de información sobre el rendimiento del usuario. Para ello, se debe proponer un procedimiento para los sistemas de auto-curación mediante el cual evaluar automáticamente los problemas de las celda usando las trazas móviles. La idea es usar para ello el sistema basado en técnicas no supervisadas propuesto en esta tesis para encontrar problemas a nivel de usuario mediante las trazas móviles y luego agregar dichos problemas a nivel de celda, proporcionando a su vez su localización sobre el mapa.
- (Obj. 9) Determinar el impacto que tiene un fallo en LTE dependiendo de la existencia de otras tecnologías de acceso radio (*Radio Access Technology*, RAT). El objetivo es analizar cómo se puede utilizar la información reportada por las trazas móviles de las RATs subyacentes para diseñar sistemas de auto-curación. El objetivo es proponer un método automático que clasifique un problema específico en base a su impacto LTE y la RAT co-existente, dado que dependiendo de ello las acciones recuperadoras serán diferentes.

A.2 Métodos para sistemas de auto-curación basados en trazas y en técnicas de aprendizaje no supervisado

Dado que esta tesis es presentada en la modalidad de compendio de publicaciones, el segundo bloque de la tesis está compuesto por las publicaciones, agrupadas en distintos capítulos de acuerdo a su tema. Con el fin de guiar al lector, el segundo bloque comienza con el Capítulo 3 en el que se presenta una visión de cada artículo, detallando las relaciones entre ellos. Por lo tanto, el Capítulo 3 presenta una descripción concisa de los principales artículos y las relaciones entre ellos. Posteriormente, se esquematiza la metodología de investigación llevada a cabo, indicando las tecnologías y herramientas utilizadas. Cada uno de los artículos que componen esta tesis abordan al menos uno de los desafíos previamente presentados.

A.2.1 Metodología para el diseño y evaluación de sistemas de auto-curación en redes LTE [I] (Capítulo 4.1).

Para cumplir los objetivos de esta tesis, el punto de partida es abordar todos los desafíos relacionados con el diseño y la evaluación de los sistemas de auto-curación. Esta parte es esencial para cumplir el resto de objetivos dado que proporciona las bases para elaborar los sistemas presentados en esta tesis y evaluar su rendimiento, así como para compararlos con otras técnicas disponibles en la literatura. Por tanto, el primer artículo de esta tesis defina el método para diseñar y evaluar cualquier funcionalidad de un sistema de auto-curación. En concreto, se propone realizar un estudio de la viabilidad del sistema en un entorno controlado antes de ser desarrollado en una red real. Para ello, en este artículo también se presenta un Modelo de Fallos que describe a nivel de implementación los fallos típicos de una red LTE relacionado con las principales categorías (cobertura, movilidad e interferencia). Además, este método también detalla cómo deben calcularse en un simulador LTE los KPIs más relevantes para identificar los fallos seleccionados. Como resultado de ello, los investigadores pueden aplicar el Modelo de Fallos propuesto en sus simuladores con el fin de testear artificialmente sus sistemas de auto-curación. Otra contribución importante de este artículo es la elaboración de un sistema de diagnosis basado en lógica difusa mediante la metodología propuesta. Este sistema identifica la causa del problema aplicando un conjunto de reglas difusas a los KPIs. La principal ventaja de utilizar lógica difusa es que los expertos en resolución de problemas pueden entender e interpretar fácilmente los resultados y el comportamiento del sistema. Además, la evaluación de los resultados muestra que el sistema difuso definido proporciona una identificación de los fallos con una alta tasa de aciertos. Este sistema y el Modelo de Fallos han sido aplicados en un simulador dinámico de LTE a nivel de red, emulando el comportamiento de la red en presencia de fallos y generando así un conjunto de datos etiquetados artificiales. En concreto, para este artículo se han generado dos conjunto de datos, uno empleando un escenario hexagonal y otro empleando un escenario irregular basado en una red LTE real. Estos conjuntos de datos están compartidos online [47] con el fin de cubrir la falta de casos etiquetados y permitir a la comunidad científica replicar y comparar los experimentos. Además, se discuten algunos ejemplos de cómo funciona el sistema de diagnosis en una red real.

A.2.2 Funciones de rendimiento no supervisadas para redes auto-organizativas [II] (Capítulo 4.2)

Una vez que el objetivo para desarrollar y evaluar funcionalidades de los sistemas de auto-curación ha sido abordado, el siguiente paso es evitar la engorrosa tarea de configurar manualmente de dichos sistemas. Por tanto, el segundo artículo de esta tesis está dedicado a automatizar el cálculo de los parámetros de configuración. Es importante destacar que dichos parámetros dependen del comportamiento de la red y las características de los indicadores, dado que los sistemas de auto-curación analizan los KPIs con el fin de detectar los problemas e identificar su causa. Como resultado de ello, en este artículo se presenta un método no supervisado para modelar el comportamiento de los KPIs (es decir, su perfil bajo condiciones normales y anómalas). Más específicamente, los indicadores son modelados mediante una pareja de funciones (denominadas funciones de rendimiento) que son fácilmente interpretables por los expertos de resolución de problemas. Para ello, las funciones de rendimiento son automáticamente diseñadas utilizando el comportamiento estadístico de los datos reales a través de algoritmos de bajo coste computacional. Esto facilita la adaptación de las funciones de rendimiento a medida que las características de la red cambian con el tiempo, sin ningún coste significativo. La fiabilidad y efectividad del método propuesto ha sido evaluado comparando los resultados con métodos de referencia.

A.2.3 Análisis automático de la causa raíz para redes LTE basados en técnicas no supervisadas [III] (Capítulo 5)

En este punto, el siguiente objetivo es diseñar un sistema de diagnosis que pueda trabajar sin utilizar datos etiquetados en su entrenamiento. Además, el objetivo también es identificar nuevos patrones de fallos cuyo comportamiento o bien no es conocido o bien no está documentado. Por lo tanto, el siguiente trabajo de investigación persigue diseñar un sistema de diagnosis utilizando los métodos no supervisados más apropiados. Durante este estudio, se han examinado distintas técnicas no supervisadas, estudiando en detalle el rendimiento y las características tanto de las redes Bayesianas no supervisadas [48] como de los mapas auto-organizativos (*Self-Organizing Maps*, SOM). El estudio realizado con las redes Bayesianas no supervisadas ha sido presentado en un congreso internacional [XV] mientras que el sistema de diagnosis basado en SOM está detallado en el artículo [III]. La razón principal por la que se ha seleccionado SOM es porque esta técnica no requiere la discretización de las entradas (a diferencia de las redes Bayesianas) lo que da lugar a pérdida de información que es crucial para garantizar un diagnóstico satisfactorio. Como resultado, el núcleo del sistema de diagnosis presentado en este artículo se basa en SOM. En concreto, este método se utiliza para modelar automáticamente los patrones de comportamiento encontrados en los datos. Posteriormente, el método jerárquico de Ward [50] se utiliza para agrupar esos patrones en clases en base a su similitud. Además, con el fin de evitar la intervención del experto, se presenta un método mediante el cual se determina automáticamente el número total de clases, combinando el índice Davies-Bouldin [51] y el test Kolmogorov-Smirnov [52]. Luego, la clases son etiquetadas en base a su comportamiento estadístico asociando cada uno a un fallo específico. Además, con el fin de aumentar la precisión de los sistemas, se presenta un proceso de ajuste basado en el índice Silhouette [53] consiguiendo así afinar la diagnosis. Finalmente, la efectividad del sistema de diagnosis desarrollado es evaluada con el Modelo de

Fallos presentados en el Capítulo 4.1, analizando su complejidad y comparando su rendimiento con mecanismos de referencia. Además, este sistema se ha puesto en práctica en una red LTE real, demostrando su usabilidad y efectividad.

A.2.4 Análisis automático de la causa raíz basado en trazas para redes LTE auto-organizativas [IV] (Capítulo 6.1).

Automatizar las tareas de resolución de problemas usando únicamente información a nivel de celda no permite evaluar los problemas que afectan directamente al rendimiento del usuario. Hasta ahora la recogida y el análisis de las medidas de usuario siguen siendo manual a través de los tradicionales *Drive Test*. En este artículo, este desafío es abordado mediante la propuesta de un método automático para diagnosticar cada interrupción indeseada del servicio mediante el análisis de las trazas móviles con el fin de determinar la causa de la desconexión. El método propuesto se caracteriza por ser un modelo descendente que identifica la naturaleza de la desconexión (normal, fallo de acceso, desconexión anormal...) y define la relación causa-indicador necesaria para determinar las condiciones radio de los usuarios en dicho momento. Además, en este estudio, la clasificación radio propuesta se ha implementado mediante un sistema basados en reglas, evaluándolo en una red LTE simulada y con datos reales, mostrando así su beneficio y potencial.

A.2.5 Analisis de datos para la diagnosis de las condiciones radio en redes auto-organizativas [V] (Capítulo 6.2)

Evaluar las condiciones radio en cada área de la red móvil es una tarea que tradicionalmente se ha realizado mediante los *Drive Test*. Por lo tanto, en este artículo se presenta un nueva metodología para automatizar el proceso mediante las trazas móviles. Para ello, cada usuario es utilizado como un muestreador del interfaz radio, recolectando automáticamente esta información a través de las trazas móviles en lugar de utilizar los tradicionales vehículos equipados. Las condiciones radio son analizadas utilizando los indicadores a nivel de usuario propuestos en el Capítulo 6.1 junto con el sistema basado en SOM propuesto en el Capítulo 5, ya que permite identificar patrones radio desconocidos. Posteriormente, las condiciones radio de los usuarios de una determinada celda son agregadas con el fin de identificar el rendimiento global de dicha celda. Además, el método de agregación tiene en cuenta la geo-localización de los usuarios con el fin de cuantificar la gravedad de los problemas radio en base a su concentración. Esto permite detectar problemas incluso si sólo afectan a una pequeña área dentro de la celda. Además, este artículo presenta un nuevo método de visualización de las condiciones radio en cada punto de la red en un único mapa, evitando así el análisis individual de cada una de las gráficas asociadas a cada indicador. El sistema propuesto se ha evaluado en dos redes LTE reales diferentes, comparando los resultados con sistemas de diagnosis disponibles en el estado del arte y utilizando los tradicionales KPIs a nivel de celda. Los resultados muestran que este sistema permite identificar los problemas radio de manera más eficiente, alcanzando áreas que no puede ser cubiertas mediante los tradicionales *Drive Test*.

A.2.6 Método para la evaluación de agujeros de cobertura en LTE [VI] (Capítulo 6.3)

El objetivo de este artículo es abordar la falta de información sobre la tecnología de acceso radio subyacente (uRAT) cuando analizamos los síntomas de los problemas en una red LTE. Esta falta de información es particularmente significativa en el caso de los agujeros de cobertura, ya que se caracterizan por causar diferentes impactos en los usuarios dependiendo de si hay o no uRATs disponibles para mantener el servicio que es liberado por LTE. Por lo tanto, con el fin de abordar dicho desafío, este artículo propone un método capaz de analizar el rendimiento de los usuarios tanto en LTE como en la coexistente uRAT al mismo tiempo. En concreto, este método detecta celdas con agujeros de cobertura, diagnostica su tipo y gravedad. Para ello, el método realiza un seguimiento inter-tecnología de los usuarios que abandonan LTE para continuar su servicio en la uRAT y luego cuantifica el efecto del agujero de cobertura mediante un nuevo indicador inter-tecnología estimado a partir de las trazas móviles de las dos tecnologías. La principal ventaja de clasificar los agujeros de cobertura en base al impacto es que el experto puede particularizar la acción recuperadora, haciéndola relativa a cada escenario concreto. El propuesto sistema ha sido validado utilizando datos de una red real LTE y su co-localizada red 3G, mostrando su eficiencia en detectar agujeros de cobertura y diagnosticar su tipo.

A.3 Conclusiones

En el Capítulo 7 se resume el trabajo realizado durante la tesis, presentando las principales contribuciones y las líneas futuras. En concreto, se ha llevado a cabo un estudio de los principales desafíos que han evitado el desarrollo de este tipo de sistemas. Como resultado, en esta tesis se establecieron 9 objetivos con los que afrontar los desafíos identificados. Siguiendo la estructura de la tesis, las contribuciones se presentan de manera separada de acuerdo al objetivo inicial definido.

- Proponer una metodología para el diseño y evaluación de sistemas de auto-curación (Obj. 3).
  - En primer lugar, se ha propuesto un esquema complete para diseñar y evaluar cualquier funcionalidad de un sistema de self-healing. Este método abarca tres tareas: 1) el análisis del comportamiento de los KPIs con el fin de modelar estadísticamente las principales características y determinas la relación entre causa y KPI; 2) el estudio de viabilidad del sistema para comparar técnicas automáticas en un entorno controlado; y, 3) la implementación del sistema en una red real usando sus datos concretos.
  - Además, se han modelado los fallos más típicos de una red móvil y los KPIs más comunes, explicando sus características e indicando cómo pueden ser simulatos. En concreto, este Modelo de Fallos describe un conjunto de fallos reducido pero específico relacionado con las principales categorías (es decir, movilidad, cobertura e interferencia). También se han modelado los KPI más apropiados para analizar dicho conjunto de fallos. Es importante destacar que no se trata de un modelo cerrado, sino que admite más fallos y KPIs. Mediante la implementando este Modelo de Fallos, todas las funcionalidades de un sistema de auto-curación pueden ser evaluadas juntas en un

mismo escenario, incluyendo tanto a nivel de usuario como a nivel de celda.

- Un factor importante del Modelo de Fallos propuesto es la posibilidad de implementarlo en cualquier simulador RAN. Por tanto, esta tesis también describe la manera práctica de añadir el Modelo de Fallos a cualquier simulador.
- Siguiendo la implementación práctica propuesta, el Modelo de Fallos ha sido implementado en un simulador dinámico a nivel de sistema. Esto permite generar fallos en algunas celdas del escenario monitoreando su impacto en las celdas afectadas y sus vecinas a través de los KPIs propuestos. Primero, se puso en práctica el Modelo de Fallos utilizando un escenario LTE regular y hexagonal para evaluar los resultados en un entorno controlado. Posteriormente, se configuró un escenario irregular usando los parámetros y características de una red real correspondiente a un área urbana.

• Proporcionar un conjunto representativos de casos de fallo (Obj. 1)
  - Como se ha comentado a lo largo de la memoria, la falta de casos etiquetados es uno de los principales desafíos que deben afrontarse a la hora de diseñar sistemas de auto-curación. Por ello, se ha implementado el Modelo de Fallos en un simulador dinámico de nivel de sistema con el que generar un conjunto de datos en un escenario LTE regular. Gracias a ello, mediante el dataset obtenido se pueden probar artificialmente las distintas funcionalidades del sistema de auto-curación. Este conjunto de datos está disponible online in [47] para facilitar el estudio y la comparación con futuros experimentos.
  - Usando el Modelo de Fallos implementado en este simulador, también se ha recopilado información de nivel de usuario. Esto ha permitido analizar los sistemas de diagnosis a nivel de usuario.

• Proponer un sistema de diagnosis que siga el razonamiento experto (Obj. 4).
  - Para hacer frente a este desafío, se ha propuesto un sistema de diagnosis basado en reglas difusas. El principal beneficio es que permite seguir el pensamiento y razonamiento humano, facilitando así su interpretación.
  - El sistema propuesto se ha evaluado en el simulador utilizando tanto el escenario regular como irregular. En concreto, el rendimiento del sistema ha sido estudiado, demostrando que además de ser simple proporciona una buena tasa de aciertos incluso mejor que los sistemas de diagnosis basados en reglas de lógica tradicional. Además, se ha realizado una versión particularizando para las características de cada celda, demostrando así que el sistema de diagnosis propuesto puede ser diseñado tanto para la red entera como de manera particular para cada celda. Por último, se ha demostrado el funcionamiento del sistema con datos reales.

• Proponer un sistema automático para configurar sistemas de auto-curación (Obj. 5).
  - Aunque los sistemas de auto-curación sean diseñados automáticamente, normalmente requieren de la configuración de determinados parámetros en base a las características de los KPIs, para poder determinar, por ejemplo, si el valor de un KPI es considerado normal o no. Por tanto, otra contribución de esta tesis es la automatización del proceso de configuración. Es decir, en esta tesis se propone un nuevo método para modelar las características de los KPIs a través de las funciones de rendimiento. Este

método automáticamente diseña las funciones de rendimiento utilizando datos reales obtenidos de la red, de manera no supervisada y evitando la intervención del experto. Gracias a su bajo coste computacional y su facilidad para ser implementado, este método es apropiado para configurar y reajustar los parámetros de un sistema de auto-curación a medida que la red evoluciona.

  – El método propuesto se ha evaluado utilizando el sistema de diagnosis propuesto basado en lógica difusa y los sistemas de diagnosis disponibles en la literatura. Los resultados demuestran que la utilización de las funciones de rendimiento mejora la precisión de la diagnosis en comparación con las propuestas de la bibliografía.

- Proponer un sistema de diagnosis no supervisado para redes móviles (Obj. 2). Automatizar la identificación de fallos cuyo comportamiento es desconocido (Obj. 6).
- Automatizar la diagnosis de problemas de usuario (Obj. 7). Para alcanzar este objetivo, en esta tesis se propone una estrategia completa para automatizar la recogida y automatización de los datos de usuario, consiguiendo varias contribuciones significativas:
  – Primero, este método propone sacar provecho de las trazas móviles o los MDT para automatizar la recogida de información a nivel de usuario. En concreto, se ha realizado un estudio detallado, determinando los eventos y medidas más importantes para analizar las desconexiones de los usuarios.
  – Segundo, se ha propuesto un modelo descendente para identificar automáticamente las razones de las desconexiones de los usuarios, distinguiendo entre desconexiones problemáticas y no problemáticas. Además, en el último caso, el método clasifica las desconexiones anormales en fallos de acceso o conexiones caídas dependiendo de la fase de la conexión en la que sucede el problema. Gracias a este nuevo enfoque se puede analizar la conexión del usuario proporcionando muchos detalles.
  – Tercero, para los radio link failures o conexiones que se caen debido a malas condiciones radio, se presenta una técnica para automatizar la diagnosis de las causas radio, dado que esta información no es reportada por los eventos de las trazas móviles. En este estudio, se ha proporcionado el conocimiento necesario para construir un sistema automático de diagnosis de causas radio. Por ello, se presenta un estudio de los métodos que hay disponibles para automatizar el proceso junto con los indicadores necesarios y el conjunto de causas radio que pueden identificarse. En concreto, se han analizado en detalle las causas radio, determinando cómo pueden ser identificadas mediante el conjunto de indicadores de usuario propuesto. Como resultado, mediante el uso de éste método se puede clasificar la causa de la desconexión en agujero de cobertura, borde de celda, falta de celda dominante, problemas de movilidad o interferencia. Es importante resaltar que un gran beneficio de la estratégica propuesta es que sólo se requieren un conjunto reducido de indicadores estándar para determinar el problema radio.
  – Finalmente, se ha evaluado el método para determinar las causas radio mediante un sistema basado en reglas empleando tanto un simulador LTE como datos tomados de una red real. Los resultados revelan los beneficios de identificar y geo-localizar las causas radio que afectan negativamente a la experiencia del usuario.

- Geo-localizar problemas de celda evitando la pérdida de información sobre el rendimiento del usuario (Obj. 8).

  La diagnosis de una celda problemática a partir del análisis de los KPIs tradicionales a nivel de celda no permite identificar la localización de los fallos ni conocer el impacto que tienen en el usuario final. Este problema es particularmente importante cuando se intenta evaluar las condiciones radio de una celda. Por ello, para hacer frente a este desafío, en esta tesis se propone utilizar las trazas móviles para diagnosticar problemas a nivel de celda. Para ello, se propone un método completo con el que automatizar la identificación de las condiciones radio en una celda sin perder la información de localización. En concreto, las principales contribuciones de esta tesis en este contexto son las siguientes:

  – Se propone un método no supervisado para analizar automáticamente el rendimiento del usuario a partir de la información reportada en las trazas. En concreto, este método propone utilizar un sistema no supervisado para identificar patrones desconocidos de condiciones radio. Además, entre los sistemas no supervisados disponibles, se ha utilizado SOM para realizar la clasificación dado que: permite visualizar los resultados en base a las variables de entrada; permite identificar patrones desconocidos sin necesidad de determinar el número de clases de antemano; y, pueden trabajar con datos incompletos.
  – Se ha propuesto un procedimiento para evaluar los problemas radio de una celda en base a las condiciones radio diagnosticadas para cada uno de sus usuarios. Es decir, el método propone agregar los diagnósticos a nivel de usuario con el fin de determinar si la celda tiene un problema radio, y en dicho caso, indicar cuál es el más relevante.
  – En esta tesis también se ha abordado la diagnosis de problemas radio que sólo afectan a pequeñas zonas localizadas. En concreto, en este escenario, el número de usuarios afectados es relativamente bajo en comparación al número total de usuarios en la celda, pero se caracterizan por estar concentrados en un área problemática. El método propuesto identifica los problemas radio en base al nivel de concentración que presentan los usuarios afectados. Esto permite la detección de problemas densamente concentrados, diferenciándolos así de aquellos usuarios problemáticos que están dispersos y aislados.
  – Otra contribución son los mapas de condiciones radio que permiten mostrar y analizar en un único mapa la condición radio en cada punto de la red. Un aspecto importante de este mapa es su legibilidad dado que muestra directamente la condición radio en cada área evitando que los expertos tengan que analizar individualmente los cinco mapas típicos (uno asociado a cada medida: RSRPserving, RSRQserving, RSRPstrongest, NumCells, TARelative) con el fin de combinar manualmente toda esa información e identificar la condición RF final. Como resultado, este mapa facilita el proceso para chequear el interfaz radio, reduciendo el tiempo y esfuerzo requerido por los expertos en resolución de problemas.
  – El sistema complete ha sido evaluado en dos redes móviles reales mostrando su valor a la hora de chequear las condiciones radio de una celda y demostrando su validez y efectividad con datos reales. Además, los resultados han sido comparados con los obtenidos al utilizar sistemas de diagnosis del estado del arte, concluyendo que

el sistema propuesto no sólo proporciona una tasa de error total menor sino que permite localizar problemas radio y diagnosticar aquellos que están concentrados en un pequeño área.

- Determinar el impacto que tiene un fallo en LTE dependiendo de la existencia de otras RAT (Obj. 9).

  Considerando sólo la información de las redes LTE no es posible identificar el tipo de impacto que un fallo tiene en toda la red móvil. Por ello, las contribuciones realizadas en esta tesis para abordar este problema son las siguientes:

  - Se ha propuesto un método que combina datos de la red LTE junto con datos tomados de la tecnología subyacente. En concreto, en esta propuesta se usan las trazas móviles para analizar el rendimiento de los usuarios cuando abandonan la tecnología LTE y continúan su sesión en la tecnología coexistente. A través de este método se realiza un seguimiento inter-tecnología proporcionando el flujo de eventos inter-tecnología del usuario.
  - Otra contribución clave es una métrica inter-tecnología que permite estimar el tiempo activo que un usuario esté en LTE. Esta métrica se calcula a nivel de usuario a partir del flujo de eventos inter-technología para posteriormente ser agregados a nivel de celda.
  - La métrica propuesta se ha utilizado para diseñar las fases de detección y diagnosis. El principal beneficio es la posibilidad de identificar agujeros de cobertura y clasificarlos dependiendo de su impacto en LTE y la tecnología coexistente. Esta clasificación es muy útil y necesaria, ya que las acciones correctoras o de compensación son específicas para cada situación.
  - Este sistema se ha evaluado en una red móvil donde coexisten LTE y 3G. Los resultados obtenidos muestran el correcto funcionamiento de las fases de detección u diagnosis.

A.3.1 Revistas

Publicaciones derivadas de esta tesis

	Publicaciones derivadas de esta tesis	IF	Journal Rank
I	A. Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano and R. Barco, "Methodology for the Design and Evaluation of Self-Healing LTE Networks" *IEEE Transactions on Vehicular Technology*, Online 2015	2.243	Q1 (14/82) Telecommunications
II	A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano, "Unsupervised Performance Functions for Wireless Self-Organizing Networks" *Wireless Personal Communications*, Accepted 2016	0.701	Q4 (63/82) Telecommunications
III	A. Gómez-Andrades, P. Muñoz, I. Serrano and R. Barco, "Automatic root cause analysis for LTE networks based on unsupervised techniques" *IEEE Transactions on Vehicular Technology*, vol. 65 , pp. 2369 - 2386, 2016	2.243	Q1 (14/82) Telecommunications
IV	A. Gómez-Andrades, R. Barco, I. Serrano, P. Delgado, P. Caro-Oliver and P. Muñoz, "Automatic Root Cause Analysis based on traces for LTE Self-Organizing Networks" *IEEE Wireless Communication*, Accepted 2015	4.148	Q1 (3/82) Telecommunications
V	A. Gómez-Andrades, R. Barco, P. Muñoz and I. Serrano "Data analytics for diagnosing the RF condition in Self-Organizing Networks" *IEEE Transactions on Mobile Computing*, Under review (Minor revision) 2016	2.456	Q1 (11/82) Telecommunications
VI	A. Gómez-Andrades, R. Barco and I. Serrano "A method of assessment of LTE coverage holes", EURASIP Journal *on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic

Patente relacionada con esta tesis

	Patente relacionada con esta tesis
XIV	P. Muñoz, R. Barco, I. Serrano and A. Gómez-Andrades, "First network node, method therein, computer program and computer-readable medium comprising the computer program for determining whether a performance of a cell is degraded or not". , 2016

Publicaciones relacionadas con esta tesis

	Publicaciones relacionadas con esta tesis	IF	Ranking
VII	E. J. Khatib, R. Barco, A. Gómez-Andrades, P. Muñoz and I. Serrano, "Data mining for fuzzy diagnosis systems in LTE networks" *Expert Systems with Applications*, 42 (21), 7549-7559, 2015	2.981	Q1 (19/130) Computer science, artificial intelligence
VIII	E. J. Khatib, R. Barco, A. Gómez-Andrades and I. Serrano, "Diagnosis based on genetic fuzzy algorithms for LTE Self-Healing" *IEEE Transactions on Vehicular Technology*, vol. 65 (3), pp. 1639 - 1651, 2016	2.243	Q1 (14/82) Telecommunications
IX	P. Muñoz, R. Barco, I. Serrano and A. Gómez-Andrades, "Correlation-Based Time-Series Analysis for Cell Degradation Detection in SON" *IEEE Communications Letters*, vol. 20 (2) , pp. 396 - 399, 2016	1.291	Q2 (34/82) Telecommunications
X	E. J. Khatib, R. Barco, A. Gómez-Andrades and I. Serrano "Modelling LTE solved troubleshooting cases" *Journal of Network and Systems Management*, Under review 2016	1.078	Q3 (77/143) Computer science, information systems
XI	P. Muñoz, R. Barco, I. de-la-Bandera, E. J. Khatib, A. Gómez-Andrades and I. Serrano, "Root Cause Analysis based on Temporal Analysis of Metrics toward Self-Organizing 5G Networks" *IEEE Transactions on Vehicular Technology*, Under review 2016	2.243	Q1 (14/82) Telecommunications
XII	P. Muñoz, R. Barco, E. Cruz, A. Gómez-Andrades, E. J. Khatib and N. Faour, "A method for identifying faulty cells using a classification tree-based UE diagnosis in LTE", EURASIP Journal *on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic
XIII	I. de-la-Bandera, R. Barco, P. Muñoz and A. Gómez-Andrades, "Fault Compensation Algorithm based on Handover Margins in LTE Networks", *EURASIP Journal on Wireless Communications and Networking*, Under review 2016	0.627	Q3 (191/255) Engineering, electrical & electronic

A.3.2 Conferencias

Conferencias derivadas de esta tesis

	Conferences arising from this thesis
XV	L. Flores-Martos, A. Gómez-Andrades, R. Barco and I. Serrano, “Unsupervised system for diagnosis in LTE networks using Bayesian networks” *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1-5, Glasgow 2015
XVI	A. Gómez-Andrades, P. Muñoz, E. J. Khatib, I. de-la-Bandera, I. Serrano and R. Barco “Simulador de fallos en una red LTE para sistemas de diagnosis”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014
XVII	L. Flores-Martos, A. Gómez-Andrades and R. Barco, “Diagnosis no supervisada basada en Redes Bayesianas para redes LTE”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014

Conferencias relacionadas con esta tesis

	Conferencias relacionadas con esta tesis
XVIII	I. de-la-Bandera, R. Barco, A. Gómez-Andrades, P. Muñoz and I. Serrano, “Compensación de Celdas Degradadas en Redes LTE”, *XXIV SImposium nacional de la Unión Científica Internacional de Radio, Valencia*, 2014
XIX	E. J. Khatib, A. Gómez-Andrades and R. Barco, “Diagnosis en LTE Self-Optimizing Networks basada en algoritmos genéticos”, *XXIII SImposium nacional de la Unión Científica Internacional de Radio, Santiago de Compostela*, 2013

A.3.3 Proyectos relacionados

Esta tesis ha estado parcialmente financiada por los siguientes proyectos:

- Optimi-Ericsson, ref.59288, Junta de Andalucía (Agencia IDEA, Consejería de Ciencia, Innovacion y Empresa) y ERFD.
- Proyecto de Investigación de Excelencia P12-TIC-2905, Junta de Andalucía.

A.3.4 Estancia de investigación

Durante la elaboración de esta tesis se ha realizado una estancia de tres meses en Plano, Texas (EEUU), colaborando con GCD RAN Ericsson y AT&TCU en la automatización del análisis de la causa de fallos.

Bibliography

[1] Information and Communication Technology (ICT) Statistics, International Telecommunication Union (ITU). The World in 2011: ICT Facts and Figures, 2015.

[2] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, version 9.1.0 (2009-09). TS 36.300.

[3] 3GPP. Self-Organizing Networks (SON); Concepts and requirements. TS 32.500, 3rd Generation Partnership Project (3GPP).

[4] S. Hamalainen, H. Sanneck, and C. Sartori. *LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency.* Wiley, 2011.

[5] NGMN, Recommendation on SON and O&M Requirements, 2008.

[6] NGMN, Use Cases related to Self Organising Network, Overall Description, 2008.

[7] R.Barco, P.Lázaro, and P.Muñoz. A unified framework for self-healing in wireless networks. *IEEE Communications Magazine*, pages 134–142, 2012.

[8] SOCRATES. Use Cases for Self-Organising Networks. Deliverable D2.1, v. 1.0, 2008 , www.fp7-socrates.org.

[9] 3GPP. Self-configuring and self-optimizing network use cases and solutions, version 9.0.0 (2009-09). TR 36.902.

[10] 3GPP. Self-Organizing Networks (SON); Self-healing concepts and requirements, version 11.0.0 (2012-09). TS 32.541.

[11] Z. Altman et al. The Celtic Gandalf framework. In *Proc. of IEEE Mediterranean Electrotechnical Conference (MELECON)*, 2006.

[12] ICT-2007-216248 E3. Project Presentation Report. Technical Report Deliverable D0.2, Version 1.0, May, 2008.

[13] INFSO-ICT-216284 SOCRATES. Use Cases for Self-Organising Networks. Technical Report Deliverable D2.1, Version 1.0, March, 2008.

[14] INFSO-ICT-224344 Self-NET. System Deployment Scenarios and Use Cases for Cognitive Management of Future Internet Elements. Technical Report Deliverable D1.1, Version 1.0, October, 2008.

[15] FP7-257513 UniverSelf. Self-diagnosis and self-healing for IMS VoIP and VPN services.

Technical Report Case study - Part I, Version 1.0, September, 2012.

[16] INFSO-ICT-316384 SEMAFOUR. Self-Management for Unified Heterogeneous Radio Access Networks. Technical Report Deliverable D6.1, Version 1.0, October, 2012.

[17] CP08-004 COMMUNE. Outline Requirements for COMMUNE. Technical Report Deliverable D2.1, Version 1.0, April, 2012.

[18] G. A. Barreto, J. C. M. Mota, L. G. M. Souza, R. A. Frota, and L. Aguayo. A new approach to fault detection and diagnosis in cellular systems using competitive learning. In *Proceedings of the VII Brazilian Symposium on Neural Networks*, 2004.

[19] G. A. Barreto, J. C. M. Mota, Luís G. M. Souza, R. A. Frota, and L. Aguayo. Condition monitoring of 3g cellular networks through competitive neural models. *IEEE Transactions on Neural Networks*, pages 1064 – 1075, 2005.

[20] G. A. Barreto, J. C. M. Mota, L. G. M. Souza, R. A. Frota, L. Aguayo, J.S. Yamamoto, and P. Oliveira. Competitive neural networks for fault detection and diagnosis in 3g cellular systems. In *Lecture Notes in Computer Science*, pages 207 – 213, 2004.

[21] J. Laiho, M. Kylväjä, and A. Höglund. Utilisation of advanced analysis methods in umts networks. In *Proc. IEEE Vehicular Technology Conference VTC*, pages 726 – 730, 2002.

[22] J. Laiho, K. Raivio, P. Lehtimäki, K. Hatonen, and O. Simula. Advanced analysis methods for 3G cellular networks. *IEEE Transactions on Wireless Communications*, 4(3):930 – 942, 2005.

[23] P. Lehtimäki and K. Raivio. A knowledge-based model for analyzing gsm network performance. In *Proc. International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, Bari, Italy*, 2005.

[24] P. Lehtimäki and K. Raivio. A som based approach for visualization of gsm network performance data. In *Proc. International Symposium on Intelligent Data Analysis, Madrid, Spain*, 2005.

[25] M. Kylvaja, K. Hatonen, P. Kumpulainen, J. Laiho, P. Lehtimäki, K. Raivio, and P. Vehvilainen. Trial Report on Self-Organizing Map Based Analysis Tool for Radio Networks. In *IEEE 59th Vehicular Technology Conference, 2004. VTC 2004-Spring.*, volume 4, pages 2365 – 2369, May 2004.

[26] J Puttonen, J. Turkka, O. Alanen, and J Kurjenniemi. Coverage optimization for Minimization of Drive Tests in LTE with extended RLF reporting. In *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1764–1768, 2010.

[27] Y. Ma, M. Peng, W. Xue, and X. Ji. A dynamic affinity propagation clustering algorithm for cell outage detection in self-healing networks. *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2266–2270, 2013.

[28] S. Chernov, M. Cochez, and T. Ristaniemi. Anomaly Detection Algorithms for the Sleeping Cell Detection in LTE Networks. *IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2015.

[29] P. Szilágyi and C. Vulkán. LTE user plane congestion detection and analysis. *IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1819–1824, 2015.

[30] R. Barco, V. Wille, and L. Díez. System for automated diagnosis in cellular networks

based on performance indicators. In *European Transactions on Telecommunications, 2005.*, volume 16, pages 399 –409, 2005.

[31] R.M.Khanafer, B.Solana, J.Triola, R.Barco, L.Moltsen, Z.Altman, and P.Lázaro. Automated Diagnosis for UMTS Networks Using Bayesian Network Approach. *IEEE Transactions on vehicular technology*, 57(4):2451 –2461, 2008.

[32] R.Barco, V.Wille, L.Díez, and M. Toril. Learning of model parameters for fault diagnosis in wireless networks. *Wireless Networks*, 16(1):255–271, 2010.

[33] P. Szilágyi and S. Nováczki. An Automatic Detection and Diagnosis Framework for Mobile Communication Systems. *IEEE Transactions on network and service management*, 9(2):184–197, 2012.

[34] S. Novaczki. An Improved Anomaly Detection and Diagnosis Framework for Mobile Network Operators. In *Proc. of 9th International Conference on the Design of Reliable Communication Networks (DRCN)*, 2013.

[35] A. Imran, A. Zoha, and A. Abu-Dayya. Challenges in 5G: how to empower SON with big data for enabling 5G. *IEEE Network*, 28:27 – 33, 2014.

[36] 3GPP. Telecommunication management; Subscriber and equipment trace: Trace concepts and requirements. TS 32.421, 3rd Generation Partnership Project (3GPP).

[37] 3GPP. Universal Mobile Telecommunications System (UMTS); LTE; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2. TS 37.320, 3rd Generation Partnership Project (3GPP).

[38] 3GPP. Technical Specification Group Radio Access Network; Study on minimization of drive-tests in Next Generation Networks. TR 36.805, 3rd Generation Partnership Project (3GPP).

[39] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog. Minimization of drive tests in 3GPP release 11. *IEEE Communications Magazine*, 50(11):36–43, 2012.

[40] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sébire. Minimization of drive tests solution in 3GPP. *IEEE Communications Magazine*, 50(6):28 – 36, 2012.

[41] A. J. Hoglund, K. Hatonen, and A. S. Sorvari. A computer host-based user anomaly detection system using the self-organizing map. *Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 5:411 – 416, 2000.

[42] K. Raivio, O. Simula, and J. Laiho. Neural Analysis of Mobile Radio Access Network. In *IEEE International Conference on Data Mining*, pages 457–464, December 2001.

[43] K. Raivio, O. Simula, J. Laiho, and P. Lehtimäki. Analysis of mobile radio access network using the Self-Organizing Map. *IEEE Eighth International Symposium on Integrated Network Management*, pages 439 – 451, 2003.

[44] M. Amirijoo, L. Jorguseski, R. Litjens, and R. Nascimento. Effectiveness of Cell Outage Compensation in LTE Networks. In *Proc. of IEEE Consumer Communications and Networking Conference (CCNC)*, 2011.

[45] A. Saeed, O. Glenn Aliu, and M. Ali Imran. Controlling Self Healing Cellular Networks using Fuzzy Logic. In *Proc. of IEEE Wireless Communications and Networking Conference: Mobile and Wireless Networks*, 2012.

[46] 3GPP. Physical Layer, Measurements. TS 36.214, 3rd Generation Partnership Project (3GPP).

[47] A. Gómez-Andrades P.Muñoz, R.Barco, E. J. Kathib, I. de-la Bancera, and I.Serrano. Labelled cases of LTE problems, http://webpersonal.uma.es/de/rbarco/, 2014.

[48] D. T. Pham and G. A. Ruz. Unsupervised training of bayesian networks for data clustering. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 465(2109):2927–2948, 2009.

[49] T. Kohonen, M. R. Schroeder, and T. S. Huang, editors. *Self-Organizing Maps*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3rd edition, 2001.

[50] F. Murtagh and P. Legendre. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, 2013.

[51] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 1(2):224–227, 1979.

[52] F. J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[53] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[54] P. Muñoz, I. de la Bandera, F. Ruiz, S. Luna-Ramírez, R. Barco, M. Toril, P. Lázaro, and J. Rodríguez. Computationally-Efficient Design of a Dynamic System-Level LTE Simulator. *Intl Journal of Electronics and Telecommunications*, 57(3):347 –358, 2011.