UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A NEW PARADIGM FOR PROACTIVE SELF-HEALING IN FUTURE SELF-ORGANIZING MOBILE CELLULAR NETWORKS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

BY

AHMAD ASGHAR
Norman, Oklahoma
2018

A NEW PARADIGM FOR PROACTIVE SELF-HEALING IN FUTURE SELF-ORGANIZING MOBILE CELLULAR NETWORKS

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Ali Imran, Chair

Dr. Timothy G. Ford, Outside Member

Dr. Thordur Runolfsson, Memeber

Dr. Kam Wai Clifford Chan, Member

Dr. Samuel Cheng, Member

Acknowledgments

In the name of Allah, the most merciful, the most beneficent.

I thank Almighty Allah for his blessings upon me, for letting me pursue my passion and for surrounding me with people who care about me and who help get the best out of me. This dissertation would not have been possible without His mercy and the sacrifices my family made for me and the support they gave. I am here because my mother sacrificed her own safety and comfort to allow me to pursue my passion. I am here because my wife gave up her life in California to move to Oklahoma for me. I am here because my sisters Madiha, Urooj and Sarah never let me get away without making me finish my schoolwork, always showed me the utmost affection, always shared a laugh with me and always gave me an outlet for venting. And most importantly, I am here because my father, the first Ph.D. I ever knew, gave up his health and comfort to help me achieve my potential.

I would like to thank my advisor, Dr. Ali Imran who gave me the opportunity to come to the University of Oklahoma to pursue my passion. He guided me all these years through my work and helped me develop my research and writing skills while also imparting in me the strong work ethic required to write a doctoral dissertation. His support and leadership has been the primary contributing factor in my becoming the researcher that I am today. I would also like to thank my committee members Dr. Thordur Runolfsson, Dr. Timothy Ford, Dr. Samuel Cheng and Dr. Cliff Chan whose advice was instrumental in my completing my doctoral studies. I would also like to extend my thanks to all of my lab members, past and present with special mention for Dr. Hasan Farooq, who helped me understand and work through my research problems while also creating a great environment for research and collaboration.

Finally, this dissertation is dedicated to everyone who is currently working on completing or has in the past completed their doctoral studies.

Table of Contents

List of Figures

List of Tables

Abstract

Mobile cellular network operators spend nearly a quarter of their revenue on network management and maintenance. Remarkably, a significant proportion of that budget is spent on resolving outages that degrade or disrupt cellular services. Historically, operators have mainly relied on human expertise to identify, diagnose and resolve such outages while also compensating for them in the short-term. However, with ambitious quality of experience expectations from 5th generation and beyond mobile cellular networks spurring research towards technologies such as ultra-dense heterogeneous networks and millimeter wave spectrum utilization, discovering and compensating coverage lapses in future networks will be a major challenge. Numerous studies have explored heuristic, analytical and machine learning-based solutions to autonomously detect, diagnose and compensate cell outages in legacy mobile cellular networks, a branch of research known as self-healing. This dissertation focuses on self-healing techniques for future mobile cellular networks, with special focus on outage detection and avoidance components of self-healing.

Network outages can be classified into two primary types: 1) *full* and 2) *partial*. Full outages result from failed soft or hard components of network entities while partial outages are generally a consequence of parametric misconfiguration. To this end, chapter 2 of this dissertation is dedicated to a detailed survey of research on detecting, diagnosing and compensating full outages as well as a detailed analysis of studies on proactive outage avoidance schemes and their challenges.

A key observation from the analysis of the state-of-the-art outage detection techniques is their dependence on full network coverage data, susceptibility to noise or randomness in the data and inability to characterize outages in both spacial domain and temporal domain. To overcome these limitations, chapters 3 and 4 present two unique and novel outage detection techniques. Chapter 3 presents an outage detection technique based on entropy field decomposition which combines information field theory and entropy

spectrum pathways theory and is robust to noise variance. Chapter 4 presents a deep learning neural network algorithm which is robust to data sparsity and compares it with entropy field decomposition and other state-of-the-art machine learning-based outage detection algorithms including support vector machines, K-means clustering, independent component analysis and deep auto-encoders.

Based on the insights obtained regarding the impact of partial outages, chapter 5 presents a complete framework for 5th generation and beyond mobile cellular networks that is designed to avoid partial outages caused by parametric misconfiguration. The power of the proposed framework is demonstrated by leveraging it to design a solution that tackles one of the most common problems associated with ultra-dense heterogeneous networks, namely imbalanced load among small and macro cells, and poor resource utilization as a consequence. The optimization problem is formulated as a function of two hard parameters namely antenna tilt and transmit power, and a soft parameter, cell individual offset, that affect the coverage, capacity and load directly. The resulting solution is a combination of the otherwise conflicting coverage and capacity optimization and load balancing self-organizing network functions.

CHAPTER 1

Introduction

At a time when mobile cellular network operators are competing for customers demanding higher data rates and greater data capacity at lower costs, keeping revenue margins up is proving increasingly difficult. Furthermore, the rising network operating expenses add to the stress on network operator revenues. Mobile cellular network expenditures are divided into two primary categories: 1) *capital expenditure* which is spent on acquiring and updating network entities, and 2) *operational expenditure* which is spent on managing and maintaining existing network resources. Based on industry estimates, mobile cellular network operators spend between 23% and 26% of their total revenue on mobile cellular network operation [1, 2]. A breakdown of operational expenses reveals that a significant proportion of it is spent on managing mobile cellular network outages and resolving performance degradation. Such service interruptions require human intervention and may sometimes go unnoticed leading to poor customer experience, eventually resulting in high customer churn. According to one survey estimate [3], mobile cellular network operators worldwide spent nearly $20 billion in the year 2015 to counter issues caused by network outages and service degradation which accounts for almost 1.7% of total revenue and nearly 7% of total operational expenses.

The primary solution to this challenge proposed by researchers and the mobile cellular network standardization body, 3rd generation partnership project (3GPP), is the deployment of self-organizing network (SON) solutions to automate processes that would otherwise require skilled human input. SON are broken down into three key areas: self-configuration, self-optimization [4] and self-healing [5]. Self-configuration refers to solutions that autonomously configure mobile cellular network nodes for plug and play. Self-optimization is related to solutions that target mobile cellular network performance optimization based on operator specifications. Self-healing is focused on solutions that identify performance issues in the mobile cellular network such as cell outages and key performance indicator (KPI) degradation. On top of the three components of SON mentioned above, self-coordination was also introduced by the 3GPP as part of Release 10 specifications for 4th generation mobile cellular networks [6] to address the potential conflicts arising between SON solutions that would lead to KPI degradation and partial outage. Therefore, for the purpose and scope of this dissertation, SON coordination is considered a sub-type of self-healing and referred to as proactive outage avoidance.

To understand how the four SON components are related, a generalized SON framework is given in Fig. 1.1. While self-configuration and self-optimization represent more implicit areas of operational expenditure reduction, self-healing provides the clearest quantifiable path towards operational expenditure reduction by minimizing the impact of mobile cellular network outages [3]. These include outages caused due to failure of physical or soft components of the network entities, rendering them non-functional and causing complete or *full outages*, or significant service degradation leading to *partial outages* that may not necessarily generate any system level alarms.

1.1 Self-Healing: Background

1.1.1 Self-Organizing Networks in Cellular Mobile Networks

SON functions gained popularity with the introduction of 4th generation mobile cellular networks, primarily due to the increased network complexity. The efficacy of a SON function depends on four key design components [7]: *Autonomy:* SON functions must be independent of human input, *Scalability:* Any SON functions deployed in the mobile cellular network must be scalable in terms of both time and space, *Adaptability*: The functions must be able to adapt to outside influences and internal failures. Additionally, it has been proposed that future SON networks must be *intelligent* [8], which implies

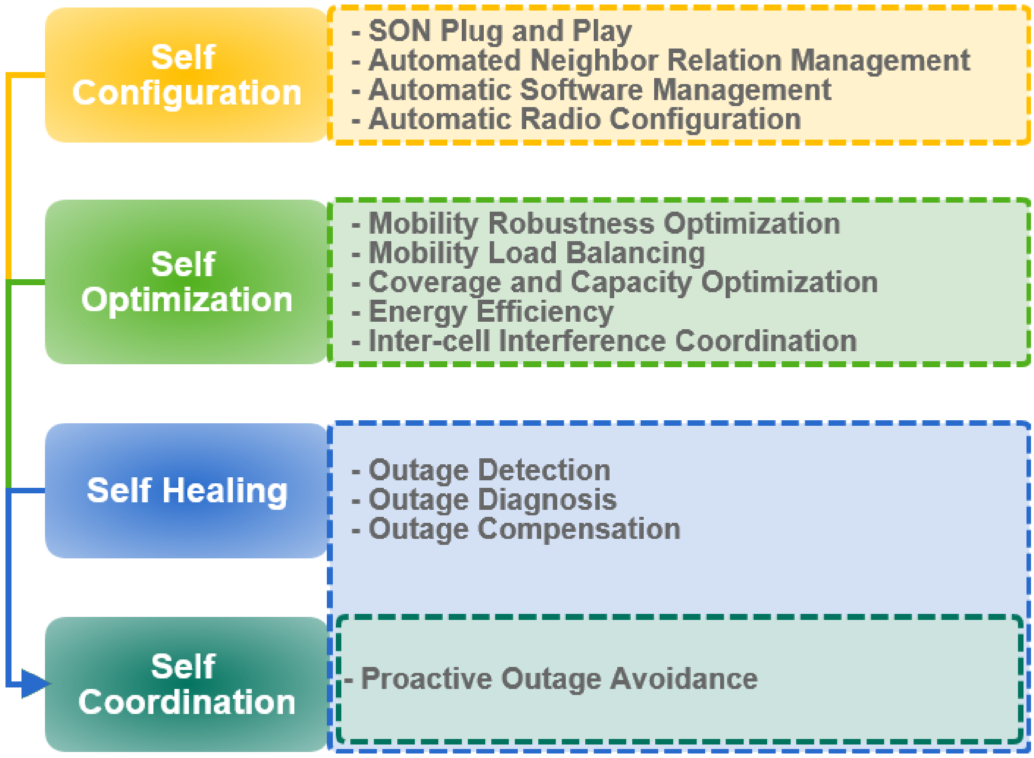

Fig. 1.1: Self-Organizing Network Framework for Mobile Cellular Networks

that, they must be able to learn from the information generated by the users and mobile cellular network entities to become completely independent in terms of adapting network parameters based on the primary goals of the operator.

1.1.2 Self-Healing in Mobile Cellular Networks

Traditionally, mobile cellular network operators have employed human experts to detect, diagnose and resolve outages in the network. As per the standard fault management framework defined by the 3GPP [9], faults and outages include issues such as hardware failures of mobile cellular network nodes, software failure issues at the nodes, failures of functional resources in which case no hardware component is responsible for the fault, loss of node functionality due to system overloading, and communication failure between two nodes due to internal or external influence. In such cases, the node will become completely dysfunctional leading to a *full outage*. As per 3GPP specifications, faults must be accompanied by the generation of an alarm that identifies the node and the type of failure that has occurred. The alarm may contain additional information to aid the recovery of the system but that is dependent on the equipment manufacturer.

Conversely, many service affecting issues in mobile cellular networks do not generate alarms or may not specifically be classified as faults or failures. Such issues are labeled *partial outages*. One such example is the degradation of a performance metric due to sudden changes in the mobile cellular network environment. Partial outages may include service degradation due to environmental effects, sudden variations in traffic, the presence of man-made interference sources that hinder normal operation of the network, or the result of misconfigured network parameters. Thus, mobile cellular network operators are dependent on human experts to monitor the network data to identify any such anomalies and to execute recovery actions to counter them. However, with the advent of 5th generation networks and the growth in network sizes and subscribers, network operators can no longer rely purely on human experts to sift through the vast amounts of network performance data generated consequently in search of anomalies.

Research in Self-Healing

Self-healing specifically for cellular networks has been studied as part of several research projects focusing on SON for cellular networks including the EUREKA Gandalf project [10] which explored the parametric interactions in 2nd and 3rd generation networks with the environment and studied the impact of automation in wireless networks, especially 3rd generation mobile cellular and Wi-Fi networks. The key deliverable of the project was bayesian networks based fault identification and diagnosis toolkit.

Similarly, the SOCRATES project [11] was aimed at investigating the impact of automation particularly in 4th generation mobile cellular networks, while the QSON project [12] investigated SON solutions primarily for self-optimization and self-healing along with preliminary analysis of the interactions of parameters and metrics as part of SON coordination. The project investigated new techniques, especially the exploitation of

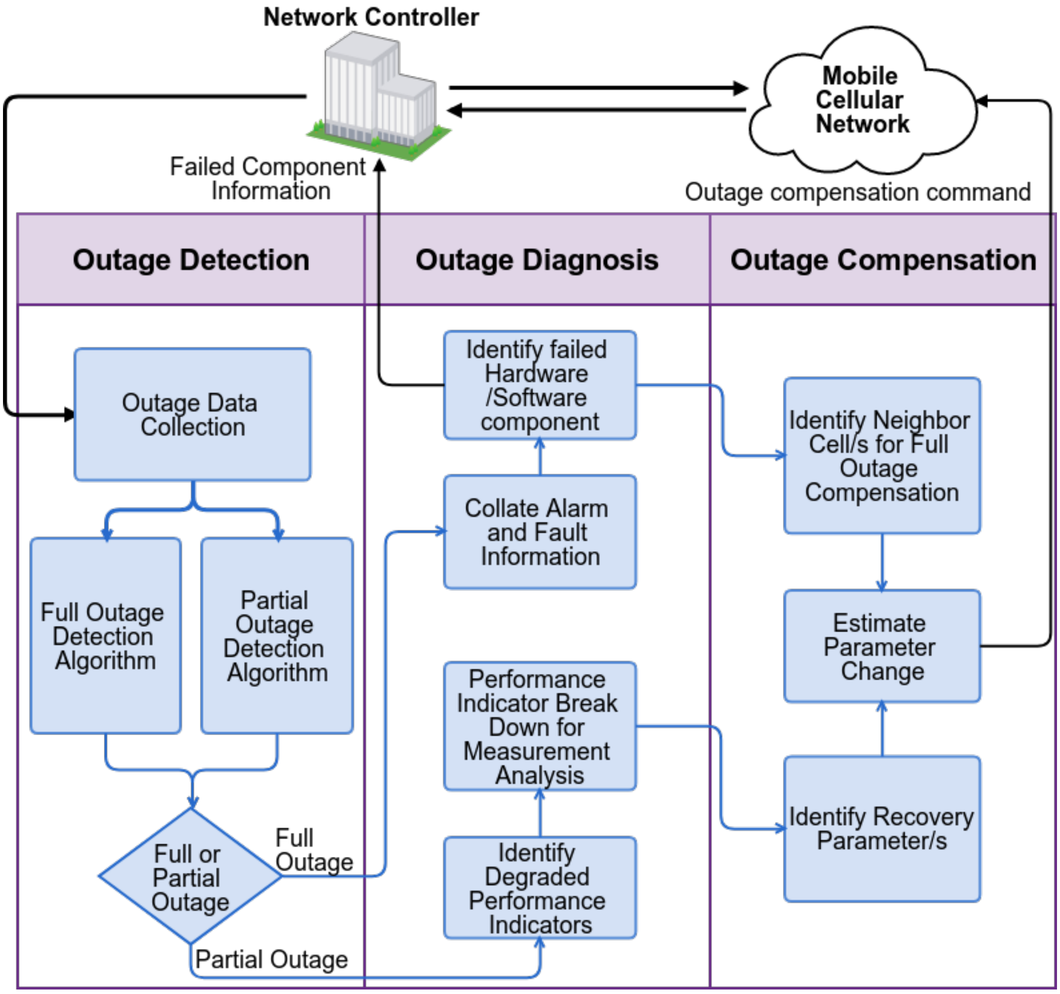

Fig. 1.2: Reactive Self-Healing Framework

big data analytics [8], to empower existing SON solutions. Recently, the SEMAFOUR project [13] has been launched which aims to develop a unified self-management system for heterogeneous radio access networks, comprising multiple radio access technologies and SON solutions including solutions for network outage detection, diagnosis and compensation for 4th generation standards and possible solutions for future mobile cellular networks.

Self-Healing Framework for Mobile Cellular Networks

As the number of physical entities in a network increases, the probability of network outages, both full and partial, may increases proportionally. In order to respond to these network outages, typical self-healing solutions employ a reactive three-stage framework. The first stage is detecting network outages for which *outage detection* algorithms are deployed. For effective self-healing, the outage detection solution must be able to detect both full and partial outages. In case a network outage is detected, the outage detection solution flags the affected network node for further actions, depending on the outage type. For example, in case a cell experiences hardware failure and is no longer able to send and receive data, it will be flagged for self-healing.

Once the outage has been detected, diagnostic algorithms will execute routines to identify the exact cause of network outage. For the sample case of hardware failure, the detection algorithm will examine alarms and fault codes to pinpoint the hardware component whose failure led to the outage. This information will then be relayed to the network controller which will either command field teams to replace the failed component or activate the redundancy elements to take over operations of failed entity. Conversely, if the outage is partial, the diagnosis algorithm will break down the degraded KPIs in order to identify the reason for the outage.

Upon completion of outage diagnosis, the information is passed along to the final stage of the self-healing function, that is, outage compensation. The outage compensation stage determines the impact of outage on neighboring entities and the network subscribers which is then used to execute changes to mitigate the outage. For example, in the case of hardware failure, outage compensation solution will identify the coverage hole created as a result of the outage and execute changes in neighboring cells to provide temporary coverage to affected subscribers. Alternatively, in the case of partial outage, the outage compensation solution may execute emergency parameter changes at either the affected cell or its neighbors or both to recover the degraded KPI/s. The framework for reactive self-healing is demonstrated in Fig. 1.2.

1.2 Key Research Drivers for Self-Healing

An overview of the key research drivers for self-healing are as follows.

Reduction of network operating expenses As mentioned already, mobile cellular network operators can spend as much as 1.7% of total revenue on fixing issues due to network outages. Network outages have the potential to disrupt service to millions of subscribers, as recently observed in case studies [14] and [15]. Overwhelming reliance on manual outage detection, diagnosis and compensation not only slows down the recovery process, but is also more expensive than autonomous solutions. Thus, autonomous self-healing solutions are one of the most inviting areas for mobile cellular network operators to cut down their operational costs for managing network outages.

Increase in network data The limited capability of human experts to absorb large amounts of network information at the same time and coming to conclusions about the existence of outages or KPI degradation in mobile cellular network means that as the number of entities in the network grows, the number of experts to monitor the network would grow proportionally. This will put further strain on the operators' already inflated operating expenses. Self-healing can reduce the load on human experts by providing solutions for the detection of service degradation and disruptions.

Complexity of network architecture With small cells expected to make up a significant part of future cellular network infrastructure [16], solutions specifically focusing on them must be developed. This concern is further fueled by the fact that small cells are subject to sparse reporting due to the low percentage of users associated with them and a more packed mobile cellular network topology in terms of inter-node distances. This makes it more difficult to identify service disruptions at small cells through traditional means.

Increase in network density The increasing number of radio nodes in the future mobile cellular networks can result in an increase in node failures [17]. This is demonstrated in Fig. 1.3, which shows the outage probability of a cell as mobile cellular network density increases, obtained using a Poisson distribution based method for estimating node failures derived from [17]. Fig. 1.3 shows the probability of a single node failure in one day (lower line chart), in three days (middle line chart), and seven days (top line chart). It is clear that probability of node failures is relatively low in a low density network such as a 2nd generation mobile cellular network. However, as the network density increases, the probability of node failure increases, so much so that on any given day the probability of node failure could be anywhere between 60% and 99.8%.

Hardware failures are already a significant area of concern for network operators. In [18] the authors present an analysis of customer complaints over a period of nine months in an enterprise network. The authors conclude that nearly 39% of all customer complaints are due to hardware failures. Therefore, it is safe to assume that if the number of network nodes is increased significantly, the corresponding probability of hardware failure will also increase. In wake of increasing number of nodes per unit area, dealing with such high rates of node failures will be very difficult if mobile cellular network operators continue the practice of manual outage management. In short, self-healing solutions will be less of a luxury and more of a necessity in future 5th generation and beyond networks.

Increase in network parameters With the introduction of 5th generation services and the associated technologies discussed above, the number of configuration and optimization parameters are expected to grow significantly [8]. The increasing number of network control parameters and entities can raise the probability of parameter misconfiguration significantly. The frequency and impact of parametric misconfiguration have been noted by Yin et al. [19]. Based on an analysis of a large number of customer

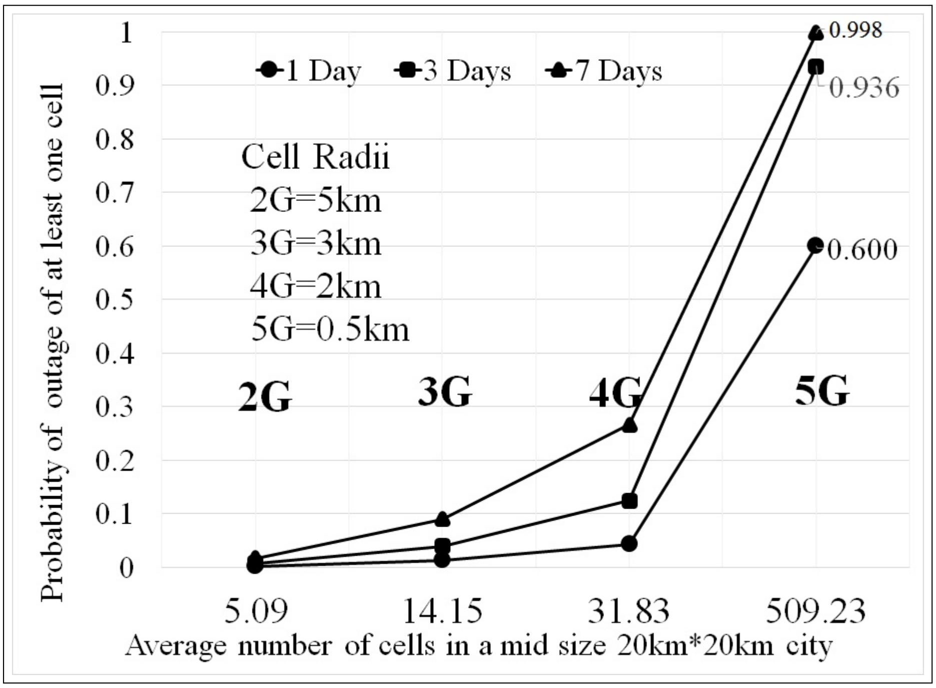

Fig. 1.3: Outage Probability of One Cell with Increase in Cell Density

complaints, the authors conclude that nearly 31% of high-severity customer complaints are due to misconfigured parameters. Out of this, 85.5% issues were due to mistakes in parameter configuration and in only 15% of the cases does a misconfiguration lead to an actual alarm. Otherwise, the misconfiguration is only identified when a customer complains about service outage. Though the actual count of customer complaints is not shared in [19], if it is assumed that there are 2000 parameters in the network and 10,000 complaints are received over a period of two years, the probability of a parametric misconfiguration every 100 days is 1.5%

A quantitative analysis of parameter misconfiguration in 5th generation mobile cellular networks is presented in Fig. 1.4 which shows the probability of misconfiguration of one parameter per cell every 100 days as the total number of configurable parameters per cell increases. The parameter misconfiguration probability is also derived using the Poisson distribution based method of failure estimation presented in [17]. In Fig. 1.4, three different probabilities, 0.01% (bottom line chart), 0.05% (middle line chart), and 0.1% (top line chart), of parametric misconfiguration per 100 days are assumed. These probabilities are well below the parameter misconfiguration probability estimated from [19]. Furthermore, since the data in [19] comes from an analysis of customer complaints, it is safe to argue that parametric misconfiguration does lead to a disruption of service. From Fig. 1.4 it is clear that parametric misconfiguration will become a major concern for mobile network operators in future networks.

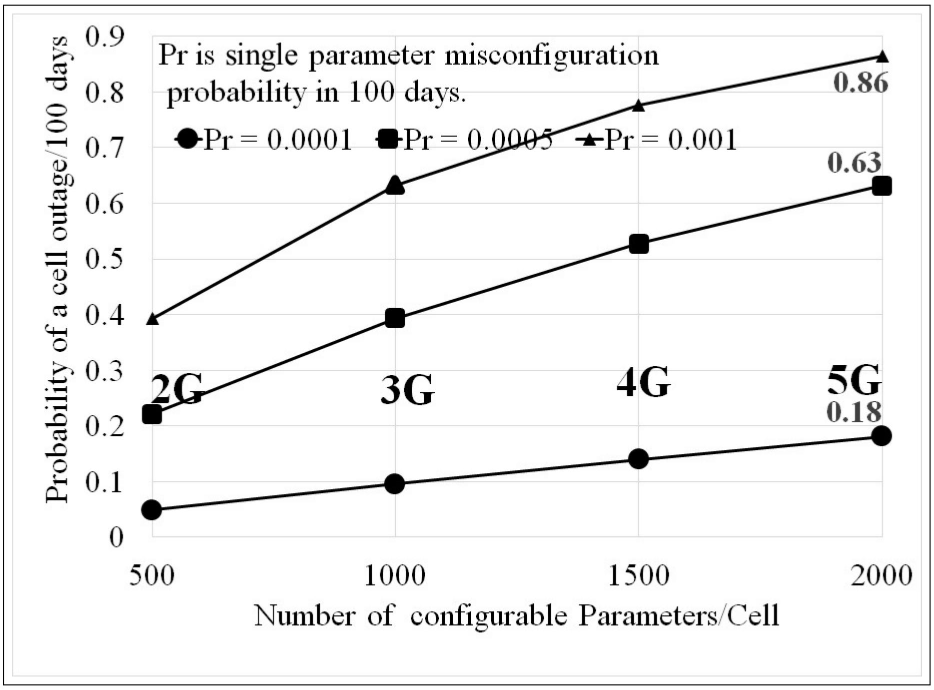

Fig. 1.4: Probability of Single Parameter Misconfiguration with Increase in Configurable Parameters

<u>**Increased focus on subscriber quality of experience calls for increased focus on self-healing**</u> Very high user quality of experience requirements in 5th generation mobile cellular networks mean near ubiquitous spatial and temporal network availability for various 5th generation use cases. State-of-the-art network availability estimation process depends on classic drive test based methods. However, the process is time and resource consuming while lacking comprehensiveness due to inaccessibility of a major portion of the network including all areas other than paved roads. Therefore, better methods are needed for network availability estimation and outage detection for 5th generation networks.

Additionally, low latency requirements for several 5th generation use cases mean that classic methods of manual outage diagnosis and manual outage compensation will not suffice. To address this challenge autonomous mechanisms to compensate outages quickly and seamlessly need to be developed. Furthermore, the vast majority of self-healing solutions for mobile cellular networks are reactive by nature. This means that the solutions come into effect once an outage has occurred. However, with the increased emphasis on user quality of experience in 5th generation networks, it is imperative that proactive self-healing solutions are developed to predict the impact of network outages and identify compensatory actions to resolve them before they effect user quality of experience.

1.3 Previous Work

In terms of mobile cellular networks, SON and self-optimization have received significant attention, with comprehensive studies published highlighting the contributions in both areas. Aliu et al. [20] present an overview of the recent studies carried out under the scope of SON for cellular networks, while Peng et al. [21] have presented an overview of state-of-the-art in self-configuration and self-optimization in mobile cellular networks.

Another area of automation in wireless networks are cognitive radio technologies. Cognitive radio technologies refer to dynamic spectrum access techniques that enable need based bandwidth allocation to mobile users via heterogeneous physical layer resource usage [22]. A survey of cognitive radio technologies has been presented by Akyildiz et al. [23]. Discussion on state-of-the-art and future challenges of cognitive radio technologies has been presented by Akyildiz et al. [24] while Akhtar et al. [25] have discussed the exploitation of unlicensed and unused spectral resources for dynamic spectrum allocation. Furthermore, Zhang et al. [26] have presented a survey of the research studies on self-optimization for cognitive radio technologies.

In terms of self-healing, a survey of applications from natural systems to software engineering has been presented in [27] where analogies between self-rectifying software systems and natural systems have been studied. Psaier and Dustdar [28] discuss the applications of self-healing in autonomous systems pertaining to the fields of information technology and communications. Furthermore, Paradis and Han [29] have surveyed studies on self-healing capabilities in wireless sensor networks.

Self-healing techniques in mobile cellular networks have briefly been discussed in [20] in the larger context of SON. The authors have presented description of self-healing in mobile cellular networks accompanied by a review of four outstanding works in the area. Since the publication of [20], research on self-healing techniques for mobile cellular networks has grown significantly. With the efforts to propose and standardize SON solutions for 5th generation technologies reaching their climax, the need for a comprehensive study on self-healing highlighting the efforts of research groups, equipment manufacturers and standardization bodies could not be higher. Furthermore, this dissertation aims to go well beyond the limited contributions of [20] towards surveying self-healing techniques for mobile cellular networks by breaking down the studies in terms of the type of outages, the measurements and methodologies used, and their results, while also proposing methods to detect and resolve outages with particular focus on future networks.

1.4 Research Objectives

In light of the discussion in sections 1.2 and 1.3, this dissertation explores the following research questions:

- What is the state-of-the-art in self-healing for legacy and 4th generation mobile cellular networks?
- What are the limitations of the state-of-the-art self-healing techniques that would make them nonviable given the direction of technological evolution for 5th gener-

ation and beyond mobile cellular networks?

- What are the potential solutions to the limitations of the state-of-the-art self-healing techniques and how can such solutions be implemented?
- What are the gains that could be achieved by solving the challenges faced by state-of-the-art self-healing techniques vis-a-vis emerging technologies for future mobile cellular networks?
- Are there any potential means of actually avoiding the situations when self-healing solutions would need to be invoked, simply by preventing the causes of outages in the first place?

1.5 Contributions

The primary contributions of this dissertation are summarized as follows:

- First of all, this dissertation takes an in-depth look at the state-of-the-art in self-healing techniques for mobile cellular networks and organizes them into the three primary areas of self-healing namely, detection, diagnosis and compensation which is the intrinsic flow of self-healing in nature and in practical applications. This literature survey is categorized in terms of the network topology, performance metrics, control mechanisms, and methodologies used for detection, diagnosis and compensation of full and partial outages, as well as proactive prevention of partial outages in a mobile cellular network. This allows easy understanding and comparison of studies within each particular area of self-healing. To complement this review of the state-of-the-art, this dissertation also presents a comprehensive discussion on the challenges in self-healing and identifies the research directions therein. Notably, it discusses the two primary types of challenges faced by existing self-healing solutions to adapt to future network requirements: 1) challenges that stem from ambitious quality of experience and low latency requirements of emerging networks, and 2) challenges that arise from the idiosyncrasies of anticipated future network enabling technologies such as ultra-dense cell deployments, millimeter wave cells (in which outage is the norm, not anomaly) and increased rate of emergence of sudden traffic hot-spots due to higher data rate per users leading to sudden change in KPIs (partial outage).

- Two of the primary limitations of the state-of-the-art self-healing techniques in general and outage detection techniques in particular learned from the in-depth review are their susceptibility to noise variance in the source data to be used for self-healing algorithms and the sparsity of such data. This dissertation explores why these two issues will be felt even more acutely in future mobile cellular networks and how they can be overcome. Based on the insights thus obtained, this dissertation presents an entropy field decomposition based outage detection solution that can not only detect outages in both space and time, but is also robust to the noise variance in the data. As well as presenting an algorithm to implement the proposed outage detection solution in future mobile cellular networks, this dissertation also provides an analysis of the time complexity of the proposed solution along with a comparison with other relevant techniques to demonstrate its efficacy.

- The proposed spatio-temporal outage detection solution is excellent for solving the noise variance problem when full network state information is available. However, if the network data are sparse, the proposed solution is not as effective. Therefore, to overcome the issue of low outage detection success rate due to sparse network coverage data, this dissertation presents an alternative outage detection solution that employs the hidden feature extraction capabilities of deep neural networks. Additionally, the proposed outage detection solution is compared against other state-of-the-art outage detection techniques as well as the entropy field decomposition based outage detection technique to demonstrate its efficacy when dealing with sparse network state data and variable noise.

- Next, using the insights gained from state-of-the-art partial outage avoidance techniques, this dissertation presents a comprehensive SON function coordination framework that uses the plethora of network and user level data to create network state information knowledge database and exploits it to identify the optimal parameter values that would take the network to the operator policy based performance levels.

- Lastly, this dissertation presents a concurrent coverage, capacity and load optimization solution to prevent partial outages caused by misconfiguration of the three key parameters, namely, cell transmit powers, antenna tilts and cell individual offsets. The solution leverages the framework for partial outage avoidance by using the network state prediction to optimize user data rates such that cell load constraints and user satisfaction constraints are fully satisfied. This solution provides a guiding pathway for how the proposed framework can be leveraged to optimize other network performance measures such as mobility management, energy efficiency and capacity exploitation efficiency.

1.6 Dissemination and Publications

The preparation of this dissertation has resulted in two provisional patent applications, several peer reviewed articles (accepted and pending) as well as conference papers and presentations. These include:

Patents

P1. A. Imran, A. Asghar and H. Farooq, "Method for enhancement of capacity and user Quality of Service in Mobile Cellular Networks", provisional patent application number: 62681320 filed June 06, 2018.

P2. A. Imran, H. Farooq and A. Asghar, "Method and apparatus for proactive self-optimization using data about network user behavior, mobility and measurements", 2018 (pending provisional patent application).

Journal Articles

J1. A. Asghar, H. Farooq and A. Imran, "Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions", IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1682-1709, 2018.

J2. A. Asghar, H. Farooq and A. Imran, "Concurrent Optimization of Coverage, Capacity, and Load Balance in HetNets Through Soft and Hard Cell Association Parameters", IEEE Transactions on Vehicular Technology, vol. 67, no. 9, pp. 8781-8795, Sept. 2018.

J3. H. Farooq, A. Asghar and A. Imran, "Mobility Prediction based Automated Proactive Energy Saving (AURORA) Framework for Emerging Ultra-Dense Networks", IEEE Transactions on Green Communications and Networking, 2018, DOI: 10.1109 /TGCN.2018.2858011.

J4. A. Asghar and A. Imran, "Proactive SON Function Coordination for 5th Generation Networks and Beyond", IEEE Network Magazine (revision submitted)

J5. A. Asghar, H. Farooq and A. Imran, "Entropy Field Decomposition Based Time and Space Aware Outage Detection Solution for Ultra-Dense Millimeter Wave Heterogeneous Networks", IEEE/ACM Transactions on Networking (under review).

J6. A. Asghar, U. Masood, H. Farooq and A. Imran, "Analysis and Comparison of Deep Neural Network Based Outage Detection in Variable Noise Environments and Sparse Network State Data", IEEE Transactions on Communication (under review).

J7. H. Farooq, A. Asghar and A. Imran, "Mobility Prediction based Proactive Dynamic Network Orchestration for Load balancing with QoS Constraint (OPERA)", IEEE/ACM Transactions on Networking (under review).

Conference Publications

C1. A. Asghar, H. Farooq, and A. Imran, "Concurrent CCO and LB Optimization in Emerging HetNets: A Novel Solution and Comparative Analysis", in Proc. IEEE PIMRC'18, Bologna, Italy, September 2018.

C2. A. Asghar, H. Farooq, and A. Imran, "A Novel Load-Aware Cell Association for Simultaneous Network Capacity and User QoS Optimization in Emerging HetNets", in Proc. IEEE PIMRC'17, Montreal, Canada, pp. 1-7, October 2017.

C3. U. Masood, A. Asghar, A. Imran, and A. N. Mian, "Deep learning based detection of sleeping cells in next generation cellular networks", in Proc. IEEE GLOBECOM'18, Abu Dhabi, United Arab Emirates, 2018.

C4. H. Farooq, A. Asghar, and A. Imran, "Mobility Prediction empowered Proactive Energy Saving Framework for 5G Ultra-Dense HetNets", accepted in Proc. IEEE GLOBECOM'18, Abu Dhabi, United Arab Emirates, 2018.

C5. A. Asghar, H. Farooq and A. Imran, "Outage Detection for Millimeter Wave Ultra-Dense HetNets in High Fading Environments", (under review).

C6. H. Farooq, A. Asghar, and A. Imran, "Mobility Prediction based Proactive Dynamic Network Orchestration for QoS aware Load balancing", (under review).

1.7 Organization

This dissertation is organized as follows. Chapter 2 presents a comprehensive survey of the state-of-the-art self-healing techniques for mobile cellular networks along with a summary of challenges they face, particularly in wake of emerging mobile cellular network technologies. Based on the lessons learned from the challenges facing the state-of-the-art outage detection solutions, chapter 3 presents an entropy field decomposition based solution that is tailored for future networks. Chapter 4 presents another solution for outage detection that is based on deep neural networks which is able to detect outages accurately even in networks with sparse coverage information. Leveraging the lessons learned from the state-of-the-art solutions for partial outage prevention, chapter 5 presents a comprehensive framework that employs data analytics, mathematical analysis and machine learning techniques to avoid the issue of parametric misconfiguration while also obtaining the optimal parameter set for given network configuration and usage statistics. The chapter also presents a solution that leverages said framework to optimize network coverage, capacity and load using cell transmission powers, antenna tilts and cell individual offsets, thus preventing partial outages due to their misconfiguration. Finally, chapter 6 presents a conclusion of the work incorporated in this dissertation and summarizes the future work that can be undertaken as a result of the findings presented herein.

CHAPTER 2

Literature Survey of Self-Healing Techniques

As identified in the introduction of this study, self-healing techniques for mobile cellular networks can be broken down into two broad categories that is, 1) *reactive*, and 2) *proactive* self-healing techniques. In this chapter, a comprehensive review of state-of-the-art reactive self-healing techniques is presented. This includes detection, diagnosis and compensation of both full and partial outages once they have occurred, as well as proactive avoidance of partial outages.

However, before presenting a survey of self-healing solutions for mobile cellular networks, a collection of key definitions is presented here that will enable the reader to quickly comprehend the nuances of the techniques reviewed in this chapter.

2.1 Key Components of Self-Healing Techniques for Mobile Cellular Networks

The five core components that constitute the logical structure of these studies are: 1) *methodology*, 2) *network topology*, 3) *performance metrics*, 4) *control mechanism*, and 5) *direction of control*. Fig. 2.1 shows the taxonomical distribution of the studies included in this chapter based on the self-healing framework and these core components.

2.1.1 Methodology

Each study presenting a solution for detection, diagnosis or compensation of outages follows an underlying methodology. These can be split into three broad categories: 1) *heuristic*, 2) *analytical* and 3) *learning based*. Heuristic solutions follow a set of

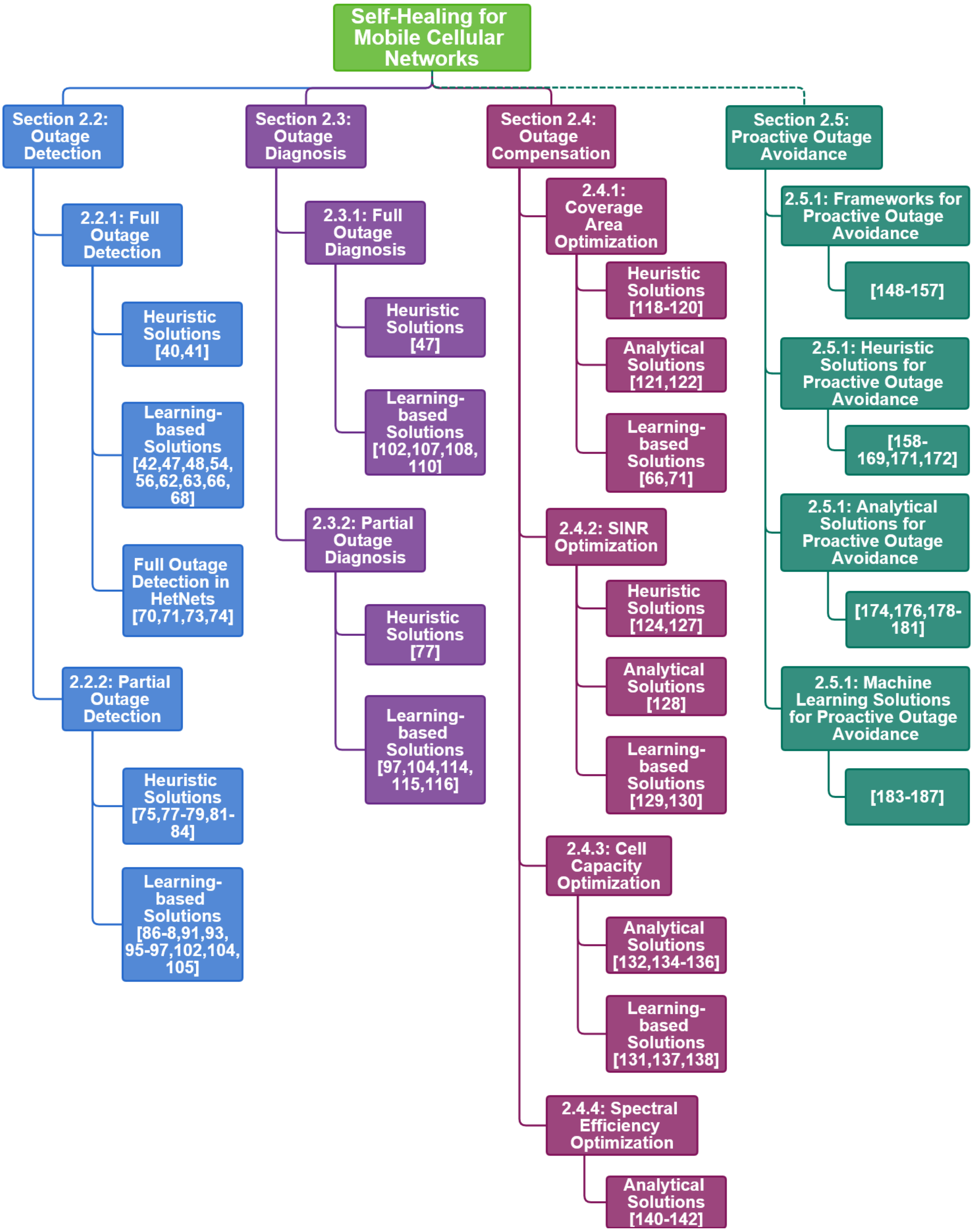

Fig. 2.1: Taxonomical Breakdown of Self-Healing Solutions

pre-defined rules and are built upon intuition or prior knowledge gained from existing literature or experience. Two heuristic solutions commonly found in literature are rule based algorithms, which follow a set of if-else rules, and frameworks, which mostly consist of guidelines. Analytical solutions break down a given problem into its mathematical components which are then solved to achieve an optimal or close to optimal solution. Analytical solution methodologies include techniques such as convex optimization [30], non-convex optimization (such as pattern search [31], genetic algorithms [32], simulated annealing [33]), multi-objective optimization [30], and game theory [34]. Learning based solutions are built on machine learning techniques popularized by the field of computer science. These algorithms rely overwhelmingly on user and network data and very little on expert knowledge [35]. Machine learning techniques are generally split into three overarching techniques [36, 37] that is, supervised, unsupervised and reinforcement learning.

2.1.2 Network Topology

The term network topology is defined as the architecture or layout of the network in terms of cell deployments. More specifically, network topology is used to describe the tiered structure of the network. There are two main types of network topologies used in literature. Homogeneous networks consist of only one tier of cells. These cells may be only macro cells with large coverage areas or only small cells which have lower power, and consequently lower coverage. Conversely, a combination of macro and small cells forming a multi-tier cellular network is referred to as a heterogeneous network (HetNet). While most studies on legacy mobile cellular networks employ homogeneous network topology as the baseline, HetNets are quickly gaining popularity due to their flexibility and their potential to achieve the goals set out for 5th generation and future cellular networks [38].

2.1.3 Performance Metrics

Performance metrics are the benchmark measurements used to evaluate network performance and can be obtained from network entities and user-generated reports. The solutions and algorithms presented in any study rely heavily on the choice of performance metrics employed in the study to construct and evaluate them. The performance metrics most relevant to studies on self-healing can be classified under the umbrella term *network health*.

Network health is a broad term used to describe the performance of the network in terms of universally accepted KPIs such as Accessibility, Retainability and Mobility [39]. *Accessibility* is the ability of subscribers to access the network resources for data transmission and includes KPIs such as attach success rate, radio resource control setup success rate, connection setup success rate, random access success rate etc. *Retainability* is the ability of the network to carry a data session to its completion without drop and is characterized by the session drop rate KPI. *Mobility* is the ability of the network to allow successful transition of a subscriber from one cell to another with minimal impact on services and is generally represented by handover attempt, success and failure rate KPIs.

Additionally, measurements signifying network coverage including reference signal received power (RSRP), and network quality including spectral efficiency, signal-to-interference and noise ratio (SINR), reference signal received quality (RSRQ), network and user data throughput, channel quality indicators and data latency are also often employed in the design and analysis of self-healing solutions.

2.1.4 Control Mechanism

Control mechanism is defined as the method of controlling SON solution functionality and can be categorized by the following methods: 1) *centralized*, 2) *distributed*, and

3) *hybrid*. Centralized control implies that the SON functions are controlled from one central controller connected to every node in the network, whereas distributed control implies that the control of SON functions resides within the network nodes. Hybrid control is a combination of central and distributed control and implies that while some SON functions may reside inside a centralized SON controller, other less computationally heavy functions which do not directly impact neighboring nodes, can be distributed to the nodes.

2.1.5 Direction of Control

Direction of control defines whether a SON function is designed to optimize the node-to-user link, user-to-node link, or both. Solutions designed to optimize the node-to-user link are *downlink* controlled, whereas the solutions optimizing the user-to-node link are *uplink* controlled. Some solutions optimize both downlink and uplink and thus, offer *bidirectional* control of network performance.

2.2 Outage Detection in Cellular Mobile Networks

While the standardized self-healing framework [5] does present a roadmap to a fully integrated self-healing framework, the precise inner workings of each component have been deliberately left open-ended. This has allowed researchers and network equipment manufacturers to come up with proprietary algorithms to suit the needs of evolving mobile cellular networks. In this and the following sections the research done in each of the self-healing framework components is described, beginning with a review of outage detection techniques. The studies in this section are ordered based on the type of outage and methodology employed within.

2.2.1 Full Outage Detection in Mobile Cellular Networks

The following subsections describe techniques and methodologies proposed for full outage detection in mobile cellular networks. The studies included in this section have been summarized in Table 2.1 in terms of techniques, network architectures, measurements and tools used within them.

Table 2.1: Qualitative Comparison of Cell Outage Detection Algorithms

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Full Outage Detection	[40]	Heuristic	Rule based	Homogeneous	Retainability, Mobility, Quality	Centralized	DL
	[41]				Accessbility, Quality		
	[42]	Learning Based	Supervised Learning		Coverage		
	[45]				Retainability, Mobility, Quality		UL/DL
	[73]			HetNet	Coverage	Hybrid	DL
	[48]		Unsupervised Learning	Homogeneous	Accessibility, Mobility, Coverage	Centralized	
	[56]				Coverage		
	[59,62,63, 66]				Mobility, Coverage		
	[68]				Retainability, Mobility		
	[54]				Coverage	Distributed	
	[71]			HetNet	Coverage	Centralized	
	[74]				Retainability		
	[70]				Coverage	Hybrid	

Heuristic Solutions for Full Outage Detection

Heuristic algorithms and frameworks for cell outage detection are heavily reliant on pre-existing knowledge of domain experts which makes them extremely useful for deployment in existing mobile cellular networks. One such framework has been proposed by Amirijoo et al. [40] which employs rule based decision tree algorithm for full outage detection in mobile cellular networks. The framework derives its rules from expert knowledge to create full outage detection trigger thresholds for performance metrics such as cell load, radio link failures, handover failures, user throughputs and cell coverage. A more comprehensive approach to rule based outage detection has been proposed by Liao et al. [41] that uses variations in user performance metric distributions to detect outages. The authors propose the construction of a weighted cost function composed of channel quality indicator distribution, the time correlation of channel quality differential and radio resource connection re-establishment requests. The cost function is treated as a hypothesis of normal cell performance. A cell is considered in outage if its neighboring cells fail this hypothesis that is, their targeted KPIs deviate from normal. The authors demonstrate that, using measurements from cell edge users, the proposed algorithm can detect neighbor cell outages almost instantaneously.

Learning based Solutions for Full Outage Detection

Beyond the heuristic methodologies of identifying outages in the network, machine learning based algorithms have been the prevailing method for full outage detection in research. Most of the studies on full outage detection that employ learning based algorithms can be split into two categories that is, supervised learning techniques for full outage detection solutions and unsupervised learning techniques for full outage detection.

Supervised Learning Techniques for Full Outage Detection Supervised algorithms are a popular choice in terms of full outage detection due to their reliance on pre-classified data. In the study by Mueller et al. [42], the authors have compared the performance of a rule based heuristic algorithm against a decision tree algorithm [43] and a linear discriminant binary classification function [44] to identify complete cell outages. The algorithms use user reports containing downlink signal power measurements to detect when a cell stops featuring in neighbor cell lists due to outage. The results show that the expert system is faster but less successful in detecting neighbor cell outages while the linear discriminant binary classification function performs the best in terms of true positive detection rate.

Another supervised learning approach for full outage detection is developing cell profiles for outage detection. Alias et al. [45] have proposed to develop performance profiles of cells in mobile cellular networks using hidden Markov chains [46] which track the state progression of network nodes that undergo outages. The proposed framework requires execution of controlled outages to build state profiles using signal quality and signal strength measurements of the outage affected cell and its neighbors. These measurements are then used to identify cell performance in real-time to predict if a cell has experienced an outage. The results show that the proposed approach can reach an accuracy of up to 90% in low fading environments.

Since the idea of executing controlled outages to build cell profiles may be prohibitive for live mobile cellular networks, Szilágyi and Novaczki [47] have proposed to construct default activity profiles of cells using simulated network data to detect when a cell faces an outage. The proposed algorithm uses level functions which continuously monitor downlink signal metrics such as channel quality, call drop rate and handover timing advance to detect when a cell falls below the acceptable threshold set by human experts. The authors have demonstrated that the proposed approach can act in near-real time by detecting outages within a few minutes of occurrence, which is a significant improvement over the detection time by human experts, especially in very large networks.

Unsupervised Learning Techniques for Full Outage Detection The unique ability of unsupervised learning algorithms to cluster data into distinct groups without any pre-classification makes them highly popular in outage detection applications. A major application of unsupervised learning is the detection of cells that are in outage but do not generate any alarms, otherwise known as *sleeping cells*. Detection of such cells is not immediately possible manually due to the lack of alarms accompanying the outage which makes their detection a highly useful application of unsupervised learning.

An extensive comparison of clustering algorithms for sleeping cell detection has been presented by Chernov et al. [48] where they have compared the performance of k-Nearest Neighbors (kNN) [49], self-organizing maps (SOM) [50], Local-Sensitive Hashing [51] and Probabilistic Anomaly Detection. The authors use random access channel access failure measurements in addition to the high-dimensional minimization-of-drive-test (MDT) data [52] as input data for clustering algorithms. To compare the performance of individual clustering algorithms, receiver operating characteristics and precision-recall curves are used. The results show that Probabilistic Anomaly Detection has the best receiver operating characteristics out of the four algorithms and a higher precision-recall curve compared to the other algorithms. Additionally, the authors have compared the training time of the four clustering algorithms which shows that Local Sensitivity Hashing has a training time of linear order, whereas Probabilistic Anomaly Detection takes the least amount of time to detect sleeping cells compared to the other algorithms.

Another clustering algorithm, Dynamic Affinity Propagation [53], has been utilized for sleeping cell detection by Ma et al. [54]. The proposed algorithm uses Dynamic Affinity Propagation to calculate user clusters based on received power values of neighboring and serving cells reported by users, while Silhouette index [55] is used as clustering quality criterion to estimate the number of significant user clusters. The resultant clustering is mapped to physical data including user location to identify cells in outage. While the approach clearly succeeds in identifying sleeping cells using simulated outages, it is possible that in a live network, some users suffering deep fade may be wrongly clustered.

Dimensionality Reduction for Unsupervised Learning While the above unsupervised learning solutions have a high degree of accuracy, their computational cost is equally high because network and user data can have very high dimensions. In addition to being resource hungry, the highly dimensional network and user data may cause increased detection latency as well as over-fitting. As the implications of these caveats are likely to surface in large scale real network, they are not exclusively addressed in above studies that rely on simulated small-scale network and user population for performance evaluation.

To tackle high dimensional network and user data, Chernogorov et al. [56] have proposed to construct diffusion maps [57] of user handover attempts and successes data. These diffusion maps are obtained through Eigen decomposition of Markov matrix obtained from the diffusion maps of network and user data. The resulting low-dimensional data are used to create cell coverage dominance maps which are then used to detect sleeping cells through k-means clustering [58] of cells into normal and sleeping cell clusters. Alternatively, Chernogorov et al. [59] have employed principal component analysis [60] to reduce the dimensionality of network and user data. The lower dimension data are then used to identify sleeping cell using the FindCBLOF algorithm [61] which separates clusters of normal cells from sleeping cells. Although a direct comparison of the results of the approaches in [56] and [59] has not been presented, the authors separately demonstrate that the proposed algorithms in [56] and [59] can identify sleeping cells and the affected neighboring cells as a result of the outage with high level of accuracy and also quantify the impact of the outages in terms of failed handover and call events.

Alternatively, Zoha et al. [62, 63] have addressed the challenges posed by high dimensionality through multi-dimensional scaling [64]. Multi-dimensional scaling allows easy visualization of the high dimensional network and user data by translating them into fewer dimensions using kernel transformations. This reduces the convergence time of clustering algorithms. In [62], the resulting low dimensional data are passed to Local Outlier Factor (LOF) [65] algorithm for sleeping cell identification, whereas kNN and LOF are compared with each other in [63]. It is observed that kNN outperforms LOF in terms of speed and reliability since LOF can sometimes misclassify normal cells.

The concepts from [62] and [63] are further extended by Zoha et al. [66] to include comparison of LOF with One-class Support Vector Machine (OCSVM) algorithm [67] under different shadowing scenarios. The results show that like kNN, OCSVM algorithm also outperforms LOF. Since LOF is limited to identifying localized outliers to cell clusters, the algorithm is prone to identifying normal cells as sleeping cells. This is avoided in both kNN and OCSVM because of the global approach adopted by both algorithms which only identifies global outliers. However, OCSVM takes significantly longer to train compared to either k-NN or LOF algorithms.

Full Outage Detection in HetNets

In the studies described above, the target topology for outage detection was invariably a homogeneous mobile cellular network of macro cells. Due to the large serving radii of macro cells and high subscriber count associated with them, generating measurements for full outage detection is not a primary concern.

What makes outage detection in HetNets different than homogeneous networks? Cell outage detection in HetNets differs compared to homogeneous networks due to the architectural difference between the two topologies. The low computational ability of small cells, sparse network information due to fewer connected users and proposed future 5th generation solutions such as network densification means that outage detection algorithms for HetNets must be designed separately. The influences of sparse network data on outage detection algorithms plays an extremely important role in the accuracy of the algorithm. Less data can mean less accurate outage detection and an increase in false positive rate.

This fact is demonstrated by Chernov et al. [68] who compare the performance of several learning based outage detection algorithms using radio link and handover failure metrics under different subscriber density levels. The results demonstrate that as the number of subscribers per cell, and consequently samples of performance metric report, starts to decrease, the area under the curve of true positive rate plot decreases exponentially. The authors also demonstrate that this result is true regardless of the outage detection algorithm, which makes it a universal issue. Similar evidence is also implicit in the results presented in [45, 62, 63, 66].

Outage Detection in Sparse Data Environment In a sparse data environment such as a HetNet with control-data separation architecture [69], Onireti et al. [70, 71] have proposed to use Grey first order one variable prediction model [72] to predict downlink received power of the cell at locations where no such data are reported. Outage detection is triggered when sudden changes in user associations are observed. The Grey prediction model predicts the downlink received power of the cells if user associations had remained the same. The predicted information is then compared to actual downlink measurement reports to identify cells in outage. For this purpose, the authors use k-NN and LOF algorithms with k-NN demonstrating higher prediction accuracy just as it did for the case of homogeneous networks [63]. The choice of Grey prediction models in this study stems from the fact that these models have been shown to have higher prediction accuracy in sparse data environments compared to other prediction algorithms such as linear regression.

The algorithm proposed by Wang et al. [73] also refers to a HetNet with control-data separation and outages in small cells are detected through a comparison of predicted versus actual measurements. Measurement prediction is made using collaborative filtering where data collected during normal circumstances from highly correlated users is used to generate predictions for normal cell performance. The predicted data is then passed through sequential hypothesis testing which measures the likelihood of a hypothesis being true and returns the hypothesis with maximum likelihood to be true that is, whether a cell is in outage or not. The proposed algorithm is accurate nearly 75% of the time even in very low user density (1 user per 10000m^2) and very high fading (8 dB).

Finally, Xue et al. [74] have proposed to use simulated radio link failure data of normal and outage-hit cells to overcome the lack of data generated per cell in an ultra-dense HetNet. The authors propose to use kNN clustering to detect outages in HetNets using simulated outages in the network to train the algorithm.

2.2.2 Partial Outage Detection in Cellular Networks

Partial outage detection has historically been the domain of network optimization experts since, unlike full outage, KPI degradation generally does not generate network alarms. Degradation of network performance can lead to poor user quality of experience and may go unnoticed not only because no alarms are generated, but also because unlike full outage, the effect of partial outage may not manifest itself right away in the form of customer complaints. Therefore, it is integral to include partial outage detection in the autonomous self-healing framework. In this subsection recently proposed solutions for partial outage detection in mobile cellular networks are discussed, while Table 2.2 presents a qualitative comparison of the studies included in this subsection. Before presenting techniques for partial outage detection, it is clarified that the terms partial outage and performance degradation are used interchangeably in this subsection.

Heuristic Solutions for Partial Outage Detection

Heuristic solutions leveraging large-scale network data for Partial Outage Detection Karatepe and Zeydan [75] have proposed a heuristic rule based algorithm for network misconfiguration detection due to its scalability and speed of operation compared to learning based approaches especially when dealing with large-scale network data. The authors deploy a Hadoop [76] based data processing cluster to process large amounts of customer call detail record data which contain timestamps, handover

Table 2.2: Qualitative Comparison of Partial Outage Detection Algorithms

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Partial Outage Detection	[75]	Heuristic	Rule based	Homogeneous	Mobility	Centralized	DL
	[78,84]				Quality		
	[79]				Retainability	Distributed	UL
	[77]		Framework		Accessibility, Retainability, Quality	Centralized	DL
	[81]				Quality		
	[82]				Retainability, Mobility		
	[83]				Retainability, Mobility, Quality		
	[86]	Learning Based	Supervised Learning		Retainability, Quality		
	[87,88]				Accessibility, Retainability, Mobility, Quality		
	[91,93]		Unsupervised Learning		Quality		
	[95]				Accessibility, Retainability, Coverage, Quality		DL/UL
	[96]				Quality	Distributed	DL
	[97]				Retainability, Mobility	Centralized	
	[102,104]				Accessibility, Retainability		
	[105]				Retainability		

attempts and successes, and all the cells a user is associated with during the call. After data processing, the information is forwarded to a heuristic algorithm that matches user location with the associated cells and returns any misconfigurations observed during the call. The authors claim that the proposed algorithm can detect misconfigured cells over 82% of the time.

Similarly, Shafiq et al. [77] have proposed to compare cell profiles during routine network operation with performance during heavy traffic situations to identify partial outages. The authors use data from a large mobile cellular network operator to study the trend of several network performance metrics including radio link setup failures, user counts, dropped calls, blocked calls, data session count, data session duration and the average time between consecutive data sessions of a user. The resulting time series profiles of cells during routine operation is compared with their operation during an unusual traffic activity period such as a sporting event. The authors demonstrate that if the normal cell performance during routine operations is known, it is possible to predict the level of cell performance degradation during non-routine events with a high degree of accuracy.

Comparative analysis based Heuristic solutions for Partial Outage Detection In order to facilitate partial outage detection through comparative analysis of normal and degraded cell behavior, Novaczki and Szilagyi [78] propose construction of faultless network performance profiles by fitting network performance metrics such as channel quality to a β-distribution. The detection algorithm compares the parameters of real time cell performance distribution with the faultless performance distribution parameters. In case the real-time parameters differ from faultless profile parameters by a threshold decided by experts, the cell is considered to be suffering partial outage.

Comparison of time-series distribution has also been explored by D'Alconzo et al. [79] who propose to construct univariate probability distribution functions of performance metrics including number of synchronization packets and number of distinct network addresses contacted. The baseline distribution functions are constructed for different temporal resolutions to avoid false detections. The approach in [79] differs from that in [78] since the proposal is to identify partial outages using the Kullback-Leibler divergence [80] or relative entropy of current behavior distribution from baseline behavior distribution, while the behavior distribution modeling is not limited to β-distributions.

Correlational comparison of time-series is an alternative methodology of comparative analysis based techniques for partial outage detection. An example of correlational comparison has been presented by Asghar et al. [81] who have proposed to utilize *Pearson's correlation factor* to match cells based on cell load estimated through the number of active users associated with the cell. The algorithm states that if a cell falls below an arbitrary correlation threshold with multiple cells with which it was previously well correlated, it is considered to be degraded. The authors demonstrate that not only is the proposed method effective for detecting slow partial outages, it is also effective for full outage detection. However, the performance of this algorithm is highly dependent on correlated cells that is, if multiple correlated cells suffer same degradation, it may go undetected. To avoid this pitfall, Muñoz et al. [82] have proposed to correlate successful handover count and call drop count time series of a cell with a synthesized data series that represent partial outage and a reference data series of the cell itself during normal behavior as a preventive measure for false flags. High correlation with synthesized data and low correlation with reference data signifies partial outage. The authors advocate use of time-series correlations over cumulative data correlations since cumulative correlation may hide any short-term degradations in cell performance. However, time-series correlation requires higher and faster computations especially if more performance metrics are included in the comparison process.

Other heuristic solutions for Partial Outage Detection In their work on partial outage detection, Sanchez-Gonzalez et al. [83] propose a decision tree based solution to identify partial outages in a mobile cellular network. The proposed algorithm applies a set of expert-defined rules separating normal and degraded behavior on the uplink and downlink received power measurements, handover failures, and radio link failures to categorize the performance of each cell. If a cell fails said rules, it is considered to be in partial outage and diagnostic functions are initiated. The solution is validated using real-network data where it is able to effectively identify the degraded cells.

Merging heuristic and learning based methodologies, Kumpulainen et al. [84] have proposed a hybrid solution for partial outage detection. The proposed solution evaluates channel quality measurements of a cell over one day and categorizes the quality samples as good, medium and bad based on a heuristic algorithm developed using expert knowledge. Additionally, the solution utilizes fuzzy C-means clustering [85] to generate cell clusters based on the commonality of their profiles in terms of channel quality data distribution over a day [84, Fig. 7]. Based on the similarity of channel quality measurement distribution of a cell over a day with fuzzy clusters, the solution decides if it is degraded. The authors have demonstrated that the proposed solution can not only identify degraded cell performance but also the amount of time it spends as degraded. However, scalability of the solution requires further investigation since the proposed approach is limited to evaluation of one performance metric over a period of a whole day.

Learning based Solutions for Partial Outage Detection

One of the application areas of machine learning is the estimation of network reliability explored by Sattiraju et al. [86]. The authors capture long-term reliability data such as link availability and apply semi-Markov transition process to construct renewal models for normal and degraded network link states. Link reliability is defined as the amount of time network links spend in normal states and two transition actions that is, failure and repair exist in the network. The authors find that lower reliability states are highly absorbing states that is, once a link is sufficiently degraded, its recovery probability approaches zero.

Ciocarlie et al. [87], have also explored the feasibility of deploying time-series averaging based anomaly detection algorithms over variable window lengths. However, unlike the heuristic approaches presented in [78] and [79], the proposed algorithm uses autoregressive integrated moving average to compute predicted KPI values for a cell which are then compared with an ensemble of models for different unspecified KPIs. The authors propose to construct normal and anomalous KPI models using different techniques including empirical cumulative distribution function and SVM with radial basis function kernel. The proposed solution is validated against human experts using visualization tools. Results show that while the proposed approach is able to accurately predict a partial outage, the detection delay between outage occurring and being detected was never less than five hours. Another important concern raised by the authors is the exponential training time of the machine learning algorithms which can make the proposed methodology prohibitive in live networks. The authors have provided further refinement of this approach in [88] by including the utilization of the Kolmogorov-Smirnov test [89] to identify the sliding window size for data streams used to train the SVM models. Another key distinction of [88] over [87] is that the authors use seasonal trend decomposition based on Loess [90] to identify and remove outliers from the original training data to create true performance models.

A key commonality among [78, 79, 87, 88] is the use of individual data streams for input to outage detection algorithms. However, Barreto et al. [91] postulate that using single variable data streams for anomaly detection, though simple, is not always effective. Therefore, the authors have proposed a joint neural network that takes univariate and multivariate data containing channel quality measurements, traffic loads and user throughputs from the network as inputs to generate global and local network performance profiles which are used to detect anomalous cells via percentile based confidence intervals computed over global and local network profiles. The authors demonstrate the efficacy of training a multivariate neural algorithm by presenting a comparison with a single-threshold neural algorithm using several neural network based algorithms including winner-take-all, frequency sensitive competitive learning [92], SOM and neural gas algorithm. Results show that the proposed multivariate partial outage detection algorithm consistently outperforms single-threshold method in terms of false positive alarm rate by 0.6% to over 5.5%.

Frota et al. [93] have presented an extension to the work in [91] where the authors combine the originally proposed multivariate neural networks with Gaussian distribution based SOM clustering algorithm to create a partial outage detection algorithm. The authors use network core traffic statistics to train the Gaussian distribution based SOM clustering algorithm which is compared with multivariate heuristic anomaly detection methods. It is demonstrated that the proposed technique can lower false partial outage detection rate by nearly 30% when trained over 10% of dataset compared to the algorithm proposed in [94] for fault diagnosis in rotating machines. However, the solution proposed in [93] builds on an underlying assumption that network performance metrics such as user count, throughput, noise levels and interference levels are normally distributed which may not hold always true in typical real networks.

Partial Outage Detection using Self-Organizing Maps Self-Organizing Maps are a popular neural networks based clustering technique. SOMs work by projecting input vectors of large size onto a 2-dimensional space using weights obtained by training the underlying neural network. A number of studies have proposed SOM based algorithms for partial outage detection including [91, 93, 95, 96, 97].

As already discussed, Barreto et al. [91] and Frota et al. [93] have used SOMs for comparison based partial outage detection. On the other hand Lehtimäki and Raivio [95] harness the capability of SOMs to arrange similar input vectors of network measurements including call request blocking, traffic channel availability, channel quality, voice call traffic, and uplink/downlink signal strength together. The authors use this arrangement to identify cells with partial outage through k-means clustering algorithm. The proposed scheme is compared with principal component analysis and independent component analysis [98] to detect partial outages in control signaling and traffic channel statistics of a real 2G network. Results show that SOM and principal component analysis performed equally well while outperforming independent component analysis.

Kumpulainen and Hätönen [96] also use SOM based clustering to detect localized partial outages compared to the general global partial outage detection models. The proposed algorithm first creates SOM which is then used to identify best matching units for each node in the map and distance (quantization error) between the two units is calculated. A cell is considered in partial outage if its best matching unit is also in outage and the distance between the two is less than a pre-defined threshold. The authors compare the usage of local partial outage detection model using SOM with Gaussian Mixture Models and k-means clustering with results showing that the local anomaly detection scheme not only detects all the outages but also whenever the activity level of a cell changes.

Gómez-Andrades et al. [97] employ a similar approach to [96] in their work where SOM is used to arrange the cells based on signal strength, quality, call drop and handover failure metrics, and then clustered using Ward's hierarchical clustering [99]. The authors use the Davies-Bouldin index [100] and the Kolmogorov-Smirnov test [101] to set the number of clusters to be created in the SOM. The clusters are labeled as normal or faulty based on expert knowledge. A comparison of the proposed methodology with a rule based algorithm and a Bayesian network classifier shows that the proposed approach outperforms them by 31% and 12% respectively.

Partial Outage Detection using clustering techniques Apart from SOMs, other unsupervised clustering technique such as k-means, density based and hierarchical clustering, topic modeling and LOF clustering have also been explored in literature for partial outage detection. Rezaei et al. [102] have presented a comparison of several supervised partial outage detection schemes in a 2G network. The study uses input data including call blocking and drops, as well as signal quality measurements. Classification techniques explored by the authors for partial outage detection include chi-squared automatic interaction detection [103], quick unbiased efficient statistical tree, Bayesian networks, SVM, and classification and regression trees. The authors find that SVM has the best detection rate among supervised learning techniques (94%) but requires from longer training time while quick unbiased efficient statistical tree has the shortest training time with relatively high accuracy (93%).

Ciocarlie et al. [104] use topic modeling to detect partial outages in a cellular network. The method resembles other clustering techniques with the difference that it assigns a probability to the presence of commonality within the cluster of cells. Once the clusters have been developed, the framework uses domain knowledge to identify which cluster represents anomalous behavior. The approach is tested on real-network data with verification of results performed using visual analysis of data by experts. Alternatively, Dandan et al. [105] have used kernel based LOF anomaly detection which is simply LOF with kernel based distance calculation. The authors propose using kernel based LOF to identify cells in partial outage by associating a degree of anomaly to each cell in a density map for LOF based on kernel Gaussian distance. Normal cells are characterized by having a kGD of 1 and any cells with kernel Gaussian distance above are outliers. The authors also suggest that kernel based LOF can better deal with non-uniform distributions of cells in real datasets compared to typical LOF algorithm. The proposed method has a 91% success rate in detecting outages compared to 70% for normal LOF.

2.2.3 Summary and Insights

Outage detection is one of the most labor intensive process in a mobile cellular network. Researchers have devoted a lot of attention to autonomous full and partial outage detection solutions. Majority of these solutions attempt to detect outages based on coverage metrics such as received signal strength. For outage detection in future 5th generation networks with millimeter wave cell deployment, researchers will need to consider additional metrics. This is because millimeter wave cells have a very high pathloss leading to natural loss of coverage even at a distance of a few hundred meters [106]. A challenge for future studies is to come up with solutions that can detect outages in spite of the coverage limitations of millimeter wave cells.

A common theme among the studies for full and partial outage detection is the growing use of machine learning techniques in general, and unsupervised clustering techniques in particular, for outage detection. This reduces the chances of outages due to unconventional reasons, such as weather anomalies, to be missed. This is not the case for heuristic and supervised machine learning based solutions since they are only trained to look for evidence of outage based on human expert knowledge. This does not mean that unsupervised learning solutions for outage detection can become industry standard as is. Some of the major issues concerning unsupervised learning solutions include:

1 Machine learning techniques in general are prone to errors due to noise in the recorded dataset, as demonstrated in [45, 62, 63, 66]. This means that unsupervised learning solutions deployed for outage detection in areas with high shadowing and multipaths, such as metro hubs, can result in higher false negatives. Future solutions for outage detection must address this issue before they can become practically viable.

2 Majority of techniques for outage detection discussed above only consider spatial data for outage detection purposes. This means that the KPI data used for outage detection is gathered over a set of spatial points representing user locations for one time instance. Therefore, outages detected by these solutions are instantaneous. This raises the issue of outages that are extremely short-lived, have little impact on subscriber quality of experience, and may be gone by the time they can be

compensated. To address this issue, future solutions for outage detection must consider the temporal dimension as well as the spatial dimension of user reported data to differentiate between temporary and long-term outages.

3 Most of the approaches for outage detection reviewed above require a secondary analysis by human expert to confirm the existence of the outage which can add some delay before outage compensation is triggered. This can be an issue in 5th generation networks where low latency and high quality of experience requirements mean that the outages would have to be detected and compensated as quickly as possible.

In addition to addressing the above issues, future studies for outage detection must also incorporate the effects of millimeter wave propagation and capacity enhancement solutions such as massive MIMO. Additionally, detecting partial outages in massive MIMO cells such as failure of some beams will also need to be addressed. Based on the review of existing literature, there are no current studies that expressly include either of these two features which makes them prime candidates for future research in outage detection.

2.3 Outage Diagnosis in Cellular Mobile Networks

Once a network outage (full or partial) is detected, the next phase is to diagnose the underlying cause of the outage. In this section, the literature on Outage Diagnosis is analyzed. Some full outages can trigger fault alarms, thus eliminating the need for full outage detection in those particular cases. However, the exact cause of the failure still needs to be diagnosed. Conversely, the key difficulty in diagnosis with partial outage is the lack of fault alarms associated with the anomalies which makes their diagnosis more difficult, thus requiring sophisticated diagnostic techniques. Table 2.3 provides the qualitative comparison of studies describing full and partial outage diagnosis techniques.

Table 2.3: Qualitative Comparison of Outage Diagnosis Algorithms

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Full Outage Diagnosis	[47]	Heuristic	Rule based	Homogeneous	Retainability, Mobility, Quality	Centralized	UL/DL
	[107]	Learning Based	Supervised Learning		Accessibility, Retainability, Mobility, Quality		
	[108,110]				Retainability		DL
	[102]		Unsupervised Learning		Accessibility, Retainability		
Partial Outage Diagnosis	[77]	Heuristic	Framework		Accessibility, Retainability, Quality		
	[104]	Learning Based	Supervised Learning		Accessibility, Retainability		
	[114]				Accessibility, Retainability, Mobility, Quality		
	[115]				Retainability		
	[97]		Unsupervised Learning		Retainability, Mobility		
	[116]				Quality		UL

2.3.1 Diagnosis of Full Outages in Cellular Networks

A starting point towards full outage diagnosis is building the knowledge-base of possible faults. A quite extensive description of standard faults in cellular networks has been presented in [9] which are applicable to 2G, 3G and 4G networks. The standard documentation also provides alarm descriptions for faults associated with hardware failure, software failure, functionality failure or any other faults that cause the network node to stop performing its routine operations. However, outage diagnostics have remained in the domain of human experts who use their knowledge to identify outage causes. While this method is effective, it cannot remain as the method of choice going forward towards ultra-dense networks.

To this end, some studies have proposed techniques combining expert knowledge with mobile cellular network data to create autonomous outage diagnosis algorithms. One such approach has been demonstrated by Szilágyi and Novaczki [47] which utilizes expert knowledge to create targets for network performance such as channel quality, dropped calls and handover failures. The solution uses weighted sums of the difference of actual KPI value to the target value to calculate a diagnostic score. The algorithm then uses expert knowledge to associate a range of scores with different fault causes to complete the diagnosis process. The proposed technique is validated using real data, with results showing that the algorithm was able to diagnose each outage correctly.

Learning based solutions for Full Outage Diagnosis

Solutions for outage diagnosis using stationary KPI targets derived from expert knowledge can become obsolete quickly in the face of changing network dynamics. Khanafer et al. [107] argue this point and propose an alternate learning based solution using Naïve Bayes Classifier (NBC) to predict possible causes of hardware faults and KPI degradations in the network given the symptoms (failures). The algorithm uses discretized value ranges for various KPIs including blocked calls, dropped calls, connection request failures, and HO failures to indicate normal and faulty performance states. The authors compare two different techniques of KPI value discretization namely percentile based discretization and entropy minimization discretization. Results show that outage diagnoses are over 10% more accurate when entropy minimization discretization is used compared to percentile based discretization.

Barco et al. [108] compare the performance of a NBC for outage diagnosis with a modified NBC which assumes the independence of causal influence [109]. The two methods are compared using data from a live network containing faults such as call drops, handover failures and call blocking with results showing modified NBC to be more efficient in terms of simplicity with the same level of accuracy as regular NBC. However, in order for modified NBC to diagnose outages accurately, it needs knowledge of prior KPI distributions in the event of an outage. Barco et al. [110] have discussed the process of developing this knowledge using a knowledge acquisition tool. The tool combines past diagnoses performed by experts with fault data from the mobile cellular network. The tool takes faults such as high network congestion or high call drops, possible causes such as high interference, observed performance metrics at the time of the fault such as handovers due to high interference, and cell parameter settings. Combining this information, the tool outputs the prior probabilities of different diagnoses.

Unlike other techniques for full outage diagnosis, Rezaei et al. [102] propose to use unsupervised clustering techniques for fault diagnosis and present a comparison of several such techniques including expectation minimization, density based spatial clustering of applications with noise [111], agglomerative hierarchical clustering [112], X-means and k-means clustering. The authors use clustering algorithms to split cells based on their call drops and blocking values. Diagnosis is done by comparing cells in clusters to faulty cells with known diagnosis. Validation is done using expert knowledge to confirm the result of fault diagnosis through clustering. The clustering results are verified using the Silhouette Coefficient [113] and show that expectation minimization is the most successful technique in terms of data clustering with clearest cluster divisions between different sets of faulty cells.

2.3.2 Partial Outage Diagnosis in Cellular Networks

Diagnostic techniques are primarily needed in mobile cellular network for performance degradations scenarios that is, partial outages which generally do not generate any alarms. The operators can define thresholds for KPI values to generate customized alarms; however, apart from being useful only for KPI degradation detection, this technique cannot help in diagnosis or root cause analysis. For this reason, partial outage diagnosis carries great importance in autonomous self-healing solutions for SON.

Shafiq et al. [77] have presented an analysis of real-time measurements from some cells of a large mobile cellular network before, during and after two abnormally high traffic events. The results have been used to present heuristic detection and diagnosis schemes for network congestion and dropped calls during such events along with suggestions on how to rectify these problems. The authors analyze network performance measurement for call connections, link performance and data service performances, and suggest that major issues in terms of call drops and congestion occur when users access the network without coordination. While this would not pose problems during routine network operations since the network is designed to handle such traffic, it becomes an issue during major events or gatherings if additional capacity is not deployed. The analysis presented in the paper solely relies on expert knowledge to derive diagnostic inferences from the real data.

Partial Outage Diagnosis using Learning based Techniques

Other than heuristic techniques, learning based techniques have also been exploited in literature [97, 104, 114, 115, 116] for KPI degradation diagnosis.

Supervised Learning Techniques for Partial Outage Diagnosis Ciocarlie et al. [104] propose to use Markov Logic Networks and Principal Component Analysis to diagnose weather-related and parameter misconfiguration-related partial outages from real network data. The proposed technique generates clusters of degraded cells using Principal Component Analysis which are then passed through a Markov Logic Network for diagnosis. The Markov Logic Network generates a sequence of events that would lead to a degradation in call drop rate, throughput or handover failures, thus leading to the diagnosis. Weights for each sequence of events in the Markov Logic Network leading to a diagnosis are initialized using expert knowledge and updated with each successful and unsuccessful diagnosis. The diagnostic results of the proposed approach have been validated against expert diagnoses. The proposed approach also relies heavily on expert knowledge to generate the event sequences used in the Markov Logic Networks.

Barco et al. [114] present a comparison of the impact of continuous versus discretized data models for auto-diagnostic systems in cellular network using Bayesian network classifier. The authors use β-distributions to construct continuous models from KPI data streams, and selective entropy minimization discretization [117] to construct discrete KPI models. The study uses dropped call rate, blocked call rate, handover blocking, throughput, and active neighbor set update rate KPIs to generate probability of degradation in the network given a set of symptomatic KPI distributions. The results show that continuous models exhibit nearly 10% higher diagnosis accuracy when the training set size is sufficiently large ($\sim$2000 examples) while the discrete models are more accurate ($\sim$20%) when the training data are sparse ($\sim$50 examples).

The results from [114] have been used by Barco et al. [115] to propose a hybrid KPI modeling methodology called Smoothed Bayesian Networks which can decrease the sensitivity of diagnosis accuracy to imprecision in the model parameters. The posterior probabilities of the causes follow a smoother transition near the boundaries between states given their related symptoms in Smoothed Bayesian Networks than in traditional Bayesian networks. The authors compare the accuracy of diagnoses for both Smoothed Bayesian Networks and Discrete Bayesian Networks on real network data for diagnosis of call drop rate. The results suggest that Smoothed Bayesian Networks perform better by almost 10% when there was a certain degree of inaccuracy in the model brought about by sparseness in data. However, Discrete Bayesian Networks perform better on a larger dataset resulting in a more accurate KPI model.

Unsupervised Learning Based Solutions for Partial Outage Diagnosis SOMs have been used frequently not only to detect KPI degradations [93, 95, 96, 97], but also to diagnose them [97, 116]. Gómez-Andrades et al. [97] have used SOM based clustering cell in 4G networks based on call drop rate, channel interference, handover failures, received signal strength, channel quality, and throughput to diagnose the possible cause of performance degradations in the eNBs. The clustering algorithm arranges cells based on their degree of association with other degraded cells by finding the best matching unit for each cell. If a cell is experiencing KPI degradations, it will be clustered with pre-existing degraded cells with known diagnosis. The authors demonstrate that the proposed scheme can outperform rule based algorithms and Bayesian Network Classifiers by ∼32% and ∼12% respectively but takes longer to train compared to the other two techniques. Laiho et al. [116] have proposed a similar solution to diagnose degradations in channel quality and frame error rate in 3G networks with the exception that the cells are clustered using k-means clustering. Cells are diagnosed by taking the diagnosis of the nearest known degraded cell and the results are validated using real-network data and comparing expert diagnoses with the diagnoses generated by the technique.

2.3.3 Summary and Insights

Outage diagnosis is a relatively under-explored aspect of self-healing in mobile cellular networks compared to outage detection and compensation techniques. Part of the reason for this are the standardized fault and alarm codes that are automatically generated in the event of a full outage due to hardware/software failure. However, no such standardized diagnostics exist for partial outages. This is because the same partial outage may be caused by two different sets of circumstances. For this reason, majority of studies on outage diagnosis use supervised learning solutions such as Bayesian networks and Markov logic networks which can associate a probability with each known cause leading to an outage. However, the use of such solutions can be challenging in practical networks since training them would require constructing a database of every root cause resulting in an outage.

To address this issue, future studies on outage diagnosis should focus on how this database of root causes can be created without creating artificial outages. In addition, causes of full and partial outages in future 5th generation networks with millimeter wave cells, massive MIMO and ultra-dense cell deployment must also be explored since they are an uncharted territory as yet.

2.4 Outage Compensation in Cellular Mobile Networks

Outage compensation forms the core element of the self-healing framework; therefore, it is no surprise that, among the three components of self-healing, outage compensation has received the most attention from the research community. Compensation actions and algorithms are designed specifically to provide temporary service to users in case of a full outage or partial outage since both events are not immediately recoverable. While detection and diagnosis of full outage and partial outage in a mobile cellular network require different methodologies, compensatory actions for both events involve similar techniques. The majority of studies on compensation algorithms are presented as a solution for full outage but lend themselves seamlessly to compensation for partial outages.

The key principle of outage compensation is to leverage resources from neighboring cells of outage-affected cells to provide temporary services in affected area. These resources include cell bandwidth and user associations which can be modified using primary parameters such as cell/user equipment transmit powers, and antenna parameters as well as secondary parameters such as neighbor lists and cell selection parameters [40]. In the following subsections, compensation algorithms are presented based on the optimization objective with description of their methodology of optimization along with parameters of choice and other taxonomically significant insights.

2.4.1 Coverage Area Optimization for Outage Compensation

One of the key consequences of network outages and KPI degradations is the loss of network coverage near effected network entity. Several studies [66, 71, 118, 119, 120, 121, 122] have presented outage compensation algorithms that focus on coverage optimization. A list of these studies along with their proposed techniques is presented in Table 2.4.

Table 2.4: Qualitative Comparison of Coverage Optimization Algorithms for Compensation

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Coverage Optimization	[118]	Heuristic	Framework	Homogeneous	Retainability, Coverage, Quality	Centralized	DL
	[119]				Coverage, Quality		UL/DL
	[120]				Coverage, Quality		DL
	[121]	Analytical	Non-convex Optimization		Coverage, Quality		UL/DL
	[122]						DL
	[66]	Learning Based	Reinforcement Learning		Coverage		
	[71]			HetNet			

Choosing the right neighboring cells, optimization parameters, and recovery action

Choice of neighboring cells, optimization parameters, and recovery action plays an important role in the effectiveness of an outage compensation solution and has been investigated in [118], [119], and [120] respectively. The self-healing framework proposed by Asghar et al. [118] defines an outage compensation algorithm that uses received power measurements from users of outage-affected cell to create coverage polygons for neighboring cells. The algorithm then iterates through different antenna configurations of key neighboring cells with potential coverage overlap to outage cell until coverage constraints of all users are met. Additionally, the algorithm monitors downlink throughputs and radio link failures of the neighboring cells to benchmark network recovery. A demonstration of the algorithm by the authors on real network outages shows it can effectively compensate for outages within 2 hours of their occurrence.

The outage compensation framework proposed by Amirijoo et al. [119] compares compensation potential of different control parameters suggested in [40] that is, reference signal power, uplink target received power level P0 and antenna tilt in mitigating outage-induced performance degradations. An iterative algorithm is used to update the parameters of neighboring cells and their results are benchmarked. Results in terms of cell coverage and user throughput indicate that uplink target received power level P0 and antenna tilt are the most effective parameters for improving coverage, while P0 is most effective for improving throughput.

Frenzel et al. [120] discuss choice of optimal recover action based on three inputs that is, the probability of effectiveness of a solution which depends on the outage cause, the preference of the network operator for a recovery action, and the preference of the network operator for a degradation resolution. The authors propose a weighted-sum function which returns the cost of selecting a solution, action and resolution tuple. The proposed framework is flexible to changing network technology as more tuples can be added for future networks; however, the determination of probabilities and preferences requires manual input by experts.

Non-convex Coverage Optimization Techniques for Outage Compensation

Several studies have explored the use of non-convex optimization methods for outage compensation based on the analysis that in a large network with a diverse set of optimization parameters, outage compensation can be a NP-hard non-convex problem. Conversion of the outage compensation problem into a convex problem requires too many generalizations and assumptions which can make the result unsuitable for practical implementation. Jiang et al. [121] and Wenjing et al. [122] base their solutions on this premise and use non-convex optimization techniques to solve the problem of coverage optimization.

Jiang et al. [121] have proposed a cost function minimization approach which uses weighted sum of downlink channel quality and received signal strength. The authors state that the problem is a large scale non-convex optimization problem. Outage compensation is carried out by calculating the optimal uplink target received power P0 using a non-convex optimization technique called immune algorithm [123] for cost function maximization. The authors show that the immune algorithm improves both coverage and channel quality after optimization and can converge in a very short time period. The results, compared against two other techniques [124], [125], show that the proposed methodology can significantly improve coverage post-optimization by 10% without significantly sacrificing cell edge throughput. However, it is observed that the immune algorithm is highly sensitive to initial parameters that is, it may not be able to escape the infeasible solution set if initial parameters are not set correctly.

Similarly, Wenjing et al. [122] propose that the minimization of coverage holes and pilot pollution using downlink pilot powers of neighboring cells for outage compensation is also a non-convex problem. In this study, the authors propose to use a non-convex optimization technique called particle swarm algorithm [126]. Results on the analysis of the algorithm indicate that it is highly efficient in terms of execution time while also recovering over 98% of the coverage area in terms of signal strength without significantly degrading link quality. However, like immune algorithm, the particle swarm algorithm is also highly dependent on initialization parameters for convergence.

Learning based Coverage Optimization Solutions for Outage Compensation

Examples of learning based algorithms for outage detection and diagnosis covered in the previous sections mostly employed classification and clustering techniques. However, reinforcement learning [37] represents the most effective learning based solution for outage compensation algorithms, primarily due to its ability to identify maximum reward strategies over a learning period. One reinforcement learning based solution for outage compensation has been proposed by Zoha et al. [66] within a complete learning based self-healing framework. The outage compensation component of the framework is built upon fuzzy-logic based reinforcement learning which adjusts antenna tilts and cell transmit powers to achieve the desirable compensated performance in terms of cell coverage. The compensation algorithm makes incremental or decremental step changes in optimization parameters after an outage using exploration of new rewards or exploitation of past rewards. The resulting network state from the reinforcement learning database is interpreted through the fuzzy-logic regulator as better or worse than the previous state which then dictates the next step of the reinforcement learning algorithm. The authors demonstrate that the proposed solution can improve post-outage cell edge coverage by 5 dB while also helping to regain mean data rate to pre-outage levels.

A similar approach to [66] has been presented by Onireti et al. [71] for heterogeneous networks with the difference that the fuzzy logic component has been replaced with an actor-critic module for enabling reinforcement learning. The actor-critic module executes an exploratory or exploitative actions such as changing antenna tilt or transmit power of a neighboring cell based on probability of reward learned over time. The critic then evaluates the reward associated with the action taken and updates past rewards and probabilities. The solution is compared against the one presented in [66] with results showing it improves cell coverage and channel quality, particularly for cell edge users, and brings them closer to pre-outage levels.

Table 2.5: Qualitative Comparison of SINR Optimization Algorithms for Compensation

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
SINR Optimization	[124]	Heuristic	Rule Based	Homogeneous	Coverage, Quality	Centralized	UL/DL
	[127]			HetNet	Quality	Distributed	DL
	[128]	Analytical	Convex Optimization				
	[129]	Learning Based	Supervised Learning			Centralized	
	[130]			Homogeneous		Distributed	

2.4.2 SINR Optimization for Outage Compensation

A secondary consequence of outage compensation can be the degradation of SINR of existing users in neighboring cells due to parameter reconfiguration. Therefore, some studies [124, 127, 128, 129, 130] use SINR as the objective to be optimized while including the existing and outage-affected users into the optimization process. This allows them to avoid or minimize the degradation of SINR in areas not affected by outage. Table 2.5 lists a qualitative comparison of the studies targeting SINR optimization for outage compensation.

Heuristic SINR Optimization Solutions for Outage Compensation

Wang et al. [127] present a distributed heuristic outage compensation algorithm for SINR optimization in HetNets. The proposed algorithm minimizes the number of neighboring cells to be reconfigured to achieve desired post-outage SINR. This is done by calculating an inner group of femtocells that can recover the outage-affected femtocell through reconfiguration of transmit powers, and by creating a second outer group of femtocells beyond which no further outage compensation actions can be propagated to prevent the effects of reconfigurations from rippling outwards. The authors demonstrate that the proposed technique requires fewer neighboring cells for SINR optimization compared to other solutions such as [131] while also reducing the number of cells with negative differential SINR compared to pre-outage values. However, the authors also show that as the density of the mobile cellular network increases, the grouping algorithms takes longer to converge.

While the solution in [127] endeavors to find the optimal set of compensating neighbors, the solution put forth by Amirijoo et al. [124] focuses on optimization parameters of the neighboring cells for outage compensation. The algorithm iterates through values of uplink target received power P0 and the antenna tilts of neighboring cells in a homogeneous network. The optimal set is obtained when cell coverage can no longer be improved without affecting SINR. Results indicate that the algorithm can regain pre-outage SINR and coverage values in low network load scenario. Moreover, the compensation potential of the solution in terms of SINR improves as the network load decreases while quality degradation is most visible for high and medium loads.

Convex SINR Optimization Solution for Outage Compensation

Lee et al. [132] present an outage compensation solution using the concept of collaborative resource allocation strategy. The solution is based on reallocation of dedicated bandwidth called Healing Channels (HCs) to provide physical channel resources to users affected by an outage. The concept has been used in associated studies for outage compensation, such as the one by Lee et al. [128] who use a fairness-aware collaborative resource allocation algorithm with the objective of maximizing the sum of logarithmic user rates. The maximization process guarantees user fairness in terms of resource allocation while maximizing user throughput which is directly related to bandwidth and user SINR. Use of log-rate removes the possibility of outage facing users not being allocated any resources and ensures that the rate maximization algorithm treats all users fairly. The proposed scheme is compared with a number of competing resource allocation solutions for outage compensation including regular collaborative resource allocation [128], non-cooperative resource allocation, and the outage compensation solution for wireless sensor networks proposed in [133]. Results show that even though regular collaborative resource allocation offers nearly 10% more mean throughput gains, those gains are overshadowed by large disparity between maximum and minimum throughput levels. On the other hand the fairness aware-collaborative resource allocation algorithm offers a fairer throughput distribution between users.

Learning based SINR Optimization Algorithms for Outage Compensation

Saeed et al. [129], and Moysen and Giupponi [130] employ reinforcement learning techniques to optimize SINR for outage compensation. Saeed et al. [129] propose a fuzzy Q-learning algorithm for compensation of SINR loss due to outage. The algorithm configures transmit power and antenna tilts of neighboring cells iteratively using fuzzy logic control and records the rewards in terms of change in downlink SINR of affected users. The rewards are used by the reinforcement learning algorithm for learning future actions which might lead to better outage compensation in terms of overall DL SINR. Simulation results indicate around 40% of effected users are restored to their original SINR under low load conditions. Similarly, Moysen and Giupponi [130] propose reinforcement learning technique for adjusting neighbor cell coverage using antenna tilt and the downlink transmission power. The approach differs from the one in [129] such that the actions and rewards are calculated using the actor-critic approach discussed previously in [71] for coverage optimization instead of fuzzy logic. To make the algorithm in [130] work, each cell reserves a certain amount of frequency bandwidth for users effected by the outage. Neighboring cells are informed of this bandwidth through the inter-cell interface so that a distributed and cooperative outage compensation solution

Table 2.6: Qualitative Comparison of Capacity Optimization Algorithms for Compensation

<table>
<tr><th>Solution</th><th>Reference</th><th>Methodology</th><th>Sub-Method</th><th>Network Topology</th><th>Performance Metrics</th><th>Control Mechanism</th><th>Direction of Control</th></tr>
<tr><td rowspan="5">Cell Capacity Optimization</td><td>[132,134]</td><td rowspan="3">Analytical</td><td>Convex Optimization</td><td>HetNet</td><td rowspan="5">Accessibility, Quality</td><td>Distributed</td><td rowspan="5">DL</td></tr>
<tr><td>[135]</td><td rowspan="2">Non-convex Optimization</td><td>Homogeneous</td><td>Centralized</td></tr>
<tr><td>[136]</td><td rowspan="2">HetNet</td><td>Distributed</td></tr>
<tr><td>[137]</td><td rowspan="2">Learning Based</td><td rowspan="2">Supervised Learning</td><td rowspan="2">Centralized</td></tr>
<tr><td>[131,138]</td><td>Homogeneous</td></tr>
</table>

can be achieved. The algorithm modifies cell power and antenna tilts in fixed step sizes to exploit the reward of each change which is based on the SINR of users effected by outage. Simulation results indicate that compensation delay is around 500 ms and the approach can compensate 98% of outage users.

One key observation regarding reinforcement learning solutions is that solutions such as the ones presented in [66], [71], [129] and [131] require considerable number of training examples, or outages, before their actions can become effective. This can make effective deployment of such solutions a challenge for mobile cellular network operators.

2.4.3 Cell Capacity Optimization for Outage Compensation

Like degradation in SINR, cell overloading is another consequence of network outages resulting from re-association of affected users to neighboring cells. Moreover, compensatory actions to achieve another objective, such as coverage optimization, can also result in overloading of neighboring cells. This can lead to users being blocked and service requests being discarded, which affects subscriber quality of experience. To circumvent these problems, some studies [131, 132, 134, 135, 136, 137, 138] have focused on outage compensation solutions that focus on optimizing user associations so that the load is fairly distributed among neighboring cells. Table 2.6 presents a qualitative comparison of these studies.

Convex Capacity Optimization Solution for Outage Compensation

As mentioned previously, Lee et al. [132] have proposed an outage compensation solution for HetNets based on collaborative resource allocation. The authors state that users in faulty femtocells cannot be served reliably by the macro cells due to power imbalance between macro cell and small cells, and cell edge performance limitations of macro cells. Therefore, only normal small cells can support users in a faulty small cell. To this end, the reserved HCs of healthy small cells are allocated cooperatively to users of the outage-affected cell. The proposed scheme finds adaptable set of HCs, sub-channels and power allocation to maximize network capacity through convex optimization implemented via an iterative gradient descent algorithm. The solution is quick and improves the total capacity utilization of neighboring cells by nearly 30% while also ensuring fairness in terms of user throughputs.

The collaborative resource allocation solution [132] is further extended by Lee et al. [134] to include collaborative beamforming strategy along with HC allocation for outage compensation. The proposed cooperative beamforming strategy can be performed without power cooperation between nodes, and is also the optimal transmission strategy under individual power constraints. The proposed algorithm performs HC selection through convex optimization based on maximizing system capacity in outage scenario, and then carries out sub-channel allocation and power allocation based on an iterative algorithm. The proposed solution is compared against several resource allocation schemes including regular collaborative resource allocation, equal power allocation [133] and multi-user iterative water filling [139] schemes, with the results showing that for 10 HCs, the proposed algorithm improves the average cell capacity by 5% and user fairness by 10%.

Non-convex Capacity Optimization Solutions for Outage Compensation

As already discussed, a diverse set of problem constraints and parameters can result in the outage compensation problem becoming non-convex. To solve these problems researchers must resort to non-convex optimization methods. One such solution presented by Xia et al. [135] uses genetic algorithm [32] to solve the capacity optimization problem for outage compensation. The problem objective is to minimize the sum of squared difference between capacity utilization of a compensated cell and average network capacity utilization in a homogeneous network. In this study, the genetic algorithm searches over the user association sets including users affected by the outage to find the set that minimizes the capacity utilization objective. Results show that the proposed methodology can improve average resource utilization by at least 5% compared to non-optimized cell capacity utilization. The key advantage of using genetic algorithms is their immunity to initialization point and their ability to get out of the non-feasible zones in the solution set. However, as the size of a system grows larger, the genetic algorithm takes longer to converge.

Rohde and Wietfeld [136] propose to use probabilistic network performance estimation to compensate network outages through ad-hoc deployment of unmanned aerial vehicles (UAVs) mounted relays. Aerial relays can help to exploit unused local capacities of nearby macro cells which cannot be used optimally for connectivity by users or ground based relays when no line of sight link is available. The proposed algorithm builds probabilistic estimation models of interference and throughputs through iterative modification of relay positions to achieve stable cell loads. The authors have compared results using 1 to 6 aerial relays at different distances from outage cell under stationary user locations with results showing that as the number of relays increases and distance from outage cell center decreases, average resource utilization on neighboring cells decreases.

Table 2.7: Qualitative Comparison of Spectral Efficiency Optimization Algorithms for Compensation

Solution	Reference	Methodology	Sub-Method	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Spectral Efficiency Optimization	[140]	Analytical	Convex Optimization	HetNet	Quality	Distributed	DL
	[141]		Game Theory				
	[142]		Multi-objective Optimization			Centralized	

Learning based Capacity Optimization Solutions for Outage Compensation

Aráuz and McClure [137] utilize probabilistic graphic models derived from Bayesian Networks to detect sleeping cells in HetNets and compensate for their outage. Probabilistic graphic models are used to predict user distribution in the outage-affected cell as well. It also allows the categorization of incoming load based on the user distribution and the active cell load without the need to store lengthy baseline data. Each neighboring cell of the faulty cell arranges the predicted load probabilities in increasing order and decides the expansion of its coverage. The authors report that the probabilistic graphic model can successfully predict the expected user distribution and incoming loads for majority of the cases which results in 91.1% of the cases in total coverage recovery with just two sectors cooperating by expanding their footprint. Total recovery is reported for 96% of the cases with three sectors cooperating. The key advantage of proposed approach is that instead of using all neighboring sites or sectors it can yield substantial recovery using only two or three neighboring sectors.

In another study based on supervised learning, Tiwana et al. [131] use statistical learning with constrained optimization for outage compensation. The study utilizes logistic regression to extract the functional relationships between the noisy KPIs including file transfer time, block call rate and drop call rate, and cell resource utilization. These relationships are then processed by an optimization engine to calculate the optimized resource allocation which improves the KPIs of a degraded cell. The process is iterative and converges to the optimum value in few iterations, which makes it suitable for large mobile cellular networks. Results using Monte Carlo simulations indicate 44% improvement in blocked call rate and ~26% improvement in file transfer time.

The algorithm in [131] has been extended by Tiwana [138] to utilize α-fair packet scheduling for radio resource allocation at neighboring cells for outage compensation. At $\alpha = 0$, the scheduler acts as max-throughput scheduler, whereas at $\alpha = 1$, the scheduler becomes proportional fair. Changing the value of α allows compromise between higher capacity (higher throughput for its mobile users) and greater coverage (serving higher number of users concurrently). The results indicate that for $\alpha = 1.3$, the average blocked call rate decreases by 61%, which is a gain of 17% compared to the scheme in [131], while average bit rate falls by 4%. However, for $\alpha = 0.8$, the average bit rate increases by 3% while blocked call rate falls by 5%.

2.4.4 Spectral Efficiency Optimization for Outage Compensation

Spectral efficiency is the ratio of data rate to the used bandwidth and depends on factors which include user distribution, interference, neighboring cell load, geographical SINR distribution, topology, spectrum reuse, modulation schemes, and the number of data links between the communicating nodes, among others. Therefore, spectral efficiency is heavily dependent on the outage compensation actions and has been used as the optimization objective in several studies [140, 141, 142] which are presented below while their qualitative comparison is given in Table 2.7.

The physical implementation of HCs, described in [132], has been discussed by Lee et al. [140] for outage compensation. The study assumes that indoor base stations or small cells can support scalable bandwidths which can be used to compensate users affected by outage in neighboring small cells. Furthermore, it is shown that the maximum spectral efficiency in the event of an outage is achieved when the minimum number of HCs, predetermined by an indoor central unit, is assigned to support users covered by the outage-affected cell. The proposed technique achieves the largest average cell capacity and user fairness in terms of spectral efficiency when compensating cells can be selected by affected users opportunistically for each HC, which is called the multi-cell diversity effect.

Fan and Tian [141] employ game theory to address outage compensation in HetNets. The authors propose a resource allocation scheme in which data transmission can be done cooperatively by the cells. Similar to the approach in [134], channel allocation and cooperation is done at sub-channel level that is, by splitting the bandwidth of healthy cells for the purpose of compensating users affected by the outage. The problem is formulated as a rate maximization coalition game with weights for individual users and is solved using equal power allocation strategy. Once coalitions are formed between users and compensating cells, the authors use Lagrangian multipliers to solve for the optimal power set with the objective function of maximizing rate over a coalition. The approach requires users to go through multiple iterations of cell coalitions until the Pareto-optimal coalition is found which may require significant time expense.

Finally, He et al. [142] present a multi-objective optimization based approach for outage compensation in Cloud-RAN architecture. The optimization objective is the weighted sum of spectral efficiency of edge users of outage-affected remote radio units, and average spectral efficiency of users in outage and compensating remote radio units. Optimization parameters that is, antenna tilt of adjacent remote radio units, are adjusted to expand the coverage in an online-iterative manner. The algorithm is designed to maximize spectral efficiency of compensating cells and users affected by the outage but does not guarantee global maximization. Results show that the solution can recover spectral efficiency of users affected by an outage by 90%.

2.4.5 Summary and Insights

A review of techniques for outage compensation in self-healing mobile cellular networks suggests four basic metrics are targeted in the event of an outage. These are: 1) *coverage area*, 2) *SINR*, 3) *cell capacity/load*, and 4) *spectral efficiency.* The optimization of these metrics is suitable for legacy mobile cellular networks. However, future 5th generation cellular networks will be more complex and quality of experience-focused. This means that outage compensation solutions of the future will have to focus on more than just these basic metrics. Some examples of potential metrics which will be important in 5th generation cellular networks include energy efficiency, service latency, and throughput fluctuations [38].

Ensuring service latency by itself will be a major challenge for network operators in 5th generation mobile cellular networks due to the complex nature of these networks. A review of outage compensation studies suggests that the most popular techniques for outage compensation are convex and non-convex optimization. Both of these techniques are computationally tedious and require far more time than would be acceptable in a 5th generation network. Furthermore, as these networks become denser, and the number of tunable parameters increases, the optimization process will get slower and more complex. Thus, one of the foremost challenges for future outage compensation solutions will be to reduce the time it takes for an optimization algorithm to reach its solution. Exploring trade-offs between different metrics for outage compensation in 5th generation networks will also be an interesting future area of study.

Another important research area in terms of outage compensation solutions is their integration into the larger SON framework. The SON framework includes technique for self-optimization which oftentimes use the same parameters as outage compensation techniques. For example, coverage and capacity optimization solutions use transmit powers, antenna tilts and beam-forming parameters which are also key for outage compensation techniques, as evidenced by the review of studies above. To avoid this issue, network operators will need to incorporate a self-coordination entity to resolve such conflicts. Additionally, coordination will be important to avoid the triggering of self-optimization as a result of some outage compensation action. For example, changing the azimuth of a cell to provide coverage to subscribers of a cell affected by a full outage might trigger coverage and capacity optimization in a neighboring cell. This could, in turn, trigger a cascade of changes in neighboring cells. While some studies have proposed the use of exclusion zones to reduce the impact of outage compensation on other cells [127], this area needs further research.

Finally, like existing outage detection and outage diagnosis techniques, outage compensation techniques do not incorporate technologies such as massive multi-input multi-output antennas and millimeter spectrum utilization. To enable self-healing in 5th generation networks, more solutions must be explored which focus on these technologies, making this a key area of research.

2.5 Proactive Self-Healing Techniques for Partial Outage Avoidance

So far the discussion in this chapter has revolved around reactive techniques for dealing with full and partial outages. However, partial outages caused by parametric misconfiguration, which may happen due to human error or due to multiple SON functions executing conflicting parametric changes, are of equal if not higher importance especially in light of growing outage probabilities in future mobile cellular networks as shown in Figs. 1.3 and 1.4. As existing mobile cellular networks upgrade to more emerging technologies, the use of SON functions will become inevitable [143], and without any proper coordination between them, so too will partial outages due to their conflicts [7].

A wide variety of these conflicts have been described by Lateef et al. [144] who have also discussed the potential causes of these conflicts including due to parametric dependencies, KPI dependencies, and logical dependencies of different SON functions among others. To avoid such conflicts, Kemptner and Tsvetkov [145] and Galani et al. [146]

have separately presented conceptual road maps for how a SON coordination system can be created in the network which will have the authority to control and streamline SON function actions for smooth network operation.

Since the official addition of SON coordination into the SON framework by the 3GPP [147], research on SON coordination solutions has grown significantly with a diverse array of solution methodologies being proposed. These methodologies can be grouped together into four overarching techniques:

- Frameworks for SON Coordination
- Heuristic Solutions for SON Coordination
- Analytical Solutions for SON Coordination
- Machine Learning Solutions for SON Coordination

Note that due to the nature of solutions for SON coordination, the following revised definitions for frameworks and heuristic solutions are used for all subsequent discussion. Frameworks refer to general guidelines for the implementation of a solution and as such do not have well defined internal structures. Conversely, heuristic solutions refer to strictly guided algorithms that follow a set of rules to reach upon an optimal value of some objective function. In the following subsections a discussion of the studies that fall within each of the methodologies for SON coordination, as shown in the self-healing taxonomy in Fig. 2.1, is presented. Additionally, individual characteristics of each study based on the key components of self-healing techniques are presented in tabular form in each subsection.

2.5.1 Frameworks for SON Function Coordination

Frameworks have historically been the most common approach towards solving the SON coordination problem. A list of all framework based SON coordination solutions is given

Table 2.8: Qualitative Comparison of Frameworks for SON Coordination

Solution	Reference	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
SON Coordination Frameworks	[148]	Heterogeneous	Coverage, Quality, Accessibility	Centralized	DL
	[149]	Homogeneous	Coverage, Quality		
	[150, 151, 152, 153, 154, 155, 156]		Coverage, Quality, Mobility		
	[157]		Quality	Distributed	

in Table 2.8. One of earliest frameworks for SON coordination is from Tsagkaris et al. [148] who leverage the concept of SON coordinator presented in [146] to present a SON coordination framework for heterogeneous networks. The authors propose four modes of SON coordination: 1) *hierarchical coordination* for SON functions operating at same entities but different time scales, 2) *parallel coordination* for SON functions operating at different entities but at same times scales, 3) *synchronous coordination* where the actions of different SON functions are synchronized to avoid overlap, and 4) *asynchronous coordination* where the SON functions initiate actions without waiting for the impact of other actions. It is postulated that a combination of synchronous and hierarchical coordination would be the way forward for future heterogeneous networks and demonstrate the efficacy of the proposed approach by coordinating between load balancing of macro cells and backhaul resource allocation of small cells. The authors compare network coverage, downlink throughput and backhaul link capacity as a result of the proposed technique against a non-coordinated network and show that their approach achieves better performance with fewer SON function conflicts.

In [149], the authors present a SON coordination framework that has two means of avoiding conflicts between SON function. First it implements a three step SON function coordination mechanism which can: 1) *enable*, 2) *disable*, and 3) *suspend* a SON function for conflict prevention. Secondly, it implements a three step SON action coordination mechanism that can: 1) *stop*, 2) *suspend*, and 3) *modify* a SON action to avoid potential conflicts with other SON actions. The authors state that the priorities for different SON functions and their actions will be provided by the operator in the form of high-level KPIs which will be broken down by the coordinator to decide how to navigate between different SON function conflicts.

Schmelz et al. [150] present a brief description of the different types of conflicts that could occur in a 3rd and 4th generation mobile cellular network. The authors leverage this information to propose priority based SON coordination which operates using three unique sub-systems: 1) *policy translator* converts high level KPI priorities by the network operator into lower level KPI and SON function priorities, 2) *alignment function* aligns the actions of different SON functions into a queue based on the priority of their respective KPIs and actions, and 3) *data analyzer* collects the data for results of SON action execution for feedback to the network operator for future priority adjustment.

Ali-Tolppa and Tsvetkov [151] have proposed a SON coordination framework that uses a similar approach to the one presented in [150] with the assumption that the impact of a SON action is known beforehand. The first stage of the proposed framework streamlines all SON actions by weeding out all conflicting actions. In the next phase the framework executes all non-conflicting actions and verifies their impact for future action executions. Similarly, Bandh et al. [152] have proposed a policy based SON coordination framework which takes KPI priorities from the network operator and translates them into priorities for different SON functions and their actions. Additionally, the proposed framework measures the impact time of each SON action including preparation, execution and monitoring, and prevents any other SON actions from interfering during that time period. The authors demonstrate the efficacy of their proposed algorithm by coordinating the antenna tilt and transmit power changes initiated by a coverage and capacity optimization function to optimize network coverage. Results show that the proposed approach achieves better network coverage and fewer outages than an uncoordinated system.

Bandh and Schmelz [153] have carried the concept of separating SON function actions in time a step further by proposing that the execution cycle of a SON action should be divided into four phases: 1) *enforcement time* is the time a SON action takes to be fully implemented, 2) *visibility delay* is the time it takes for a SON action to become visible in terms of updated parameter values, 3) *protection time* which is required to fully observe the impact of a SON action on the network, and 4) *relevance time* during which the impact of SON action remains relevant. The authors suggest that SON actions can be scheduled in such a way that they do not interfere with the execution cycle of other SON actions, thus removing the possibility of a conflict.

Similary, Romeikat et al. [154] have proposed to use the information regarding the priority, impact area and time duration of a SON function action to coordinate between different SON functions. The proposed framework has three states: 1) *acknowledge* where a SON action request is acknowledged, 2) *reject* where a SON action request is rejected, and 3) *rollback* where a previously implemented SON action is rolled back. Concurrent requests to the coordinator are handled based on a dynamic priority allocation scheme. If request from a function is accepted, its priority is lowered and if it is rejected, its priority is raised. The authors demonstrate the effectiveness of their framework by coordinating the antenna tilt and transmission power changes by a coverage and capacity optimization function and show that the framework allows fewer call drops and higher data rates for users compared to an uncoordinated system.

Another similar approach is proposed by Mwanje and Mitschele-Thiel [155] who propose to use the knowledge about the targeted network entities and the execution time of different SON function actions to schedule their implementation in the network. The authors postulate that conflicting SON actions can be allowed to execute concurrently if they are separated in space by at least one tier of cells, or if they are separated in terms of time scales for action visibility. The authors use the example of handover optimization and load balancing to show that their proposed scheduling system can achieve more balanced cell loads, fewer handover failures and fewer radio link failures compared to uncoordinated systems.

Yet another SON action scheduling framework is proposed by Stamatelatos et al. [156] who suggest that all SON actions must go through a four-stage process that: 1) *monitors*, 2) *analyzes*, 3) *plans*, and 4) *executes* each SON action based on a set of priorities provided by the network operator. Based on the proposed framework, no two SON functions can execute actions at the same time at any spatial or temporal level regardless of the fact that there may not be any conflict between them.

Finally, Moysen and Giupponi [157] propose a SON coordination framework that leverages past information about interactions of SON functions to predict future conflicts. The proposed framework then implements a conflict prevention mechanism that takes advantage of game theory to coordinate between different SON functions. The authors give the example of coverage and capacity optimization and intercell interference coordination to show that using past data about SON conflicts allows the framework to learn pareto optimal values of antenna tilts and transmission powers so that no cell can improve its average downlink throughput without degrading the performance of another cell.

2.5.2 Heuristic Solutions for SON Function Coordination

Heuristic solutions offer a more explicit method of coordinating between SON functions by defining the objectives of their algorithms clearly and the parameters that would be optimized. This makes heuristic solutions another popular choice for SON function coordination. Table 2.9 provides a quantitative comparison of heuristic solutions for SON function coordination.

One such solution is proposed by Vlacheas et al. [158] who have used operator priorities

Table 2.9: Qualitative Comparison of Heuristic Solutions for SON Coordination

Solution	Reference	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Heuristic Solutions for SON Coordination	[158]	Homogeneous	Coverage, Quality, Accessibility	Centralized	DL
	[159]		Quality, Mobility, Accessibility		
	[160]	Heterogeneous	Coverage, Quality		
	[161]	Homogeneous	Coverage, Quality, Accessibility		
	[162]	Heterogeneous	Retainability, Mobility, Accessibility	Distributed	
	[163]	Homogeneous	Coverage, Retainability, Mobility, Accessibility		
	[164]		Retainability, Mobility		
	[165]		Coverage, Retainability		
	[166]		Retainability, Mobility		
	[167]		Coverage, Retainability, Mobility, Accessibility		
	[168]		Retainability, Mobility, Accessibility		
	[169]	Heterogeneous	Coverage, Quality	Centralized	
	[171]		Quality		
	[172]	Homogeneous	Quality	Distributed	

to optimize a multi-objective optimization problem involving multiple SON functions for their coordination. The authors define the objective function as weighted sum of the coverage and capacity optimization and intercell interference control SON functions. The authors then propose a heuristic algorithm that solves the multi-objective problem to obtain the optimal physical resources to be allocated to each user and cell transmit powers. The authors compare the resulting spectral efficiency and cell capacity utilization to demonstrate that the proposed algorithm performs better than an uncoordinated system.

Multi-objective optimization is a common approach for solving the SON coordination problem and majority of the heuristic solutions solve some form of it through a set of pre-defined steps. One such approach is presented in [159] where the authors propose to coordinate between coverage and capacity optimization, load balancing and handover optimization SON functions by solving the maximization problem of weighted sum of channel quality, ping-pong handover, and cell load KPIs. The authors propose a heuristic algorithm that searches for the optimal values of antenna tilts and cell individual offset parameters for which the objective function is maximized.

Bjornson et al. [160] discuss the effectiveness of multi-objective optimization for solving the SON coordination problem and suggest that this is the best possible approach going forward. The authors show that it is possible to solve a weighted sum of KPIs by looking for the pareto optimal values of network parameters. To back this statement, the authors present a case study of spectral efficiency, downlink throughput, and energy efficiency optimization by finding the pareto optimal values of serving antenna count, number of allowed users per cell and cell transmit powers. Results show that the proposed algorithm performs better than an uncoordinated system in terms of overall downlink throughput and blocked user requests.

In a similar approach to [159], Dinh and Kuklinski [161] propose to use heuristic algorithms on conflicting SON functions. However, the authors propose to exploit the time sensitivity of changes from different SON function to enable their coordination. The authors propose two algorithms for coordination between handover optimization, load balancing and admission control SON functions. The first algorithm optimizes handover offset parameter to minimize the weighted sum radio link failure, ping-pong handover, handover failure and outage KPIs for load balancing. The authors assume that load balancing will not interfere with handover optimization due to bigger time scale. The other algorithm is designed to minimize the same objective function using time to trigger and cell hysteresis parameter for handover optimization on a much finer time scale. The authors use real network data to demonstrate how the two algorithms can operate concurrently without interfering with each other.

In [162], the authors propose a heuristic solution for SON coordination that works by leveraging SON function priorities defined by the network operator, and defining two event types: 1) *undershot events*, where the coordination algorithm initiates changes when the current target KPI value is below a certain threshold, and 2) *overshot events*, where the changes are initiated as a result of target KPI values exceeding set thresholds. The authors demonstrate the results for a heuristic algorithm by optimizing handover count, call request, handover failure and blocked call KPIs using dynamic cell capacity and reserved handover capacity parameters, and show that the proposed algorithm outperforms a system with no SON coordination.

Two of the most popular SON functions for heuristic coordination solutions are mobility robustness optimization (MRO) and mobility load balancing (MLB) functions. Coordination between the two functions has been explored in several studies including [163, 164, 165, 166, 167, 168]. In [163], the authors have proposed a heuristic solution to coordinate between the two SON functions which polls both cells and users for degradation in network coverage, blocked calls, dropped calls, handover failures, ping-pong handovers and capacity demand KPIs. If a degradation is detected, the proposed algorithm initiates search for the optimal values of cell individual offset and cell hysteresis parameters that minimize the weighted sum of these KPIs.

Similarly, Lobinger and Stefanski [164] propose a priority based heuristic solution to coordinate between MRO and MLB where MRO is given higher priority over MLB. The authors demonstrate the efficacy of this proposition by presenting three versions of their algorithm: 1) where MLB is switched off and only MRO is active, 2) where MRO is switched off and MLB is active, and 3) where MLB and MRO are enabled but MLB is suspended to allows actions from MRO to execute in order to avoid any KPI degradations. The authors show that the proposed solution with mixed MRO and MLB achieves better network reliability with fewer dropped calls and handover failures compared to an uncoordinated system.

A similar approach is proposed in [165] where the authors prioritize MRO over MLB by defining ranges of cell individual offset, time to trigger and cell hysteresis parameters within which the MLB function can operate without effecting MRO. The heuristic algorithm proposed by the authors monitors network coverage and radio link failure KPIs to dynamically update the parameter ranges for MLB. Liu et al. [166] also prioritize MRO over MLB in their proposed approach to tune handover offset, cell individual offset and time to trigger parameters for optimization of handover failure, call drop rate and ping-pong handover KPIs. The algorithm assumes that majority of the user are stationary which means MLB does not act frequently. Conversely, MRO is a frequent actor in modifying the optimization parameters and is given a higher priority over MLB for its actions. The authors demonstrate that by keeping cell individual offset static and tuning the other two parameters using MRO only, the solution can achieve better KPIs than an uncoordinated system.

In [167], the authors propose to resolve the conflict between MRO and MLB by combining the actions of the two functions into one unique action. The authors use the example of handover margin parameter to demonstrate their solution. The proposed algorithm combines the optimal values from the two algorithms to produce one final handover margin value. For example, if MLB proposed a negative value and MRO proposes a positive value, the final action is the sum of the two values. The authors show that proposed algorithm achieves better network coverage, handover success rate, call drop rate and blocked call rate than uncoordinated actions of the two functions.

Mwanje and Mitschele-Thiel [168] propose to use a game theory based heuristic algorithm that searches for the pareto optimal values of time to trigger, cell hysteresis and cell individual offset for coordination between MRO and MLB. The proposed algorithm initiates bargaining games between sets of adjacent cells such that the values for radio link failure, ping-pong handover, handover count and cell load KPIs become quasi-stationary meaning values for one cell do not improve without degrading KPIs for an adjacent cell.

Heuristic algorithms are also a popular choice for coordinating between other SON functions. One such example is the coordination between handover optimization and energy efficiency which is explored in [169]. The authors suggest that there is a parametric conflict between the two functions in terms of cell transmission powers and on/off states. The authors use integer linear programming [170] to optimize the values of the two parameters for maximization of a weighted sum of the two SON functions. The authors show that using the proposed solution, network operators can achieve better energy savings without degrading handover KPIs compared to an uncoordinated system.

Coordination between coverage and capacity optimization and energy efficiency is the focus of [171] where the authors use cell on/off switching to reduce network energy consumption with constraints on minimum assured average downlink throughput and network coverage reliability. The proposed algorithm monitors average downlink SINR values reported by all users compared to a pre-defined threshold to identify when a cell can be switched on or off. In contrast to other heuristic SON coordination techniques, spatial SON coordination is explored in [172] where the authors propose to coordinate the resource allocation SON function between multiple adjacent cells. The authors propose a heuristic solution that creates bargaining games between adjacent cells to optimize their transmission powers in order to maximize the joint downlink throughput of all adjacent cells. The result of the algorithm is a set of pareto optimal transmission powers for which the average cell throughput is maximized.

2.5.3 Analytical Solutions for SON Function Coordination

Table 2.10: Qualitative Comparison of Analytical Solutions for SON Coordination

<table>
<tr><th>Solution</th><th>Reference</th><th>Network Topology</th><th>Performance Metrics</th><th>Control Mechanism</th><th>Direction of Control</th></tr>
<tr><td rowspan="6">Analytical Solutions for SON Coordination</td><td>[174]</td><td rowspan="3">Homogeneous</td><td>Coverage, Accessibility</td><td>Distributed</td><td>DL/UL</td></tr>
<tr><td>[176]</td><td>Coverage, Accessibility, Quality</td><td rowspan="4">Centralized</td><td>DL</td></tr>
<tr><td>[178]</td><td>Quality</td><td>UL/DL</td></tr>
<tr><td>[179]</td><td rowspan="3">Heterogeneous</td><td>Coverage, Quality, Accessibility</td><td rowspan="3">DL</td></tr>
<tr><td>[180]</td><td rowspan="2">Quality</td></tr>
<tr><td>[181]</td><td>Distributed</td></tr>
</table>

Analytical solutions for SON coordination are generally a product of translating the relationship between KPIs targeted by SON functions and the optimization parameters into analytical expressions which are either tractable or can be solved using common stochastic optimization techniques such as genetic algorithms, simulated annealing, and pattern search among others [173]. Analytical solutions allow network operators in-depth visibility of the workings of SON coordination and offer greater control over SON functions. Table 2.10 gives a list of SON coordination solution that leverage analytical techniques.

An example of an analytical solution for SON coordination is presented by Tall et al. [174] who propose to model the relationships between KPIs and their optimization parameters in the form of sets of linear equations which can be solved using convex optimization. The authors establish linearity of these relationships by employing the Hartman-Großman theorem [175] for finding the point of the function near a hyperbolic equilibrium point. The authors demonstrate the effectiveness of their solution using the example of admission control and load balancing SON functions. Using transmission power and bandwidth as the optimization parameters, the authors show that the proposed solution results in fewer blocked requests, more balanced cell loads and lower outage probability compared to an uncoordinated system.

In [176] the authors have presented an analytical solution for joint coordination of coverage and capacity optimization and load balancing SON functions in a homogeneous macro cell network using antenna tilts and cell individual offset parameters. The authors propose to minimize the log sum of cell loads which is calculated by adding the number of resources allocated to each user. User association is done using a combination of offered downlink throughput at the cell and the downlink received signal strength. The load minimization problem is solved using a variation of Nelder-Mead pattern search algorithm [177] over the antenna tilts and cell individual offsets of all cells. The authors show that compared to coverage and capacity optimization only and load balancing only, the proposed solution performs much better in terms of load distribution and overall downlink throughput.

[178] is an extension of [176] where the authors use a similar pattern search algorithm to optimize antenna tilts for maximization of uplink and downlink throughputs with sparse network knowledge. The sparsity of knowledge comes in the form of missing antenna tilt and throughput relationship and user location information. Another extension of [176] is [179] where the authors optimize cell transmission powers and the allocation of sub-band resources for coordination between MLB and inter-cell interference control. Users are associated with cells based on the offered throughput and received signal strength. The objective of the proposed solution is to minimize the log sum of cell loads for given transmission powers and sub-band allocations. The optimal values of the

parameters are obtained using the pattern search algorithm described in the previous studies. The authors suggest that the proposed algorithm can also achieve energy saving implicitly by optimally allocating bandwidth resources to users allowing network operator to potentially switch off unused cells.

A similar problem to the one described in [179] is solved in [180] where the authors propose an analytical solution for maximizing downlink SINR across all the users using transmit power control which results in coordination between admission control and inter-cell interference control SON functions. The proposed solution transforms non-convex power control into convex problem using the log transform and then solves it using convex optimization.

In [181] the authors propose to jointly formulate admission control and load balancing SON functions into a single tractable expression. This is done by assuming that small and macro cells in a network are distributed according to a Poisson point process [182]. This allows the authors to exploit the properties of Poisson point processes for the formulation of a tractable expression for downlink SINR with cell individual offset and transmission power parameters. The objective of the problem is to maximize the average downlink SINR for all users and the authors use it to demonstrate the upper-bound of SINR that can be achieved for different cell densities across the network.

2.5.4 Machine Learning Solutions for SON Function Coordination

Machine learning techniques have seen a sharp uptick in usage for solving optimization problems in different domains and SON function coordination is no different. Table 2.11 presents a list of all machine learning based SON coordination solutions available in literature. One of the earliest examples of machine learning being used for SON coordination is by Deb and Monogioudis [183] who compare an analytical and machine learning solution for maximizing uplink SINR across the network by coordinating the uplink transmission powers of users in the network. The authors formulate the up-

Table 2.11: Qualitative Comparison of Machine Learning Solutions for SON Coordination

Solution	Reference	Network Topology	Performance Metrics	Control Mechanism	Direction of Control
Machine Learning Solutions for SON Coordination	[183]	Homogeneous	Quality	Distributed	UL
	[184]		Coverage, Quality		DL
	[185]		Accessibility		
	[186]	Heterogeneous	Quality		
	[187]	Homogeneous	Coverage, Mobility, Quality, Accessibility		

link SINR as a function of randomly overlapping uplink transmissions and use convex optimization, stochastic optimization and linear regression to solve the optimization problem. The authors show that the proposed solution improves uplink SINR across all the user compared to an uncoordinated system with stochastic optimization achieving better results than linear regression.

The most commonly used machine learning technique for SON coordination is reinforcement learning [37] which has been used in numerous studies due to its exploration and exploitation properties for reward calculation. In [184], the authors use reinforcement learning with a Markov decision process to coordinate between several SON functions. The authors propose that each SON function can be modeled as a Markov decision process. Within each decision process there are states which point to good or bad network KPI values. To explore these states and to exploit knowledge of past rewards in terms of KPI improvement, the authors use reinforcement learning. Thus, the final SON function action is decided by the reinforcement learning algorithm to maximize cumulative reward for all SON functions. The approach is demonstrated using the example of channel quality, SINR and network coverage KPIs which are jointly optimized using cell individual offset, time to trigger, handover offset, antenna tilt and antenna azimuth parameters.

A similar approach is presented in [185] with the difference that the authors use the example of spatial coordination between different cells to optimize cell loads between them using cell individual offset and transmit power parameters. The authors demonstrate that the proposed solution can improve cell load distribution and downlink throughput compared to an uncoordinated system. Another spatial SON coordination solution leveraging reinforcement learning is presented by Simsek et al. [186] who propose to use reinforcement learning for coordinated selection of antenna beams from neighboring small cells for maximization of average downlink throughput. The reinforcement learning algorithm exploits knowledge of channel quality matrices for different antenna beamwidth configuration and explores new configurations to achieve the optimal beamwidth selection. The authors compare their solution against the minimization of mean square error based beam selection method which is the most commonly used approach in real network. Results show that the proposed scheme can achieve greater average throughput compared to the state-of-the-art technique.

A variation of reinforcement learning is Q-learning which does not require a complete analytical breakdown of relation between states, actions and rewards and instead relies on the long term average rewards obtained using repeated exploration and exploitation. The approach is used by Mwanje et al. [187] for coordination between MRO and MLB SON functions. The reward which is to be maximized by the reinforcement learning solution is a weighted sum of radio link failure, handover count, and ping-pong handover for MRO, while for MLB it is the cumulative change in cell loads. The solution defines actions as changes in time to trigger, cell individual offset and handover offset parameters. The authors demonstrate that when the two functions are deployed together using the action scheduling algorithm defined in [155], the overall reward, that is, the target KPIs are much higher than the KPI values for an uncoordinated system.

2.5.5 *Summary and Insights*

The survey of SON coordination techniques presented above shows how far SON coordination has come from being a nascent idea to a reality and necessity in future mobile cellular networks of 5th generation and beyond. However, there is still a long way to go until the problem of SON coordination can be considered solved. This is because each of the methodologies listed above has certain limitations that must be overcome before SON coordination will become universally employed.

A quick overview of SON coordination frameworks shows that these are usually highly abstract solutions that are meant to be more like guiding principles rather than practical solutions. Furthermore, some of the assumption in frameworks such as the stationarity of the network, and prioritization of some SON functions over others will clearly not be optimal when faced with fast changing environments of ultra-dense multi-technology heterogeneous networks that will form the backbone of future networks.

On the other hand, heuristic solutions are efficient and provide quick solutions to small coordination problems. However, as the number of SON functions, optimization parameters and number of nodes will increase, heuristic solutions will find it more difficult to come up with solution in acceptable amount of time, thus violating the latency requirements of future mobile cellular networks.

The two best approaches for comprehensive SON coordination and proactive partial outage avoidance are using analytical and machine learning solutions since analytical solutions fully characterize SON functions, KPIs and their relationships with optimization parameters while machine learning solutions can exploit large amounts of past data to attain optimal SON coordination. However, analytical solutions are complex to derive and often require significant assumptions that may not hold true in practical networks. Similarly, machine learning algorithms need large amounts of training data and training time until they become practically viable, and even then they need to be periodically updated lest they fall victims to the problem of concept drift [8].

Based on this analysis, the best potential solution for SON coordination appears to be a combination of all four methodologies, that is, a framework for SON coordination that leverages big data analytics and machine learning techniques to create semi-analytical models for KPI and optimization parameter relationships with the choice of SON function dictated by network state extracted from user and cell level data. One such framework along with a solution for partial outage avoidance leveraging it is presented in chapter 5 of this dissertation.

2.6 Challenges and Future Prospects in Self-healing for 5G and beyond

In order for future 5G mobile cellular networks to achieve the desired gains laid out by the research and standardization community [38], SON solutions must play a far greater role than ever before [8]. This means that future mobile cellular networks must be intelligent, proactive, knowledge-rich and interactive at the same time. To achieve this goal, researchers must develop solutions which enable the network to achieve self-reliance, and harness the power of vast quantities of data generated by the users and network nodes to empower such solutions. However, Self-healing in future mobile cellular networks must cope with several research challenges which have been discussed below.

Challenge 1: Coping with increased outages due to increased network density

Network densification, driven by the need to meet capacity and data rate requirements of 5G mobile cellular networks, means that future mobile cellular networks will have to handle far more network nodes than before. Higher cell densities coupled with technologies such as millimeter wave spectrum utilization, and more configurable parameters will result in frequent network outages driven by both parametric misconfigurations and routine equipment failures as demonstrated in Figs. 1.3 and 1.4.

Possible Solution and Future Direction

A number of research areas have been highlighted in recent studies that can aid in dealing with network outages quickly and efficiently, especially in the context of dense and ultra-dense HetNets. One such approach is the control-data separation architecture (CDSA) [69] where the control functionality lies with macro cells while data transmission is handled by small cells. This adds redundancy to the network architecture. For example, in the event of a small cell failure, the macro cell can handle both control and data transmissions to the affected users.

Furthermore, with the development of UAV technology for enabling 5G mobile cellular networks, UAV based outage compensation techniques, such as the one presented in [136], can become ubiquitous. Additionally, decreasing cost of small cell deployment will mean network densification itself can be used to create redundancies within the network such that the UE-to-cell ratio becomes less than 1. This will mean that in the event of a small cell failure, there will be additional small cells ready to serve the users without effecting their quality of experience. Network densification will play an especially significant role in the context of millimeter wave cells where coverage will be limited to line of sight links and outages due to link obstruction will be frequent.

Challenge 2: Coping with sparsity of data due to smaller number of users per cell

With network densification, another challenge arises in the form of data sparsity due to fewer users per cell. This will make full outage detection and partial outage detection extremely difficult since there will not be enough measurements to accurately distinguish between cell edge users and outage scenarios. Moreover, even though the expected throughput per user will increase, decreasing user density per cell will mean fewer users will consume more data, hence data sparsity will stay an issue for Self-healing in 5G mobile cellular networks.

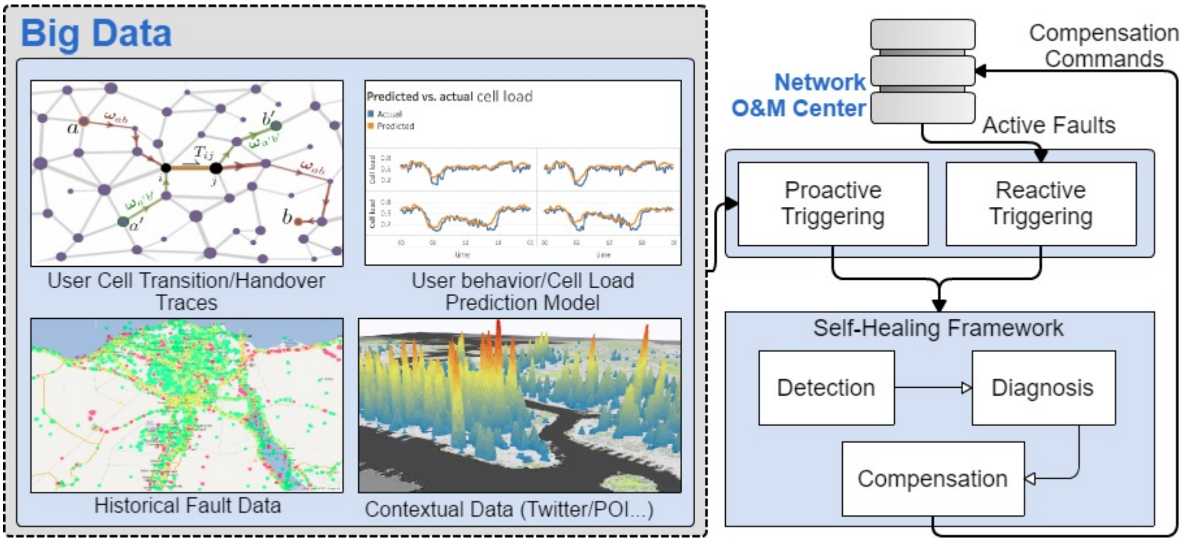

Fig. 2.2: Proactive Self-Healing Framework for Future Cellular Networks

Possible Solution and Research Direction

As was shown in section 2.2, the overwhelming majority of full outage detection and partial outage detection solutions relied on machine learning techniques. However, unlike analytical or heuristic techniques, learning based algorithms are overwhelmingly dependent on data from the network, which can be sparse especially in the case of ultra-dense small cell deployment. To improve the accuracy of learning based outage detection solutions and to counter data sparsity in future mobile cellular networks, measurement prediction techniques can be used. Predictive techniques such as Grey prediction model [72], and smoothing techniques such as Witten-Bell smoothing [188] and Good-Turing smoothing [189] can be used to remove knowledge gaps in the measurement data.

Challenge 3: Meeting 5G latency requirements in self-healing

5G mobile cellular networks are expected to have end-to-end data latency of 1 ms. This means that any Self-healing solution deployed in the network must be able to detect, diagnose and compensate any outage in far less time than state-of-the-art solutions.

Possible Solution and Research Direction

While solutions to proactively avoid outages due to parametric misconfiguration have been discussed extensively in section 2.5, future self-healing solutions must be fully proactive in nature. This implies that the self-healing framework will predict when and where an outage might occur with some probability, and execute changes in neighboring cells proactively. Despite the seemingly random nature of outages, especially full outages, outage prediction is possible and has been demonstrated by Kumar et al. [190] who have used different machine learning techniques such as neural networks, NBC and SVM to predict the next fault from real network data. Similarly, Kogeda and Agbinya [191] have predicted fault occurrences by collecting the past data and calculating maximum likelihood of next fault location using Bayesian Network prediction models.

All of the above-mentioned techniques rely on exploitation of big data [8] to identify key patterns in cell and user performance data and associating the information with previous outage information and data. This will allow the proactive self-healing algorithms to identify changes in network performance that lead to a failure or an outage. Fig. 2.2 illustrates the concept of exploiting big data resources for prediction of faults in a future mobile cellular network. The definition of big data in the context of Self-healing framework includes historical fault data, user transition and handover data, network traffic and cell load data, and contextual data mined from sources such as social media. A more comprehensive solution for proactive outage prevention in also discussed in chapter 5 of this dissertation.

Challenge 4: Meeting quality of experience requirements in self-healing

The combination of requirements for 5G mobile cellular networks including low latency, high capacity, high throughput and low energy consumption means 5G networks will be user quality of experience centric compared to legacy networks which were user quality of service centric. This implies that meeting user quality of experience requirements will be the utmost priority in future mobile cellular networks, even in the event of an outage. Given that outages due to failures and parameter misconfigurations are likely to increase, meeting user quality of experience will be a key challenge for Self-healing solutions.

Possible Solution and Research Direction

The solution to meeting user quality of experience requirements despite outages is to deploy intelligence-rich proactive Self-healing framework such as the one shown in Fig. 2.2. The user-centricity of the framework will be driven by spatio-temporal user activity models. These include user mobility models derived from user transition data in the form of MDT reports [52] along with user location information which can easily be harvested from the positioning sensors inside modern cell phones. Additionally, user behavior load prediction models can be generated using machine learning techniques shown in Fig. 2.3 while contextual data from social media sources such as Twitter and Facebook can be mapped to network topology which would help to identify potential traffic hotspots and failures. Historical fault data collection can be done by setting up databases that would include network failure records as well the KPI data immediately preceding the failure. All this information will be fed to the proactive fault prediction algorithms which would sit alongside a reactive Self-healing triggering algorithm which monitors fault data from live network.

Challenge 5: Coping with bandwidth constraints for Self-healing

Bandwidth constraints are one of the greatest limiting factors for mobile cellular network capacity. Limited bandwidth means extra capacity can only be added by adding more cells into the network. However, as discussed previously, network densification can lead to a rise in network outages itself. Furthermore, bandwidth limitation becomes even more acute in the event of an outage when already strained neighboring cell resources

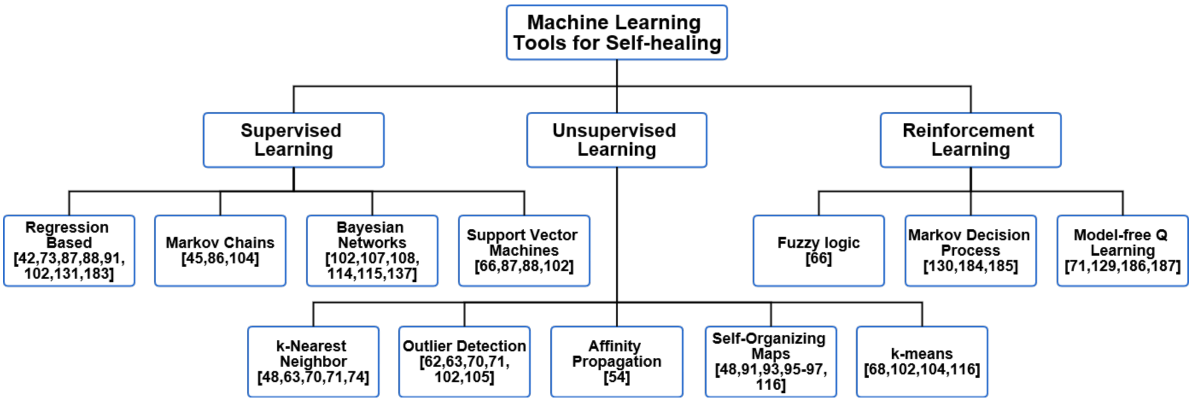

Fig. 2.3: Machine Learning Tools to Enable Proactive Self-Healing Framework for Future Cellular Networks

can become completely choked causing partial outages.

Possible Solution and Research Direction

While millimeter wave spectrum utilization has been promoted as the primary solution to bandwidth limitation [106], it is still in exploratory phases. In addition, the limited range of millimeter wave cells does not make them the ideal candidates for outage compensation solutions unless they are deployed in very high densities. One possible solution to the issue of bandwidth limitation for Self-healing is to deploy spectrum sensing or cognitive radio solutions [25, 24]. Some outage compensation solutions based on spectrum splitting have been proposed in [132, 134, 140, 141] but these solutions propose to reserve HCs specifically for outage compensation. Given that mobile cellular networks are already facing bandwidth shortage, this approach may not be suitable especially when there are no outages. To avoid dedicating bandwidth for outage compensation, cognitive radio technologies can be explored to split the spectrum between HCs and normal bandwidth specifically in the event of an outage. Not only would this improve radio resource utilization under normal circumstances, it can also improve the service provided to outage-affected users by assigning them low interference resources.

2.7 Conclusion

Self-healing is potentially the most powerful SON component in terms of reducing mobile cellular network operational expenses, especially for future 5th generation and beyond networks. This chapter explores the full breadth of studies pertaining to self-healing techniques for mobile cellular networks. The studies on self-healing have been broken down by the area of self-healing framework they belong to in addition to their methodologies, topologies, design metrics and control mechanisms. Additionally, the survey includes the studies for avoiding outages due to SON function conflicts, also known as self-coordination which has been discussed under the umbrella of self-healing. This is done in order to consolidate the work done on outage prevention, detection, diagnosis and compensation under into a comprehensive survey on mobile cellular network outages.

In addition to the review of existing literature supporting self-healing for mobile cellular networks, this chapter presents and elaborates the challenges faced by self-healing functions in terms of future 5th generation and beyond mobile cellular networks while also presenting possible solutions and future research directions. This literature survey and the insights gained from it pave the way for solutions presented for outage detection and prevention in the proceeding chapters.

CHAPTER 3

Outage Detection Using Entropy Field Decomposition for High Noise Environments

Quality of experience (QoE) enhancement compared to legacy mobile cellular networks is the primary drive of 5th generation mobile cellular networks [192]. To enable this QoE enhancement, 5th generation networks will rely on a combination of factors including 10x more throughput, less than 1 ms latency, and 10x more battery life than 4th generation mobile cellular networks [193]. To meet these expansive requirements, several solutions have been proposed [194], with network densification [195] and millimeter wave (mmWave) spectrum utilization [196] among the most popular. It has been demonstrated that a combination of network densification and mmWave spectrum deployment could potentially yield exponential increase in area spectral efficiency [197].

However, network densification and mmWave spectrum utilization are not without their own limitations. Ultra-dense heterogeneous networks (UDHNs) are prone to generating sparse network coverage information due to low user density per cell [68]. On the other hand, mmWave cells are subject to very high pathloss due to their operation in 30GHz - 300GHz band. One solution to reduce the impact of high pathloss in mmWave cells is the deployment of highly directional antennas with beam-widths as low as 7° [198]. Unfortunately, this opens mmWave cell networks to the problem of very large coverage gaps that must be filled by additional antennas per cell compared to traditional macro cells or by employing umbrella macro cells.

The challenges above highlight the difficulties of ensuring reliable and omnipresent coverage in mmWave-UDHNs, especially considering the QoE requirements for 5th generation networks. Even without these challenges, ensuring coverage in mobile cellular networks requires continuous network performance testing and monitoring. Gaps in network coverage result from following three causes: 1) poor network planning, 2) changes in radio environment, and 3) cell outages —partial or full. Coverage gaps due to first two causes, also known as coverage holes, require mobile cellular network operators to invest heavily in regular network coverage testing, usually through drive-tests. However, the process is time and resource consuming while lacking comprehensiveness due to inaccessibility of a major portion of the network that is, all areas other than paved roads. On the other hand, identifying and resolving cellular network outages, as described through the literature survey in chapter 2, requires highly trained engineers parsing gigabytes of network health logs and network performance indicator data looking for outages. Given the continuous growth in cell density and increasing pressure to reduce operational costs [199], both of the above approaches are quickly becoming impracticable [200, 201].

3.0.1 Related Work

To address the coverage hole and outage detection problem, 3GPP has introduced the minimization of drive test (MDT) reports feature [202] that is used for baseline coverage data in the proposed solution. MDT reports are a solution to the challenges of periodic drive testing, high carbon footprint, and rising operational costs of mobile cellular networks. These reports consist of serving and neighboring cells identities, downlink received power levels, and channel quality measurements. The measurements are collected periodically by user equipment (UE) in both connected and idle modes and are reported back to the network along with their location tags. This offers network providers a more detailed view of network coverage, including indoor network coverage, compared to drive tests. MDT reports have been used for designing coverage hole [203, 204, 205], and outage detection solutions [62, 63, 71, 66, 70, 71], as well as other SON functions [206, 59, 207].

Coverage hole and outage detection in mobile cellular networks

Given the significance of the problem in network service management, in the last few years, full and partial outage detection has been studied extensively as was presented in chapter 2. A small subset of those studies that is most relevant to the solution proposed in this chapter are discussed here.

Most prior studies on outage detection have focused on homogeneous macro cellular networks [62, 63, 71, 104, 88, 56], as large number of users per cell in macro cell provide enough data for classic machine learning methods to be trained for coverage anomaly detection. For example, in [62, 63] the authors employ kNN [208] and LOF [209] techniques to detect coverage anomalies in macro cell environment. The authors use the two clustering techniques to separate cells into normal and anomalous based on their received power measurements from MDT data, and use expert analysis to determine the accuracy of anomaly detection. In [66, 210] the authors compare LOF with one-class SVM [67] using MDT data with different levels of shadowing and inter-site distances. The authors compare the Receiver Operating Characteristic curves of the two techniques and demonstrate that for the same shadowing and inter-site distance, one-class SVM outperforms LOF considerably.

Apart from the use of MDT data for outage detection, some studies such as [104, 88, 56] also propose to use cell level performance metrics such as uplink and downlink data rates, radio link failures, and handover failures to detect network outages. In [104, 88] the authors use SVM and auto-regressive integrated moving average to identify network anomalies using network throughput data. The authors construct healthy network performance models using the two techniques and predict future cell performance data from those models. If there is a significant deviation between actual and predicted data, the algorithm determines that the cell is in outage. In [56] the authors use handover and radio link failure data to predict network outages in the network. The authors create a diffusion map of changes in user associations due to handover and call failure events.

The solution then maps the user association changes to cell dominance areas in the diffusion map. Finally, the solution runs k-means clustering algorithm [58] to detect the cell with abnormal changes in user associations.

In contrast to the plethora of studies on cell-wide outage detection, only a few studies have addressed the problem of coverage hole detection [203, 204, 205]. In [203] the authors argue that MDT report data does not necessarily represent a complete network coverage picture. Therefore, they apply Bayesian prediction framework on a real network MDT report data to predict missing data and generate radio network coverage maps. These maps are then used to predict the presence of coverage holes by estimating their likelihood based on neighboring pixel. The algorithm in [203] is extended in [204] to use four neighboring pixels, instead of one, to construct the network coverage environment map. Moreover, the authors use Bayesian kriging and interference cartography to generate these maps, which improves the prediction accuracy. In [205] the authors propose to use a combination of radio link failures along with a piece-wise deterministic coverage model to predict the boundaries of coverage and, consequently, coverage holes. The use of deterministic coverage model in [205] eliminates the uncertainty in coverage estimation due to shadowing.

Despite advancements in coverage hole and outage detection described above, there exist several key issues that need to be addressed for such solutions to be applicable to mmWave UDHNs:

Sensitivity to shadowing Most of the existing cell outage detection solutions are highly sensitive to shadowing. This is a key observation made in [66, 210] where the authors have investigated the impact of shadow fading on the accuracy of several outage detection algorithms. The authors have shown that as the standard deviation of shadowing increases, machine learning based outage detection models become less accurate ultimately becoming analogous to a coin toss. The same is true for accuracy of coverage hole detection solutions which rely on neighboring pixel data to predict the existence of a coverage hole [203, 204]. Since heavy shadowing is an intrinsic feature of mmWave-UDHNs, a practical solution for coverage hole and cell outage detection in such networks must be robust to the effects of shadowing.

Inclusion of Spatio-Temporal Domains Network coverage is susceptible to randomness due to shadowing in both spatial and temporal domains. However, prior studies, such as the ones discussed above, only consider coverage information in terms of spatial snapshots taken at certain time instants. This implies that the solutions dependent on this information run at each time instant with no information cascading to subsequent time snapshots. This leads to the possibility of instantaneous coverage holes or outages triggering outage compensation algorithms if they occur at the time when the snapshot of the coverage is being built, or coverage holes or outages being missed out between the snapshots. Even the solutions based on time-averaged models such as [104, 88, 211] do not offer the temporal depth to deal with these instantaneous coverage fluctuations. Similarly, while some studies employ temporal coverage data to identify network anomalies [212, 213, 214, 77], they do not explicate the complete spatial impact of these anomalies on the subscribers or their QoE. Given these issues, any coverage hole or outage detection solution must consider both spatial as well as temporal domains in the detection process.

Sensitivity to data distribution model Many prior studies discussed above leverage tools such as k-means, LOF, SVM, Bayesian kriging, or Bayesian prediction frameworks that implicitly assume some specific data distribution, usually Gaussian, exponential or some variation of the two. However, acute dynamics of mobile environment, particularly those associated with mmWave UDHN can make these assumptions invalid. Therefore, an outage detection solution that does not rely on a-priori assumptions about distribution of data is highly desirable, particularly to be applicable for mmWave UDHN.

Coverage Hole and Outage Detection in mmWave UDHNs As mentioned in chapter 1, the likelihood of outages increases with cell density as well as complexity of the cell hardware which will be the case in mmWave-UDHNs. While some outage detection solutions have been proposed to incorporate heterogeneous network topology, to the best of the author's knowledge, a coverage hole or outage detection solution that explicitly targets mmWave-UDHNs while addressing idiosyncrasies of such networks does not exist.

3.0.2 Proposed Approach and Contributions

In this chapter a new approach for anomaly detection is proposed, one that is based on entropy field decomposition (EFD), an algorithm first introduced by Frank and Galinsky [215]. EFD has previously been used successfully for brain activity mode detection in biomedical engineering [216] and the study of severe weather phenomenon [217]. However, this is the first use of EFD in the context of wireless communication networks. The motivation for leveraging EFD to solve cell outage and coverage hole detection problem is its ability to identify the flow of information in data over both space and time by combining information field theory [218] and entropy spectrum pathways theory [219]. Furthermore, EFD is independent of baseline data model/distribution and it actively suppresses the effects of noise in activity mode detection process which makes it a natural solution for coverage hole and outage detection in mmWave-UDHN environments marked by heavy shadowing.

3.1 System Model

Consider a system with single-cell connectivity for the sake of simplicity and assume that the user association changes immediately after an outage which is a reasonable assumption for mmWave - UDHNs with highly overlapping cell coverage.

3.1.1 Network Coverage

For the purpose of developing this solution, it is assumed that a user is in outage when the downlink received power of that user from its associated cell $P^c_{r,u}$ falls below a threshold $P^{th}_{r_{out}}$ that is,:

$$\text{Outage} := P^c_{r,u} \leq P^{th}_{r_{out}} \tag{3.1}$$

and in a coverage hole when $P^c_{r,u}$ falls below a threshold $P^{th}_{r_{ch}}$ that is, :

$$\text{Coverage Hole} := P^c_{r,u} \leq P^{th}_{r_{ch}} \tag{3.2}$$

where $P^{th}_{r_{ch}} > P^{th}_{r_{out}}$. The rest of the discussion in this and subsequent section will focus on outage detection but it lends itself directly to coverage hole detection as well without any changes.

To calculate $P^c_{r,u}$, the standard exponential pathloss model is considered, the log of which can be written as:

$$P^c_{r,u_{dBm}} = f(P^c_t, G_u, G^c_u, a, d^c_u, \beta) + \epsilon^c_u \tag{3.3}$$

where P^c_t is the transmit power of cell c, G_u is the gain of user equipment, G^c_u is the channel gain of cell c, a is the pathloss constant and depends on the clutter, ϵ^c_u is the shadowing at the location of user u from cell c and usually assumed to be log-normally distributed, d^c_u is the distance of subscriber u from cell c, and β is the pathloss exponent. Assuming each of G_u, G^c_u, a, d^c_u and β remains constant, (3.3) can simply be re-written as:

$$P^c_{r,u_{dBm}} = f(P^c_t) + \epsilon \tag{3.4}$$

Thus, each subset $\hat{\boldsymbol{P_t}}$ of the set of cell transmit powers $\boldsymbol{P_t}$ will result in a different set of received powers $\boldsymbol{P_r}$. As such, the likelihood of getting a set of downlink received powers $\boldsymbol{P_r}$ given some set of transmit powers $\hat{\boldsymbol{P_t}}$ can be defined as:

$$p(\boldsymbol{P_r}) = \int p(\boldsymbol{P_r}|\boldsymbol{\hat{P}_t})p(\boldsymbol{\hat{P}_t})d\boldsymbol{\hat{P}_t} \tag{3.5}$$

In the event of a cell outage, the loss of transmission from the affected cell will result in a unique set of downlink received powers $\boldsymbol{P_{r_{out}}}$. Given this set of received powers, the set of transmit powers including the affected cell transmit power can be estimated using Baye's rule as:

$$p(\boldsymbol{\hat{P}_t}|\boldsymbol{P_{r_{out}}}) = \frac{p(\boldsymbol{P_{r_{out}}}, \boldsymbol{\hat{P}_t})}{p(\boldsymbol{P_{r_{out}}})} \tag{3.6}$$

3.2 Entropy Field Decomposition

For a deterministic system with a fixed signal propagation model and no shadowing, the estimation of (3.6) is simply a question of going through all the subsets $\boldsymbol{\hat{P}_t}$ and calculating the resulting sets $\boldsymbol{P_r}$. However, for a system with random variations in the signal, this estimation becomes more complex. Furthermore, if these random variations affect the system both spatially and temporally, as is the case in real mobile cellular networks, obtaining the conditional probabilities in (3.6) becomes intractable. However, one method of obtaining an estimate of these probabilities which has been explored in [218] is to use information field theory which represents the probability distributions in terms of an information field. In this case, the transmit power or signal data must first be represented as an information field such that:

$$P_t(x_l, y_l, t_l) \equiv P_t(\zeta_l) = \int \boldsymbol{\vartheta}\delta(\zeta - \zeta_l)d\vartheta \tag{3.7}$$

where $\zeta_l = x_l, y_l, t_l$ represents the transformation of spatial coordinates $x_i, y_i, i = 1, ..., NM$ and temporal coordinate $t_j, j = 1, ..., O$ as a point on the information field $\boldsymbol{\vartheta}$. The key descriptor of an information field is the Hamiltonian $\mathcal{H}$ which corresponds to the total energy of the field [220] and is defined as:

$$\mathcal{H}(\boldsymbol{P_r}, \boldsymbol{\vartheta}) = -\ln\left(p(\boldsymbol{P_r}, \boldsymbol{\vartheta})\right) \tag{3.8}$$

Using the above transformations, (3.6) can be re-written as:

$$p(\boldsymbol{\vartheta}|\boldsymbol{P_r}) = \frac{e^{\mathcal{H}(\boldsymbol{P_r},\boldsymbol{\vartheta})}}{\mathcal{Z}(\boldsymbol{P_r})} \tag{3.9}$$

where $\mathcal{Z}(\boldsymbol{P_r}) = \int e^{\mathcal{H}(\boldsymbol{P_r},\boldsymbol{\vartheta})} d\boldsymbol{\vartheta}$ is called the partition function. Since the spatio-temporal received powers in a real network are not independent of each other, $P_t(\zeta_l)$ is considered an interacting field [218] whose Hamiltonian can be derived through Taylor series expansion of (3.8) as given in [215]:

$$\mathcal{H}(\boldsymbol{P_r},\boldsymbol{\vartheta}) = \mathcal{H}_0 + \frac{1}{2}\boldsymbol{\vartheta}^{\dagger}\boldsymbol{\mathcal{D}}^{-1}\boldsymbol{\vartheta} - \boldsymbol{j}^{\dagger}\boldsymbol{\vartheta} + \sum_{n=1}^{\infty}\frac{1}{n!}\int \dots \int V^{(n)}_{\zeta_1 \dots \zeta_n} \vartheta(\zeta_1)\dots\vartheta(\zeta_n) d\zeta_1 \dots d\zeta_n \tag{3.10}$$

where $\boldsymbol{\mathcal{D}}$ matrix is the information propagator, the vector $\boldsymbol{j}$ is the information source, the $(.)^{\dagger}$ notation represents the adjoint of a matrix, and $\mathcal{H}_0$ is the free energy Hamiltonian [220] which can be obtained by integrating the joint probability $p(\boldsymbol{P_r},\boldsymbol{\vartheta})$ over $\boldsymbol{P_r}$ and $\boldsymbol{\vartheta}$. Since $\mathcal{H}_0$ is a consequence of an interaction-less field, it simply acts as a scaling factor for an interacting field. Also, since it is assumed that the received power at each point is not independent of other points due to shadowing and fading effects, $\mathcal{H}_0$ can be safely ignored. The terms $V^{(n)}_{\zeta_1 \dots \zeta_n}$ represent the interactions of up to n field components and are integrated over each coordinate. The matrix $\boldsymbol{\mathcal{D}}$ and vector $\boldsymbol{j}$ can be obtained by using the free theory formalism for a Gaussian signal [218] and are given as:

$$\boldsymbol{\mathcal{D}} = \left[\boldsymbol{\sigma^2_{\hat{P}_t}}^{-1} + f(\boldsymbol{\hat{P}_t})^{\dagger}\boldsymbol{\mathcal{N}}^{-1}f(\boldsymbol{\hat{P}_t})\right]^{-1} \tag{3.11a}$$

$$\boldsymbol{j} = f(\boldsymbol{\hat{P}_t})^{\dagger}\boldsymbol{\mathcal{N}}^{-1}\boldsymbol{P_r} \tag{3.11b}$$

where $\boldsymbol{\sigma^2_{\hat{P}_t}} = \langle \boldsymbol{\hat{P}_t}\boldsymbol{\hat{P}_t}^{\dagger} \rangle$ is the covariance of the transmit powers and $\boldsymbol{\mathcal{N}}$ is the covariance of noise in the data.

The interaction terms $V^{(n)}_{\zeta_1 \dots \zeta_n}$ can be obtained using entropy spectrum pathways theory which ranks the optimal pathways within a disordered lattice according to their path entropy [219]. To construct the entropy pathways, a coupling matrix $\boldsymbol{\mathcal{Q}}$ must be constructed from points on the information field lattice. This can be done by generating an adjacency matrix $\boldsymbol{\mathcal{A}_{ij}}$ of spatio-temporal points in the dataset $\boldsymbol{P_r}$ and using the transformation $\zeta_l = x_l, y_l, t_l$ to obtain the components of $\boldsymbol{\mathcal{Q}}$ matrix as follows:

$$\mathcal{Q}(\zeta_i, \zeta_j) = P_r(i)P_r(j)\mathcal{A}_{ij} \tag{3.12}$$

It is important to highlight here that the $\boldsymbol{\mathcal{Q}}$ matrix can be used to represent any relationship between two or more points in the network regardless of the signal propagation model and the distributions of shadowing and fading. This is a major advantage compared to other techniques such as Bayesian classification that rely on some underlying assumptions regarding data and noise distributions which can lead to very high misclassification error if the actual distribution differs from the assumed one.

The information field can be reconstructed via entropy spectrum pathways that allow the representation of the field in terms of the eigenmodes of $\boldsymbol{\mathcal{Q}}$ using Fourier expansion. In mathematical terms, field components are given as:

$$\vartheta(\zeta_l) = \sum_{k}^{K}[b_k\varphi^{(k)}\zeta_l + b_k^*\varphi^{*(k)}\zeta_l] \tag{3.13}$$

where $\varphi^{(k)}$ is the k^{th} eigenmode, b_k is the mode amplitude of k^{th} eigenmode and the $*$ operator refers to the conjugate of a number, while K is the number of significant eigenmodes considered for field transformation. A key insight here is that by only considering the most important eigenmodes and keeping K to a reasonably small value compared to the total number of eigenmodes, a decent estimate of the information field can be obtained, thus reducing the problem complexity significantly. To test the importance of an eigenmode, the corresponding eigenvalue λ_k can be compared with the determinant of the noise covariance matrix $\boldsymbol{\mathcal{N}}$.

Using the above information, the transformed information Hamiltonian $\mathcal{H}(\boldsymbol{P_r}, \boldsymbol{b_k})$ can be obtained by:

$$\mathcal{H}(\boldsymbol{P_r}, \boldsymbol{b_k}) = \frac{1}{2}\boldsymbol{a_k}^{\dagger}\boldsymbol{\Lambda}\boldsymbol{a_k} - \boldsymbol{j_k}^{\dagger}\boldsymbol{a_k} +$$

$$\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k_1}^{K} \cdots \sum_{k_n}^{K} \tilde{V}^{(n)}_{k_1 \dots k_n} a_{k_1} \dots a_{k_n} \tag{3.14}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\mathcal{Q}}$, $\tilde{V}^{(n)}_{k_1 \dots k_n}$ represent the interaction terms of the eigenmodes, and $\boldsymbol{j_k}$ is the amplitude of the k^{th} eigenmode in expansion of the information source $\boldsymbol{j}$:

$$\boldsymbol{j_k} = \int \boldsymbol{j}\varphi^{(k)} d\zeta \tag{3.15}$$

To calculate the values of mode amplitudes, the Hamiltonian is minimized with respect to the field and the field is replaced with its transformation in terms of its eigenmodes [215] which gives:

$$\boldsymbol{\Lambda}\boldsymbol{b_k} = \boldsymbol{j_k} - \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k_1}^{K} \cdots \sum_{k_n}^{K} \tilde{V}^{(n)}_{kk_1 \dots k_n} b_{k_1} \dots b_{k_n} \tag{3.16}$$

If the field interaction terms $V^{(n)}_{\zeta_1 \dots \zeta_n}$ are defined as powers of the coupling matrix $\boldsymbol{\mathcal{Q}}$ such that:

$$V^{(n)}_{\zeta_1 \dots \zeta_n} = \frac{\alpha^{(n)}}{n} \sum_{p=1}^{n} \prod_{\substack{m=1 \\ m \neq n}} \mathcal{Q}_{\zeta_p} \mathcal{Q}_{\zeta_m} \tag{3.17}$$

then the mode interaction terms $\tilde{V}^{(n)}_{k_1 \dots k_n}$ are obtained by:

$$\tilde{V}^{(n)}_{k_1 \dots k_n} = \frac{\delta^{(n)}}{n} \sum_{p=1}^{n} \left(\frac{1}{\lambda_{k_p}} \prod_{m=1}^{n} \lambda_{k_m} \right) \int \left(\prod_{r=1}^{n} \varphi^{(k_r)}(\zeta) \right) d\zeta \tag{3.18}$$

where coefficients $\delta^{(n)}$ should be chosen sufficiently small to ensure the convergence of (3.18). Using the formulation presented above, the posterior probability of the field can be obtained as defined in (3.9). The decomposition of the information field into its modes provides a reflection of the field into its independent modes.

Algorithm 1 Outage Detection Using EFD

Input: $\boldsymbol{P_t}, \boldsymbol{P_{r_{norm}}}, \boldsymbol{P_{r_{RT}}}, \mathbb{U}_c$
Output: $\vartheta(\zeta_l), c$ in outage

Calculate $\boldsymbol{\mathcal{Q}}$ from (3.12) using $\boldsymbol{P_{r_{RT}}}$
Obtain eigenvalues and eigenvectors of $\boldsymbol{\mathcal{Q}}$
Calculate entropy field $\boldsymbol{\vartheta}$ using the eigenvalues and eigenvectors of $\boldsymbol{\mathcal{Q}}$ from (3.13)
for $l \in 1, ..., NMO$ **do**
  if $\vartheta_l > 0$ **then**
    $\{u(x_{out}, y_{out})\} = \{u(x_{out}, y_{out})\} + u(x_l, y_l)$
  end if
end for
if $\{u(x_{out}, y_{out})\} = \phi$ **then**
  continue
else
  for $u(x_{out}, y_{out}) \in \{u(x_{out}, y_{out})\}$ **do** $\{c_{out}\} = \{c_{out}\} + \{c = \arg\max_{\forall c \in \mathbb{C}} P^c_{r,u(x_{out},y_{out})_{norm}}\}$
  end for
end if

3.2.1 Outage Detection Using Entropy Field Decomposition

The result of applying EFD to the coverage data is an entropy field which identifies how information flows across the spatio-temporal domains. In order to make this resulting field output useful for outage detection in MCNs, algorithm 1 presents the proposed coverage hole and outage detection solution.

The algorithm takes user-cell association $\mathbb{U}_c$ information from the MDT data and the network coverage data when no outage is present in the network $\boldsymbol{P_{r_{norm}}}$, as well as real-time spatio-temporal coverage data $\boldsymbol{P_{r_{RT}}}$. The real-time data are processed using EFD which outputs the data points where high information flows are detected. In simple terms, the EFD algorithm gives the boundary between outage and non-outage effected areas. The points with high energy that is, the points at the boundary of the outage are passed to a localization module which identifies the cell with which the high energy points are associated. Once the degraded cell is identified, it is passed as an output which can be fed to an outage diagnosis and compensation algorithm.

3.2.2 Complexity Analysis of EFD based Outage and Coverage Hole Detection

The complexity of algorithm 1 depends on two key factors: 1) the size of input matrices $\boldsymbol{P}_{r_{norm}}$ and $\boldsymbol{P}_{r_{RT}}$, and 2) the duration of time over which outage detection is being done. These factor define the size of adjacency matrix $\boldsymbol{\mathcal{A}}$ which in turn defines the computation required to estimate the matrix $\boldsymbol{\mathcal{Q}}$. Let $N = M$ and $L = N^2 * O$, then $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{Q}}$ are LxL matrices. Calculating $\boldsymbol{\mathcal{A}}$ is then $\mathcal{O}(L^2)$ complex. $\boldsymbol{\mathcal{Q}}$ matrix calculation only requires to replace estimated $\mathcal{Q}_{ij}$ values at points where $\mathcal{A}_{ij} \neq 0$ which is $\mathcal{O}(log(L))$ complex. The eigenvalue decomposition of $\boldsymbol{\mathcal{Q}}$ is $\mathcal{O}(L^3)$ complex [221]. The reconstruction of $\vartheta(\zeta_l)$ is $\mathcal{O}(K)$ complex based on (3.13). Finally, identifying points with non-zero entropy is a simple search action which is also $\mathcal{O}(log(L))$ complex. Thus, the complete algorithm is $\mathcal{O}(L^3 + L^2 + 2 * log(L))$ complex. Fig. 3.1 shows the complexity of the EFD based outage and coverage hole detection algorithm when $N = M$ and $O = 2$ which confirms the estimated complexity.

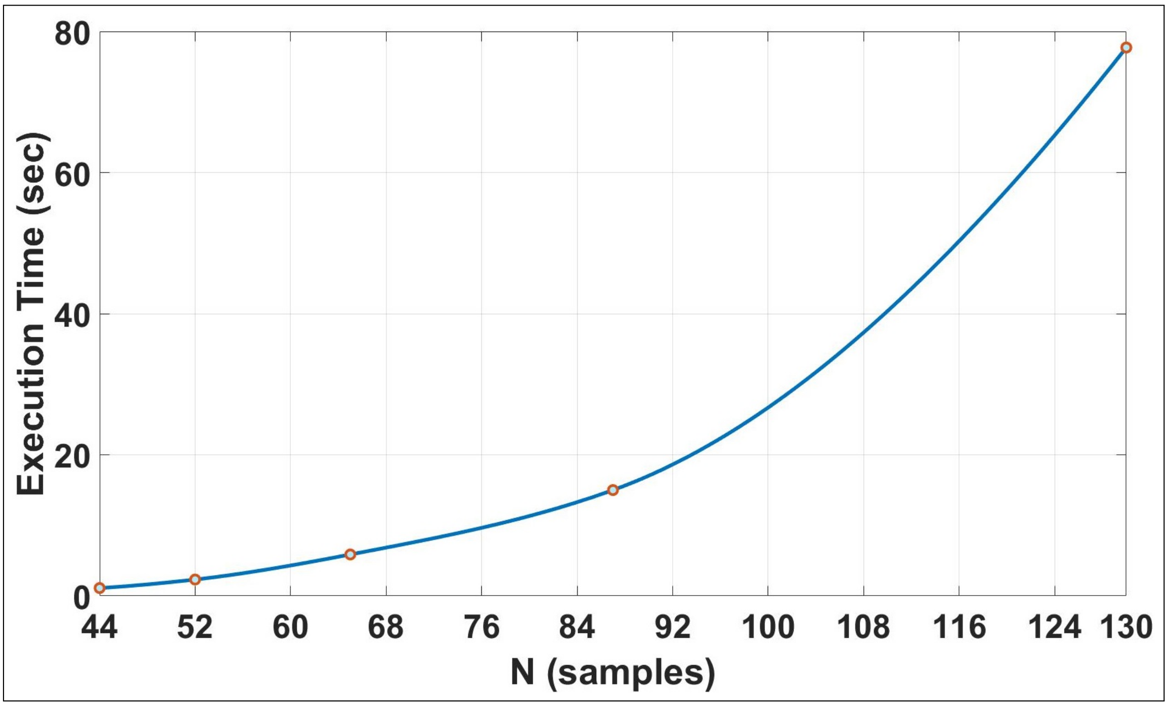

Fig. 3.1: Complexity of EFD based outage and coverage hole detection algorithm

3.2.3 Other techniques for outage detection

To evaluate the performance of EFD based outage detection, it is compared with two unsupervised machine learning techniques namely 1) *independent component analysis (ICA)* [98] and 2) *k-means clustering* [58]. Both of these techniques are briefly described here before the discussion on simulation details and results.

Independent Component Analysis

Just like EFD, ICA is an unsupervised clustering technique and is highly useful for decomposing real datasets where the noise distribution in the source signals may not always be Gaussian. ICA finds clusters in a dataset in the form of source signals which can be used to reconstruct the received signals such that:

$$\boldsymbol{P_r} = \boldsymbol{A} * \boldsymbol{P_t} \tag{3.19}$$

where $\boldsymbol{P_r}$ is the received signal vector, $\boldsymbol{P_t}$ are the source signal vectors and $\boldsymbol{A}$ is the weight matrix for signal reconstruction.

There are several methods of estimating $\boldsymbol{P_t}$ and $\boldsymbol{A}$ including projection pursuit, infomax, maximum likelihood estimation [222], and reconstruction ICA [223]. For the purpose of comparison with the proposed EFD based outage detectio solution, reconstruction ICA is employed as the algorithm of choice. Reconstruction ICA was initially developed to denoise and reconstruct images from only a few source signal vectors obtained using ICA. Since the objective here is not only to identify the source of an outage but also its impact area, reconstruction ICA is used to separate the outage source signal from noise source signal and normal source signal, which is then used to reconstruct a map of only outage affected areas.

k-means Clustering

k-means clustering is another commonly used unsupervised clustering technique. The k-means clustering algorithm splits the data into k separate clusters based on their distance from randomly distributed means. The most common implementation of k-means clustering is also referred to as Lloyd's algorithm [224] which iteratively updates the values of the k different means based on the distance between the means and data points. The algorithm begins by assigning each data point P_{r_p} to a set of points $\boldsymbol{S_i^{(t)}}$ which are closest to the mean μ_i in terms of Euclidean distance such that:

$$\boldsymbol{S_i^{(t)}} = \{P_{r_p} : ||P_{r_p} - \mu_i^{(t)}||^2 \leq ||P_{r_p} - \mu_j^{(t)}||^2 \forall j, 1 \leq j \leq k\} \tag{3.20}$$

The means are updated based on the new sets or clusters such that:

$$\boldsymbol{\mu_i^{(t+1)}} = \frac{1}{\boldsymbol{S_i^{(t)}}} \sum_{P_{r_p} \in \boldsymbol{S_i^{(t)}}} P_{r_p} \tag{3.21}$$

The algorithm continues to update the means until it converges or some stopping criteria for distance is met.

3.2.4 Simulation Setup

To obtain coverage data, an ultra-dense network of mmWave cells and small cells is created in a 260m x 260m area of downtown New York City as shown in Fig. 3.2. The coverage data are generated using Atoll Radio Planning Software [225]. The small cells (red) are configured to operate in 2GHz frequency spectrum with omni-directional antennas, while the mmWave cells (green) are configured to operate in the 28GHz spectrum with directional antennas. Antenna specifications for mmWave cells were obtained from [106]. mmWave and standard Aster propagation models [226] which use ray tracing for mmWave and small cell pathloss calculation are employed while taking factors such as clutter information, atmospheric absorption, and frequency-selective fading into

Table 3.1: Parameter Settings for Simulation Testing of EFD Based Outage Detection Solution

System Parameters	Value
Number of Sites	mmWave: 7; Small: 4
Transmission Frequency	mmWave: 28GHz; Small: 2 GHz
Transmission Bandwidth	mmWave: 100 MHz; Small: 20 MHz
Transmit Power	mmWave: 30 dBm, Small: 20 dBm
Antenna Tilt	mmWave: 7°; Small: 0°
Antenna Gain	mmWave: 18 dBi; Small: 5.7 dBi
Site Placement	Random
Shadowing	Street: 0 dB - 10 dB Open Space: 0 dB - 10 dB Grassland: 2 dB - 12 dB Low Vegetation: 4 dB - 14 dB Building $\geq$ 30m: 8 dB - 18 dB Building 12m - 30m: 10 dB - 20 dB Building $<$ 12m: 12 dB - 22 dB

account. Cell transmit power is used as the input signal for information field generation; however, any other coverage actuation parameter such as transmitter antenna tilts or transmitter antenna gains can be used just as easily if the relationship between the parameter and output data is known. To simulate outages, the transmitters of Small Site 1 and mmWave Site 1 shown in Fig. 3.2 are deactivated one by one. Rest of the details of simulation parameters are given in Table 3.1.

3.2.5 Results

The results presented below compare proposed EFD based outage detection solution with ICA and k-means clustering based outage detection algorithms. Simulations were carried out for a range of shadowing levels to assess the efficacy of EFD in mitigating the effects of noise in the data. Furthermore, results for two different scenarios are presented: 1) *no outage*, and 2) *outage* to investigate the ability of EFD to distinguish between coverage holes and outages. Note that though the algorithms for EFD, ICA and k-means clustering use data from the two spatial and one temporal dimensions, the presented results for baseline coverage data are averaged over time. To compare the

Fig. 3.2: Network Layout

results of EFD, k-means and ICA, two coverage snapshots are shown, one for 0 dB open area shadowing and one for 10 dB open area shadowing.

The true positive detection rate and false positive detection rate are calculated for each algorithm with open area shadowing standard deviation ranging from 0 dB to 10 dB giving a performance benchmark in a range of scenarios in a dense urban setting. It is important to highlight here that as the open area shadowing standard deviation is changed, shadowing standard deviation for other clutter classes such as different building heights is also changed by the same amount. Furthermore, to ensure generality, the true and false positive outage detection rates are averaged over 100 simulation iterations.

Impact of Shadowing on Coverage Hole Detection

For this study, coverage holes are defined as points where $P_r < -105$ dBm. Since all three algorithms are unsupervised, this threshold can be set to any value without the loss of generality.

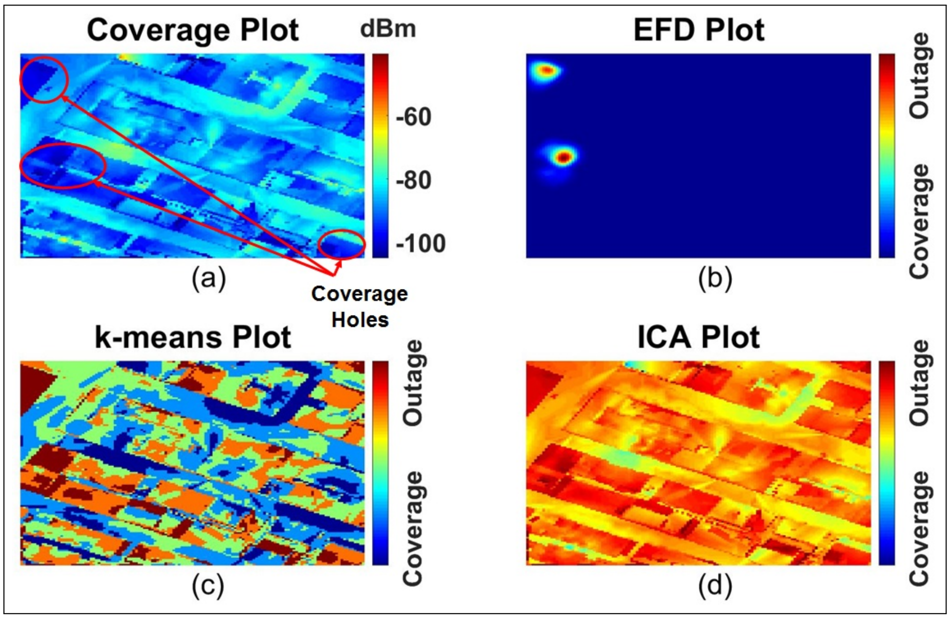

Fig. 3.3: Coverage hole detection using EFD, k-means clustering and ICA with 0 dB open space shadowing

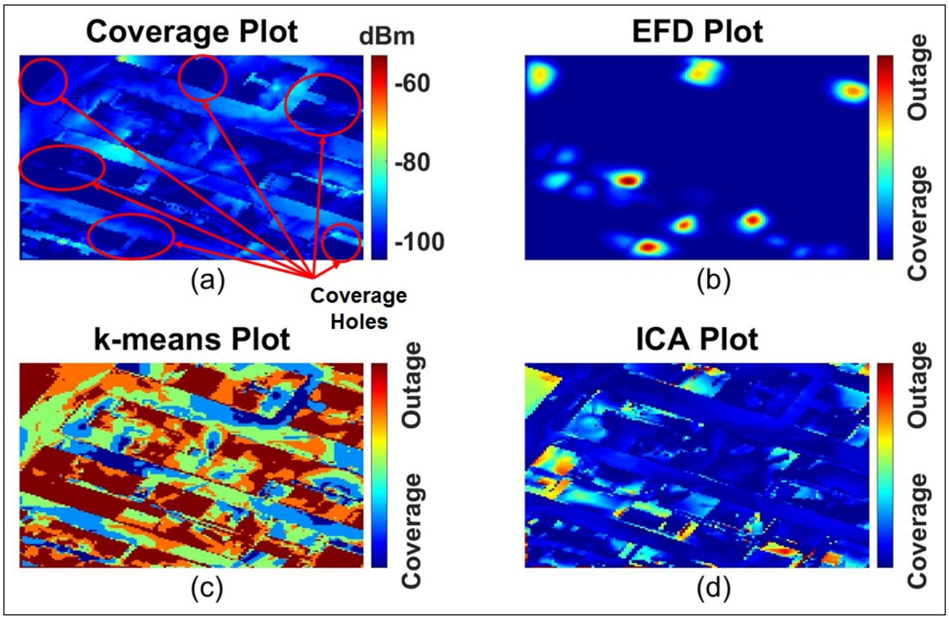

Fig. 3.4: Coverage hole detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

Figs. 3.3 and 3.4 show coverage holes in the network and the results for the three coverage hole detection algorithms when open area shadowing standard deviation is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of RSRP, Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

While the efficacy of each algorithm can be visually observed from Figs. 3.3 and 3.4, a comprehensive evaluation of the three algorithms requires more analytical comparison. To this end, Fig. 3.5 presents the true positive detection rate (TPR) of the three algorithms when open area shadowing standard deviation is varied between 0 dB and 10 dB. k-means clustering performs the best of all three algorithms in terms of TPR for low shadowing levels. However, as the level of shadowing increases, EFD begins to become more dominant as the difference between indoor and outdoor coverage becomes starker. On the other hand, TPR for k-means clustering begins to fall as shadowing increases because the predicted cluster mean values start getting closer to each other resulting in some coverage holes being misclassified.

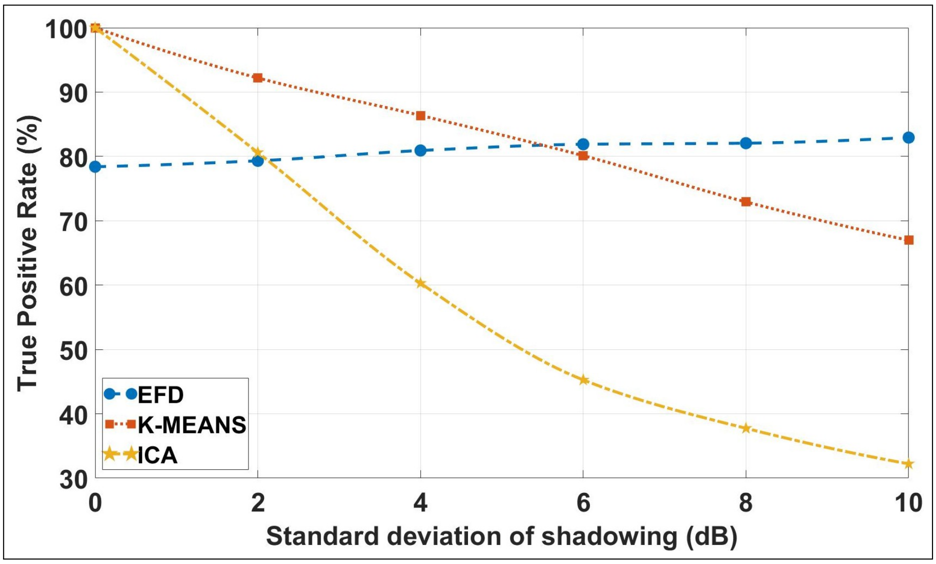

Fig. 3.5: True positive rate of coverage hole detection over varying shadowing with EFD, k-means clustering and ICA

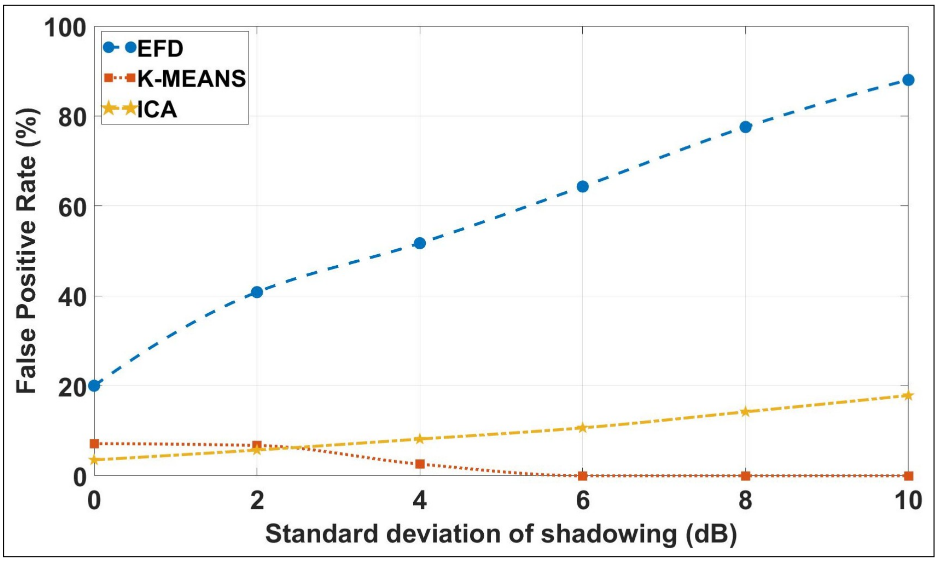

Fig. 3.6: False positive rate of coverage hole detection over varying shadowing with EFD, k-means clustering and ICA

For a complete perspective, the false positive detection rate (FPR) of the three algorithms is also compared and is given in Fig. 3.6. EFD starts off with a relatively high FPR which continues to increase with increase in shadowing standard deviation. This is because EFD identifies the maximum entropy boundary in the information field which means areas near the edge of a coverage hole will also have non-zero entropy leading to false positives. As a result, as the number of coverage holes increases, so does the FPR for EFD. Conversely, the FPR for k-means decreases as shadowing increases. For k-means this is again because of reduced mean separation which leads to fewer false predictions. For ICA, the FPR increases because of its unsupervised nature which means it clusters some areas with $P_r > -105$ dBm with coverage holes. Since the overall P_r values fluctuate more with increase in shadowing standard deviation, the potential for misclassification obviously increases.

It is pertinent to point out here that since coverage holes are not actually outages but just the absence of coverage in a particular region, using a solution that has higher TPR and higher FPR is preferable than having a solution with low TPR and low FPR. This is particularly true in the case of coverage holes detection since it is still acceptable to err on the side of caution if we can get higher detection accuracy and coverage hole detection is not as time sensitive as outage detection.

Impact of shadowing on Outage Detection

To compare the performance of the three algorithms in detecting outages in the network regardless of network coverage or shadowing, two different scenarios are employed: 1) *small cell outage* and 2) *mmWave cell outage*. Outage affected areas are defined as the spatio-temporal points where user association u_c changes after a cell becomes affected by a full or partial outage. Again, since all three algorithms are unsupervised, this outage criteria can be applied without the loss of generality.

Small Cell Outage Detection: Figs. 3.7 and 3.8 show the small cell outage and the results for the three algorithms when open area shadowing standard deviation is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of reference signal received power (RSRP), Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

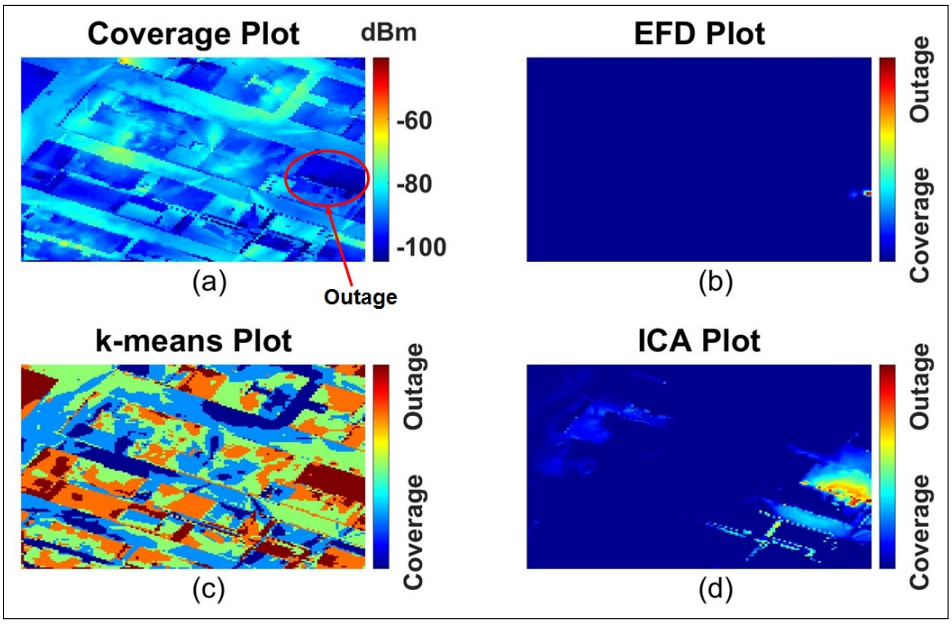

Fig. 3.7: Small cell outage detection using EFD, k-means clustering and ICA with 0 dB open space shadowing

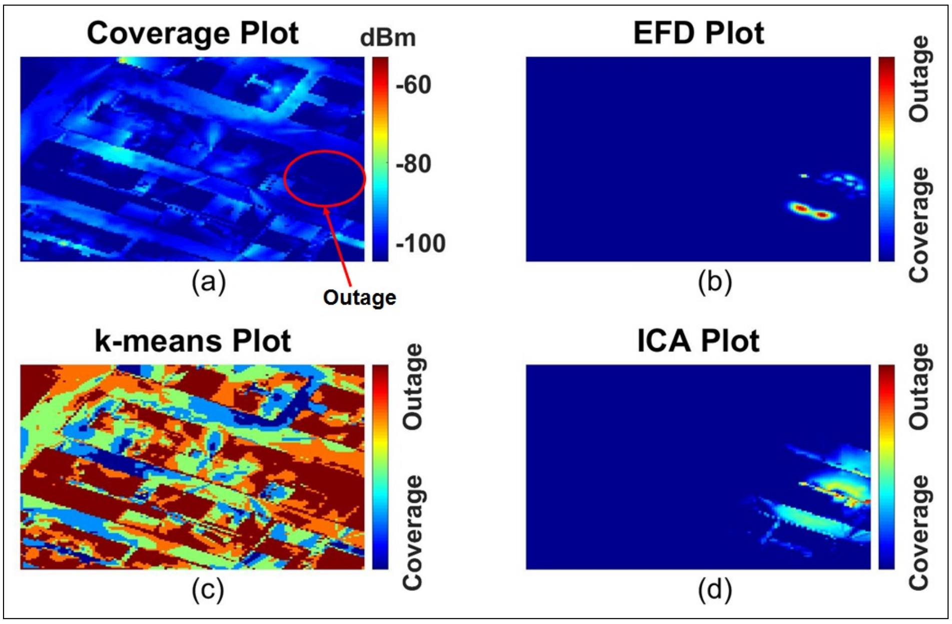

Fig. 3.8: Small cell outage detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

Fig. 3.9 compares the TPR for the three solutions for small cell outage detection and it can be seen that ICA performs the best of all three algorithms for lower levels of shadowing standard deviation. However, as the shadowing standard deviation increases,

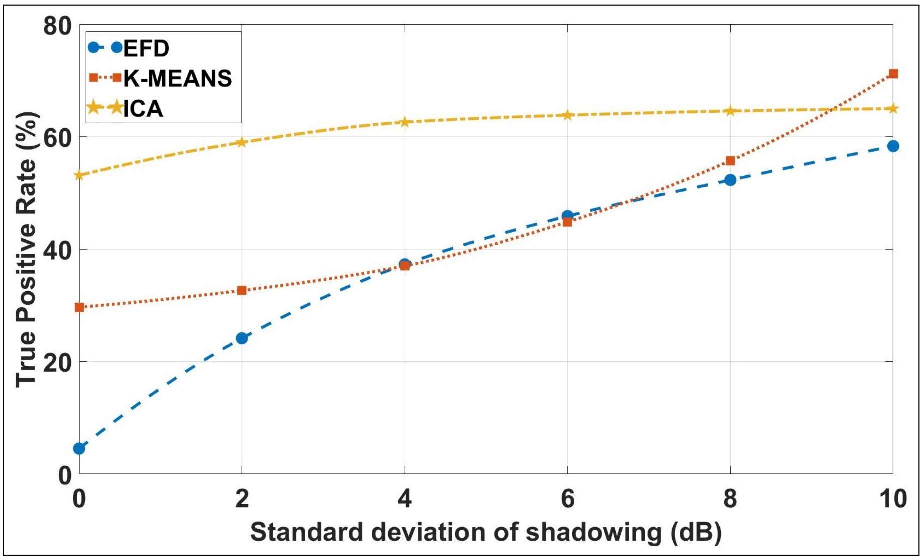

Fig. 3.9: True positive rate of small cell outage detection over varying shadowing with EFD, k-means clustering and ICA

k-means begins to become the more dominant algorithm. On the other hand EFD begins with rather low value of TPR but improves as the level of shadowing increases. The increase in TPR for k-means clustering is due to the fact that k-means does not the have the capacity to separate outages from coverage holes resulting in all of the outage points and coverage holes being lumped together as shadowing standard deviation increases. On the other hand, the increase in TPR for EFD stems from the fact that at low shadowing, the number of points associated with the outage affected small cell are very high but EFD only detects the source of the outage. However, with the increase in shadowing standard deviation, the number of points associated with the outage affected cell grows smaller which means EFD is able to identify them with higher accuracy.

Conversely, a comparison of FPR from Fig. 3.10 shows that as the shadowing increases, the ability of k-means to distinguish outages from coverage holes clearly decreases. Same is true for ICA but to a lesser degree since ICA is better at separating source of variation in data than k-means. Compared to both ICA and k-means, EFD has zero FPR for all levels of shadowing due to its ability to extract the source of an anomaly very cleanly from the data regardless of the noise variations.

It is pertinent to note here that the trend for EFD TPR is increasing with increase in

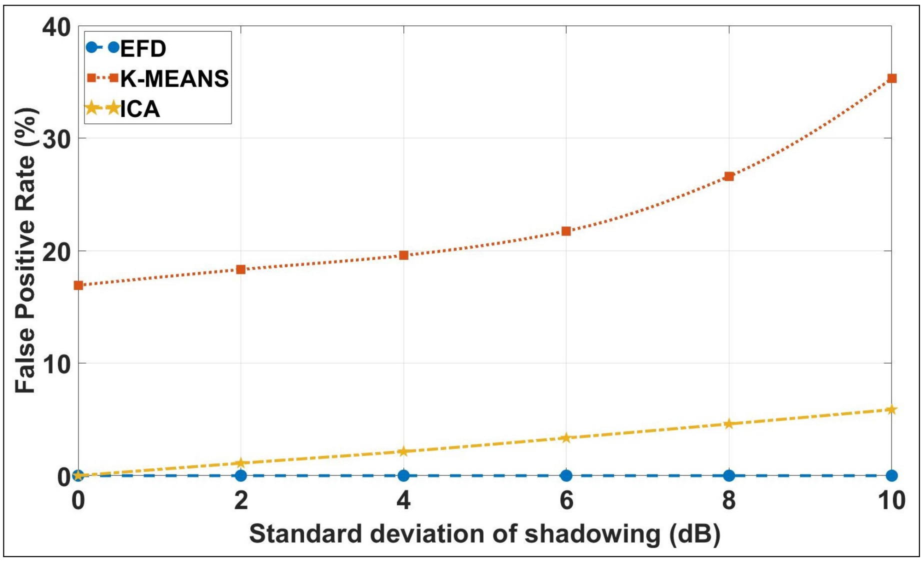

Fig. 3.10: False positive rate of small cell outage detection over varying shadowing with EFD, k-means clustering and ICA

shadowing, and when taken in context with its FPR, it clearly shows that EFD is the better algorithm for outage detection in high noise environments. Furthermore, since EFD takes both spatial and temporal data as input, not only does it identify the point where received power changes in space, it also detects where the received power changes in time.

mmWave Cell Outage Detection: Figs. 3.11 and 3.12 show the mmWave cell outage and the results for the three algorithms when open area shadowing standard deviation is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of reference signal received power (RSRP), Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

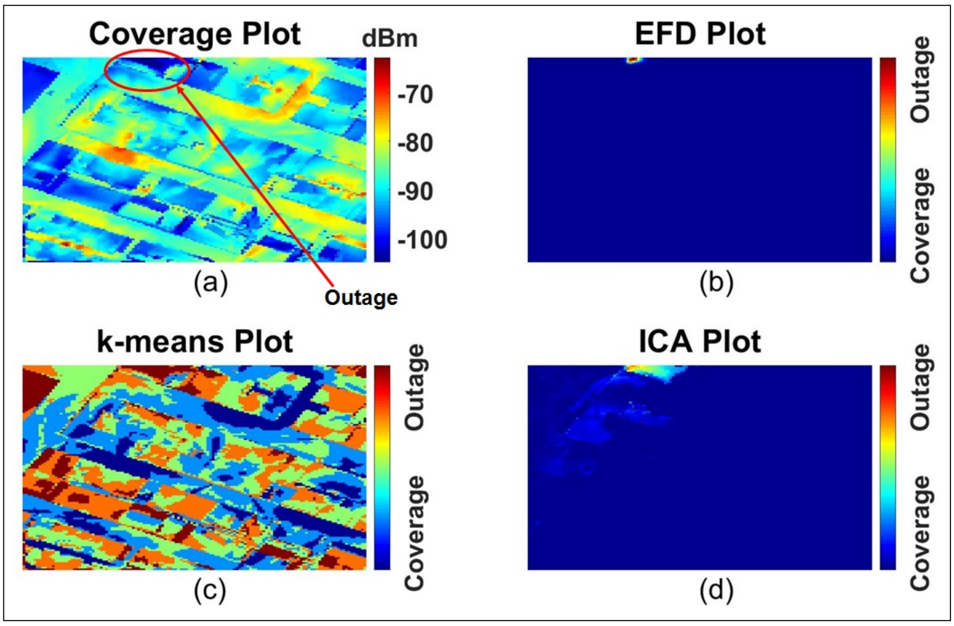

Fig. 3.11: mmWave cell outage detection using EFD, k-means clustering and ICA with 0 dB open space shadowing

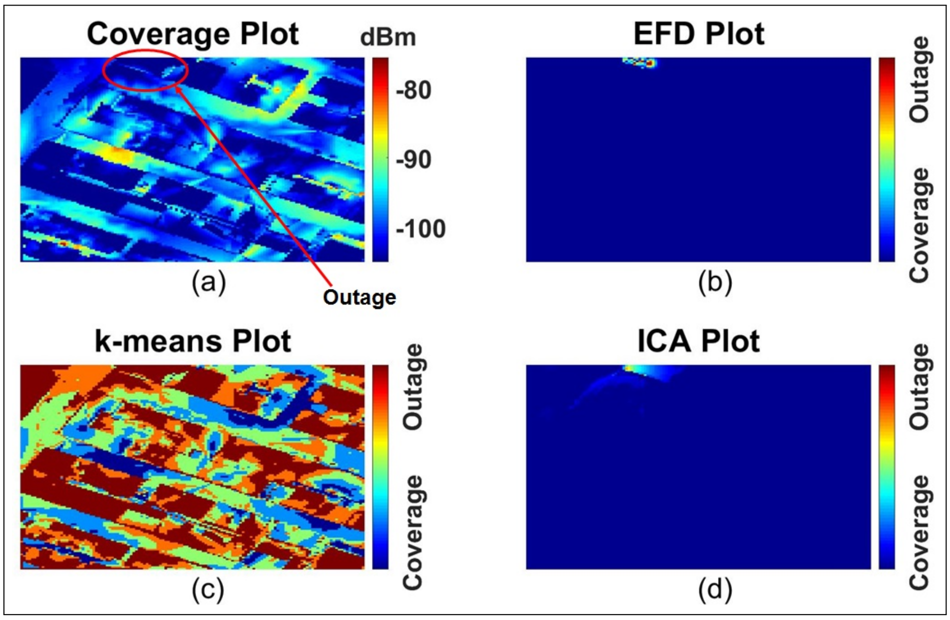

Fig. 3.12: mmWave cell outage detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

From Fig. 3.13 it can be seen that ICA performs the best of all three algorithms in terms of TPR at low to mid levels of shadowing standard deviation and its TPR remains stable with increasing shadowing. However, this increase needs to be seen in the context of its FPR given in Fig. 3.14 which also increases as the level of shadowing increases.

This is again because at higher levels of shadowing, it becomes more difficult for the algorithm to separate the source of an outage from the source of a deep coverage hole.

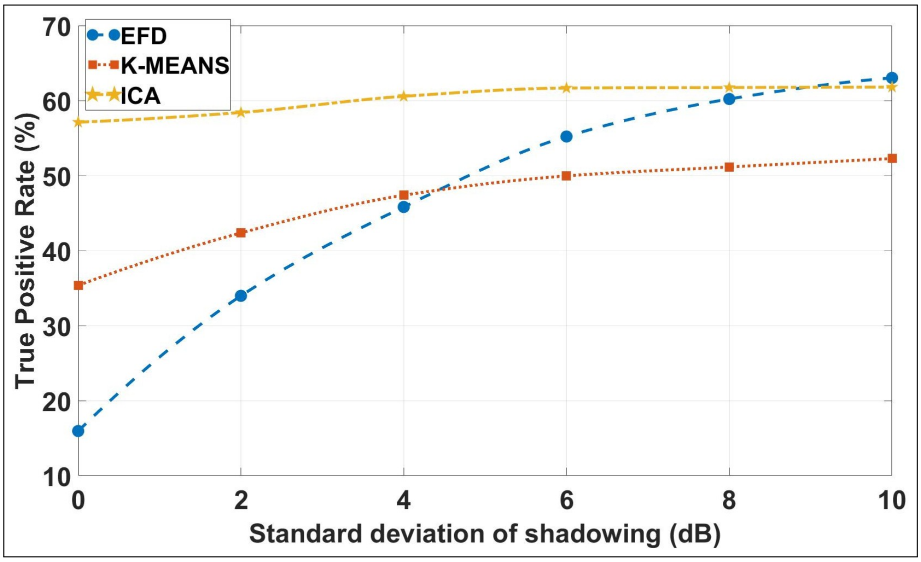

Fig. 3.13: True positive rate of mmWave cell outage detection over varying shadowing with EFD, k-means clustering and ICA

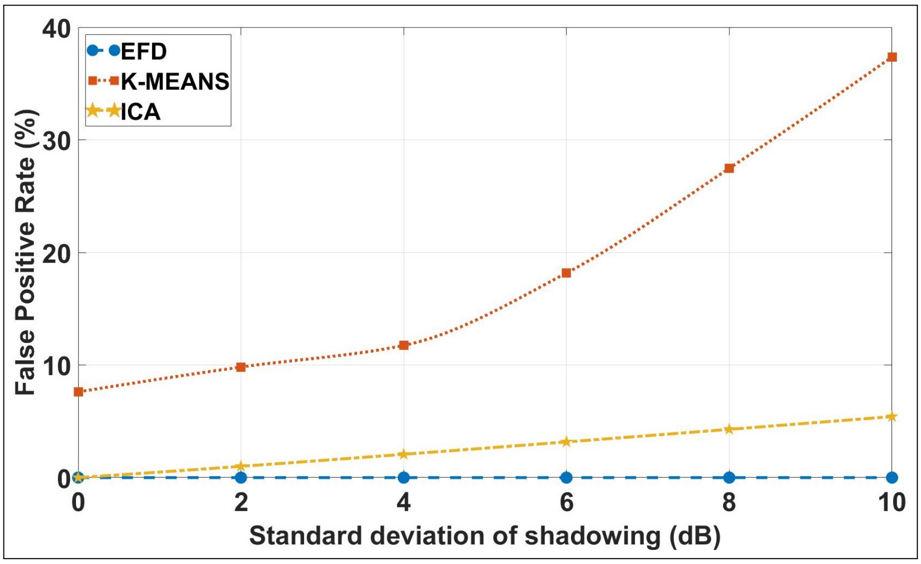

Fig. 3.14: False positive rate of mmWave cell outage detection over varying shadowing with EFD, k-means clustering and ICA

In comparison, EFD again starts from a lower TPR and gradually improves becoming the best algorithm at 10dB open space shadowing standard deviation. The reason for this is the same as before that is, it is able to extract the source of the outage very well from the data and as shadowing increases the impact of outage becomes smaller compared to coverage holes.

For k-means clustering, the TPR in the case of a mmWave cell outage improves only gradually since there are not a lot of points associated with the outage affected cell to begin with. However, as mentioned previously, it still lumps coverage holes and outages together which means its FPR increase dramatically with increase in shadowing.

Key Insights from the Results

While the relative performance of EFD, ICA and k-means is obvious from the results presented above, some of the more implicit, yet key insights are given below:

- k-means clustering is effective for coverage hole detection at lower shadowing levels while EFD is more effective at identifying coverage holes at higher levels of shadowing.
- Both EFD and ICA algorithms are more effective at spatially extracting the source of an outage than they are at identifying the impact of the outage making them better suited for identification of a cell in outage than k-means. In contrast, k-means is more suitable for identifying the impact of an outage.
- EFD is a highly effective choice for locating outages in spatio-temporal data since the $\boldsymbol{\mathcal{A}}$ matrix can be constructed to include both spatially and temporally adjacent points. This can allow EFD based outage detection solution to identify the source of an outage but also the time at which said outage occurred. This is of great value in real networks where outages can be very costly if allowed to continue for extended periods.

Practical Implementation of EFD based Coverage Hole and Outage Detection Solution

The results presented in this section have served to highlight the power of the proposed EFD based solution at overcoming the challenges posed by shadowing and temporal variations in coverage hole and outage detection. These advantages of EFD make it a very attractive deployment proposition in future mmWave UDHNs.

For implementation in a practical setup, EFD offers several design options and flexibilities including tweaking the value of K in (3.13) to obtain more complex coverage hole and outage profiles. The algorithm is also agnostic to the distribution of noise in the data which also makes it an ideal choice in unpredictable propagation environments such as UDHNs in dense urban areas. Furthermore, the algorithm sensitivity to changes in information flow can also be controlled by the parameter α in (3.18). A higher value of α makes the algorithm more sensitive to information changes while a smaller value makes it more focused.

3.3 Conclusion

In this chapter, several key issues present in the state-of-the-art coverage hole and outage detection algorithms have been presented. These issues include: 1) sensitivity to shadowing, 2) considering instantaneous spatial network coverage profile only, and 3) making assumptions regarding data and noise distributions. To overcome these limitations, a novel entropy field decomposition based solution has been proposed here which detects outages and coverage holes in spatio-temporal coverage data. The results presented in this chapter compare the efficacy of the proposed entropy field decomposition based outage detection method with the state-of-the-art outage detection methods including independent component analysis and k-means clustering in a dense urban mmWave-small cell UDHN over a range of shadowing values. Results show that proposed entropy filed decomposition based solution is a powerful tool in combating the effects of shadowing on coverage hole and outage detection while also demonstrating its efficiency at extracting spatio-temporal information flows from coverage data. However, entropy field decomposition is a computationally complex solution with exponential computation time with respect to input coverage data. Therefore, for practical implementation, a distributed execution where the network is divided into small spatial units is advisable.

CHAPTER 4

Outage Detection with Sparse User Data

From the discussion in chapters 2 and 3, it is clear that both data sparsity and variable noise will be features of mobile cellular networks of the future relying on network densification and mmWave spectrum utilization. Therefore, a solution is needed that is able to detect outages even when the network data are sparse and there is variable noise in the data.

The discussion and results in chapter 3 showed that while the proposed EFD based outage detection solution was able to detect outages in variable noise environments, it required full network coverage information to do so. This was because EFD relies heavily on identifying the entropy of information flow in the information field. However, if there are not enough spatial or temporal data, constructing a contiguous field becomes improbable and a field created from such data will not characterize an outage completely and accurately.

To this end, a new method for outage detection is presented in this chapter which uses deep neural networks (DNNs) [227] to extract hidden features from sparse, variable noise coverage data to localize a small or mmWave cell outage. DNNs have previously been demonstrated to effectively use the hidden features in the data for a variety of applications ranging from image recognition, text recognition, and classification among many other use case [227]. In this chapter, in addition to the performance analysis of the proposed DNN based outage detection solution, a comparison is presented with the state-of-the-art outage detection techniques.

The choice of DNNs for solving the outage detection with sparse data problem was motivated by the fact that DNNs can be trained to give extremely accurate results provided there is enough training data. It also understood that while the spatial coverage data is bound to be sparse in future mobile cellular networks, there will nevertheless be a healthy supply of temporal data since network coverage measurements are reported at millisecond timescales. Therefore, DNNs will still be able to detect outages with sparse spatial data since their will be no shortage of training data during the learning process.

4.1 Related Work and Contributions

One way of avoiding the issues associated with data sparsity and noise variability is to remove these two problems from the data before they are used for outage detection. This would require data augmentation and noise filtering, both of which have been explored in detail in literature, especially from the perspective of visual data. The following subsections provide a quick overview of the common techniques to deal with both these challenges along with their practical limitations in terms of outage detection with low latency. In addition, studies dealing with the use of DNNs for mobile cellular networks and anomaly detection are also discussed below.

4.1.1 Overcoming the Issue of Data Sparsity

The issue of data sparsity is not unique to UDHNs; in fact, a large number of practical systems need to deal with the issue of data sparsity which means the topic has been well researched. Overall, there are two broad methods of dealing with data sparsity which are: 1) *matrix completion* [228], and 2) *kriging* [229]. Different versions of both these techniques have been used for reconstruction of radio environment maps [230] and are briefly explained below.

Matrix Completion

Matrix completion refers to predicting values in a matrix with low rank and incomplete entries. There are several methods of achieving this with the most common approaches being singular value thresholding [231] and grey prediction models [232]. Out of the two methods, singular value thresholding based radio map prediction has been demonstrated in [230] while grey prediction has been used for completing small cell coverage matrices in [71] for outage detection.

Despite its usefulness, matrix completion is a highly complex process with very high computational complexity. In fact, the convergence time for singular value thresholding increases exponentially with the increase in matrix size. Similarly, grey prediction models have large overhead costs associated with them in terms of time complexity which increases as the number of available measurements decreases [71]. Other techniques used for matrix completion for outage prediction include collaborative filtering [73] which groups similar user profiles to predict normal cell performance, while [74] use simulations to augment sparse user data and build normal cell profiles for cell outage detection. Again, both these techniques require additional time to build full network coverage profile before detecting outages which adds to the latency of the solutions.

Kriging

Kriging is another technique that is used to interpolate missing values between reported measurements. Kriging algorithms work by calculating a probability estimate of the next point based on previous points and then predicting the next missing value [230]. One example of this is demonstrated in [204] where the authors have used Bayesian kriging to predict radio environment map for coverage hole detection.

While kriging gives very good results for systems with known noise distribution, its performance degrades if the noise distribution is unknown. For example, Bayesian kriging assumes that the noise distribution in the data is normal; however, that is not necessarily true in real networks. In addition, kriging is a computationally expensive process with exponential computational complexity.

4.1.2 Overcoming the Issue of Unpredictable Noise Variance

The issue of unpredictable noise variance is very common when dealing with image data with several approaches exist for overcoming it [233]. Among the two most popular techniques are: 1) *data denoising* and 2) *noise estimation*.

Data Denoising

Denoising technique are generally divided into two broad categories: 1) *spatial*, and 2) *transform* domain filtering methods [234]. Spatial filtering methods apply some sort of filters on the data to remove random perturbations from the data to achieve data smoothing. The complexity of such methods depends on the nature of filter, that is, is it linear or non-linear – and the size of data. Transform domain methods convert data from the spatial domain into a transform domain, commonly using cosine or fourier transform, and then separate the noise signal from the transform domain data before retransforming the data back to the spatial domain.

Noise Estimation

Most of the solutions for outage detection presented in chapter 2 assume some sort of noise distribution in the data since most predictive algorithms depend on pre-knowledge of noise distribution. However, in case noise distribution is unknown, noise estimation techniques can be used which generally use heuristic algorithms to characterize the distribution of noise in given data [235]. For example, [236] have presented a survey of methods for estimating noise in the data before applying SVM for regression or classification applications, including priori knowledge, distribution fitting, and regularization among others. However, noise estimation is never guaranteed to be accurate since it is only applicable to seen data and the distribution of noise in unseen data might be different.

4.1.3 DNNs and Outage Detection with DNNs

The stringent low latency requirements for future mobile cellular networks mean outage detection solutions with high computation complexity will not always be viable going forward. This means that solutions that require matrix completion, kriging, denoising and noise estimation will no longer prove to be viable options due to their high time complexity. Therefore, the ideal outage detection solution for future mobile cellular networks should have low computational complexity and high accuracy regardless of data density. To this end, a DNN based outage detection solution is proposed here and compared with the other state-of-the-art outage techniques described in chapters 2 and 3.

DNNs are an extension of artificial neural networks with the difference that instead of the traditional single hidden layer neural network, DNNs usually have multiple hidden layers with potentially different number of nodes within each layer. Other versions of deep neural networks include convolutional neural networks, autoencoder, and reduced Boltzmann machines among others [227].

Over the last few years, DNNs have become extremely popular for a wide variety of applications and mobile cellular networks are no different. DNNs have been used for interference control, channel estimation, traffic prediction, energy efficiency, and power allocation among others [237]. In terms of outage detection in mobile cellular networks, the number of DNN based outage detection solutions are very limited with the only relevant study being [238] where the authors have proposed to use DNN based autoencoder to detect outages in a homogeneous network.

Beyond that, several studies have used DNNs for anomalous traffic behavior detection in communication networks. In [239] the authors have used different DNN architectures to detect rogue packet in network traffic data transmitted over a wireless radio frequency channel. Similarly, [240] have used restricted Boltzmann machine based DNNs to detect anomalous packets in the network traffic data to detect suspicious activity. The same problem is the focus of [241] where a standard neural network is used to detect anomalous traffic data with comparisons against other machine learning algorithms.

4.1.4 Proposed Approach and Contributions

Given the limitations of techniques for dealing with data sparsity and noise variations in the data in terms of time complexity, and the issues faced by other machine learning techniques in dealing with spatio-temporal variations in sparse network coverage data, a DNN based outage detection solution is proposed here. The choice of DNNs for outage detection was made because of their inherent ability to extract hidden features from the data while being reasonably efficient in terms of computational complexity. The proposed solution takes in pre and post outage received signal strength at a user location and the distance of the user from serving cell and classifies whether the user is in outage or not. To evaluate the performance of the DNN based outage detection solution, it is compared with several state-of-the-art outage detection solutions including k-means clustering, SVM, EFD, autoencoder and ICA.

4.2 System Model and Solution Description

4.2.1 System Model

For the purpose of developing the DNN based outage detection solution a system with single-cell connectivity is assumed for the sake of simplicity. It also assumed that the user association changes immediately after an outage which is a reasonable assumption for mmWave UDHNs with highly overlapping cell coverage. Additionally, it is assumed that the user location along with received signal strength is frequently updated with the help of MDT reports [52]. It is also assumed that locations of the users do not change immediately before and after the outage which is also a fair assumption considering that mmWave UDHNs will be primarily deployed in dense urban environments with very slow mobility relative to the frequency of MDT reports. A user is assumed to be in outage when the downlink received power of that user from its associated cell $P^c_{r,u}$ falls below a threshold $P^{th}_{r_{out}}$ due to an outage such that:

$$\text{Outage} := P^c_{r,u} \leq P^{th}_{r_{out}} \tag{4.1}$$

and the user association changes from cell c to another cell $\hat{c}$.

To calculate $P^c_{r,u}$, the standard exponential pathloss model is considered, the log of which can be written as:

$$P^c_{r,u_{dBm}} = f(P^c_t, G_u, G^c_u, b, d^c_u, \beta) + \epsilon^c_u \tag{4.2}$$

where P^c_t is the transmit power of cell c, G_u is the gain of user equipment, G^c_u is the channel gain of cell c, b is the pathloss constant and depends on the clutter, ϵ^c_u is the random variation in the received power due to shadowing and other factors at the location of user u from cell c. d^c_u is the distance of subscriber u from cell c, and β is the pathloss exponent. Assuming each of G_u, G^c_u, b and β remains constant, (4.2) can simply be re-written as:

$$P^c_{r,u_{dBm}} = f(P^c_t, d^c_u) + \epsilon \tag{4.3}$$

4.2.2 Proposed DNN based Outage Detection Solution

DNNs are generally made up of several layers with at least one initial input layer, one internal hidden layer, and one obligatory output layer. While the inclusion of input and output layers are necessary, the number of hidden layers and their widths are flexible. A sample DNN algorithm architecture is shown in Fig. 4.1 where the circles represent neurons and the edges represents connection between neurons in different layers. Generally in DNNs neurons are fully connected, that is, each neuron in a preceding layer is connected to every neuron in the proceeding layer. Other variations such as convolutional neural networks do have sparse connectivity between neurons but for this study, a fully connected DNN is used.

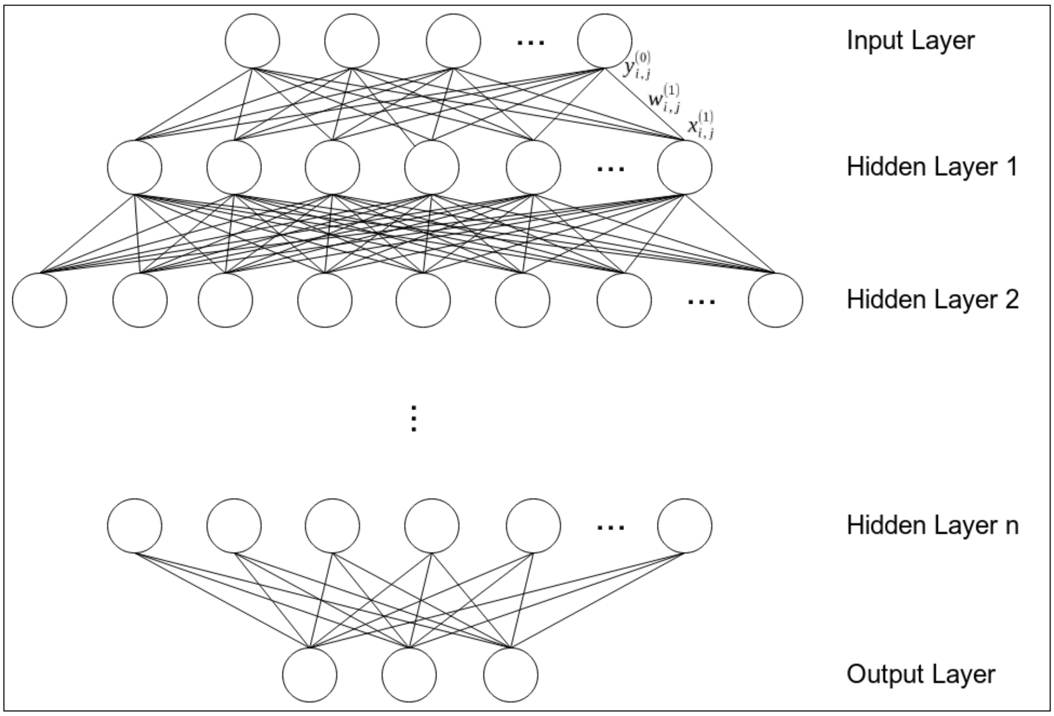

Fig. 4.1: DNN Algorithm Architecture

Each neuron in the hidden and output layers has an activation function along with three floats associated with it. These are: 1) *input weights* $\boldsymbol{w}$, 2) *output* $\hat{y}$, and *bias* b. The weights are the coefficients of each input connection to a neuron and are updated as the algorithm is trained, outputs are the result of activation function being applied to the weighted inputs, and biases are constants used to avoid overfitting. Given an input vector $\boldsymbol{x}$, activation function f, weights $\boldsymbol{w}$ and bias b, the output $\hat{y}_j^{(l)}$ of the j-th neuron of layer l will be:

$$\hat{y}_j^{(l)} = f(\boldsymbol{w_j^{(l)}}.\boldsymbol{x_j^{(l)}} + b_j^{(l)}) \tag{4.4}$$

The choice of activation function is dependent on the type of application for which the DNN is being used. A popular example of activation function for DNNs is the sigmoid function:

$$sigm(\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{x}}} \tag{4.5}$$

Sigmoid function is a popular activation function for hidden and output layer activation function because it is monotonically increasing and has a well-defined gradient function:

$$\frac{d(sigm(\boldsymbol{x}))}{d(\boldsymbol{x})} = sigm(\boldsymbol{x})(1 - sigm(\boldsymbol{x})) \tag{4.6}$$

However, sigmoid function is susceptible to the problem of exploding and vanishing gradients which can happen if the input values are too small or too big. To counter this issue, several other activation functions have been proposed including rectified linear unit, hyperbolic tangent, and softmax function among others [227]. Apart from that, regularization can also be used to avoid weights from exploding or vanishing.

Thus, the process of weight update becomes a critical one to train DNNs well. To update the weights of all the edges in the algorithm, DNNs follow a two stage process: 1) *feedforward* and 2)*backward propagation*. Feedforward refers to passing an input to the DNN so that outputs of each neuron in the first hidden layer are generated using (4.4) which are then used as input to the subsequent layer and so on until the final output is generated. The output is then compared to the actual target value using a loss function. If the aim of the DNN is to perform binary classification, as it will be used to do for outage detection problem, then binary cross-entropy can be used to calculate loss value E such that:

$$E = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)] \tag{4.7}$$

where N is the number of training samples, y_n is the target value and $\hat{y}_n$ is the predicted output of n-th training sample. Once the loss function is calculated, the backward propagation algorithm updates the weights of the DNN such that:

$$\Delta w_{ij}^{(l)} = -\chi \frac{\partial E}{\partial w_{ij}^{(l)}} \tag{4.8}$$

where $\Delta w_{ij}^{(l)}$ is the change in i-th weight of j-th neuron in layer l and χ is the learning rate which dictates how quickly the algorithm weights change. The partial derivative in the right-hand side of (4.8) can further be broken down such that:

$$\frac{\partial E}{\partial w_{ij}^{(l)}} = \begin{cases} (\hat{y}_j^{(l)} - y_j^{(l)})\hat{y}_j^{(l)}(1-\hat{y}_j^{(l)})\hat{y}_i^{(l-1)} & \text{if } j \text{ is an output neuron} \\ (\sum_{k\in K} w_{kl} \frac{\partial E}{\partial \hat{y}_k^{(l)}} \frac{\partial \hat{y}_k^{(l)}}{\partial \sum_{m=1}^{M} w_{mk}\hat{y}_m^{(l-1)}})\hat{y}_j^{(l)}(1-\hat{y}_j^{(l)})\hat{y}_i^{(l-1)} & \text{if } j \text{ is an inner neuron} \end{cases} \tag{4.9}$$

The weights of the DNN are updated for each batch of input samples until the algorithm converges to a minimum value or the training process is stopped.

Apart from the choice of activation function and learning rate, other important factors to be considered in building a DNN are the number of hidden layers and the number of neurons in each layer. More layers and more neurons will mean the number of computations to be done for each training batch will be higher but may also allow the DNN to converge faster and perform better.

For the problem of outage detection, a four layer DNN is used with four input features which are received power before the outage $P^c_{r,u_{t-\Delta t}}$, received power after the outage P^c_{r,u_t}, distance between user u and serving cell c before the outage $d^c_{u_{t-\Delta t}}$, and distance between user u and serving cell c after the outage $d^c_{u_t}$ where $\Delta t << t$. The distances can be computed using the user location information from the MDT measurement data. The DNN has three hidden layers of eight, four and two neurons each with rectified linear unit activation function. The output layer has one neuron with sigmoid activation function. The choice for layer count and widths was made after testing the performance of the algorithm from two to eight hidden layers. It was observed that the performance of the solution did not improve drastically after adding more than three hidden layers. For brevity, results for different layer counts are not presented here.

The DNN based outage detection solution is implemented using algorithm 2. A trained version of the DNN takes the four feature input vectors for all the users in the network $\mathbb{U}$ and classifies them as normal or outage affected. Once the classification is done, the algorithm updates a set of all the cell whose user were detected to be in outage $\hat{\mathbb{C}} \subset \mathbb{C}$ where $\mathbb{C}$ is the set of all cells. At the end of the algorithm, all user classifications $\hat{\boldsymbol{y}}$ along with the set of all outage affected cells is returned.

Algorithm 2 Outage Detection and Localization Using DNNs

Input: $P^c_{r,u_{t-\Delta t}}, P^c_{r,u_t}, d^c_{u_{t-\Delta t}}, d^c_{u_t}, \mathbb{C}, \mathbb{U}$, DNN
Output: $\hat{\boldsymbol{y}}, \hat{\mathbb{C}}$

for $u \in \mathbb{U}$ **do**
    Pass inputs $P^c_{r,u_{t-\Delta t}}, P^c_{r,u_t}, d^c_{u_{t-\Delta t}}, d^c_{u_t}$ to the DNN
    if then$\hat{y} == 1$
        if $c \in \hat{\mathbb{C}}$ **then** Continue
        else
            Add c to $\hat{\mathbb{C}}$
        end if
    end if
end for

4.2.3 Other Techniques for Outage Detection

To analyze the performance of DNN based outage detection solution, it is compared against state-of-the-art machine learning based anomaly detection solutions published recently. The approaches compared here can be broken down into two categories: 1) *supervised* outage detection techniques, and 2) *unsupervised* outage detection techniques. The DNN based outage detection solution presented here falls into the supervised category. SVM is another example of supervised technique for outage detection that has been used for outage detection in several recent studies [62, 63, 66]. The unsupervised techniques for outage detection used for comparison in this chapter include k-means clustering, ICA and EFD, all of which were described previously in chapter 3. In addition, the DNN based outage detection solution is also compared with the deep autoencoder based outage detection technique presented in [238]. A brief overview of SVM and deep autoencoder are presented below for reference.

Support Vector Machines

SVM algorithm belongs to the supervised family of machine learning algorithms which means it needs labeled or classified data for model training. SVM is an extremely popular technique, especially for high dimensional data. The aim of the SVM algorithm is to find the optimal hyperplane which fits the multidimensional data in the case of regression and splits the multidimensional data into separate classes in the case of classification.

To obtain the optimal decision boundary, the SVM algorithm applies a single-layer version of (4.4) to the input data. The definition of the function f in SVM is a choice and there are a variety of functions that transform multidimensional data to two dimensions for easier convergence of the optimization problem that SVM solves [236]. This optimization problem is given by:

$$\max_{\boldsymbol{w},b} \varrho, \tag{4.10}$$

$$\text{s.t.} \begin{cases} y_n(\boldsymbol{w}.\boldsymbol{x}_n + b) \geq \varrho, n = 1, \ldots, N \\ ||\boldsymbol{w}|| = 1 \end{cases} \tag{4.11}$$

where ϱ is the functional margin which separates the two classes in the classification problem, $\boldsymbol{w}$ is the weights vector and b is the bias. The optimization problem is solved using convex optimization or any other gradient descent algorithm. For the purpose of this study, SVM with radial basis function kernel is used [242].

Autoencoder

Deep autoencoders are a variation of DNNs which are used to learn lower dimensional representation of the input data. Autoencoders are actually a combination of two identical DNNs with the first DNN called the encoder and the second DNN called the decoder. The second DNN is an inverted version of the first DNN and the two DNNs are attached at the output layer of the first DNN. A simple autoencoder is shown in Fig. 4.2. Input data are passed to the encoder which reduces the input features to fewer dimensional output. That output is then used as the input to the decoder which uses it to reconstruct the original encoder input. The loss function is calculated by comparing the original and reconstructed data. In [238], the authors use this loss function to detect outages in the network. Coverage and quality measurements of top three serving cells at a user's location are passed to the autoencoder which then reconstructs the input data. An outage is detected if the reconstruction error for a user exceeds the mean loss plus two standard deviations. The same autoencoder solution is used here to identify users affected by an outage and the information is passed through algorithm 2 to localize the outage.

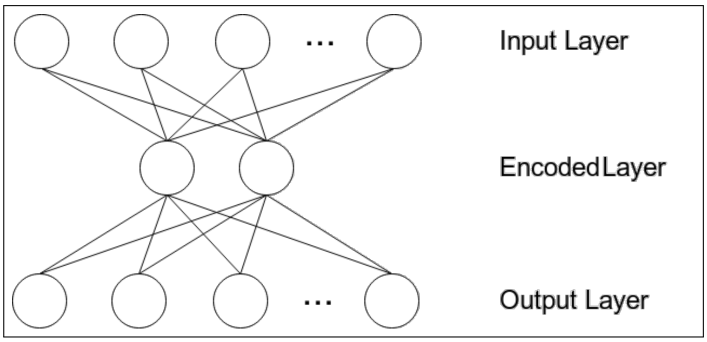

Fig. 4.2: Deep Autoencoder Architecture

4.3 Data Generation and Results

The following subsections describe the details of the algorithm evaluation methodology that was used for comparing the proposed DNN based outage detection solution with

Table 4.1: Parameter Settings for Simulation of DNN Based Outage Detection Solution

System Parameters	**Value**
Number of Sites	mmWave: 7; Small: 4
Transmission Frequency	mmWave: 28GHz; Small: 2 GHz
Transmission Bandwidth	mmWave: 100 MHz; Small: 20 MHz
Transmit Power	mmWave: 30 dBm, Small: 20 dBm
Antenna Tilt	mmWave: 7°; Small: 0°
Antenna Gain	mmWave: 18 dBi; Small: 5.7 dBi
Site Placement	Random
Shadowing	Street: 0 dB - 10 dB Open Space: 0 dB - 10 dB Grassland: 2 dB - 12 dB Low Vegetation: 4 dB - 14 dB Building $\geq$ 30m: 8 dB - 18 dB Building 12m - 30m: 10 dB - 20 dB Building $<$ 12m: 12 dB - 22 dB
User Density	10 - 35 users per cell

other solutions. Complete details of the parameters used for data generation are given in Table 4.1.

4.3.1 Network Layout Description

Data were collected through a real network coverage prediction tool Atoll [225] which is most commonly used for planning and optimization of actual networks by mobile cellular network operators. The same network topology as was shown in Fig. 3.2 was used with small cells and mmWave cells deployed randomly within the downtown New York area. Outage data were created by switching off Small Site 1 and mmWave Site 1 separately and collecting network coverage and user distance measurements before and after the outage.

4.3.2 User Distribution Description

Users are assumed to be dispersed across the network randomly but with overall fixed densities per cell. As given in Table 4.1, the comparison is done over six different user densities which vary from 35 users per cell to 10 users per cell. However, it is important to note that density per cell does not mean each cell will have that many users associated with it guaranteed. In fact, it was observed while obtaining the data that small cells generally have more users associated with them compared to mmWave cells due to the lower carrier frequency and omni-directional coverage.

4.3.3 System Shadowing Description

The performance of the outage detection solutions is compared over six different values of standard deviation of noise distribution in the data. For simplicity, the term shadowing is used for the remainder of this chapter to describe the noise in the data since the results were only compared for stationary systems. The standard deviation values of shadowing used in the comparison ranged from 0 dB to 10 dB open air shadowing with 2 dB increments for each subsequent clutter class. This means that if open air shadowing of standard deviation 0 dB is assumed, then for grassland it will be 2dB, for low vegetation it will be 4 dB, for high rise it will be 8 dB, for mid-size building it will be 10 dB, and for small building it will be 12 dB. The complete ranges of shadowing standard deviations are given in Table 4.1.

4.3.4 Training and Testing Methodology

The number of users affected by outages in a network will always be a small percentage compared to the total number of users. Therefore, if the user data are given as is to any machine learning algorithm, it will overfit on the normal user data distribution and label all users as normal. To avoid this, the training data were split into smaller randomly selected sets using random stratified sampling so that the ratio of users affected by outages remained the same. For example, for user density of 35 users per cell and 23 cells, 805 samples are randomly picked out of a total of over 100,000 user data samples. If the ratio of outage affected users in the original data is 1%, then the random stratified sampling ensures that the ratio of users in the final data is still around 1% which in this case will be 9 out of 805.

4.3.5 F1 score calculation

In datasets such as the coverage data with a small number of anomalies, it is generally recommended that the confusion matrix [243] should be used as the actual response of the algorithm instead of comparing the overall accuracy of the algorithm on the entire dataset. The confusion matrix allows the calculation of a set of other metrics such as precision and recall which are defined as follows:

$$\text{Precision} \stackrel{\text{def}}{=} \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \tag{4.12}$$

$$\text{Recall} \stackrel{\text{def}}{=} \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{4.13}$$

These metrics are used to calculate a cumulative score called F1 score which is defined as:

$$\text{F1 Score} \stackrel{\text{def}}{=} 2.\frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}} \tag{4.14}$$

F1 score gives a good measure of how accurate an algorithm is without losing important information about the detection of anomalies in the data as compared to a straight accuracy measure. Therefore, F1 score is used to compare the results for different outage detection solutions in this study.

4.3.6 Results

The following results for outage detection over varying user density and shadowing distribution are averages of the results for every combination of user density and open area shadowing standard deviation. For example, in the case of outage detection over varying user density, the results were first obtained for 200 test iterations over each user density and shadowing standard deviations. Then, the results were averaged over the 200 iterations, and finally the results were averaged over different shadowing standard deviations. The same process was applied to results for different shadowing standard deviations with the averaging being done over different user densities.

Outage Detection with Varying User Density

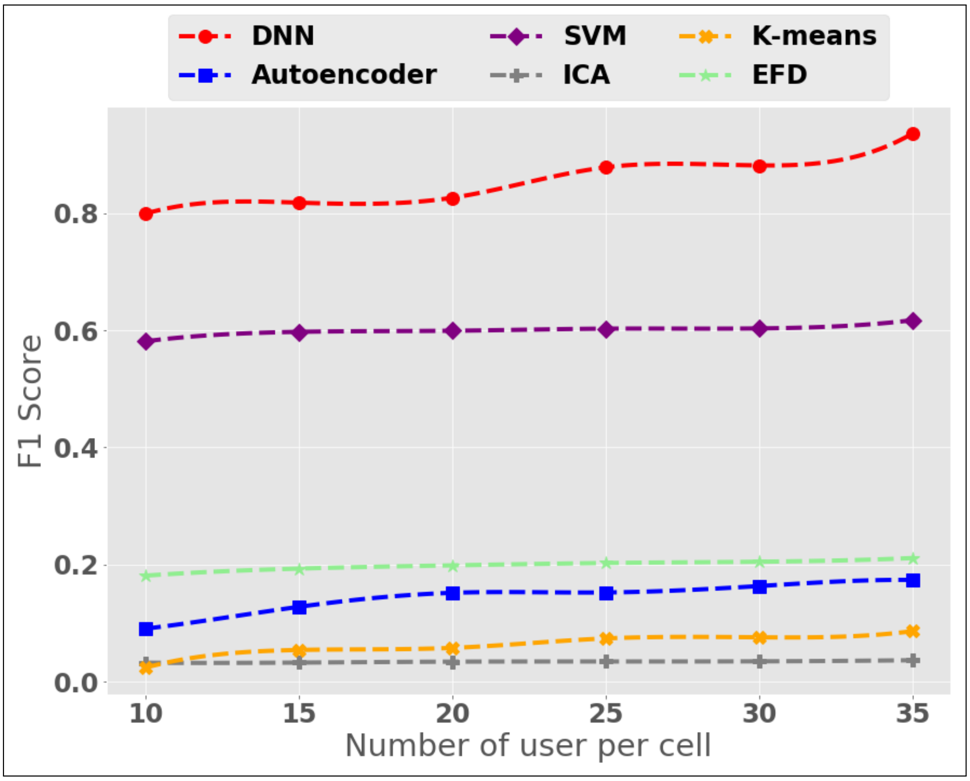

Fig. 4.3: F1 Score of mmWave cell outage detection over varying user densities with DNN, SVM, EFD, k-means clustering and ICA

As mentioned previously, the number of users per cell will be an important contributor to the sparsity of coverage data information in the network. Therefore, any outage detection solution used for mmWave UDHNs has to be able to detect outage even when the coverage data is sparse and with very little computation overhead to avoid latency issues. Fig. 4.3 shows the results for mmWave cell outage detection for different user densities. It can be clearly seen that the F1 score trend for all solutions is monotonically increasing with increase in user density with DNN based outage detection performing the best. In comparison, SVM performs better than all the other algorithms but still loses out to DNN based outage detection. The large gap in the performance of DNN and SVM compared to all the other techniques stems from the fact that the first two are supervised learning techniques and have learned the distribution of the outage data from the whole data. Compared to that, the rest of the techniques are unsupervised learning techniques which leaves them open to misidentification of users affected by shadowing as outages.

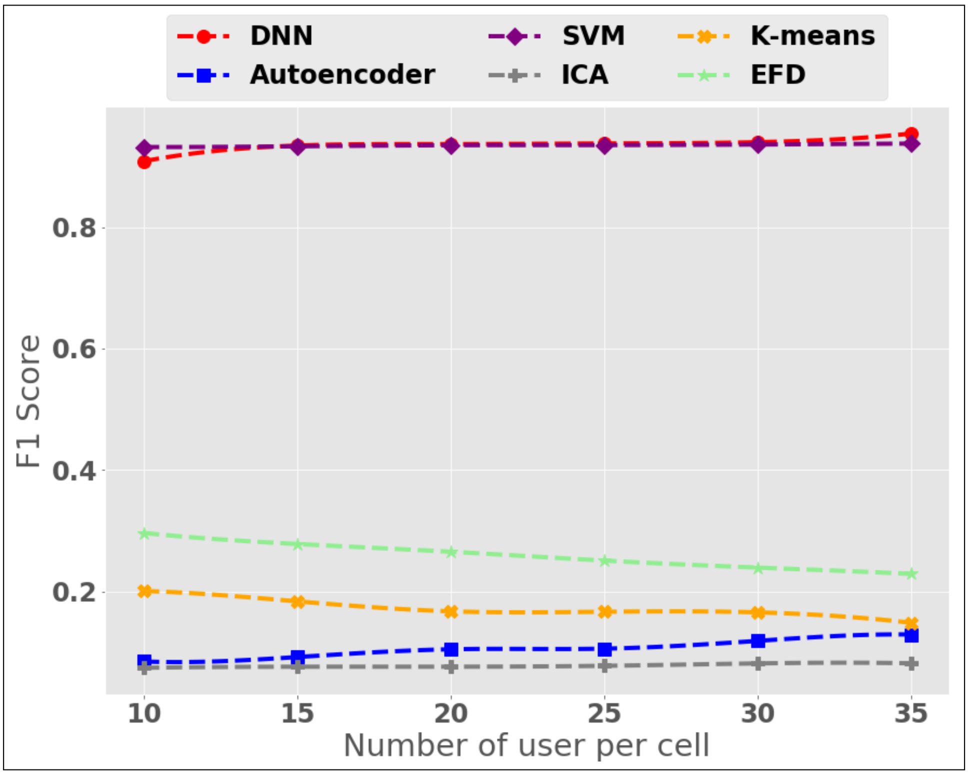

Fig. 4.4: F1 Score of small cell outage detection over varying user densities with DNN, SVM, EFD, k-means clustering and ICA

A special case is the outage detection over sparse data using EFD. Recall from chapter 3 that the data used for outage detection with EFD contained user information at every point in the network which meant that EFD could identify the precise boundaries of the source of an outage. Here, since outage detection is done per user with sparse coverage data, EFD does not perform as well, though it still manages to outperform the other unsupervised solutions. Another interesting observation is the low performance of the deep autoencoder based outage detection solution since it was shown in [238] that deep autoencoders perform very well for detecting outages. The potential causes of this are that the results presented in [238] were identifying outages with full network coverage data and fixed shadowing standard deviation and fixed user association before and after the outage. Here the performance is evaluated with sparse coverage data, changing shadowing standard deviation and changing user association.

Fig. 4.4 shows a comparison of the outage detection solutions for small cell outage. The results here show a somewhat different picture to mmWave cell outage detection. SVM performs better than DNN for smaller user densities while DNN performs better when user densities increase. This is because there are more users associated with small cells than mmWave in the data obtained from Atoll, which means that SVM is able to learn the class separating hyperplane more accurately compared to the mmWave cell outage case.

On the other side of the plot, EFD and k-means clustering now show a monotonically decreasing trend compared to mmWave cell outage detection where the trend was monotonically increasing with increasing user density. This can also be explained by the number of users affected by small cell outage. Since EFD tries to detect the information flow boundary around each user, it now has more false positives than true positives which means that precision becomes more dominant than recall in the F1 score calculation. The same is true for k-means clustering which now has more false positive detections than for the case of mmWave cell outage where the coverage of the mmWave cell was very limited.

Outage Detection with Varying Shadowing Standard Deviation

Fig. 4.5 shows a comparison of the outage detection techniques for different open area shadowing standard deviations and mmWave cell outage. Again, DNN performs better than all other algorithms with SVM a close second. An interesting observation is that

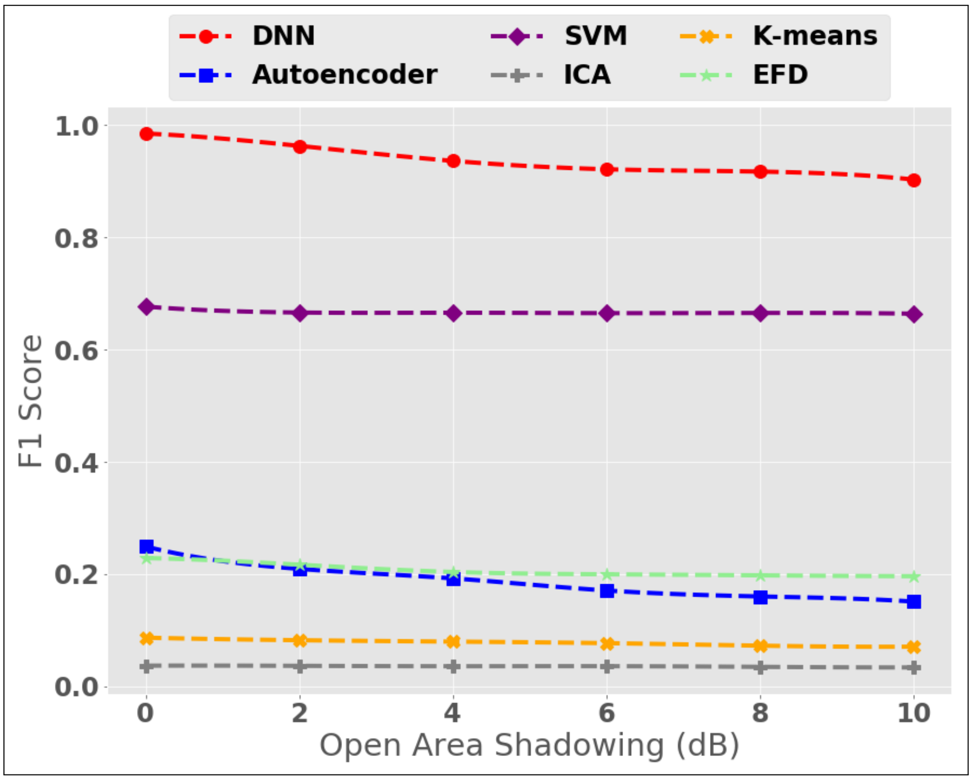

Fig. 4.5: F1 Score of mmWave cell outage detection over varying user open area shadowing standard deviations with DNN, SVM, EFD, k-means clustering and ICA

the F1 scores degrade much more steeply for DNN than for SVM which mean that for very high shadowing standard deviations, SVM actually performs better than DNN.

This is visible in the results for small cell outage detection in Fig.n4.6 which show that while DNN performs very well for low levels of shadowing, SVM performs better at shadowing standard deviations at and over 7 dB. And while the trend for SVM also shows that its performance decreases monotonically, its gradient is much lower compared to DNN.

In conclusion, DNNs perform very well for mmWave cell outage detection in general and at higher user densities and lower shadowing standard deviations. In comparison, SVM performs better than DNN at detecting small cell outages at higher shadowing standard deviations and when network coverage data is sparse. By comparison, in case of sparse data, unsupervised learning techniques do not perform as well as supervised techniques due to the fact that their precision is much lower compared to supervised techniques.

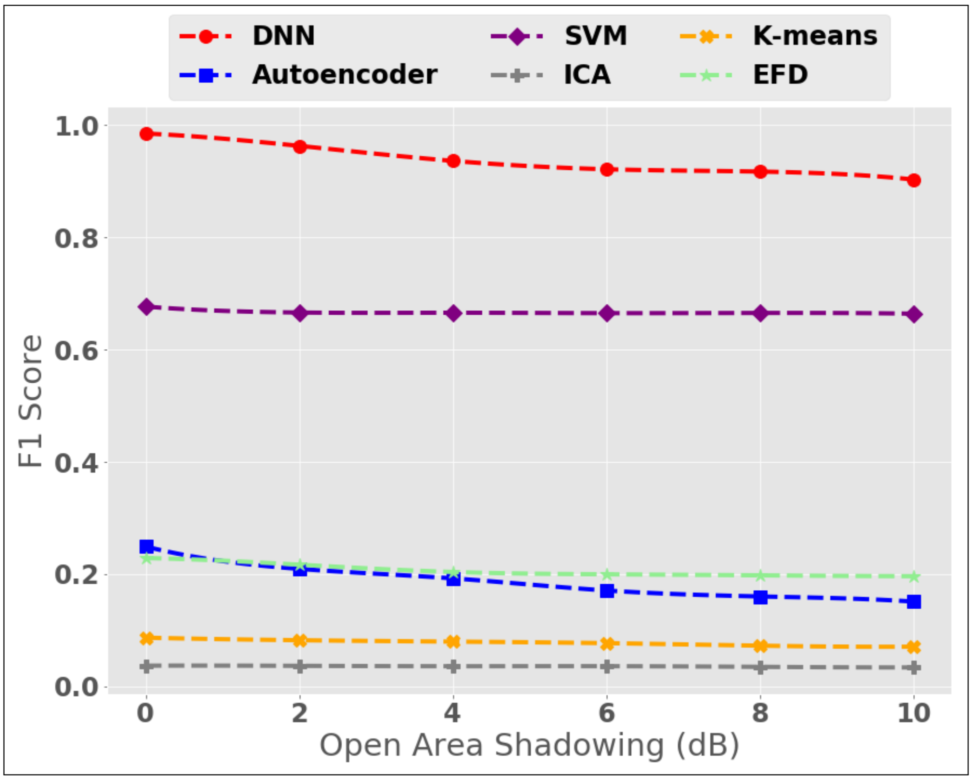

Fig. 4.6: F1 Score of small cell outage detection over varying user open area shadowing standard deviations with DNN, SVM, EFD, k-means clustering and ICA

4.4 Conclusion

This chapter has explored the potential problems which may be caused by sparse coverage data information and unpredictable noise distribution in the coverage data. In addition, some solutions for data sparsity and noise unpredictability have also been discussed along with their shortcomings, especially in terms of computational complexity which is critical for low-latency 5th generation and beyond mobile cellular networks. Based on the lessons learned from these techniques, a novel supervised deep neural network based solution for outage detection has been proposed here which is not only capable of detecting outages in unpredictable noise environments but also in sparse spatio-temporal coverage data. The results presented in this chapter compare the proposed deep neural network based outage detection method with several other outage detection methods including entropy field decomposition, support vector machines, independent component analysis, k-means clustering and deep autoencoders in a dense urban mmWave-small cell UDHN over a range of shadowing values and user densities. Results show that the proposed deep neural network based outage detection solution performs better than all other algorithms in detecting mmWave cell outage and performs comparably well with support vector machines for small cell outage. In addition, the results show that as the user density decreases, the prediction accuracy of all solutions decreases and the same result is observed for increasing shadowing standard deviation.

CHAPTER 5

Coordination for Partial Outage Avoidance: Framework and Proactive Partial Outage Avoidance Solution

So far, the solutions presented in chapters 2 and 3 have focused on reactively detecting outages in variable noise and sparse coverage data environments. However, as was explained in chapter 1, a key source of partial outages in exisitng and future mobile cellular networks are the parametric misconfigurations which would result from SON function conflicts. These parametric conflicts would be further exacerbated by the increase in network entities and OPs as explained in chapter 1 and through Figs. 1.3 and 1.4. In addition to these challenges, future networks will also have to adapt to enable the coexistence of multiple network types and layers as demonstrated in Fig. 5.1, further adding to the network complexity and potential partial outages.

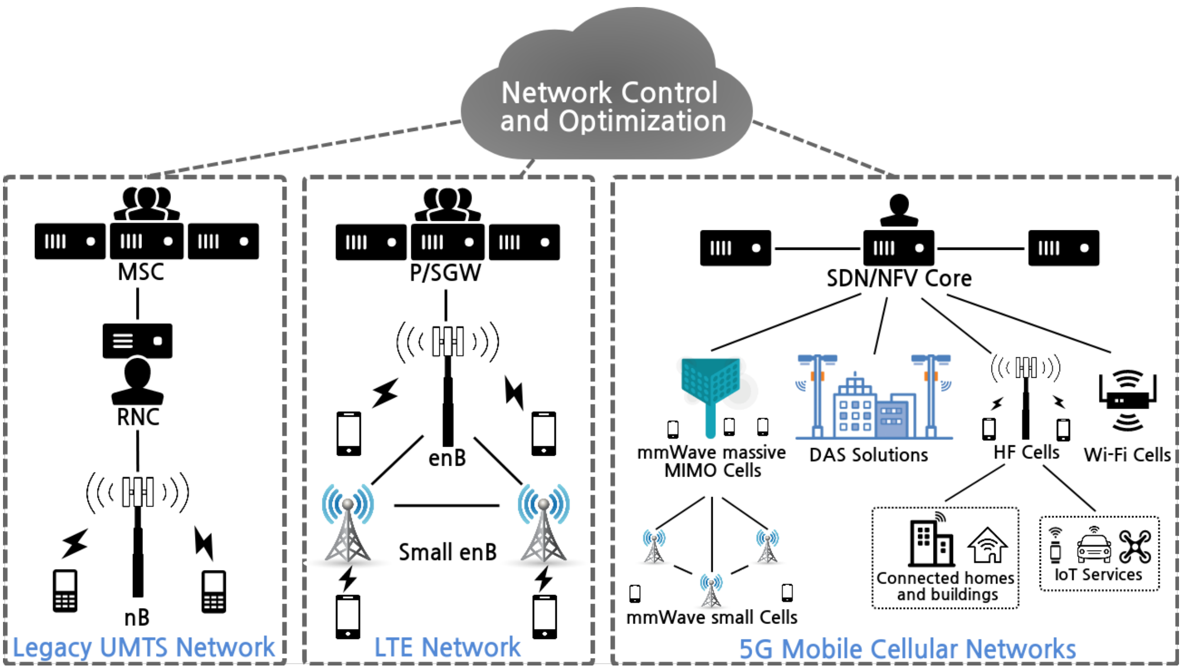

Fig. 5.1: Future Mobile Cellular Network Architecture.

Earlier versions of these challenges prompted network solution providers to shift away from traditional semi-manual methods of network control and optimization, and towards autonomous SON functions. SON functions were originally designed to reduce network operation cost and service disruptions due to human error. But over the years SON functions have become more sophisticated to the point where any future network deployment is inconceivable without them. Yet, despite their benefits, SON functions have not been as widely adopted by mobile cellular network operators as anticipated due to their potential conflicts described in chapter 2.

In this chapter, a case study is presented of how uncoordinated changes in just one optimization parameter (OP) can result in unpredictable network performance. Drawing inspiration from this case study, the need for a proactive instead of reactive approach towards SON coordination that uses both network and user behavior predictions to provide holistic coordination among SON solutions is identified. Then, a novel proactive SON coordination framework based on this network state characterization is proposed. The proposed framework is designed to resolve SON conflicts in future 5th generation and beyond networks with multitudes of network nodes, both mmWave and otherwise, and OPs by combining key aspects of big data analytics, machine learning, and human oversight. Finally, a solution leveraging the proposed framework is presented which proactively optimizes coverage, capacity and load by tuning cell transmit powers, antenna tilts and cell individual offsets to prevent partial outages due to parametric misconfiguration while maximizing user quality of experience.

5.0.1 Why is a Paradigm Shift in SON Coordination Needed?

Each of the four techniques described and discussed in chapter 2 has its own unique limitations. However, the biggest limitation of the state-of-the-art SON coordination methods stems from two underlying characteristics:

Static nature: This implies that once a state of the art SON coordination solution has been designed and deployed, its modification to reflect new realities is impossible. For example, a joint coordination solution may have been designed assuming Poisson traffic arrival distribution to maintain tractability. Such a solution would fail if hotspots develop within the network where there is very high traffic activity for short periods and very little activity otherwise.

Reactive nature: This implies that state of the art SON coordination solutions have limited visibility of only present state of the network to avoid conflicts. Here state implies current user and network behavior where user behavior is characterized by location, mobility status and traffic demand, while network behavior is characterized by the current KPI-OP relationship which in turn depends upon current user behavior.

5.1 Modelling SON Function Conflicts in a Dynamic Network with Parametric Interdependencies

To put the challenges of state of the art SON coordination solutions into perspective, a case study of modeling SON conflicts in a simple two cell network with mobile users is presented here.

Consider two cells A and B with users moving freely between the cells. In its natural state, the load imbalance between the cells would depend on the mobility of users. SON functions can avoid this by optimizing KPIs including, but not limited to 1) '$\mathfrak{A}$': difference in the number of users at cell A and cell B, 2) '$\mathfrak{B}$': the difference in cell A $\leftrightarrow$ cell B handovers (HOs), and 3) '$\mathfrak{C}$': the ratio of users bouncing between cells that is, the ratio of ping-pong HOs between cell A and cell B. To optimize these KPIs, the MCN operator may modify different soft and hard OPs including, 1) '$\mathfrak{a}$': load offset which is used to balance users between cells, 2) '$\mathfrak{b}$': HO offset which is used to control the number of HOs between cells, and 3) '$\mathfrak{c}$': inter-cell distance which is determined when planning the network.

Though these OPs effect the KPIs uniquely, the KPIs are also dependent on each other, thus forming an intertwined system. While such systems can be very complex and are certainly not limited to these KPIs or OPs, assuming that the ordinary differential equations (1) - (3), derived from [244], represent its basic example.

$$\frac{\partial \mathfrak{A}}{\partial t} = \mathfrak{a}\left(\mathfrak{B}_t - \mathfrak{A}_t\right) \tag{5.1}$$

$$\frac{\partial \mathfrak{B}}{\partial t} = \mathfrak{A}_t\left(\mathfrak{b} - \mathfrak{C}_t\right) - \mathfrak{b}_t \tag{5.2}$$

$$\frac{\partial \mathfrak{C}}{\partial t} = \mathfrak{A}_t\mathfrak{B}_t - \mathfrak{c}\mathfrak{C}_t \tag{5.3}$$

Eq. (5.1) signifies that the user count difference changes with the difference between $\mathfrak{A}$ and $\mathfrak{B}$ and can be controlled by the load offset parameter. Similarly, (5.2) implies that change in HO count difference depends on the difference in the number of users and HOs between cell A and B as well as the number of ping-pong HOs between them and can be controlled using HO offset parameter. Finally, (5.3) indicates that the rate of change of ping-pong HOs depends on the product of the difference in the number of users at cells A and B and the number of HOs between them and decreases if the distance between the two cells is increased.

Fig. 5.2 shows KPI interaction under when two OPs $\mathfrak{a}$ and $\mathfrak{c}$ are kept fixed and only $\mathfrak{b}$ is changed. For the first set of OP values, the KPIs quickly converge to a stable value. Now assume that the value of $\mathfrak{b}$ is increased from 5 to 10 by the load balancing SON function. As a result, the KPIs become less stable and take longer to converge resulting in degraded subscriber QoE. Such a situation is an example of a partial outage where the network performance degrades temporarily but eventually recovers to a stable, if not optimal, value.

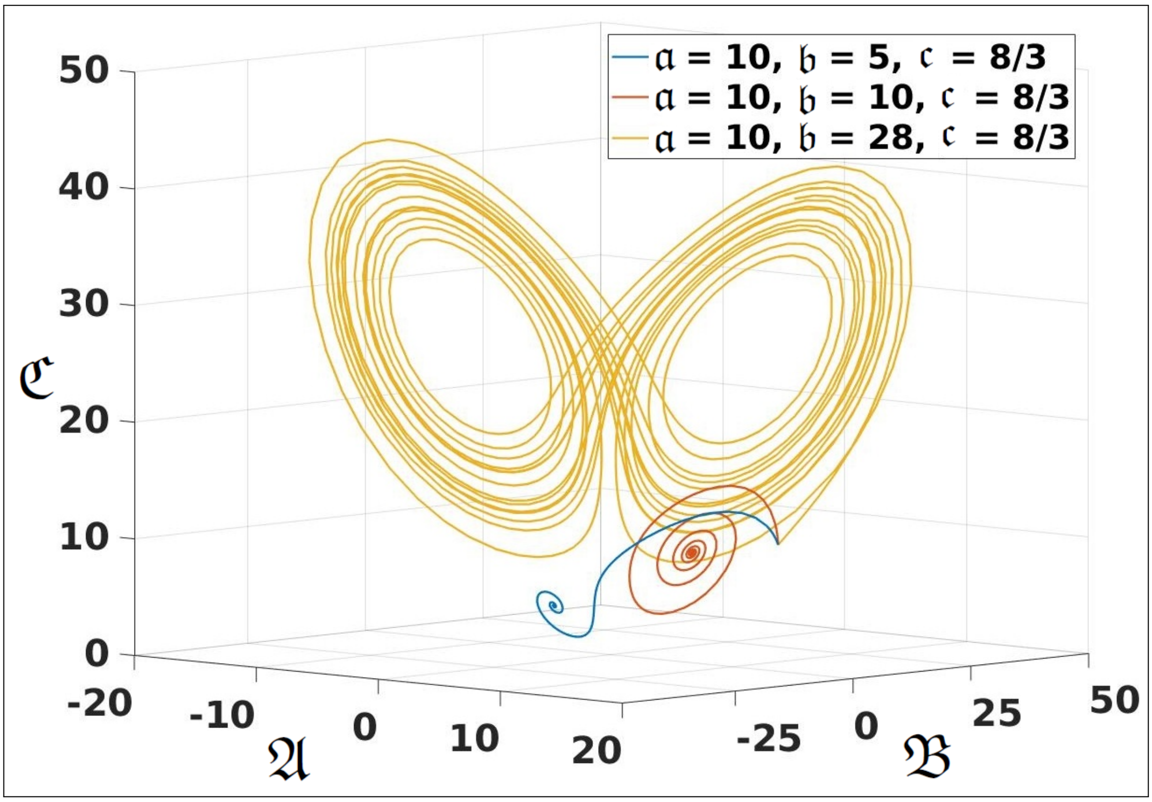

Fig. 5.2: Temporal Interaction of Three Performance Indicators.

In contrast to the first two settings, if $\mathfrak{b}$ is increased even further to 28, the KPIs become completely unstable and the system becomes chaotic. This implies that the cell loads will never reach equilibrium, and user associations will continue to change resulting in a breakdown of subscriber QoE and a complete outage.

Given the results in Fig. 5.2, in the absence of a SON coordination solution that has a holistic visibility of current and near future states of the network, the expected increase in network density and diversity in 5G MCNs may increase the likelihood of such conflicts and oscillations.

5.2 A SON Coordination Solution for 5th Generation and Beyond Mobile Cellular Networks

It is obvious from the preceding discussion that SON coordination in 5G MCNs cannot rely on static models of spatio-temporal user behavior. Therefore, future SON coordination solutions must have holistic visibility of the network state in terms of user activity and network behavior. This implies that the solutions must be able to identify how any user activity, such as mobility and traffic pattern, can influence network behavior and vice-versa. Additionally, these solutions must be able to predict network state up to some point in the foreseeable future to avoid OP settings based on current state from creating potential SON conflicts in the future.

Developing a holistic view of the network state requires understanding how the KPI-OP dependencies react to user behavior. However, as discussed previously, creating tractable models for all KPI-OP combinations for every type of user behavior is not practically possible. To overcome this challenge, a SON coordination framework is proposed here that can work with model-free characterization of network states to resolve potential SON conflicts. The proposed solution, shown in Fig. 5.3, builds on the big data-aided SON architecture in [8] by incorporating information from the network into a comprehensive state-aware SON coordination framework.

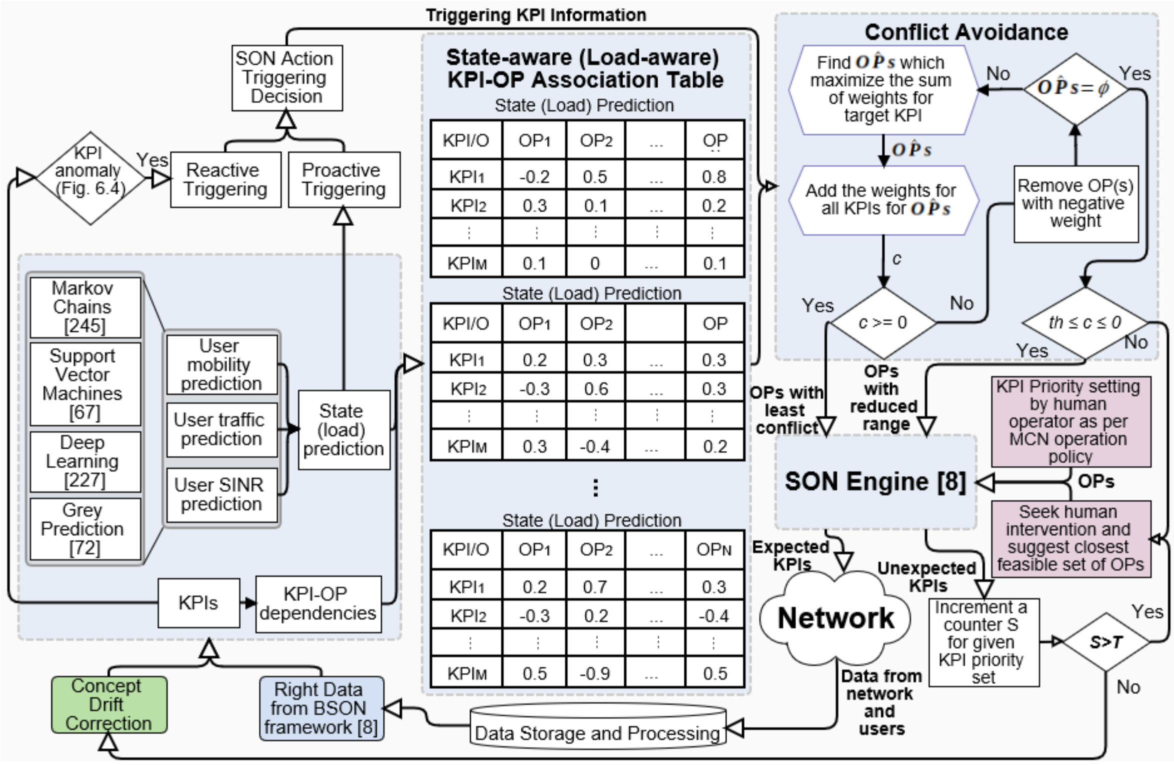

Fig. 5.3: SON Coordination Solution for 5G and Beyond.

5.2.1 Characterizing Network State

The *first step* towards enabling the proposed SON coordination framework is to characterize network state in terms of KPI-OP dependencies and user behavior. This can be accomplished using control and user plane data from the network, as well as contextual information from non-traditional sources such as social media, as described in [8].

The network behavior consists of KPI profiles and KPI-OP dependencies. KPI profiles use past and present KPI data to predict the normal behavioral trend of a KPI. These are then used in association with OP settings data to generate KPI-OP dependencies which define the impact of each OP on every KPI in terms of correlation co-efficients or weights (see KPI-OP tables in Fig. 1). The weights can be learned using techniques such as association mining or multi-modal auto-encoders which are a type of deep neural networks and can be trained to identify precisely which input OPs are associated with which KPI and what is the degree of association. Since different OPs have different units, these weights need to be normalized to reflect the impact of unit change in OP value on the value of the KPI.

The *second step* in the proposed framework is utilizing user mobility and traffic data to create spatio-temporal user behavior predictions. These include: 1) User mobility predictions which can be made using tools such as Markov chain processes as demonstrated in [245]; 2) User traffic predictions which can be made using tools such as Support Vector Machines and Deep Learning as demonstrated in [246]; and 3) User SINR predictions which predict the minimum expected SINR of the user. User SINR can be predicted using techniques such as Grey prediction as demonstrated in [71].

The *third step* in this process is to combine the above three types of predictions to determine network load or state which is basically a prediction of when, where and how users will utilize network resources. A demonstration of how such intelligence can be generated from underlying predictions is provided in [245].

5.2.2 Handling Data Dynamicity

A key challenge in practical utility of the network and user behavior prediction is the dynamicity of cellular environment. The dynamicity that is pertinent to proposed coordination framework can be boiled down to two types: that is, short and long-term dynamicity.

Addressing Short Term Dynamicity: Short-term dynamicity is a consequence of changes in user locations and traffic patterns over short periods ranging from minutes to hours. This results in variations in cell loads which in turn change the KPI-OP relationships. This type of dynamicity can be addressed by exploiting its intrinsic feature that is, its cyclic nature that stems from strong patterns in user mobility [245] and traffic generation/consumption routines [246]. Exploiting these patterns, the short-term dynamicity challenge can be resolved by creating finite set of KPI-OP association tables, where each table represents the KPI-OP dependencies for a given time window/part of the day (see

Fig. 5.3).

Addressing Long Term Dynamicity: Long-term dynamicity, also called concept drift, is a consequence of factors such as deployment of new sites, changes in topography, seasonal changes in vegetation, and significant changes in traffic pattern with public events and festivals. Therefore, long term dynamicity is difficult to predict and hence address via the approach proposed to address short term dynamicity. To re-calibrate network and user behavior predictions affected by concept drift, a weighted combinations of stored network and user behavior predictions and predictions based on current network and user data using ensemble methods such as bootstrap aggregation is created. The weighting factor of stored predictions will decrease as the number of times they result in non-optimal OP settings increases. This will prevent the SON engine from continuously producing wrong OP settings, while also preventing knee-jerk updates to the existing predictions.

5.2.3 SON Action Triggering

The *fourth step* in the proposed SON coordination framework is the design of SON function triggering mechanism. The SON triggering mechanism in the proposed coordination framework combines reactive and proactive SON triggering. The rationale behind this hybrid design is that reactive triggering is useful in slow-changing macro cell networks, while proactive SON triggering will be vital to meet the stringent QoE requirements in ultra-dense mmWave cell networks with fast user mobility and low cell sojourn times.

Reactive Triggering: SON actions are triggered in reaction to outages, either full or partial. One potential reactive outage or anomaly detection algorithm that can be employed in the SON coordination framework is illustrated in Fig. 5.4 [4]. For a comprehensive overview of other reactive anomaly detection in MCNs the reader is referred to a recent survey [247]. For example, abnormally high user traffic can lead to

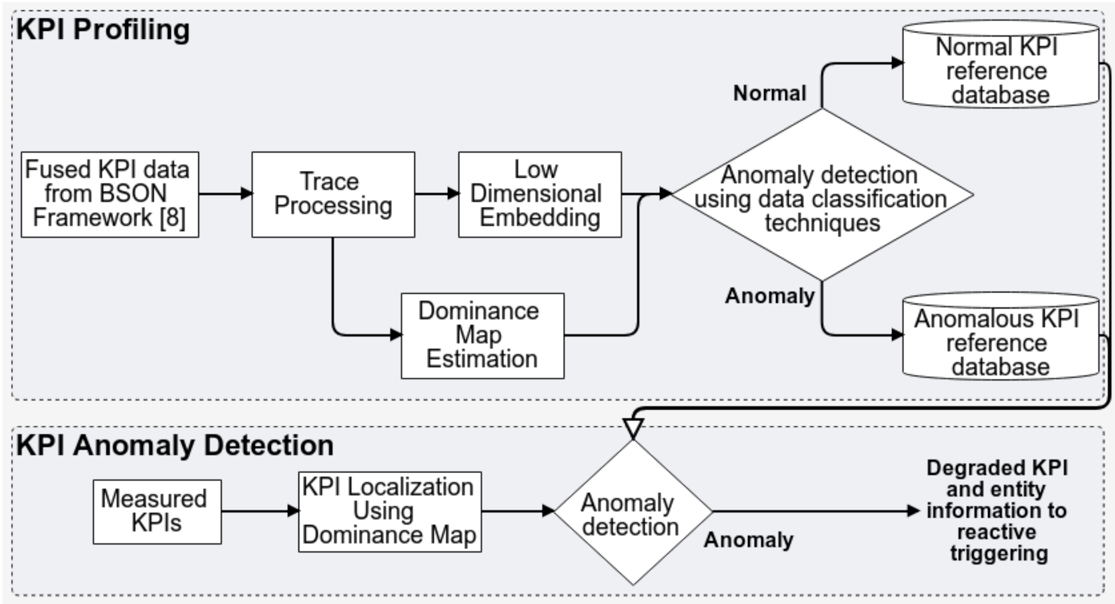

Fig. 5.4: KPI Anomaly Detection Algorithm

extreme congestion in the network which would result in unusually high call rejections. The reactive anomaly detection mechanism will detect this discrepancy in the KPI and trigger load balancing SON functions. However, while this triggering method is effective in providing self-healing/self-optimization capability in macro cell networks, it will not suffice to meet the stringent QoE requirements for 5G networks and beyond, and would only be employed as a back-up to the proactive triggering method.

Proactive Triggering: Proactive triggering uses prediction of the potential problem to preemptively trigger appropriate SON function. The enablers for proactive triggering can be cell load predictions to identify areas of the network where potential traffic congestion will most likely necessitate load balancing or additional capacity activation, or areas where network load will go down enough to allow energy saving. Other enablers include prediction of outages due to hardware/software failure, the models for which have been recently proposed in [17]. The main advantage of proactive triggering is reduced time to execution for given SON function, and hence improved QoE.

5.2.4 *Conflict Avoidance*

The *fifth and final step* in the proposed SON coordination framework is the conflict avoidance mechanism, that is, an algorithm that receives the call to reactively restore or proactively stabilize a KPI by invoking the appropriate SON function. The challenge here is to choose the OP(s) for adaptation that have only positive or no effect on other KPIs. As illustrated in [144], due to inter-dependencies among KPIs and OPs, this is a difficult task.

This problem is solved by making the KPI-OP inter-dependencies visible as explained above and illustrated in Fig. 5.3. This strips down the complex SON conflict avoidance problem into a relatively manageable two step problem: 1) Select the KPI-OP dependency table that corresponds to current network state/load; 2) In that table, choose the OPs that have maximum impact on the desired KPI and only positive or minimal negative impact on all other KPIs. The second sub-problem is analogous to a scheduling problem that can be solved using tools from operations research such as linear programming, integer programming, and dynamic programming to name a few [9]. The advantage of such a conflict avoidance mechanism over state-of-the-art SON coordination techniques lies in its ability to resolve SON function conflicts without necessitating tractable KPI-OP models for different network states.

Fig. 5.3 shows the schematic of conflict avoidance mechanism in the proposed solution where $\boldsymbol{O\hat{P}s}$ is the set of OPs for which the sum of weights ω_{i,KPI_j} with respect to the target KPI j is maximized, and c represents the sum of weights of $\boldsymbol{O\hat{P}s}$ for all the other KPIs k. In the following, the operation of the proposed conflict avoidance mechanism through is explained below with some examples.

Consider the state-aware KPI-OP association tables in Fig. 5.3 which illustrate how M different KPIs react to unit change in N different OPs for L different loads/network states. If it is assumed that a degradation in KPI_1 is the trigger to call a SON function when the network is in state 1, the OPs that maximize positive change in KPI_1 would be OP_2 and OP_N. Furthermore, these OPs do not have any negative impact on other KPIs; therefore, they can be used to obtain a KPI-OP function based on the techniques described in [8].

However, if the current load in the network corresponds to the second table that is, network is in state 2, it can be seen that the only usable OP is OP_N. This is because even though OP_1 and OP_2 have a positive impact on KPI_1, their individual impact on KPIs 2 and 3 are negative, making them unsuitable for stabilizing KPI_1.

A yet more extreme case can arise, as exemplified by state L, where none of the OPs has a non-negative impact on KPIs other than KPI_1. In such an event, the conflict avoidance algorithm will fail to converge. At this point, if c is smaller than some operator defined threshold th, the conflict avoidance mechanism can suggest the SON engine to optimize OP values within a very small range. However, if $c > th$, the conflict avoidance algorithm will trigger an alert to seek human expert intervention. The human expert can then decide on a compromise that is least deviant from the MCN operator policy in terms of QoE, energy efficiency, capacity or coverage etc.

Once the $\boldsymbol{O\hat{P}s}$ are found, they are passed on to the SON Engine which determines the optimal OP values and performs KPI optimization. An example of SON Engine implementation that can avoid risky OP settings by two-layer vetting process that is, based on system level simulator and human expert, is given in [8]. For KPI optimization purposes, SON Engine can use any one of analytical models, simulation based models, or network state predictions.

5.3 A Solution for Proactive Partial Outage Avoidance using Proposed Framework: Coordination between Coverage and Capacity Optimization (CCO) and Load Balancing (LB) SON functions

CCO and LB individually represent two of the most well researched SON functions in both academia and industry [144]. However, these SON functions represent a classic case of SON function conflicts with overlapping OPs, such as antenna tilts, transmit powers, and cell individual offsets, and conflicting KPIs that is, coverage, capacity and cell load. The following case study, derived from [248], describes an analytical solution to enable concurrent optimization of coverage, capacity and cell load using the SON framework illustrated in the previous section and Fig. 5.3 by optimizing the three most critical parameters for these KPIs, namely antenna tilts, cell transmit powers and cell individual offsets.

5.3.1 System Model

Here describe the system model employed in the formulation of the joint CCO-LB SON function is described along with the underlying assumptions.

Network and User Specifications

For formulating the joint CCO-LB problem, a network of hexagonal macro base stations is assumed with at least one randomly deployed small cell in the coverage area of each macro cell. 100% frequency reuse is considered between macro and small cells. Macro cells use directional antennas while small cells employ omni-directional antennas. An Orthogonal Frequency Division Multiple Access based system with resources divided into physical resource blocks (PRBs) of fixed bandwidth, is assumed. For conciseness, the downlink direction is chosen for the analysis as this is where most imbalance in coverage of macro and small cells occurs. It is assumed that users in the network are stationary. It is further assumed that requested user data rate is known which gives a lower-bound on the desired instantaneous user throughput. Desired user throughput can be modeled as a spatio-temporal function of subscriber behavior, subscription level, service request patterns, as well as the applications being used with the help of big data analytics as recently proposed in [249]. The formulation is not dependent on particular scheduling techniques.

Parameters and Measurements

Cell Loads: The instantaneous cell load can be defined as the ratio of PRBs occupied in a cell during a Transmission Time Interval and total PRBs available in the cell. This information is available as a standard measurement from 3GPP as "UL/DL total PRB usage" [147] and can be broadcast to the users. To define cell load η_c for this system model, first the minimum number of PRBs η_u^c to be allocated to a user is calculated:

$$\eta_u^c = \frac{1}{\omega_B}\left(\frac{\hat{\tau}_u}{f(\gamma_u^c)}\right) \tag{5.4}$$

where $\hat{\tau}_u$ represents the (desired) throughput of user $u \in \mathbb{U}_c$, where $\mathbb{U}_c$ is the set of all active users associated with cell c. γ_u^c represents the SINR of user u when associated with cell c and ω_B is the bandwidth per PRB. $f(\gamma_u^c)$ denotes the spectral efficiency of the user link for given SINR. If features such as MIMO or coding scheme gains and scheduling gains are considered, $f(\gamma_u^c)$ can be defined as $f(\gamma_u^c) := Alog_2\left(1 + B\gamma_u^c\right)$, where A and B are constants that can capture throughput gains (per PRB) achievable from various types of diversity schemes, or losses incurred by signaling overheads and hardware inefficiencies. For the sake of simplicity and without loss of generality, it can be assumed that $A = B = 1$. Thus, the residual cell capacity and cell load can be defined as:

$$\text{Residual Capacity} = \Lambda_c = N_b^c - \frac{1}{\omega_B}\left(\sum_{\mathbb{U}_c} \frac{\hat{\tau}_u}{log_2\left(1 + \gamma_u^c\right)}\right) \tag{5.5}$$

$$\text{Cell Load} = \eta_c = \frac{1}{N_b^c}\left(\frac{1}{\omega_B}\left(\sum_{\mathbb{U}_c}\frac{\hat{\tau}_u}{log_2\left(1+\gamma_u^c\right)}\right)\right) \tag{5.6}$$

where N_b^c is the total PRBs at cell c. Consequently, the range of cell load is $\eta_c \in [0, \infty)$. If the cell load exceeds 1, the cell in reality will be fully loaded and incoming users will be blocked. The value of load η_c is therefore referred to as virtual load and $\eta_c > 1$ reflects congestion in cell c.

Received Power: In LTE networks, downlink RSRP from nearby base stations is continuously monitored by the users and reported to the serving cell for a number of purposes. In the proposed CCO-LB approach the RSRP is used to calculate coverage probability in the network.

Cell Individual Offset: Cell individual offset (CIO) can be defined as a combination of multiple cell association parameters introduced by the 3GPP [250] including cell hysteresis, cell offsets and event related offsets which are used to decide user association. CIO information is by each cell and decoded by the users as part of standard operation. For the purpose of this study CIO is treated as a simple virtual boost in RSRP.

5.3.2 Problem Formulation

To incorporate user quality of experience into the joint CCO-LB optimization, the problem is formulated as a per cell per user throughput optimization problem. The first step towards this goal is to build a SINR model as function of all three optimization parameters under consideration.

User SINR as Function of Tilt, Transmit Power and CIO

Downlink SINR $\hat{\gamma}_u^c$ of a reference signal at user location u when associated with cell c can be expressed as the ratio of RSRP $P_{r,u}^c$ measured by user u from cell c to the sum of RSRP measured by user u from all interfering cells i such that $\forall i \in \mathbb{C}/c$, and the noise power κ:

$$\hat{\gamma_u^c} = \frac{P_t^c G_u G_u^c \epsilon u^c a \left(d_u^c\right)^{-\beta}}{\kappa + \sum_{\forall i \in \mathbb{C}/c} P_t^i G_u G_u^i \epsilon u^i a \left(d_u^i\right)^{-\beta}} \tag{5.7}$$

where P_t^c and P_t^i are the transmit powers of serving cell c and interfering cell i, G_u is the gain of user equipment, G_u^c and G_u^i are the gains of transmitter antenna of the cells c and i towards user u, ϵ_u^c and ϵ_u^i is the shadowing observed at the location of user u from serving cell c and interfering cell i, a is the pathloss constant, d_u^c and d_u^i represent distance of user u from cell c and i, and β is the pathloss exponent. The numerator in (5.7) is obtained from the standard exponential pathloss model while ϵ_u^c and ϵ_u^i in equation (5.7) can be modeled as random variables with either Gaussian or log-normal distribution varying over both space and time.

The expression in (5.7) is only useful when estimating the quality of reference signals which are always being transmitted by all the cells. Thus, $\hat{\gamma_u^c}$ is not a true measure of SINR on the PRBs where interference generated is dependent on utilization of that same PRB in other cells at the same time. it is assumed that user arrival in the system follows a general distribution, thus the exact interference becomes a function of time. Therefore, to obtain an SINR estimate independent of time, a reasonable low complexity substitute for average downlink interference from a cell i is to use the ratio of occupied PRBs in the cell. The expression for SINR estimate for user u in cell c can then be given as:

$$\gamma_u^c = \frac{P_t^c G_u G_u^c \epsilon_u^c a \left(d_u^c\right)^{-\beta}}{\kappa + \sum_{\forall i \in \mathbb{C}/c} \hat{\eta}_i P_t^i G_u G_u^i \epsilon_u^i a \left(d_u^i\right)^{-\beta}} \tag{5.8}$$

where $\hat{\eta}_i$ denotes actual cell load in a cell, that for a cell i can be obtained by modifying (5.6) as:

$$\hat{\eta}_i = \frac{1}{N_b^i} \left(\frac{1}{\omega_B} \left(\sum_{\hat{\mathbb{U}}_i} \frac{\hat{\tau}_u}{log_2\left(1 + \gamma_u^i\right)} \right) \right) \tag{5.9}$$

where $\hat{\mathbb{U}}_c \subseteq \mathbb{U}_c \subseteq \mathbb{U}$ is the set of all active user associated with cell c. Here $\mathbb{U}$ represents the complete set of users in the network and the difference set $\mathbb{U}_c - \hat{\mathbb{U}}_c$ represents users who requested but were denied resources by the cell c due to congestion which implies $\hat{\eta}_c \in [0, 1]$. Note that in SINR expression (5.8) above, the virtual cell load from (5.6) is not used, but the actual cell load which can never exceed 1.

As macro cells in the system under consideration use directional antennas, using the expression for 3D antenna gain from [251], the gain from base station to user G_u^c can be given as:

$$G_u^c = 10^{-1.2\left(\lambda_v\left(\frac{\psi_u^c - \psi_{tilt}^c}{B_v}\right)^2 + \lambda_h\left(\frac{\phi_u^c - \phi_{azi}^c}{B_h}\right)^2\right)} \tag{5.10}$$

where λ_h and λ_v are the weights of horizontal and vertical beam patterns of the antenna, ψ_u^c is the vertical angle between user c and the antenna of cell c, ψ_{tilt}^c is the tilt angle of serving cell antenna, ϕ_u^c is the horizontal angle of user u from cell c, ϕ_{azi}^c is the azimuth of antenna of cell c, and B_h and B_v are horizontal and vertical beam widths of the transmitter antenna of cell c. As the variable of interest in (5.10) is tilt angle and the rest of the antenna parameters can be treated as constants, for the sake of conciseness (5.10) can be simplified using the following substitution:

$$x_u^c = \frac{(B_v)^2 \lambda_h}{\lambda_v} \left(\frac{\phi_u^c - \phi_{azi}^c}{B_h}\right)^2 \tag{5.11}$$

and re-write the SINR expression from (5.8) as:

$$\gamma_u^c = \frac{P_t^c G_u 10^{\mu\left(\left(\psi_u^c - \psi_{tilt}^c\right)^2 + x_u^c\right)} \epsilon_u^c a\left(d_u^c\right)^{-\beta}}{\kappa + \sum_{\forall i \in \mathbb{C}/c} \hat{\eta}_i P_t^i G_u 10^{\mu\left(\left(\psi_u^i - \psi_{tilt}^i\right)^2 + x_u^i\right)} \epsilon_u^i a\left(d_u^i\right)^{-\beta}} \tag{5.12}$$

where μ is consolidated constant based on fixed antenna characteristics.

Finally, CIO in the SINR expression is addressed. This offset parameter is used for cell association as:

$$P_{r,u_{dBm}}^c = \acute{P}_{r,u_{dBm}}^c - P_{CIO_{dB}}^c \tag{5.13}$$

where $P^c_{r,u_{dBm}}$ is the true signal power in dBm received by user u from cell c and $\acute{P}^c_{r,u_{dBm}}$ is the received power reported back by user u to cell c in dBm. This value includes $P^c_{CIO_{dB}}$ (the CIO value of cell c in dB) which is then subtracted by the cell to retrieve $P^c_{r,u_{dBm}}$.

The motivation behind introduction of CIO was to allow load balancing among cells. However, if CIO has to be invoked to alter natural RSRP based cell association for the user, the SINR for that user is bound to be lower. Nevertheless, CIO is a necessary means to balance cell loads while capacity loss due to drop in SINR can partially be offset if the cell association takes into account cell load in addition to RSRP.

An Improved Load-aware User Association Mechanism

The state-of-the-art method of determining user associations $\mathbb{U}_c$ is to use the RSRP measurements along with CIO values as given in (5.13). However, this method overlooks the key role of user association in overall capacity and QoS through cell load and SINR distributions. To overcome this challenge, user association with cell j can be established not only based on received power but also load in that cell. More specifically, this load-aware user association with cell j can be determined as:

$$\mathbb{U}_j := \left\{ \forall u \in \mathbb{U} \mid j = \arg\max_{\forall c \in \mathbb{C}} \left(\left(\frac{1}{\eta_c} \right)^{\alpha} * \left(\acute{P}^c_{r,u_{dBm}} \right)^{(1-\alpha)} \right) \right\} \tag{5.14}$$

where $\mathbb{U}_j$ is a set of all users for whom (a scaled version of) the product of the RSRP(+CIO) in Watts $\acute{P}^c_{r,u}$ and the normalized residual cell capacity is maximized for cell j. $\alpha \in [0,1]$ is a weighting factor introduced to allow trading between the impact of RSRP and cell load measurements in the user association. As established in (5.6), cell load is dependent on the SINR of users in the cell that is, better the SINR of users in candidate cell, lesser the load in the cell for given traffic demand. Note that in (5.14), to make new user association decision with a cell the virtual load is used and not the actual load. While, using actual cell load that has range $\hat{\eta}_c \in [0,1]$ can indicate the current load in a cell, it cannot help take into account the users that are already associated with that cell but were not served. On the other hand, virtual cell load as defined in (5.6) with range $\eta_c \in [0, \infty)$, provides a truer picture of effective potential load in the candidate cell.

The expression in (5.14) gives the set $\mathbb{U}_j$ of users to be associated with the cell j and thus represents both active and idle users. On the other hand, the set $\mathbb{U}_c$ used in the expression for SINR in (5.12) represents the set of only active users associated with the cell c. With $\alpha = 1$, the user association simply becomes a function of cell load and SINR at the time of association. Consequently this cell association espouses the LB SON function only. On the other hand, if $\alpha = 0$, the proposed user association method simply represents state-of-the-art RSRP based cell association method which helps achieve coverage optimization aspect in the CCO SON function. Determining the optimal value of weighting factor α is an optimization problem worth investigating in itself. In the results, the KPIs are evaluated with a range of α.

Problem Statement

A common approach towards throughput maximization in LB or CCO is to use a problem formulation that maximizes the mean throughput per user per cell. However, if the arithmetic mean of user throughput determined by SINR expression derived above is maximized, users with no throughput and cells with no load will be equally acceptable as users with very high throughputs and cells with full loads. While such formulation will achieve the objectives of CCO, it will not perform load balancing, and hence cannot be suitable approach for joint CCO-LB. To simultaneously reflect the goals of both CCO and LB in a single objective function,the objective function to be modeled is given as:

$$\max_{\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}} \left(\prod_{\mathbb{C}} \left(\prod_{\mathbb{U}_c} \omega_u^c log_2 \left(1 + \gamma_c^u\right) \right)^{\frac{1}{|\mathbb{U}_c|}} \right)^{\frac{1}{|\mathbb{C}|}} \tag{5.15}$$

The outer geometric mean in this formulation dampens the load disparity among cells, and thus integrates LB goal into the optimization objective. This formulation is intended for scenarios where user required rates are not known or predicted. Thus, use of inner geometric mean instead of arithmetic mean for user throughput protects users with lower SINR from being unfairly treated, while maximizing the overall throughput.

However, if the framework described in section 5.3 and Fig. 5.3 is used, the desired user throughput can be predicted which allows the adoption of a more greedy approach by replacing the inner geometric mean with arithmetic mean as it is bound to provide an improved or equivalent result [252]. The new objective function with this assumption is given as:

$$\max_{\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}} \left(\prod_{\mathbb{C}} \left(\frac{\sum_{\mathbb{U}_c} \omega_u^c log_2 \left(1 + \gamma_c^u\right)}{|\mathbb{U}_c|} \right) \right)^{\frac{1}{|\mathbb{C}|}} \tag{5.16}$$

A comparison between performance of both formulations is presented in the results. The formulations in (5.15) and (5.16) inherit two basic constraints to achieve full objectives of CCO and LB SON function that is,:

i The ratio of covered users C must meet or exceed the minimum network coverage threshold ϖ that is, $C \geqslant \varpi$ where C is dependent on the number of users satisfying the equation $P_{r,u}^c \geqslant P_{th}^c$;

ii Cell load, as defined in (5.6), for every cell has to be less than or equal to the cell load thresholds set by operator policies: $\eta_c \leqslant \eta_{th}^c \forall c \in \mathbb{C}$

An additional constraint is defined in the formulation to avoid blocking any users that is,:

iii The set of served active users $\hat{\mathbb{U}}_c$ by cell c must be equal to the total set of active users $\mathbb{U}_u$ associated with the cell c: $\hat{\mathbb{U}}_c = \mathbb{U}_c$.

The satisfaction of constraint (i) depends heavily on the pathloss model employed in (5.7). Despite the assumption that user location remains the same over time, random variations in shadowing ϵ_u^c over space introduce uncertainty into the determination of $P_{r,u}^c$. Consequently C becomes a function of the distribution of ϵ_u^c such that constraint (i) becomes $Pr(C(\epsilon_u^c)) \geqslant \varpi$. This also implies that the evaluation of $P_{r,u}^c \geqslant P_{th}^c$ is a probabilistic problem rather than a deterministic one which can make the overall problem intractable. In order to overcome this issue, constraint (i) can be reformulated such that it becomes deterministic.

Proposition 6.1: *For Gaussian distributed shadowing ϵ_u^c, the probable coverage ratio $Pr(C(\epsilon_u^c))$ can be estimated using the transformation* $\frac{1}{|\mathbb{C}|}\sum_{\mathbb{C}}\frac{1}{|\mathbb{U}_c|}\sum_{\mathbb{U}_c} 1\left(P_{r,u}^c \geqslant P_{th}^c\right)$.

Proof The complete proof of Proposition 6.1 is provided in Appendix A. ■

Substituting the expression for SINR from (5.12) in (5.15) gives the fair joint CCO-LB formulation given in (5.17a), while substituting SINR from (5.12) in (5.16) gives the greedy joint CCO-LB formulation given in (5.17b). Combining the two formulations with the above problem constraint and user association expression in (5.14) gives the final formulation in (5.17).

$$\max_{\boldsymbol{P_t^c},\boldsymbol{\psi_{tilt}^c},\boldsymbol{P_{CIO}^c}} \Omega =$$

$$\max_{\boldsymbol{P_t^c},\boldsymbol{\psi_{tilt}^c},\boldsymbol{P_{CIO}^c}} \left(\prod_{\mathbb{C}}\left(\prod_{\mathbb{U}_c}\omega_u^c log_2\left(1+\frac{P_t^c G_u 10^{\mu\left((\psi_u^c-\psi_{tilt}^c)^2+x_u^c\right)}\epsilon_u^c a\,(d_u^c)^{-\beta}}{\kappa+\sum_{\forall i\in\mathbb{C}/c}\hat{\eta}_i P_t^i G_u 10^{\mu\left((\psi_u^i-\psi_{tilt}^i)^2+x_u^i\right)}\epsilon_u^i a\,(d_u^i)^{-\beta}}\right)\right)^{\frac{1}{|\mathbb{U}_c|}}\right)^{\frac{1}{|\mathbb{C}}} \tag{5.17a}$$

OR

$$\max_{\boldsymbol{P_t^c},\boldsymbol{\psi_{tilt}^c},\boldsymbol{P_{CIO}^c}} \left(\prod_{\mathbb{C}}\left(\frac{\sum_{\mathbb{U}_c}\omega_u^c log_2\left(1+\frac{P_t^c G_u 10^{\mu\left((\psi_u^c-\psi_{tilt}^c)^2+x_u^c\right)}\epsilon_u^c a(d_u^c)^{-\beta}}{\kappa+\sum_{\forall i\in\mathbb{C}/c}\hat{\eta}_i P_t^i G_u 10^{\mu\left((\psi_u^i-\psi_{tilt}^i)^2+x_u^i\right)}\epsilon_u^i a(d_u^i)^{-\beta}}\right)}{|\mathbb{U}_c|}\right)\right)^{\frac{1}{|\mathbb{C}|}} \tag{5.17b}$$

$$\textbf{subject to} = \begin{cases} \frac{1}{|\mathbb{C}|}\sum_{\mathbb{C}}\frac{1}{|\mathbb{U}_c|}\sum_{\mathbb{U}_c} 1\left(P_{r,u}^c \geqslant P_{th}^c\right) \geqslant \varpi, \\ \eta_c \leqslant \eta_{th}^c \forall c \in \mathbb{C} \\ \hat{\mathbb{U}}_c = \mathbb{U}_c \end{cases} \tag{5.17c}$$

$$\mathbb{U}_j := \left\{\forall u \in \mathbb{U} \mid j = \arg\max_{\forall c\in\mathbb{C}}\left(\left(\frac{1}{\eta_c}\right)^{\alpha} * \left(\acute{P}_{r,u_{dBm}}^c\right)^{(1-\alpha)}\right)\right\} \tag{5.17d}$$

5.3.3 Solution Methodology

Firstly, the convexity of the joint **C**CO **L**B user **As**sociation aware **S**ON function (**CLASS**) presented in (5.17) is analyzed and then methodologies to implement it are presented.

Convexity Analysis

Assuming a network of macro cells only, if the range of transmission powers can be defined as $P_t^c \in [20W, 40W]$, antenna tilts as $\psi_{tilt}^c \in 90° + [0°, 15°]$ and CIOs as $P_{CIO}^c \in [0dB, 10dB]$. Affine sets are convex sets, therefore, the first requirement for convexity for problem (5.17) that is, the constraints should be convex, is fulfilled. It has been proven that geometric and arithmetic means preserve convexity of a function. it is also known that the logarithmic function is also a convex function over the interval $(0, \infty)$. This leaves, the SINR expression in (5.12) to be examined to see if the formulation in (5.17) is convex or not.

Proposition 6.2: *SINR as a function of antenna tilts as given in* (5.12)*, is a non-convex function.*

Proof: Fig. 5.5 plots the interference (denominator of (5.12)) as function of antenna tilts of two neighboring cells. Clearly it is not a convex function implying Proposition 6.2. A more formal proof of Proposition 6.2 is provided in Appendix B. ■

Alternate Solution Methodologies

Given the non-convexity and large scale of the problem, heuristic approaches can be used that can find optimal or near optimal solution of the formulation in (5.17).

Algorithm to Implement Proposed Cell Association: Before delving into possible non-convex optimization techniques to solve (5.17), an algorithm to practically imple-

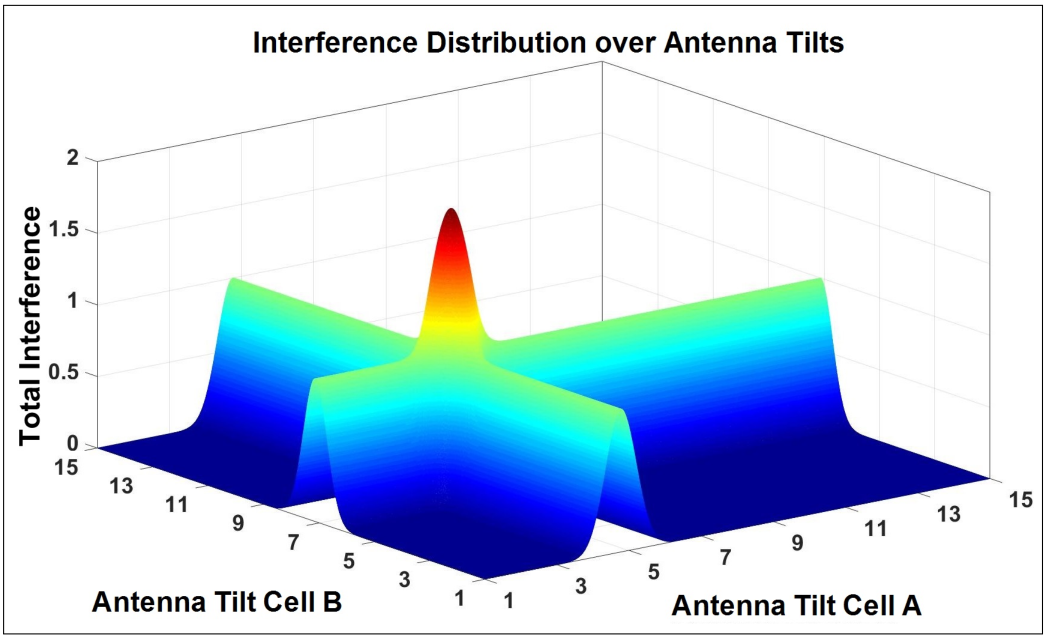

Fig. 5.5: Interference Distribution Over Antenna Tilts of Interferers

ment the proposed user associations for given values of the three optimization parameters and obtain the updated value of objective function with new user associations is presented in Algorithm 3. This routine has to be called at each iteration of the heuristic optimization techniques to be discussed in the following.

Sequential Quadratic Programming (SQP): One way to solve non-convex problems of the type (5.17) that have linear constraints is to approximate it piece-wise with a convex quadratic function and then use convex optimization to solve it, a method also known as sequential quadratic programming. To leverage SQP the problem in (5.17) is re-written as:

$$\max_{\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}} -\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \tag{5.18}$$

$$\textbf{subject to} = \begin{cases} W\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) := \\ \quad \varpi - \frac{1}{|\mathbb{C}|}\sum_{\mathbb{C}} \frac{1}{|\mathbb{U}_c|}\sum_{\mathbb{U}_c} 1\left(P_{r,u}^c \geqslant P_{th}^c\right) \geqslant 0, \\ X\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) := \eta_c - \eta_{th}^c \leqslant 0 \forall c \in \mathbb{C} \\ Y\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) := \hat{\mathbb{U}}_c - \mathbb{U}_c = 0 \end{cases}$$

Algorithm 3 Objective Function (5.17) Implementation Routine

Input: $\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}$
Output: $\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right)$ (5.17a) or (5.17b)

for $u \in \mathbb{U}$ **do**
  Find serving cell $j = \arg\max_{\forall c \in \mathbb{C}} \left(\acute{P}^c_{r,u_{dBm}}\right)$
  Calculate SINR $\hat{\gamma}_u^j$ and η_u^j
end for
for $c \in \mathbb{C}$ **do** Calculate cell load η_c
end for
for $u \in \mathbb{U}$ **do**
  Find new serving cell $j =$
  $\arg\max_{\forall c \in \mathbb{C}} \left(\left(\frac{1}{\eta_c}\right)^{\alpha} * \left(\acute{P}^c_{r,u_{dBm}}\right)^{(1-\alpha)}\right)$
  Find updated SINR γ_u^j and η_u^j
end for
for $c \in \mathbb{C}$ **do** Calculate cell load η_c
end for
if $\frac{1}{|\mathbb{C}|} \sum_{\mathbb{C}} \frac{1}{|\mathbb{U}_c|} \sum_{\mathbb{U}_c} 1\left(P^c_{r,u} \geqslant P^c_{th}\right) \geqslant \varpi$ **then**
  if $\eta_c \leqslant \eta^c_{th} \forall c \in \mathbb{C}$ **then**
    if $\hat{\mathbb{U}}_c = \mathbb{U}_c$ **then**
      Calculate $\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right)$
    end if
  end if
else
  $\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) = -\infty$
end if

$$\mathbb{U}_j := \left\{\forall u \in \mathbb{U} \mid j = \arg\max_{\forall c \in \mathbb{C}} \left(\left(\frac{1}{\eta_c}\right)^{\alpha} * \left(\acute{P}^c_{r,u_{dBm}}\right)^{(1-\alpha)}\right)\right\}$$

As compared to unconstrained problem or problem with inequality constraint, equality constraints can reduce the search space of optimization problem significantly. User association is expressed as an equality constraint such that for $u \in \mathbb{U}_c$

$$\begin{aligned} Z\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) := \sum_{i \in \mathbb{C}/c} 1 \left(\left(\frac{1}{\eta_c}\right)^{\alpha} * \left(\acute{P}^c_{r,u_{dBm}}\right)^{(1-\alpha)}\right. \\ \left. \geqslant \left(\frac{1}{\eta_c}\right)^{\alpha} * \left(\acute{P}^c_{r,u_{dBm}}\right)^{(1-\alpha)}\right) - |\mathbb{C}| + 1 = 0 \end{aligned} \tag{5.19}$$

The expression in (5.19), where 1(.) is the indicator function, means that for a user u to be associated with cell c, the association function of the user with that cell must be greater than all the other cells. Lagrangian of (5.18) can be given as:

$$\begin{aligned}
&\mathcal{L}\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}, \lambda^1, \boldsymbol{\lambda^2}, \boldsymbol{\lambda^3}, \boldsymbol{\lambda^4}, \boldsymbol{\lambda^5}, \boldsymbol{\lambda^6}, \boldsymbol{\lambda^7}\right) = \\
&\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) - \lambda^1 W\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \\
&- \sum_{c\in\mathbb{C}} \lambda_c^2 X\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) - \sum_{c\in\mathbb{C}} \lambda_c^3 Y\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \\
&- \sum_{u\in\mathbb{U}} \lambda_u^4 Z\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) - \sum_{c\in\mathbb{C}} \lambda_c^5\left(P_t^c - P_{t,min}^c\right) \\
&- \sum_{c\in\mathbb{C}} \lambda_c^6\left(\psi_{tilt}^c - 90\right) - \sum_{c\in\mathbb{C}} \lambda_c^7\left(P_{CIO}^c\right)
\end{aligned} \tag{5.20}$$

where $\boldsymbol{\lambda^x}$ represents the x-th vector of Lagrangian multipliers for the constraints in (5.18) and (5.19). Thus, the quadratic sub-problem to be solved at each iteration of SQP is given by (5.21),

$$\min_{\boldsymbol{y}}\left(\frac{1}{2}\right)\boldsymbol{y^T}\boldsymbol{\hat{H}}\left(\mathcal{L}\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}, \lambda^1, \boldsymbol{\lambda^2}, \boldsymbol{\lambda^3}, \boldsymbol{\lambda^4}, \boldsymbol{\lambda^5}, \boldsymbol{\lambda^6}, \boldsymbol{\lambda^7}\right)\right)\boldsymbol{y} + \nabla\Omega\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \tag{5.21}$$

$$\textbf{subject to} = \begin{cases}
y_i + W\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \leqslant 0, for\ i = 1 \\
y_i + X\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) \leqslant 0, for\ i = 2, \ldots, |\mathbb{C}| + 1 \\
y_i + Y\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) = 0, for\ i = |\mathbb{C}| + 2, \ldots, 2|\mathbb{C}| + 1 \\
y_i + Z\left(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}\right) = 0, for\ i = 2|\mathbb{C}| + 2, \ldots, 2|\mathbb{C}| + |\mathbb{U}| + 1 \\
y_i + P_t^c - P_{t,min}^c \leqslant 0, for\ i = 2|\mathbb{C}| + |\mathbb{U}| + 2, \ldots, 3|\mathbb{C}| + |\mathbb{U}| + 1 \\
y_i + \psi_{tilt}^c - 90^\circ \leqslant 0, for\ i = 3|\mathbb{C}| + |\mathbb{U}| + 2, \ldots, 4|\mathbb{C}| + |\mathbb{U}| + 1 \\
y_i + \psi_{tilt}^c \leqslant 0, for\ i = 4|\mathbb{C}| + |\mathbb{U}| + 2, \ldots, 5|\mathbb{C}| + |\mathbb{U}| + 1
\end{cases}$$

where $\boldsymbol{\hat{H}}$ represents the approximate Hermitian matrix, which is updated at each iteration using the Broyden-Fletcher-Goldfarb-Shanno approximation method [253].

Other Heuristic Techniques: Through results presented here, it was found that SQP returns an acceptable solution with low number of iterations in most instances at the cost of a lack of guarantee that the solution is optimal due to the large dimensions of the problem in (5.17). Furthermore, the enormous search space size of (5.17) makes validation of the results produced through brute force almost impossible. Therefore, a number of heuristic techniques were tried that are known to converge to optimal solutions given enough iterations. In the following, a discussion of two heuristics which yielded most promising results for this problem is presented.

Genetic Algorithms: Genetic algorithms are known to be one of the most suitable heuristic algorithms available for solving complex combinatorial problems of kind of (5.17). It is important to note that the genetic algorithm starts from a random parameter set in the solution space, therefore, for each run, the time to find the feasible space is different. However, once found, the algorithm can quickly move towards the optimal solution in the feasible space. Algorithm 4 represents the pseudo code for the genetic algorithm used to solve (5.17).

Pattern Search: Another effective solution methodology to solve (5.17) is Pattern Search Method, a simpler version of Powell's method [254]. Algorithm 5 presents a generic pseudo-code which describes the main elements of a pattern search method [255] where Nelder-Mead algorithm is used as the exploratory search algorithm within each iteration of pattern search [177].

5.3.4 System Level Performance Analysis

Simulation Setup

A 3GPP standard compliant network topology simulator [251] is used to generate typical macro and small cell based network and user distributions. The simulation parameters details are given in Table 5.1.

Wrap around model is used to simulate interference in an infinitely large network thus avoiding boundary effect. To model realistic networks, users are distributed non-uniformly in all the sectors such that a fraction of users are clustered around randomly located hotspots in each sector. Monte Carlo simulations are used to estimate average

Algorithm 4 Genetic Algorithm for CLASS Implementation

Input:

Algorithm 1 to solve (5.17)

Parameter set space $\mathbf{S}(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c})$,

Maximum iterations $\mathbf{G}$,

Solution space samples per iteration $\mathbf{P}$,

Key samples per iteration $\mathbf{E}$,

Mutation ratio $\mathbf{M}$.

Output:

Solution $\mathbf{X} = [\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}]$

Generate |P| parameter sets from $\mathbf{S}$ randomly;
Generate values of Ω for each set in $\mathbf{P}$
Create an empty set **Pop** and save the sets from $\mathbf{P}$ in it;
for i = 1 to G **do**
    Number of elite members in **Pop** $num_{elite} = \boldsymbol{E}$;
    Select the best num_{elite} sets in **Pop** in terms of the value of Ω and save them in $\mathbf{Pop_1}$;
    Number of crossover solutions $num_{crossover} = (|\mathrm{P}| * num_{elite})/2$;
    for j = 1 to $num_{crossover}$ **do**
        Randomly select 2 parameter sets $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ from **Pop**;
        Generate $\boldsymbol{X_3}$ and $\boldsymbol{X_4}$ by one-point crossover to $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$;
        Save $\boldsymbol{X_3}$ and $\boldsymbol{X_4}$ to $\mathbf{Pop_2}$;
    end for
    for j = 1 to $num_{crossover}$ **do**
        Select a parameter set $\boldsymbol{X_j}$ from $\mathbf{Pop_2}$;
        Mutate each element of $\boldsymbol{X_j}$ at a rate $\mathbf{M}$ and generate new solution $\acute{\boldsymbol{X}}_{\boldsymbol{j}}$;
        if $\acute{\boldsymbol{X}}_{\boldsymbol{j}}$ is non-feasible **then** Update $\acute{\boldsymbol{X}}_{\boldsymbol{j}}$ with a feasible solution by repairing $\acute{\boldsymbol{X}}_{\boldsymbol{j}}$;
        end if
        Update $\boldsymbol{X_j}$ with $\acute{\boldsymbol{X}}_{\boldsymbol{j}}$ in $\mathbf{Pop_2}$;
    end for
    Update **Pop** = $\mathbf{Pop_1}$+$\mathbf{Pop_2}$;
end for
Return the set $\mathbf{X}$ which has the best value of Ω in **Pop**;

Algorithm 5 Pattern Search Algorithm for CLASS Implementation

Input:

  Algorithm 1 to solve (5.17)

  Parameter space $\mathbf{S}(\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c})$

Output: Solution $\mathbf{X} = [\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}]$

$k = 0$;
while $k < iteration_{max}$ **do**
    Determine a step size s_k using exploratory search algorithm;
    Test Ω at parameter set $\boldsymbol{x_0}$ and two more points $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ in a triangle;
    Label best, good and worst points as $\boldsymbol{x_B}$, $\boldsymbol{x_G}$ and $\boldsymbol{x_W}$;
    Reflect $\boldsymbol{x_W}$ on the plane as $\boldsymbol{x_R}$;
    if $\Omega(\boldsymbol{x_R}) > \Omega(\boldsymbol{x_G})$ **then**
        if $\Omega(\boldsymbol{x_R}) > \Omega(\boldsymbol{x_B})$ **then** replace $\boldsymbol{x_W}$ with $\boldsymbol{x_R}$;
        else Find $\boldsymbol{x_E} | 2\boldsymbol{x_R} - (\boldsymbol{x_B} + \boldsymbol{x_G})/2$, find $\Omega(\boldsymbol{x_E})$
            if $\Omega(\boldsymbol{x_E}) > \Omega(\boldsymbol{x_B})$ **then** replace $\boldsymbol{x_W}$ with $\boldsymbol{x_E}$;
            end if
        end if
    else
        if $\Omega(\boldsymbol{x_R}) < \Omega(\boldsymbol{x_W})$ **then** replace $\boldsymbol{x_W}$ with $\boldsymbol{x_R}$;
            Compute $\boldsymbol{x_C} = ((\boldsymbol{x_B} + \boldsymbol{x_G})/2) + \boldsymbol{x_R})/2$, find $\Omega(\boldsymbol{x_C})$
        else Compute $\boldsymbol{x_C} = ((\boldsymbol{x_B} + \boldsymbol{x_G})/2) + \boldsymbol{x_W})/2$, find $\Omega(\boldsymbol{x_C})$
        end if
        if $\Omega(\boldsymbol{x_C}) < \Omega(\boldsymbol{x_W})$ **then** replace $\boldsymbol{x_W}$ with $\boldsymbol{x_C}$;
        else Compute $\boldsymbol{x_S} = (\boldsymbol{x_B} + \boldsymbol{x_W})/2$ and replace $\boldsymbol{x_W}$ with $\boldsymbol{x_S}$ and $\boldsymbol{x_G} = (\boldsymbol{x_B} + \boldsymbol{x_G})/2$
        end if
    end if
    Compute $p_k = \Omega(\boldsymbol{x_k}) - \Omega(\boldsymbol{x_k} + s_k)$
    if $p_k > 0$ **then** $\boldsymbol{x_{k+1}} = \boldsymbol{x_k} + s_k$
    else $\boldsymbol{x_{k+1}} = \boldsymbol{x_k}$
    end if
    Update pattern vectors and step size $k = k + 1$
end while
Return $\mathbf{X} = [\boldsymbol{P_t^c}, \boldsymbol{\psi_{tilt}^c}, \boldsymbol{P_{CIO}^c}]$

Table 5.1: Parameter Settings for Simulation of Joint CCO-LB SON Coordination

System Parameters	**Value**
No. of Macro Base Stations	7
Sectors per Base Station	3
Small Cells per Sector	1
Number of Users per Sector	25
Transmission Frequency	2 GHz
Transmission Bandwidth	10 MHz
Network Topology	Hexagonal
Macro Cell Transmit Power	Max: 46 dBm, Min: 40 dBm
Macro Cell Antenna Tilt	Max: 15°, Min: 0°
Small Cell Transmit Power	Max: 30 dBm, Min: 27 dBm
Small Cell CIO	Max: 10 dB, Min: 0 dB
Fixed Parameter Settings (FPSs)	Macro Transmit Power: 43 dBm; Small Transmit Power: 27 dBm; Tilt: 0° (FPS - 0), 10° (FPS - 10), 15° (FPS - 15), 20° (FPS - 20); CIO: 0 dB
Cellular System Standard	LTE
Macro Cell Height	25 m
Small Cell Height	10 m
Inter-site Distance (Macro)	500 m
Macro Cell Antenna Gain	17 dBi
Small Cell Antenna Gain	5 dBi
Coverage Threshold P^c_{th}	95%
Load Threshold η^c_{th}	100%

performance of the algorithms. Five different user traffic requirement profiles are considered corresponding to 24 kbps, 56 kbps, 128 kbps, 512 kbps and 1024 kbps desired throughput.

Results

Here the impact of different α values used in load-aware user association on CLASS is compared along with a comparison of load-aware user association with state-of-the-art-maximum RSRP and maximum SINR user association methods. Using the proposed load-aware user association with best performing α value, the results from 4 FPSs against the optimal parameter values returned by both CLASS equations using SQP, genetic algorithm and pattern search are compared to demonstrate their gain. For simplicity, the CLASS solution in equation (5.17a) is henceforth referred to as CLASS1 and solution in equation (5.17a) as CLASS2. The results of proposed solutions are further compared with the two algorithms that are most relevant to this solution that is, the distributed tilt based CCO solution presented in [249] and the tilt based CCO-LB function given in [176]. It is important to note here explicitly that due to the use of virtual loads in the system, the user association from [176] returns undefined results. Therefore, the algorithm in [176] is implemented using load-aware user association.

Impact of Load-aware User Association: The proposed load-aware user association (5.14) is dependent on three features: cell loads at the time of association, downlink received power with CIO and the association exponent α. The impact of cell loads and received powers on user association are obvious from (5.14); however, the impact of exponent value on user association requires quantitative evaluations of system KPIs for different values α. A very relevant KPI in this case is the cell load and its distribution among cells for given total traffic in the network. A lower average cell load and smaller load variance among cells for given traffic reflects a better performing user association scheme and vice versa. Though a comparison of $\alpha \in [0, 1]$ for both CLASS formulations was carried out, for brevity Fig. 5.6 only presents cell load distribution for $\alpha \in [\frac{1}{4}, \frac{1}{2}]$.

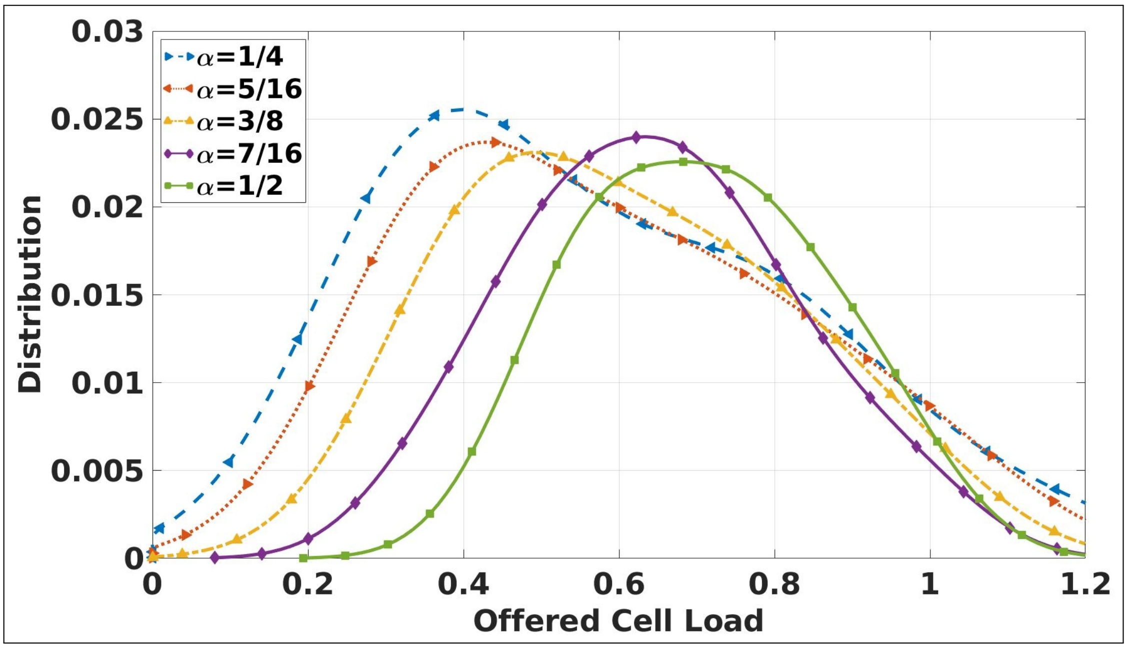

Fig. 5.6: Comparison of Offered Cell Load Distribution for α values in Load-aware User Association

From the results in Fig. 5.6 it can be seen that the load distribution improves and becomes the most compact at $\alpha = \frac{7}{16}$ and starts to spread beyond it. Using $\alpha = \frac{7}{16}$ Figs. 5.7 and 5.8 present a comparison of the proposed load-aware user association with coverage based Max RSRP user association and quality based Max SINR user association techniques for macro and small cells.

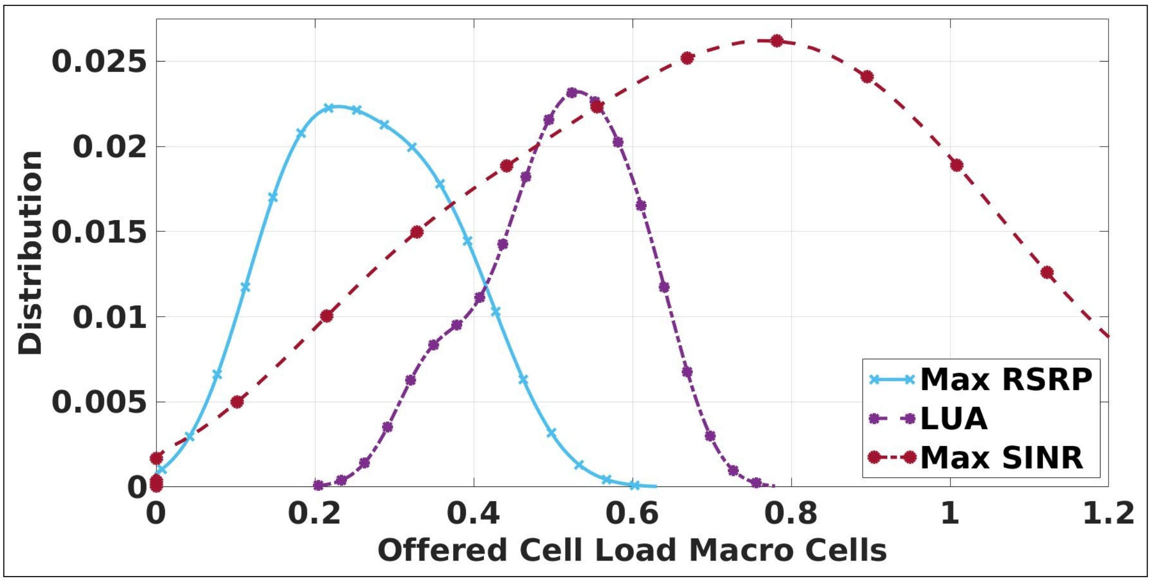

Fig. 5.7: Comparison of Offered Macro Cell Load Distribution for load-aware (LUA) vs. Max RSRP and Max SINR user association

The results in Fig. 5.7 show that the proposed load-aware user association manages to keep macro cell loads within 80%, Max RSRP keeps macro cell loads to within 60%, while Max SINR association overloads a number of macro cells due to their stronger signals. In comparison, Fig. 5.8 shows that the proposed load-aware user association technique attempts to distribute load evenly between macro and small cells, with only a few small cells marginally overloaded. On the other hand, due to a lack of load awareness, both Max RSRP and Max SINR association overload the small cells with more than half the small cells overloaded. The even load distribution offered by the load-aware user association methodology also results in fewer unsatisfied users that is, users who are unable to achieve their desired throughput due to a lack of physical resources at the serving cell.

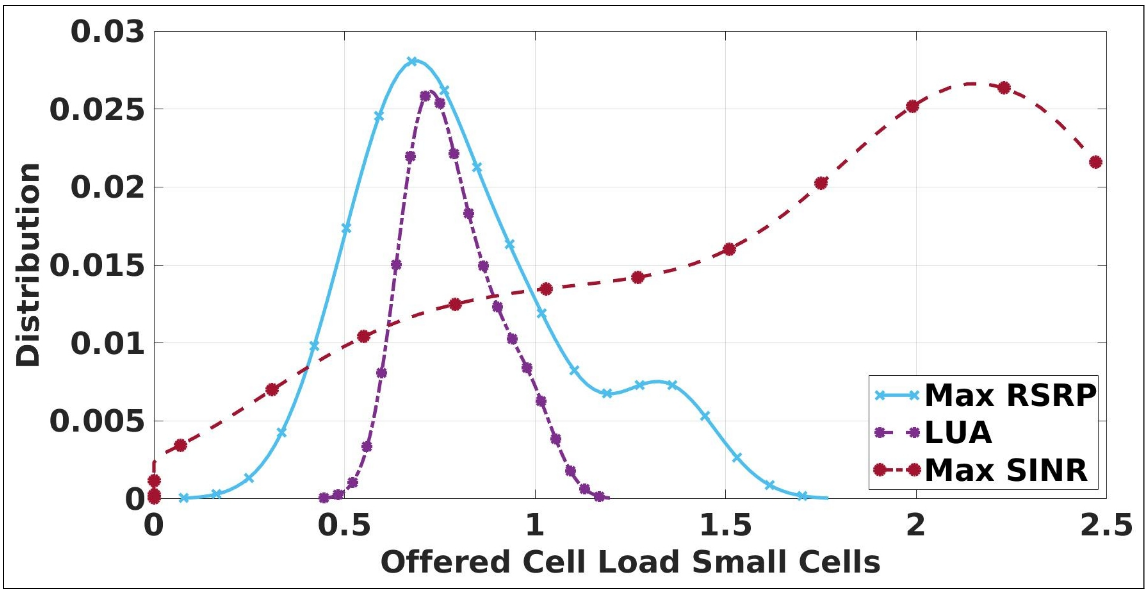

Fig. 5.8: Comparison of Offered Small Cell Load Distribution for load-aware (LUA) vs. Max RSRP and Max SINR user association

This is evidenced by the ratio of unsatisfied users in the network and the utilization of physical resources in the network given in Fig. 5.9. It can be seen that while the load-aware user association occupies more resources, it is able to minimize the ratio of unsatisfied users by evenly distributing the load between cells. On the other hand, the Max RSRP and Max SINR user association schemes are oblivious to the needs of the users and blindly associate them with cells offering best coverage and quality. This leads to cells becoming overloaded and higher ratio of unsatisfied users. The results in Figs. 5.7, 5.8, and 5.9, also demonstrate that the flexibility in the design of the proposed load-aware user association scheme allows it to be an effective coverage, capacity and load optimization solution, even when deployed independently in a cellular network.

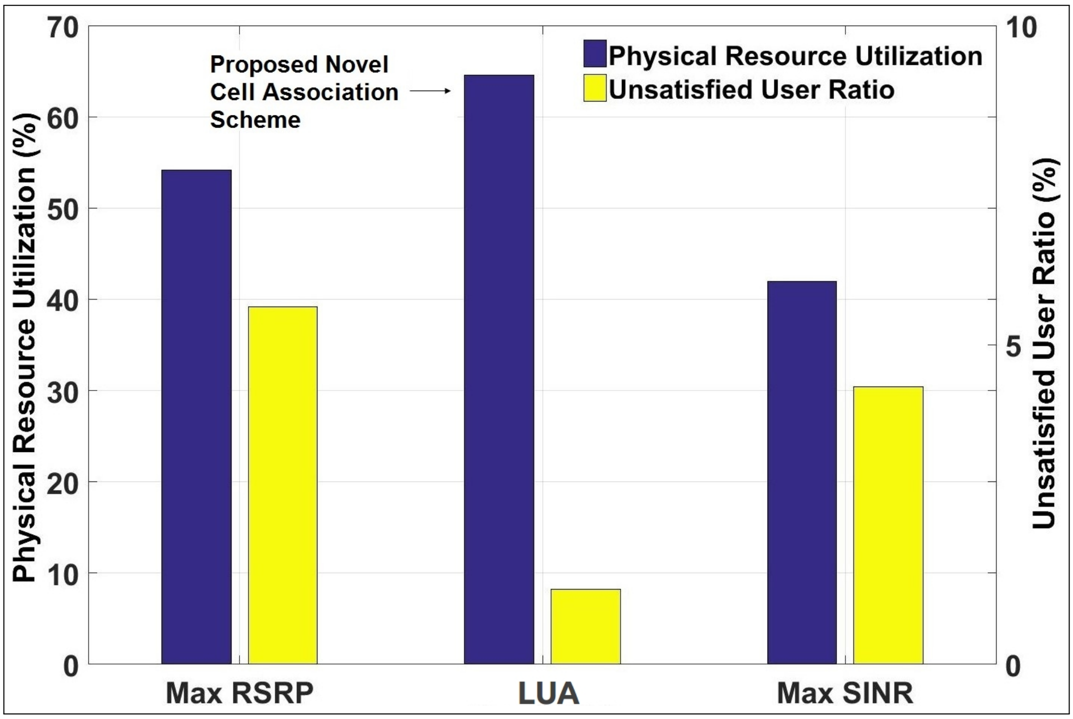

Fig. 5.9: Comparison of network utilization and unsatisfied user ratio for load-aware (LUA) vs. Max RSRP and Max SINR user association

Comparative Analysis of Proposed Solutions:

Downlink SINR: To compare the performance of the two CLASS formulations, downlink SINR is used as the benchmark performance indicator. In Fig. 5.10 the results for CLASS1 obtained using SQP, genetic algorithm and pattern search are compared against different fixed parameter settings defined in Table 5.1. The results show that 50th percentile users achieve 14dB SINR with CLASS1-PS compared to 10dB for top performing FPS-20. In Fig. 5.11, the same comparison is presented for CLASS2 which shows that 50th percentile users achieve 4.5dB higher SINR with CLASS2 compared to FPS-20. Recall that using CIOs alone for LB has negative impact on SINR. But when CIOs are adapted through the proposed load-aware user association in conjunction with transmit power and antenna tilts, a gain in SINR is still achieved. This rationalizes the need to include all three optimization parameters in the proposed CCO-LB solution, compared to existing studies which use one or two parameters at a time. Another key results to point out here is that the solutions obtained using genetic algorithm and pattern search perform better for both CLASS1 and CLASS2 compared to SQP. This is due to the fact that the genetic algorithm and pattern search attempt to find the global optimum whereas SQP is a gradient driven process that is vulnerable to convergence to local extrema.

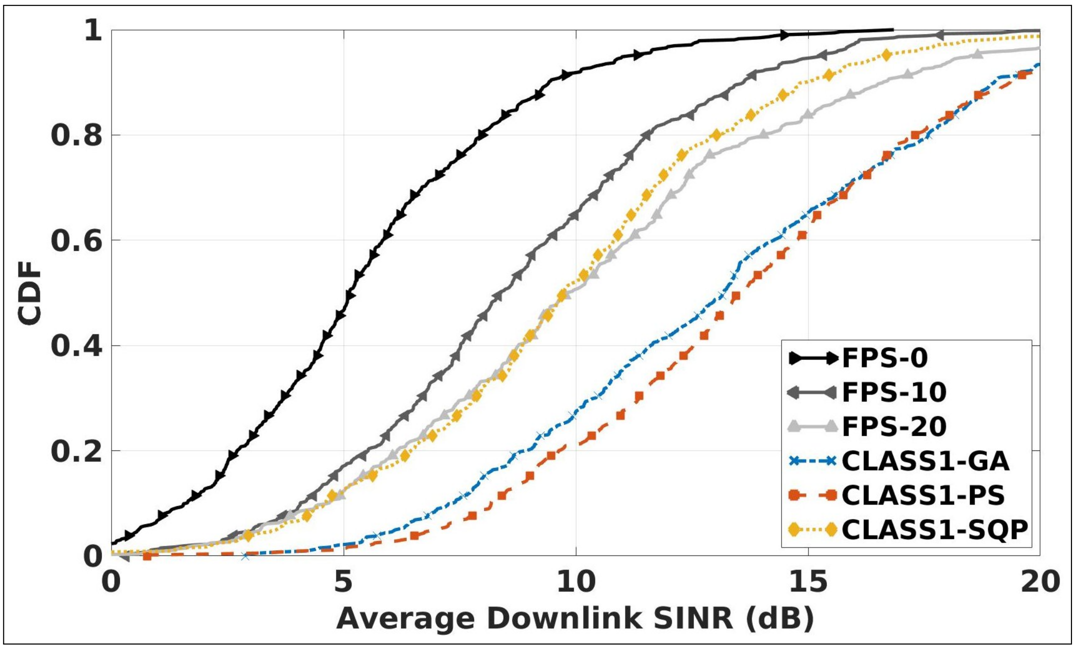

Fig. 5.10: Downlink SINR CDF - FPSs vs. CLASS1-genetic algorithm (GA), pattern search (PS) and SQP

Fig. 5.12 compares the best solution obtained for CLASS1 (pattern search) and CLASS2 (genetic algorithm) against the CCO algorithm proposed in [249] referred to by the authors as SOT, and the CCO-LB algorithm JOINT1 presented in [176]. Results show that CLASS1 and CLASS2 offer SINR > 10dB for almost 80% of users. In comparison, with SOT and JOINT1 only 20% and 30% of users have SINR above 10 dB respectively. It can also be seen that CLASS1 performs slightly better compared to CLASS2 for cell edge users that is, the lower half of users with CLASS2 giving slightly better performance for the top half. This is because of the use of geometric mean in CLASS1 which forces fairness in all user throughputs, whereas the use of arithmetic mean attempts to maximize the extreme throughput values.

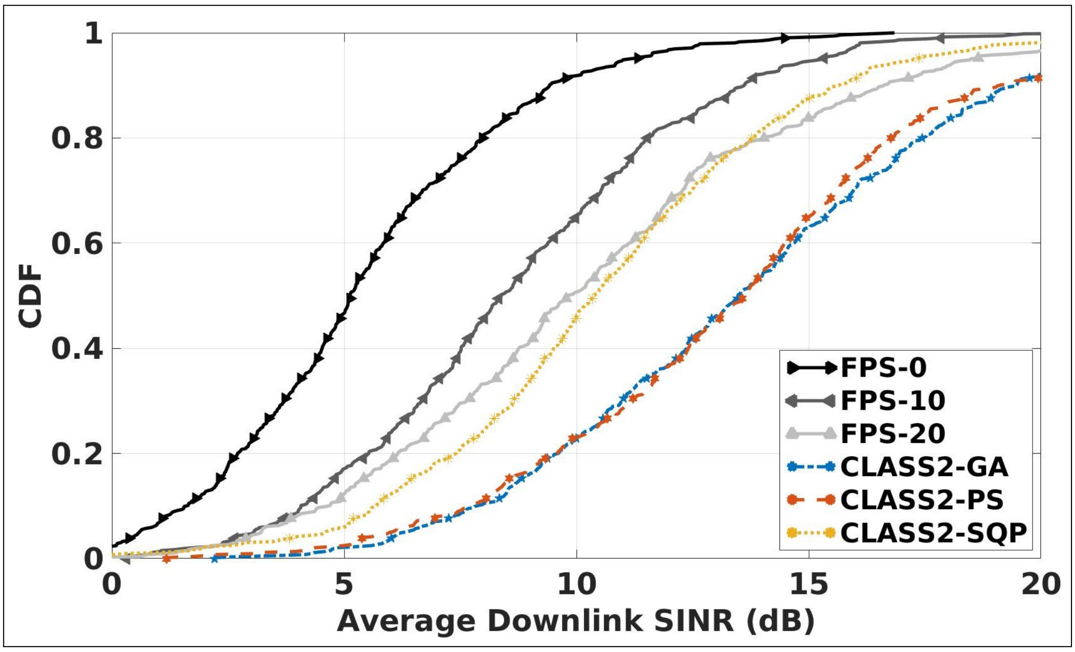

Fig. 5.11: Downlink SINR CDF - FPSs vs. CLASS2-genetic algorithm (GA), pattern search (PS) and SQP

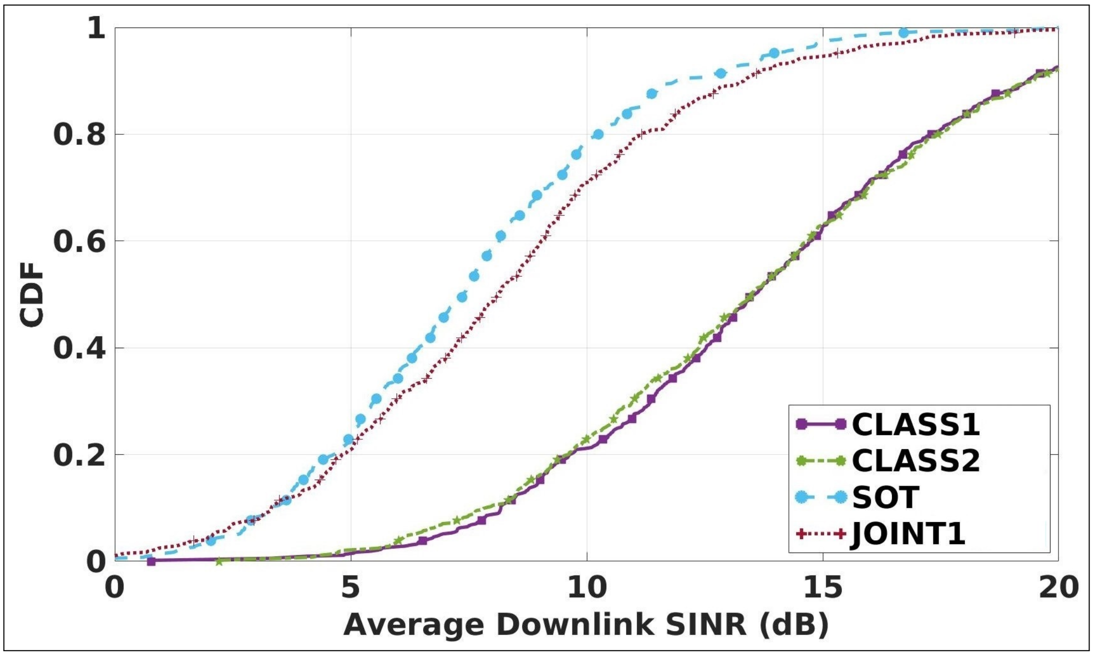

Fig. 5.12: Downlink SINR CDF - SOT, JOINT1 vs. CLASS1 and CLASS2

Offered Cell Load: Fig. 5.13 compares offered cell loads for CLASS1, CLASS2, SOT and JOINT1. The results show that for CLASS1, the cell loads range from 10% to 80%, and from 10% to 70% for CLASS2. This difference is due to the higher focus of CLASS1 on fairness which means it attempts to increase throughput of low SINR users by allocating them more resources compared to CLASS2 which only focuses on maximizing total throughput. By comparison, SOT shows the widest disparity among cell loads. This is primarily due to the fact that SOT is a CCO-only algorithm that only optimizes antenna tilts, thus highlighting the importance of formulating LB and CCO jointly with all three parameters. JOINT1 being a CCO-LB solution that incorporates two parameters that is, antenna tilts and CIOs, offers better load balancing compared to SOT, but is still significantly outperformed by both CLASS1 and CLASS2.

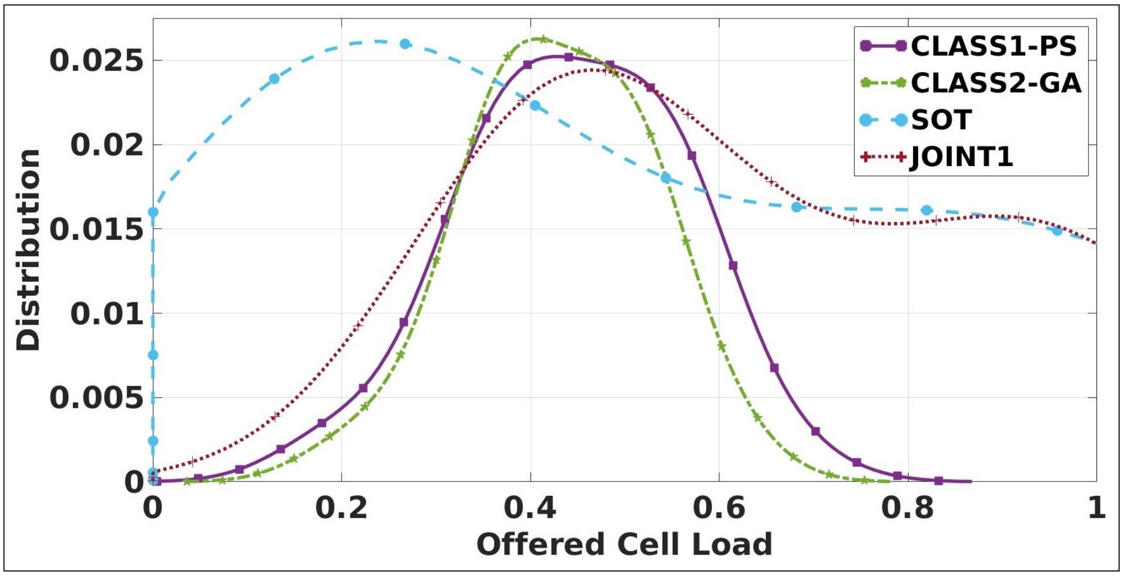

Fig. 5.13: Offered Cell Load Distribution - SOT, JOINT1 vs. CLASS1 and CLASS2

Figs. 5.14 and 5.15 show the performance of the proposed CCO-LB solution in terms of LB and QoS by showing load distributions for macro and small cells separately. While none of the macro or small cells are overloaded by the CLASS solutions, SOT heavily favors macro cells over small cells for loading causing almost 50% of the macro cells to become overloaded. Similarly, since JOINT1 only optimizes CIOs and antenna tilts, it also favors macro cells for load bearing over small cells. Another key insight here is that contrary to existing load balancing schemes [256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266], the proposed solution not only balances loads between macro and small cells but actually increases capacity in the system by jointly optimizing soft and hard parameters, thereby satisfying CCO objective at the same time.

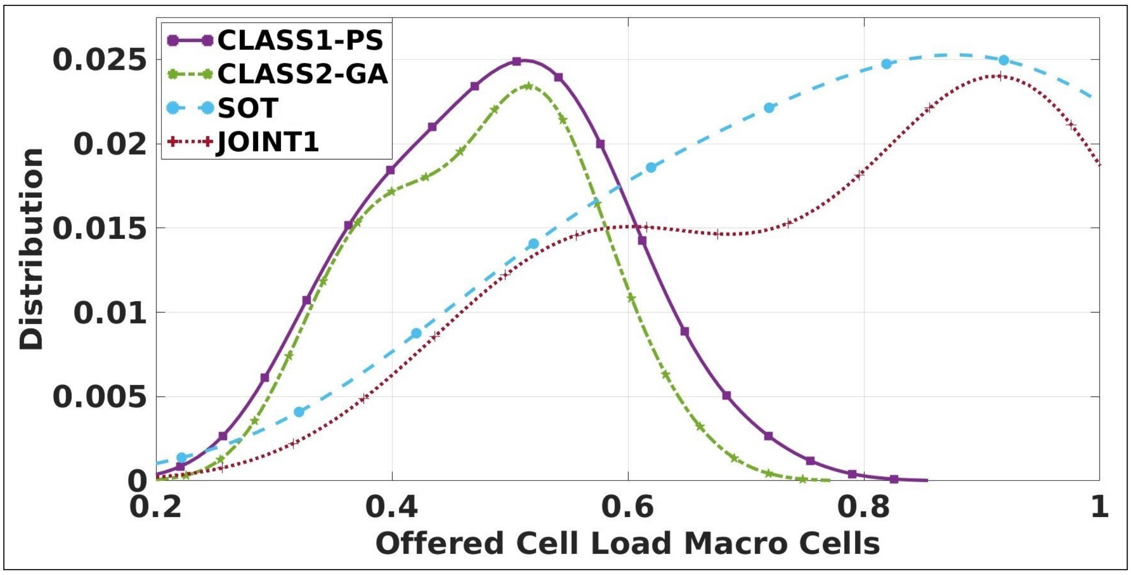

Fig. 5.14: Offered Macro Cell Load Distribution - SOT, JOINT1 vs. CLASS1 and CLASS2

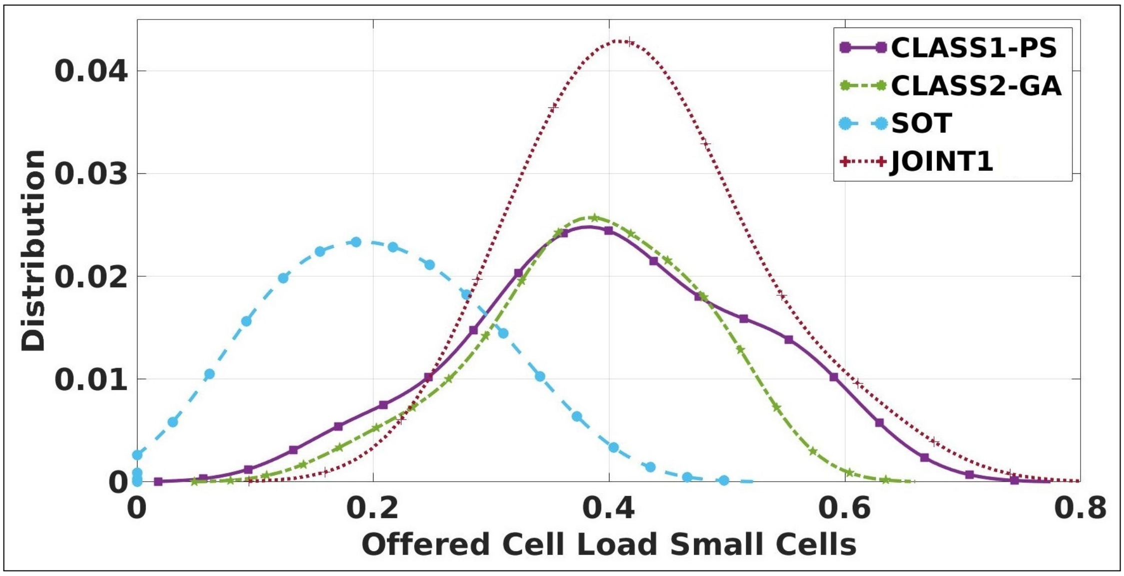

Fig. 5.15: Offered Small Cell Load Distribution - SOT, JOINT1 vs. CLASS1 and CLASS2

This is further put into perspective when the residual cell capacity across the network is observed, as shown in Fig. 5.16. The box plots show the median residual capacity value along with the distance between 1st and 3rd quartiles, whereas the points inside the box plots signify the mean residual capacity. The average residual cell capacity of the proposed CLASS1 and CLASS2 solutions are 54.8% and 55.5% respectively, which is 20% more than the average residual capacity of the algorithm in [249], and over 45% more than the residual capacity of the algorithm in [176]. However, the key observation in Fig. 5.16 is compactness of the 1st and 3rd quartile, and the outer fences for CLASS solutions compared to the residual capacities of other solutions. The increased residual capacity creates additional space for transit users within each cell, a feature that is highly desirable in ultra-dense HetNets due to the expected high user mobility.

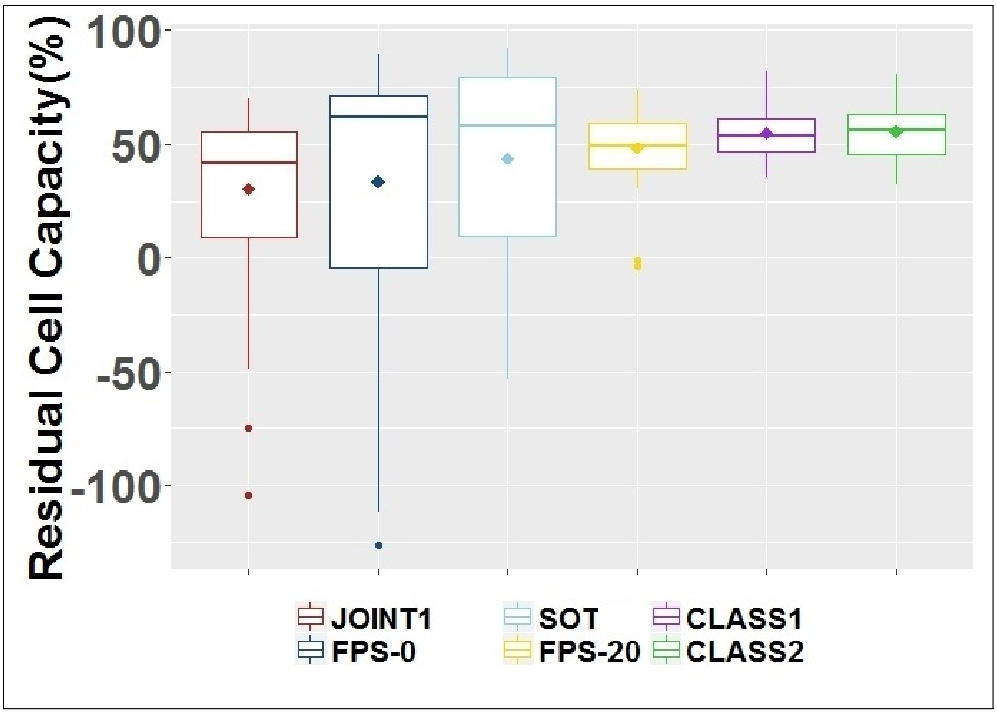

Fig. 5.16: Residual Cell Capacity - FPS-0, FPS-20, SOT, JOINT1 vs. CLASS1 and CLASS2

5.4 Conclusion

This chapter presents the case for a comprehensive SON coordination overhaul that can cater to the proposed innovations in future 5th generation and beyond mobile cellular networks including ultra-dense mmWave cell deployment and multi-technology integration. An example case study of SON function interactions is presented to highlight the chaotic consequences of uncoordinated SON function conflicts and to identify the characteristics needed in next generation SON coordination frameworks. Building on these insights, a detailed SON coordination framework has been presented that: 1) is flexible to short and long-term dynamicity that is hallmark of cellular networks; 2) implements conflict avoidance and KPI optimization in line with network operator policies; 3) can operate without analytical models of network and user behavior; and 4) renders a holistic and transparent view of all KPIs while implementing conflict avoidance.

To demonstrate the utility of the proposed framework, the chapter also includes a comprehensive solution for joint CCO and LB SON functions with transmit powers, antenna tilts and CIOs as the optimization parameters. The proposed CCO-LB solution leverages the network state prediction capability of the SON coordination framework and not only provides significant gains in terms of downlink SINR and throughput, it also provides balanced distribution of cell loads in a heterogeneous network which is key to meeting overall resource efficiency demands. The proposed solution also shows how the amount of free resources in the network after all users are satisfied, that is the residual capacity is the key to achieving temporal stability in the network optimization process due to the acute mobility dynamics of HetNets.

CHAPTER 6

Conclusion and Future Research Directions

6.1 Conclusion

Self-healing will play a frontline role in managing the costs of deploying and maintaining mobile cellular networks of the future. However, the path towards satisfying the quality of experience requirements of these future networks is fraught with challenges. These challenges stem primarily from the technologies to be employed in these networks such as network densification, millimeter spectrum utilization and massive multiple input multiple output solutions among others.

The challenges which will result from these technologies include: 1) unpredictable noise variance in the network, 2) spatio-temporal variability of network data, 3) sparsity of data, and 4) increased probability of parametric miscofiguration. An in-depth review of the state-of-the-art solutions for self-healing shows that these solutions are not yet capable of adapting to these challenges. Therefore, self-healing solutions for future mobile cellular networks will be required to address these challenges in order for them to be effective. This dissertation attempts to address these challenges for future mobile cellular networks.

To address the challenge of detecting spatio-temporal outages in variable noise environment, this dissertation presents an entropy field decomposition based outage detection solution that is robust to noise distribution. The proposed solution is demonstrably better at detecting both coverage holes and outages compared to the state-of-the-art solutions.

The entropy field decomposition based outage detection solution relies on complete network coverage information. However, in the very likely event that only sparse network coverage information is available, this dissertation presents another solution that utilizes the hidden feature extraction capabilities of deep neural networks to detect network outages. This alternative solution performs much better than other state-of-the-art solutions in highly sparse network information environments even when noise distribution is variable.

Additionally, in order to address the challenge of proactively avoiding partial outages due to parametric misconfiguration, this dissertation presents a comprehensive framework for resolving conflicts between self-organizing network functions. The proposed framework combines heuristic, analytical and machine-learning based outage avoidance methodologies under one all-encompassing umbrella.

The proposed partial outage avoidance and SON coordination framework is leveraged to develop an analytical solution for conflict avoidance between coverage and capacity optimization and load balancing functions with results showing that such a solution can perform better than state-of-the-art solutions if the user requirements are known.

6.2 Future Research Direction

The future research directions opened up by this dissertation include the development of outage diagnosis and compensation solutions for millimeter wave ultra-dense heterogeneous networks that will form the backbone of future mobile cellular networks. Another potential area of research opened up by this study is the use of the proposed outage avoidance framework proposed in this dissertation to develop coordination solutions that will address the idiosyncrasies of millimeter wave cell communication networks including the limited directional coverage, low cell sojourn time per user and higher data rate requirements.

List of Key Symbols

$P^c_{r,u}$	Received power by user u from cell c
$P^t_{r_{out}}h$	Received power threshold for outage
$P^t_{r_{ch}}h$	Received power threshold for coverage hole
P^c_t	Transmit power of cell c
G_u	User equipment antenna gain
G^c_u	Channel gain
b	Pathloss constant
ϵ^c_u	Shadowing in dBm
d^c_u	Distance of user u from cell c
$beta$	Pathloss exponent
$\hat{\boldsymbol{P}}_{\boldsymbol{t}}$	Subset of cell transmit powers
$\boldsymbol{P_t}$	Set of cell transmit powers
$\boldsymbol{P_r}$	Set of downlink received powers
$\boldsymbol{P_{r_{out}}}$	Set of downlink received powers with outage
ζ_l	Combined spatio-temporal coordinate
$\mathcal{H}$	Hamiltonian function of an interacting field
$\mathcal{Z}$	Partition function
$\boldsymbol{\vartheta}$	Field transformation of input signal
$\mathcal{D}$	Information propagator
$\boldsymbol{j}$	Information source
$\mathcal{H}_o$	Hamiltonian function of non-interacting field
V	Signal interaction terms
$\tilde{V}$	Field interaction terms
$\boldsymbol{\sigma^2_{P_t}}$	Transmit power covariance

$\boldsymbol{\mathcal{N}}$	Noise covariance
$\boldsymbol{\mathcal{A}}$	Adjacency matrix
$\boldsymbol{\mathcal{Q}}$	Information field interaction matrix
$\boldsymbol{a_k}$	Eigenmode amplitudes for information field reconstruction
$\boldsymbol{j_k}$	Eigenmode amplitudes for information source
φ	Eigenmode of information field
λ	Eingenvector of information field
$\boldsymbol{P_{r_{norm}}}$	Set of downlink received powers without outage
$\boldsymbol{P_{r_{RT}}}$	Set of real-time downlink received powers
$\mathcal{O}$	Computational complexity notation
$\boldsymbol{A}$	Weight matrix for independent component analysis
$\mu_i^{(t)}$	i-th mean at time t for k-means clustering
$\boldsymbol{S_i^{(t)}}$	i-th set of points at time t for k-means clustering
$\hat{y}_j^{(l)}$	Output of j-th neuron of layer l
$\boldsymbol{w_j^{(l)}}$	Weights of j-th neuron of layer l
$\boldsymbol{x_j^{(l)}}$	Inputs of j-th neuron of layer l
$b_j^{(l)}$	Bias of j-th neuron of layer l
E	Loss of deep neural network
$P^c_{r,u_{t-\Delta t}}$	Received power by user u from cell c before an outage
P^c_{r,u_t}	Received power by user u from cell c after an outage
$d^c_{u_{t-\Delta t}}$	Distance of user u from cell c before an outage
$d^c_{u_t}$	Distance of user u from cell c after an outage
$\mathcal{C}$	Set of all cells in the network
$\hat{\mathcal{C}}$	Set of outage affected cells in the network
$\mathcal{U}$	Set of all users in the network
ϱ	Functional margin of SVM algorithm
$\mathfrak{A}$	Difference in the number of user between cell A and B
$\mathfrak{B}$	Difference in the cell A $\Leftrightarrow$ cell B handovers

$\mathfrak{C}$	Ratio of users switching between cell A and cell B due to handovers
$\mathfrak{a}$	Load offset between cell A and Cell B
$\mathfrak{b}$	Handover offset between cell A and Cell B
$\mathfrak{c}$	Intercell distance between cell A and cell B handovers
η_u^c	Physical resource blocks allocated to user u at cell c
ω_B	Bandwidth per physical resource block
$\hat{\tau}_u$	Desired user throughput
$\mathbb{U}$	Set of all active users
γ_u^c	SINR of user u at cell c
$\hat{\mathbb{U}}$	Set of all active satisfied users
N_b^c	Total physical resource blocks at cell c
κ	Thermal noise
ϖ	Network coverage threshold
P_{th}^c	Downlink Rx power threshold
η_{th}^c	Cell load threshold
α	User association exponent
μ	Antenna gain constant
Ω	Objective value for CLASS solutions
ψ_u^c	Vertical angle between user u and cell c
ψ_{tilt}^c	Antenna tilt of cell c
ϕ_u^c	Horizontal angle between user u and cell c
ψ_{azi}^c	Azimuth of cell c
$P_{CIO_{dB}}$	Cell individual offset of cell c

Bibliography

[1] Capgemini, "Quest for Margins: Operational Cost Strategies for Mobile Operators in Europe," *Telecom Media and Insights*, no. 42, 2009.

[2] A. Networks, "Top Ten Pain Points of Operating Networks," Report, 2011.

[3] P. Donegan, "Mobile Network Outages and Service Degradations," Heavy Reading, Report, 2016.

[4] 3GPP, "TR 36.902 - V9.3.1 Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions," 2011.

[5] ——, "TS 32.541 - V10.0.0 Telecommunication management; Self-Organizing Networks (SON); Self-healing concepts and requirements," 2011.

[6] ——, "TS 32.522 - V11.7.0 Telecommunication management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP) Information Service (IS)," 2013.

[7] M. Marwangi, N. Fisal, S. Yusof, R. A. Rashid, A. S. Ghafar, F. A. Saparudin, and N. Katiran, "Challenges and practical implementation of self-organizing networks in LTE/LTE-Advanced systems," in *Proc. International Conference on Information Technology and Multimedia (ICIM)*, 2011, pp. 1–5.

[8] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, 2014.

[9] 3GPP, "TS 32.111-1 - V13.0.0 Telecommunication management; Fault Management; Part 1: 3G fault management requirements," 2016.

[10] P. Stuckmann, Z. Altman, H. Dubreil, A. Ortega, R. Barco, M. Toril, M. Fernandez, M. Barry, S. McGrath, G. Blyth, P. Saidha, and L. M. Nielsen, "The EUREKA Gandalf project: monitoring and self-tuning techniques for heterogeneous radio access networks," in *Proc. IEEE 61st Vehicular Technology Conference (VTC)*, 2005, pp. 2570–2574.

[11] L. Schmelz, J. Van Den Berg, R. Litjens, K. Zetterberg, M. Amirijoo, K. Spaey, I. Balan, N. Scully, and S. Stefanski, "Self-organisation in wireless networks use cases and their interrelation," in *Proc. Wireless World Res. Forum Meeting*, vol. 22, 2009, pp. 1–5.

[12] A. Imran, "Quality of Service Aware Energy Efficient Self Organizing Future Cellular Networks." [Online]. Available: http://qson.org/

[13] R. Litjens, F. Gunnarsson, B. Sayrac, K. Spaey, C. Willcock, A. Eisenblatter, B. G. Rodriguez, and T. Kurner, "Self-management for unified heterogeneous radio access networks," in *Proc. IEEE 77th Vehicular Technology Conference (VTC)*, 2013, pp. 1–5.

[14] CircleID, "Misconfiguration brings down entire .se domain in Sweden. [Online] www.circleid.com/posts/misconfiguration_brings_down _entire_se_domain_in_sweden."

[15] R. Johnson, "More details on today's outage. [Online] http://www.facebook.com/notes/facebook-engineering/more-details-on-todays-outage/ 431441338919."

[16] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, 2014.

[17] S.-I. Yang, D. M. Frangopol, and L. C. Neves, "Service life prediction of structural systems using lifetime functions with emphasis on bridges," *Reliability Engineering and System Safety*, vol. 86, no. 1, pp. 39–51, 2004.

[18] D. Turner, K. Levchenko, J. C. Mogul, S. Savage, and A. C. Snoeren, "On failure in managed enterprise networks," *HP Labs HPL-2012-101*, 2012.

[19] Z. Yin, X. Ma, J. Zheng, Y. Zhou, L. N. Bairavasundaram, and S. Pasupathy, "An Empirical Study on Configuration Errors in Commercial and Open Source Systems," in *Proc. 23rd ACM Symposium on Operating Systems Principles*, 2011, pp. 159–172.

[20] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A Survey of Self Organisa-

tion in Future Cellular Networks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 336–361, 2013.

[21] M. Peng, D. Liang, Y. Wei, J. Li, and H. H. Chen, "Self-configuration and self-optimization in LTE-advanced heterogeneous networks," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 36–45, May 2013.

[22] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury, "CRAHNs: Cognitive radio ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 810 – 836, 2009.

[23] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127 – 2159, 2006.

[24] I. F. Akyildiz, W. y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, April 2008.

[25] F. Akhtar, M. H. Rehmani, and M. Reisslein, "White space: Definitional perspectives and their role in exploiting spectrum opportunities," *Telecommunications Policy*, vol. 40, no. 4, pp. 319 – 331, 2016.

[26] Z. Zhang, K. Long, and J. Wang, "Self-organization paradigms and optimization approaches for cognitive radio technologies: A survey," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 36–42, 2013.

[27] D. Ghosh, R. Sharman, H. Raghav Rao, and S. Upadhyaya, "Self-healing systems - survey and synthesis," *Decision Support Systems*, vol. 42, no. 4, pp. 2164–2185, 2007.

[28] H. Psaier and S. Dustdar, "A survey on self-healing systems: approaches and systems," *Computing*, vol. 91, no. 1, pp. 43–73, 2011.

[29] L. Paradis and Q. Han, "A Survey of Fault Management in Wireless Sensor Networks," *Journal of Network and Systems Management*, vol. 15, no. 2, pp. 171–190, 2007.

[30] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[31] R. M. Lewis, V. Torczon, and M. W. Trosset, "Direct search methods: then and now," *Journal of Computational and Applied Mathematics*, vol. 124, no. 1, pp. 191–207, 2000.

[32] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[33] P. J. Van Laarhoven and E. H. Aarts, *Simulated annealing.* Springer, 1987, pp. 7–15.

[34] J. F. Nash Jr., "The bargaining problem," *Econometrica: Journal of the Econometric Society*, pp. 155–162, 1950.

[35] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach.* Springer Science and Business Media, 2013.

[36] K. P. Murphy, *Machine learning: A probabilistic perspective.* MIT Press Cambridge, 2012.

[37] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT Press Cambridge, 1998, vol. 1.

[38] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[39] 3GPP, "TS 32.450 - V13.0.0 Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Definitions," 2016.

[40] M. Amirijoo, L. Jorguseski, T. Kurner, R. Litjens, M. Neuland, L. Schmelz, and U. Turke, "Cell outage management in LTE networks," in *Proc. 6th International Symposium on Wireless Communication Systems, 2009. ISWCS*, 2009, pp. 600–604.

[41] Q. Liao, M. Wiczanowski, and S. Stańczak, "Toward cell outage detection with composite hypothesis testing," in *Proc. IEEE International Conference on Communications (ICC)*, 2012, pp. 4883–4887.

[42] C. M. Mueller, M. Kaschub, C. Blankenhorn, and S. Wanke, "A cell outage detection algorithm using neighbor cell list reports," in *Proc. International Workshop on Self-Organizing Systems*, 2008, pp. 218–229.

[43] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees.* CRC press, 1984.

[44] D. Ro and H. Pe, *Pattern classification and scene analysis.* John Wiley., 1973.

[45] M. Alias, N. Saxena, and A. Roy, "Efficient Cell Outage Detection in 5G HetNets Using Hidden Markov Model," *IEEE Communications Letters*, vol. 20, no. 3, pp. 562–565, 2016.

[46] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[47] P. Szilágyi and S. Novaczki, "An automatic detection and diagnosis framework for mobile communication systems," *IEEE Transactions on Network and Service Management*, vol. 9, no. 2, pp. 184–197, 2012.

[48] S. Chernov, M. Cochez, and T. Ristaniemi, "Anomaly detection algorithms for the sleeping cell detection in LTE networks," in *Proc. IEEE 81st Vehicular Technology Conference (VTC)*, 2015, pp. 1–5.

[49] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. 100, no. 7, pp. 750–753, 1975.

[50] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1, pp. 1–6, 1998.

[51] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proc. 20th Annual Symposium on Computational Geometry*, 2004, pp. 253–262.

[52] 3GPP, "TS 37.320 - V10.0.0 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," 2010.

[53] J. Zhang, X. Tuo, Z. Yuan, W. Liao, and H. Chen, "Analysis of fMRI data using an integrated principal component analysis and supervised affinity propagation clustering approach," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 11, pp. 3184–3196, 2011.

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

[55] P. K. Velamuru, R. A. Renaut, H. Guo, and K. Chen, "Robust clustering of positron emission tomography data," *Joint Interfce CSNA*, 2005.

[56] F. Chernogorov, J. Turkka, T. Ristaniemi, and A. Averbuch, "Detection of sleeping cells in LTE networks using diffusion maps," in *Proc. IEEE 73rd Vehicular Technology Conference (VTC)*, 2011, pp. 1–5.

[57] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.

[58] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[59] F. Chernogorov, T. Ristaniemi, K. Brigatti, and S. Chernov, "N-gram analysis for sleeping cell detection in LTE networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 4439–4443.

[60] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[61] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1641–1650, 2003.

[62] A. Zoha, A. Imran, A. Abu-Dayya, and A. Saeed, "A Machine Learning Framework for Detection of Sleeping Cells in LTE Network," in *Proc. Machine Learning and Data Analysis Symposium*, 2014.

[63] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A SON solution for sleeping cell detection using low-dimensional embedding of MDT mea-

surements," in *Proc. IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2014, pp. 1626–1630.

[64] J. B. Kruskal and M. Wish, *Multidimensional scaling.* Sage, 1978, vol. 11.

[65] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *Proc. ACM Sigmod Record.* ACM, 2000, pp. 93–104.

[66] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A learning-based approach for autonomous outage detection and coverage optimization," *Transactions on Emerging Telecommunications Technologies*, 2015.

[67] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Report, 1998.

[68] S. Chernov, M. Pechenizkiy, and T. Ristaniemi, "The influence of dataset size on the performance of cell outage detection approach in LTE-A networks," in *Proc. 10th International Conference on Information, Communications and Signal Processing (ICICS)*, 2015, pp. 1–5.

[69] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 446–465, 2016.

[70] O. Onireti, A. Imran, M. A. Imran, and R. Tafazolli, "Cell outage detection in heterogeneous networks with separated control and data plane," in *Proc. 20th European Wireless Conference*, 2014, pp. 1–6.

[71] O. Onireti, A. Zoha, J. Moysen, A. Imran, L. Giupponi, M. A. Imran, and A. Abu-Dayya, "A cell outage management framework for dense heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2097–2113, 2016.

[72] D. Julong, "Introduction to grey system theory," *The Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.

[73] W. Wang, J. Zhang, and Q. Zhang, "Cooperative cell outage detection in self-organizing femtocell networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, 2013, pp. 782–790.

[74] W. Xue, H. Zhang, Y. Li, D. Liang, and M. Peng, "Cell outage detection and compensation in two-tier heterogeneous networks," *International Journal of Antennas and Propagation*, vol. 2014, 2014.

[75] I. A. Karatepe and E. Zeydan, "Anomaly detection in cellular network data using big data analytics," in *Proc. 20th European Wireless Conference*, 2014, pp. 1–5.

[76] Apache, "http://hadoop.apache.org/."

[77] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "Characterizing and optimizing cellular network performance during crowded events," *Biological Cybernetics*, vol. 24, no. 3, pp. 1308–1321, 2016.

[78] S. Novaczki and P. Szilagyi, "Radio channel degradation detection and diagnosis based on statistical analysis," in *Proc. IEEE 73rd Vehicular Technology Conference (VTC)*, 2011, pp. 1–2.

[79] A. D'Alconzo, A. Coluccia, F. Ricciato, and P. Romirer-Maierhofer, "A distribution-based approach to anomaly detection and application to 3G mobile traffic," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–8.

[80] J. M. Joyce, *Kullback-leibler divergence.* Springer, 2011, pp. 720–722.

[81] M. Z. Asghar, R. Fehlmann, and T. Ristaniemi, "Correlation-based cell degradation detection for operational fault detection in cellular wireless base-stations," in *Proc. International Conference on Mobile Networks and Management*, 2013, pp. 83–93.

[82] P. Muñoz, R. Barco, I. Serrano, and A. Gómez-Andrades, "Correlation-Based Time-Series Analysis for Cell Degradation Detection in SON," *IEEE Communications Letters*, 2016.

[83] J. Sanchez-Gonzalez, O. Sallent, J. Pérez-Romero, R. Agusti, M. Díaz-Guerra, J. A. Moreno, and D. Paul, "A new methodology for RF failure detection in UMTS networks," in *Proc. IEEE Network Operations and Management Symposium*, 2008, pp. 718–721.

[84] P. Kumpulainen, M. Särkioja, M. Kylväjä, and K. Hätönen, "Analysing 3G radio

network performance with fuzzy methods," *Neurocomputing*, vol. 107, pp. 49–58, 2013.

[85] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[86] R. Sattiraju, P. Chakraborty, and H. D. Schotten, "Reliability analysis of a wireless transmission as a repairable system," in *Proc. Globecom Workshops (GC Wkshps)*, 2014, pp. 1397–1401.

[87] G. Ciocarlie, U. Lindqvist, K. Nitz, S. Nováczki, and H. Sanneck, "On the feasibility of deploying cell anomaly detection in operational cellular networks," in *Proc. IEEE Network Operations and Management Symposium (NOMS)*, 2014, pp. 1–6.

[88] G. F. Ciocarlie, U. Lindqvist, S. Nováczki, and H. Sanneck, "Detecting anomalies in cellular networks using an ensemble method," in *Proc. 9th international conference on Network and service management (CNSM)*, 2013, pp. 171–174.

[89] F. J. Massey Jr., "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[90] R. B. Cleveland, W. S. Cleveland, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, p. 3, 1990.

[91] G. A. Barreto, J. C. M. Mota, L. G. M. Souza, R. A. Frota, and L. Aguayo, "Condition monitoring of 3G cellular networks through competitive neural models," *IEEE Transactions on Neural Networks*, vol. 16, no. 5, pp. 1064–1075, 2005.

[92] C.-C. Hung, "Competitive learning networks for unsupervised training," *International Journal of Remote Sensing*, vol. 14, no. 12, pp. 2411–2415, 1993.

[93] R. A. Frota, G. A. Barreto, and J. Mota, "Anomaly detection in mobile communication networks using the self-organizing map," *Journal of Intelligent and Fuzzy Systems*, vol. 18, no. 5, pp. 493–500, 2007.

[94] M. Tanaka, M. Sakawa, I. Shiromaru, and T. Matsumoto, "Application of Kohonen's self-organizing network to the diagnosis system for rotating machin-

ery," in *Proc. IEEE International Conference on Systems, Man and Cybernetics/Intelligent Systems for the 21st Century*, 1995, pp. 4039–4044.

[95] P. Lehtimäki and K. Raivio, "A SOM based approach for visualization of GSM network performance data," in *Proc. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2005, pp. 588–598.

[96] P. Kumpulainen and K. Hätönen, "Local anomaly detection for mobile network monitoring," *Information Sciences*, vol. 178, no. 20, pp. 3840–3859, 2008.

[97] 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, no. 4, pp. 2369–2386, 2016.

[98] T.-W. Lee, *Independent component analysis.* Springer, 1998, pp. 27–66.

[99] J. H. Ward Jr., "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[100] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 224–227, 1979.

[101] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 3, pp. 301–315, 1998.

[102] S. Rezaei, H. Radmanesh, P. Alavizadeh, H. Nikoofar, and F. Lahouti, "Automatic fault detection and diagnosis in cellular networks using operations support systems data," in *Proc. IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2016, pp. 468–473.

[103] D. A. Hill, L. M. Delaney, and S. Roncal, "A chi-square automatic interaction detection (CHAID) analysis of factors determining trauma outcomes," *Journal of Trauma and Acute Care Surgery*, vol. 42, no. 1, pp. 62–66, 1997.

[104] G. F. Ciocarlie, C. Connolly, C.-C. Cheng, U. Lindqvist, S. Nováczki, H. Sanneck, and M. Naseer-ul Islam, "Anomaly detection and diagnosis for automatic radio network verification," in *Proc. International Conference on Mobile Networks and Management*, 2014, pp. 163–176.

[105] M. Dandan, Q. Xiaowei, and W. Weidong, "Anomalous cell detection with kernel density-based local outlier factor," *China Communications*, vol. 12, no. 9, pp. 64–75, 2015.

[106] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[107] R. M. Khanafer, B. Solana, J. Triola, R. Barco, L. Moltsen, Z. Altman, and P. Lazaro, "Automated diagnosis for UMTS networks using Bayesian network approach," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2451–2461, 2008.

[108] R. Barco, V. Wille, L. Diez, and P. Laizaro, "Comparison of probabilistic models used for diagnosis in cellular networks," in *Proc. IEEE 63rd Vehicular Technology Conference (VTC)*, 2006, pp. 981–985.

[109] D. Heckerman and J. S. Breese, "Causal independence for probability assessment and inference using Bayesian networks," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 26, no. 6, pp. 826–831, 1996.

[110] R. Barco, P. Lázaro, V. Wille, and L. Díez, "Knowledge acquisition for diagnosis in cellular networks based on bayesian networks," in *Proc. International Conference on Knowledge Science, Engineering and Management*, 2006, pp. 55–65.

[111] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," pp. 226–231, 1996.

[112] W. H. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of classification*, vol. 1, no. 1, pp. 7–24, 1984.

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

[114] R. Barco, P. Lázaro, L. Díez, and V. Wille, "Continuous versus discrete model in autodiagnosis systems for wireless networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 6, pp. 673–681, 2008.

[115] R. Barco, L. Díez, V. Wille, and P. Lázaro, "Automatic diagnosis of mobile communication networks under imprecise parameters," *Expert systems with Applications*, vol. 36, no. 1, pp. 489–500, 2009.

[116] J. Laiho, K. Raivio, P. Lehtimaki, K. Hatonen, and O. Simula, "Advanced analysis methods for 3G cellular networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 930–942, 2005.

[117] U. Fayyad and K. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," 1993.

[118] M. Z. Asghar, S. Hämäläinen, and T. Ristaniemi, "Self-healing framework for LTE networks," in *Proc. IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2012, pp. 159–161.

[119] M. Amirijoo, L. Jorguseski, R. Litjens, and R. Nascimento, "Effectiveness of cell outage compensation in LTE networks," in *Proc. IEEE Consumer Communications and Networking Conference (CCNC)*, 2011, pp. 642–647.

[120] C. Frenzel, H. Sanneck, and B. Bauer, "Automated rational recovery selection for self-healing in mobile networks," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 41–45.

[121] Z. Jiang, Y. Peng, Y. Su, W. Li, and X. Qiu, "A cell outage compensation scheme based on immune algorithm in LTE networks," in *Proc. 15th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2013, pp. 1–6.

[122] L. Wenjing, Y. Peng, J. Zhengxin, and L. Zifan, "Centralized management mechanism for cell outage compensation in LTE networks," *International Journal of Distributed Sensor Networks*, 2012.

[123] G. Hong and M. Zong-Yuan, "Immune algorithm," in *Proc. 4th World Congress on Intelligent Control and Automation*, 2002, pp. 1784–1788.

[124] M. Amirijoo, L. Jorguseski, R. Litjens, and L.-C. Schmelz, "Cell outage compensation in LTE networks: algorithms and performance assessment," in *Proc. IEEE 73rd Vehicular Technology Conference (VTC)*, 2011, pp. 1–5.

[125] F. Li, X. Qiu, W. Li, and H.-L. Wan, "High load cell outage compensation method in TD-SCDMA wireless access network," *Journal of Beijing University of Posts and Telecommunications*, vol. 35, no. 1, pp. 32–35, 2012.

[126] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th International Symposium on Micro Machine and Human Science*, 1995, pp. 39–43.

[127] W. Wang, J. Zhang, and Q. Zhang, "LOGA: Local grouping architecture for self-healing femtocell networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 5136–5141.

[128] K. Lee, H. Lee, and D.-H. Cho, "Fairness-aware cooperative resource allocation for self-healing in SON-based indoor system," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1030–1033, 2012.

[129] A. Saeed, O. G. Aliu, and M. A. Imran, "Controlling self healing cellular networks using fuzzy logic," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 3080–3084.

[130] J. Moysen and L. Giupponi, "A Reinforcement Learning based solution for Self-Healing in LTE networks," in *Proc. IEEE 80th Vehicular Technology Conference (VTC)*, 2014, pp. 1–6.

[131] M. I. Tiwana, B. Sayrac, and Z. Altman, "Statistical learning in automated troubleshooting: Application to LTE interference mitigation," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 7, pp. 3651–3656, 2010.

[132] K. Lee, H. Lee, and D.-H. Cho, "Collaborative resource allocation for self-healing in self-organizing networks," in *Proc. IEEE International Conference on Communications (ICC)*, 2011, pp. 1–5.

[133] H. Lee and K. Lee, "Resource allocation considering fault management in indoor mobile-WiMAX system," in *Proc. IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2009, pp. 1492–1496.

[134] K. Lee, H. Lee, Y.-U. Jang, and D.-H. Cho, "CoBRA: Cooperative beamforming-based resource allocation for self-healing in SON-based indoor mobile communication system," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5520–5528, 2013.

[135] L. Xia, W. Li, H. Zhang, and Z. Wang, "A cell outage compensation mechanism in self-organizing RAN," in *Proc. 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, 2011, pp. 1–4.

[136] S. Rohde and C. Wietfeld, "Interference aware positioning of aerial relays for cell overload and outage compensation," in *Proc. IEEE Vehicular Technology Conference (VTC)*, 2012, pp. 1–5.

[137] J. Aráuz and W. McClure, "PGM structures in self-organized healing for small cell networks," in *Proc. International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, 2013, pp. 7–12.

[138] M. I. Tiwana, "Enhancemant of the Statistical Learning Automated Healing (SLAH) technique using packet scheduling," in *Proc. International Conference on Emerging Technologies (ICET).* IEEE, 2012, Conference Proceedings, pp. 1–5.

[139] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Proc. Information Theory and Applications Workshop*, 2007, pp. 414–420.

[140] H. Lee, H. Kim, and K. Lee, "Collaborative self-healing with opportunistic IBS selection in indoor wireless communication systems," *IEEE Communications Letters*, vol. 18, no. 12, pp. 2209–2212, 2014.

[141] S. Fan and H. Tian, "Cooperative resource allocation for self-healing in small cell networks," *IEEE Communications Letters*, vol. 19, no. 7, pp. 1221–1224, 2015.

[142] L. He, X. Su, J. Zeng, X. Xu, and Y. Kuang, "Automated healing approach in cloud base-station with high load using RRUs cooperation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2012, pp. 285–290.

[143] O. Østerbø and O. Grondalen, "Benefits of self-organizing networks (SON) for mobile operators," *Journal of Computer Networks and Communications*, vol. 2012, 2012.

[144] H. Y. Lateef, A. Imran, and A. Abu-Dayya, "A framework for classification of Self-Organising network conflicts and coordination algorithms," in *Proc. IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2013, pp. 2898–2903.

[145] T. Kemptner and B. T. Tsvetkov, "Lte son-function coordination concept," *Network*, vol. 101, 2013.

[146] A. Galani, K. Tsagkaris, P. Demestichas, G. Nguengang, I. G. BenYahia, M. Stamatelatos, E. Kosmatos, A. Kaloxylos, and L. Ciavaglia, "Core functional and network empower mechanisms of an operator-driven, framework for unifying autonomic network and service management," in *Proc. IEEE 17th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2012, pp. 191–195.

[147] 3GPP, "TR 36.902 - V9.2.0 Self-configuring and self-optimizing network (SON) use cases and solutions," no. 36.902, 2010.

[148] K. Tsagkaris, N. Koutsouris, P. Demestichas, R. Combes, and Z. Altman, "SON Coordination in a Unified Management Framework," in *Proc. IEEE 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.

[149] Z. Altman, M. Amirijoo, F. Gunnarsson, H. Hoffmann, I. Z. Kovacs, D. Laselva, B. Sas, K. Spaey, A. Tall, and H. van den Berg, "On design principles for self-organizing network functions," in *Proc. 11th International Symposium on Wireless Communications Systems (ISWCS)*. IEEE, 2014, Conference Proceedings, pp. 454–459.

[150] L. C. Schmelz, M. Amirijoo, A. Eisenblaetter, R. Litjens, M. Neuland, and J. Turk, "A coordination framework for self-organisation in LTE networks," in *Proc. 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, 2011, pp. 193–200.

[151] J. Ali-Tolppa and T. Tsvetkov, "Optimistic concurrency control in self-organizing networks using automatic coordination and verification," in *Proc. IEEE/IFIP Network Operations and Management Symposium*, 2016, pp. 618–624.

[152] T. Bandh, R. Romeikat, H. Sanneck, and H. Tang, "Policy-based coordination and management of SON functions," in *Proc. International Symposium on Integrated Network Management (IM)*. IEEE, 2011, Conference Proceedings, pp. 827–840.

[153] T. Bandh and L. C. Schmelz, "Impact-time concept for SON-Function coordination," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 16–20.

[154] R. Romeikat, H. Sanneck, and T. Bandh, "Efficient, Dynamic Coordination of Request Batches in C-SON Systems," in *Proc. IEEE 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–6.

[155] S. S. Mwanje and A. Mitschele-Thiel, "STS: Space-time scheduling for coordinating self-organization network functions in LTE," in *Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, pp. 357–362.

[156] M. Stamatelatos, A. Paraskevopoulos, A. Kaloxylos, and N. Alonistioti, "A generic scheme and a validation model for SON coordination in 5G networks," in *Proc. 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 2015, pp. 1–6.

[157] J. Moysen and L. Giupponi, "A Functional Architecture for Self-Coordination in LTE networks," in *Proc. 20th European Wireless Conference*, 2014, pp. 1–5.

[158] P. Vlacheas, E. Thomatos, K. Tsagkaris, and P. Demestichas, "Operator-governed SON coordination in downlink LTE networks," in *Proc. Future Network & Mobile Summit (FutureNetw)*, 20, pp. 1–9.

[159] C. Frenzel, B. Bauer, C. Schmelz, and H. Sanneck, "Objective-driven coordination in self-organizing networks," in *Proc. IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 1453–1458.

[160] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting metrics in 5g systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, 2014.

[161] K. T. Dinh and S. Kuklinski, "Joint implementation of several LTE-SON functions," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 953–957.

[162] X. Gelabert, B. Sayrac, and S. B. Jemaa, "A heuristic coordination framework for self-optimizing mechanisms in lte hetnets," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1320–1334, 2014.

[163] J. Chen, H. Zhuang, B. Andrian, and Y. Li, "Difference-Based Joint Parameter Configuration for MRO and MLB," in *Proc. IEEE 75th Vehicular Technology Conference (VTC Spring)*, 2012, pp. 1–5.

[164] A. Lobinger and S. Stefanski, "Coordinating handover parameter optimization and load balancing in lte self-optimization networks," in *Proc. Vehicular Technology Conference (VTC Spring)*, 2011.

[165] P. Muñoz, R. Barco, and S. Fortes, "Conflict resolution between load balancing and handover optimization in lte networks," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1795–1798, 2014.

[166] Z. Liu, P. Hong, K. Xue, and M. Peng, "Conflict avoidance between mobility robustness optimization and mobility load balancing," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–5.

[167] N. Zia, S. S. Mwanje, and A. Mitschele-Thiel, "A policy based conflict resolution mechanism for MLB and MRO in LTE self-optimizing networks," in *Proc. IEEE Symposium on Computers and Communications (ISCC)*, 2014, pp. 1–6.

[168] S. S. Mwanje and A. Mitschele-Thiel, "Concurrent cooperative games for coordinating SON functions in cognitive cellular networks," in *Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015, pp. 1298–1303.

[169] H. Y. Lateef, A. Imran, M. A. Imran, L. Giupponi, and M. Dohler, "LTE-advanced self-organizing network conflicts and coordination algorithms," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 108–117, 2015.

[170] A. Schrijver, *Theory of linear and integer programming.* John Wiley & Sons, 1998.

[171] R. Sabouni, R. M. Hafez, and M. St-Hilaire, "A centralized algorithm for adaptive coverage in next generation cellular networks," in *Proc. 6th International Conference on the Network of the Future (NOF)*, 2015, pp. 1–6.

[172] D. Gesbert, S. G. Kiani, A. Gjendemsjo, and G. E. Oien, "Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2393–2409, 2007.

[173] D. P. Heyman and M. J. Sobel, *Stochastic models in operations research: stochastic optimization.* Courier Corporation, 2004, vol. 2.

[174] A. Tall, R. Combes, Z. Altman, and E. Altman, "Distributed coordination of

self-organizing mechanisms in communication networks," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 328–337, 2014.

[175] S. Hilger, "Generalized theorem of hartman-grobman on measure chains," *Journal of the Australian Mathematical Society*, vol. 60, no. 2, pp. 157–191, 1996.

[176] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1974–1988, 2013.

[177] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[178] S. Berger, M. Simsek, A. Fehske, P. Zanier, I. Viering, and G. Fettweis, "Joint downlink and uplink tilt-based self-organization of coverage and capacity under sparse system knowledge," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2259–2273, 2016.

[179] J. Bartelt, A. Fehske, H. Klessig, G. Fettweis, and J. Voigt, "Joint Bandwidth Allocation and Small Cell Switching in Heterogeneous Networks," in *Proc. IEEE 78th Vehicular Technology Conference (VTC Fall)*, 2013, pp. 1–5.

[180] B. Maaz, K. Khawam, S. Tohme, S. Martin, S. Lahoud, and J. Nasreddine, "Joint scheduling and power control in multi-cell networks for inter-cell interference coordination," in *Proc. IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2015, pp. 778–785.

[181] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 888–901, 2014.

[182] R. E. Miles, "On the homogeneous planar poisson point process," *Mathematical Biosciences*, vol. 6, pp. 85–127, 1970.

[183] S. Deb and P. Monogioudis, "Learning-based uplink interference management in 4g lte cellular systems," *IEEE/ACM Transactions on Networking*, vol. 23, no. 2, pp. 398–411, 2015.

[184] J. Moysen and L. Giupponi, "Self-coordination of parameter conflicts in d-son

architectures: a markov decision process framework," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 82, Mar. 2015. [Online]. Available: https://doi.org/10.1186/s13638-015-0312-7

[185] O. Iacoboaiea, B. Sayrac, S. B. Jemaa, and P. Bianchi, "Coordinating SON instances: Reinforcement learning with distributed value function," in *Proc. IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2014, pp. 1642–1646.

[186] M. Simsek, M. Bennis, and A. Czylwik, "Coordinated beam selection in LTE-Advanced HetNets: A reinforcement learning approach," in *Proc. IEEE Globecom Workshops*, 2012, pp. 603–607.

[187] S. S. Mwanje, L. C. Schmelz, and A. Mitschele-Thiel, "Cognitive cellular networks: A q-learning framework for self-organizing networks," *IEEE Transactions on Network and Service Management*, vol. 13, no. 1, pp. 85–98, 2016.

[188] I. H. Witten, T. C. Bell, M.-E. Harrison, M. L. James, and A. Moffat, "Textual image compression," in *Proc. IEEE Data Compression Conference*, 1992.

[189] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, pp. 237–264, 1953.

[190] Y. Kumar, H. Farooq, and A. Imran, "Fault Prediction and Reliability Analysis in a Real Cellular Network," in *Proc. 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017.

[191] O. P. Kogeda and J. I. Agbinya, "Proactive Cellular Network Faults Prediction Through Mobile Intelligent Agent Technology," in *Proc. 2nd International Conference on Wireless Broadband and Ultra Wideband Communications*, 2007, pp. 55–55.

[192] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[193] L. Pierucci, "The quality of experience perspective toward 5G technology," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 10–16, August 2015.

[194] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18–25, 2014.

[195] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, June 2011.

[196] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[197] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, October 2016.

[198] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends in dense urban mmWave cellular networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 3809–3814.

[199] A. Ericsson, "Ericsson mobility report: On the pulse of the Networked Society," Tech. Rep., 2015.

[200] R. K. Sahoo, M. S. Squillante, A. Sivasubramaniam, and Y. Zhang, "Failure data analysis of a large-scale heterogeneous server environment," in *Proc. International Conference on Dependable Systems and Networks*, 2004, pp. 772–781.

[201] F. Xing and W. Wang, "On the survivability of wireless ad hoc networks with node misbehaviors and failures," *IEEE Transactions on Dependable and Secure Computing*, vol. 7, no. 3, pp. 284–299, 2010.

[202] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28–36, June 2012.

[203] B. Sayrac, J. Riihijärvi, P. Mähönen, S. Ben Jemaa, E. Moulines, and S. Grimoud, "Improving Coverage Estimation for Cellular Networks with Spatial Bayesian Prediction Based on Measurements," in *Proc. ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, 2012, pp. 43–48.

[204] A. Galindo-Serrano, B. Sayrac, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," in *Proc. 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, May 2013, pp. 35–40.

[205] H. W. Liang, C. H. Ho, L. S. Chen, W. H. Chung, S. Y. Yuan, and S. Y. Kuo, "Coverage hole detection in cellular networks with deterministic propagation model," in *Proc. 2nd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, June 2016, pp. 1–6.

[206] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36–43, November 2012.

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

[208] D. T. Larose, *k-Nearest Neighbor Algorithm*. John Wiley & Sons, Inc., 2005, pp. 90–106.

[209] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-based Local Outliers," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, May 2000.

[210] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "Data-driven analytics for automated cell outage detection in Self-Organizing Networks," in *Proc. 11th International Conference on the Design of Reliable Communication Networks (DRCN)*, March 2015, pp. 203–210.

[211] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1533–1547, 2016.

[212] I. Nevat, D. M. Divakaran, S. G. Nagarajan, P. Zhang, L. Su, L. L. Ko, and V. L. L. Thing, "Anomaly detection and attribution in networks with temporally correlated traffic," *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–14, 2017.

[213] Y. Xie and S. Z. Yu, "A large-scale hidden semi-markov model for anomaly detection on user browsing behaviors," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 54–65, 2009.

[214] I. C. Paschalidis and G. Smaragdakis, "Spatio-temporal network anomaly detection by assessing deviations of empirical measures," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 685–697, 2009.

[215] L. R. Frank and V. L. Galinsky, "Detecting spatio-temporal modes in multivariate data by entropy field decomposition," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 39, p. 395001, 2016.

[216] ——, "Dynamic Multiscale Modes of Resting State Brain Activity Detected by Entropy Field Decomposition," *Neural Computation*, vol. 28, no. 9, pp. 1769–1811, 2016, pMID: 27391678.

[217] L. R. Frank, V. L. Galinsky, L. Orf, and J. Wurman, "Dynamic Multiscale Modes of Severe Storm Structure Detected in Mobile Doppler Radar Data by Entropy Field Decomposition," *Journal of Atmospheric Sciences*, vol. 75, no. 3, pp. 709–730, Mar. 2018. [Online]. Available: https://doi.org/10.1175/JAS-D-17-0117.1

[218] T. A. Enßlin, M. Frommert, and F. S. Kitaura, "Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis," *Physical Review D*, vol. 80, Nov 2009.

[219] L. R. Frank and V. L. Galinsky, "Information pathways in a disordered lattice," *Physical Review E*, vol. 89, no. 3, p. 032142, Mar. 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.89.032142

[220] W. R. Hamilton, "On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function," *Philosophical transactions of the Royal Society of London*, vol. 124, pp. 247–308, 1834.

[221] V. Y. Pan and Z. Q. Chen, "The Complexity of the Matrix Eigenproblem," in *Proc. 31st Annual ACM Symposium on Theory of Computing*, ser. STOC '99, New York, NY, USA, 1999, pp. 507–516.

[222] J. V. Stone, *Independent component analysis: a tutorial introduction.* MIT press, 2004.

[223] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[224] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[225] Forsk. (Accessed: 12-Oct-2018) Atoll Overview. [Online]. Available: http://www.forsk.com/atoll-overview

[226] ——. (Accessed: 12-Oct-2018) Aster Propagation Model. [Online]. Available: http://www.forsk.com/aster-propagation-model

[227] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, May 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[228] C. R. Johnson, "Matrix completion problems: a survey," vol. 40, pp. 171–198, 1990.

[229] N. Cressie, "The origins of kriging," *Mathematical Geology*, vol. 22, no. 3, pp. 239–252, Apr. 1990. [Online]. Available: https://doi.org/10.1007/BF00889887

[230] M. Pesko, T. Javornik, A. Košir, M. Štular, and M. Mohorčič, "Radio environment maps: The survey of construction methods." *KSII Transactions on Internet & Information Systems*, vol. 8, no. 11, 2014.

[231] J. Cai, E. Candès, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, Nov. 2018. [Online]. Available: https://doi.org/10.1137/080738970

[232] T.-L. Tien, "A new grey prediction model FGM(1, 1)," *Mathematical and Computer Modelling*, vol. 49, no. 7, pp. 1416 – 1426, 2009.

[233] S. I. Olsen, "Noise Variance Estimation in Images," in *Proc. 8th Scandinavian Conference on Image Analysis (SCIA)*, 1993, pp. 25–28.

[234] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, "Survey of image denoising techniques," in *Proc. International Signal Processing Conference and Expo (GSPX)*, 2004, pp. 27–30.

[235] E. J. Kostelich and T. Schreiber, “Noise reduction in chaotic time-series data: A survey of common methods,” *Physical Review E*, vol. 48, no. 3, pp. 1752–1763, Sep. 1993. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.48.1752

[236] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, “High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning,” *Pattern Recognition*, vol. 58, pp. 121–134, Oct. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320316300267

[237] Q. Mao, F. Hu, and Q. Hao, “Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2595–2621, Fourthquarter 2018.

[238] U. Masood, A. Asghar, A. Imran, and A. N. Mian, “Deep learning based detection of sleeping cells in next generation cellular networks,” in *Proc. IEEE Global Communications Conference (Globecom)*, 2018.

[239] N. Tandiya, A. Jauhar, V. Marojevic, and J. H. Reed, “Deep predictive coding neural network for rf anomaly detection in wireless networks,” *arXiv preprint arXiv:1803.06054*, 2018.

[240] U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis, “Network anomaly detection with the restricted boltzmann machine,” *Neurocomputing*, vol. 122, pp. 13–23, Dec. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231213005547

[241] C. Callegari, S. Giordano, and M. Pagano, “Neural network based anomaly detection,” in *Proc. IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Dec 2014, pp. 310–314.

[242] J. Park and I. W. Sandberg, “Universal approximation using radial-basis-function networks,” *Neural computation*, vol. 3, no. 2, pp. 246–257, 1991.

[243] M. Kuhn and K. Johnson, *Applied predictive modeling.* Springer, 2013, vol. 26.

[244] E. N. Lorenz, *The essence of chaos.* University of Washington Press, 1995.

[245] H. Farooq, A. Asghar, and A. Imran, “Mobility Prediction based Autonomous

Proactive Energy Saving (AURORA) Framework for Emerging Ultra-Dense Networks," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2018.

[246] A. Zoha, A. Saeed, H. Farooq, A. Rizwan, A. Imran, and M. A. Imran, "Leveraging Intelligence from Network CDR Data for Interference aware Energy Consumption Minimization," *IEEE Transactions on Mobile Computing (Accepted for publication)*, 2017.

[247] A. Asghar, H. Farooq, and A. Imran, "Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1682–1709, thirdquarter 2018.

[248] ——, "Concurrent Optimization of Coverage, Capacity, and Load Balance in HetNets Through Soft and Hard Cell Association Parameters," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8781–8795, 2018.

[249] A. Imran, M. A. Imran, A. Abu-Dayya, and R. Tafazolli, "Self organization of tilts in relay enhanced networks: A distributed solution," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 764–779, 2014.

[250] 3GPP, "TS 36.331 - V8.6.0 Radio resource control (RRC); protocol specification," no. 36.331, 2009.

[251] ——, "TR 36.814 - V9.0.0 Further advancements for E-UTRA physical layer aspects," no. 36.814, 2010.

[252] R. Muirhead, "Proofs that the arithmetic mean is greater than the geometric mean," *The Mathematical Gazette*, vol. 2, no. 39, pp. 283–287, 1903.

[253] D. F. Shanno, "On broyden-fletcher-goldfarb-shanno method," *Journal of Optimization Theory and Applications*, vol. 46, no. 1, pp. 87–94, 1985.

[254] M. J. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The computer journal*, vol. 7, no. 2, pp. 155–162, 1964.

[255] W. E. Hart, "Evolutionary pattern search algorithms," *Sandia National Laboratories report SAND95-2293, printed September*, 1995.

[256] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[257] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2484–2497, 2013.

[258] X. Zhao, W. Zhang, and C. Wang, "A load prediction based virtual cell breathing scheme for LTE-A system," in *Proc. IEEE Military Communications Conference (MILCOM)*, 2013, pp. 1296–1301.

[259] I. Siomina and D. Yuan, "Load balancing in heterogeneous LTE: Range optimization via cell offset and load-coupling characterization," in *Proc. IEEE International Conference on Communications (ICC)*, 2012, pp. 1357–1361.

[260] H. Du, Y. Zhou, L. Tian, X. Wang, Z. Pan, J. Shi, and Y. Yuan, "A load fairness aware cell association for centralized heterogeneous networks," in *Proc. IEEE International Conference on Communications (ICC)*, 2015, pp. 2178–2183.

[261] C. A. S. Franco and J. R. B. de Marca, "Load balancing in self-organized heterogeneous LTE networks: A statistical learning approach," in *Proc. IEEE 7th Latin-American Conference on Communications (LATINCOM)*, 2015, pp. 1–5.

[262] Z. Arslan, M. Erel, Y. Özcevik, and B. Canberk, "SDoff: A software-defined offloading controller for heterogeneous networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 2827–2832.

[263] I. Siomina and D. Yuan, "Analysis of cell load coupling for LTE network planning and optimization," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2287–2297, 2012.

[264] W. Song, W. Zhuang, and Y. Cheng, "Load balancing for cellular/WLAN integrated networks," *IEEE network*, vol. 21, no. 1, pp. 27–33, 2007.

[265] Q. Song and A. Jamalipour, "A network selection mechanism for next generation networks," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 2, 2005, pp. 1418–1422.

[266] A. Imran, E. Yaacoub, M. A. Imran, and R. Tafazolli, "Distributed load balancing through self organisation of cell size in cellular systems," in *Proc. IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012, pp. 1114–1119.

Appendix A

The dowlink received power based on standard exponential pathloss model with Gaussian distributed shadowing for user u associated with cell c is:

$$P_{r,u}^{c} = P_t^c G_u G_u^c \epsilon_u^c a\left(d_u^c\right)^{-\beta}$$

Due to the randomness of ϵ_u^c, the coverage constraint of user u i.e. $C_u := P_{r,u}^{c} \geqslant P_{th}^{c}$ becomes a function of ϵ_u^c such that for user u, the coverage constraint will be satisfied with some probability i.e. $Pr(P_{r,u}^{c}(\epsilon_u^c) \geqslant P_{th}^{c})$. We can then calculate the minimum value of ϵ_u^c above which the coverage constraint for an individual user will be satisfied.

$$\begin{aligned} &Pr(P_t^c G_u G_u^c \epsilon_u^c a\left(d_u^c\right)^{-\beta} \geqslant P_{th}^{c}) \\ &Pr(\epsilon_u^c \geqslant \frac{P_{th}^{c}}{P_t^c G_u G_u^c a\left(d_u^c\right)^{-\beta}}) \end{aligned} \tag{6.1}$$

Based on the assumption that ϵ_u^c has a Gaussian distribution, the value of ϵ_u^c inside the parentheses in (6.1) gives the Z-score below which the coverage constraint of an individual user will be violated. If p gives the probability $Pr(\epsilon_u^c \geqslant \frac{P_{th}^{c}}{P_t^c G_u G_u^c a(d_u^c)^{-\beta}})$, we can remodel the event that a user is inside the coverage of its serving cell as a Bernoulli variable with probability p.

The consequence of modeling $Pr(P_{r,u}^{c} \geqslant P_{th}^{c})$ as a Bernoulli random variable is that the network coverage C can be modeled as a Binomial random variable with chance of success p per user. Thus the probability of having $\varpi * |\mathbb{U}|$ users or more in coverage can be given as:

$$Pr(k \geqslant \varpi * |\mathbb{U}|) \geqslant \sum_{i=\varpi*|\mathbb{U}|}^{|\mathbb{U}|} \binom{|\mathbb{U}|}{i} p^i (1-p)^{|\mathbb{U}|-i} \tag{6.2}$$

If the desired value for $Pr(k \geqslant \varpi * |\mathbb{U}|) \to 1$, and $\varpi \to 1$, then

$$\lim_{\substack{\varpi \to 1 \\ Pr(k \geqslant \varpi * |\mathbb{U}|) \to 1}} p = 1$$

In such a scenario, we can substitute the probability p with the indicator function $1(.)$, thus giving us the following formulation for constraint (i):

$$\frac{1}{|\mathbb{C}|} \sum_{\mathbb{C}} \frac{1}{|\mathbb{U}_c|} \sum_{\mathbb{U}_c} 1 \left(P^c_{r,u} \geqslant P^c_{th} \right)$$

Appendix B

The simplified form of the SINR function in antenna tilts is given as:

$$\gamma_u^c = \frac{10^{\mu\left(\left(\psi_u^c - \psi_{tilt}^c\right)^2 + x_u^c\right)}}{\kappa + 10^{\mu\left(\left(\psi_u^i - \psi_{tilt}^i\right)^2 + x_u^i\right)}}$$

The expression for antenna gains expressed as a function of tilts is given as:

$$G_u^c = 10^{\mu\left(\left(\psi_u^c - \psi_{tilt}^c\right)^2 + x_u^c\right)}$$

We treat x_u^c and μ as constants and assign then unit value and -1.2 respectively which gives us the resulting function of ψ_{tilt}^c:

$$f(\psi_{tilt}^c) = 0.0631 * 10^{-1.2\left(\psi_u^c - \psi_{tilt}^c\right)^2}$$

Taking derivative of $f(\psi_{tilt}^c)$ gives us:

$$f'(\psi_{tilt}^c) = 0.0631 l_n(10) * 10^{-1.2\left(\psi_u^c - \psi_{tilt}^c\right)^2} * (2.4(\psi_u^c - \psi_{tilt}^c))$$

Taking the second derivative:

$$f''(\psi_{tilt}^c) = 0.3634 l_n(10) * 10^{-1.2\left(\psi_u^c - \psi_{tilt}^c\right)^2} [(\psi_u^c - \psi_{tilt}^c) l_n(10) - 0.417]$$

For a function to be convex, the second derivative has to be non-negative which is only possible in the range $[(\psi_u^c - \psi_{tilt}^c) \leqslant -0.4254, (\psi_u^c - \psi_{tilt}^c) \geqslant 0.4254]$. Hence, the antenna gain function is a non-convex function and by extension, the SINR expression with antenna gain is non-convex.