Feature-Rich Interworking Architecture for Mobile Traffic Offloading

Asrat Mulatu

A dissertation Submitted to

The IT Doctoral Program

Presented in the Fulfillments of the Requirements for the Degree of Doctor of Philosophy in Information Technology

(Wireless Communication Systems)

Addis Ababa University

Addis Ababa, Ethiopia

April 2018

Addis Ababa University

IT Doctoral Program

This is to certify that the dissertation prepared by Asrat Mulatu, entitled: ***Feature-Rich Interworking Architecture for Mobile Traffic Offloading*** and submitted in fulfillment of the requirements for the Degree of Philosophy in Information Technology (Wireless Communication Systems) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

External Examiner: ___________________________ Signature: _____________ Date: __________

Internal Examiner: ____________________________ Signature: _____________ Date: __________

Advisor (PSV): ______________________________ Signature: _____________ Date: __________

Advisor (CSO): ______________________________ Signature: _____________ Date: __________

Dedicated to

Hailu Ayele (Dr.-Ing.),

RIP!!!

Acknowledgment

This work would not have been started at all if Dejene Ejigu (PhD) had helped me to secure the external advisor. I am indebted to you greatly!

Yalemzewd Negash (PhD) deserves great acknowledgment not only because he is my advisor but also he is the originator of the idea of the research topic. Honestly, you were not my advisor. You were just a friend!

My main advisor, Jordi Casademont Serra (Dr.Eng.), gratitude alone may not be enough for your understanding, cooperation, criticality, and guidance. You are really an advisor any PhD candidate would like to have. Thank you for everything, *gràcies senyor*! Your colleague, Prof. Oriol Sallent, deserves recognition, first, for his willingness and, second, for his brief but significant insight on the overall research roadmap.

I would like to thank AAU staffs working at the various academic units and administrative offices who full-heartedly gave their hands whenever I needed it.

I would like to extend my love and respect for my family members for their understanding and support, especially for my wife, Yewibdar Molla, for embellishing me, and my life, with two sweet kids: Kiya (Noah) and Enana (Yemariam).

I owe you God and GULS F.W.X!!!

Asrat Mulatu

Abstract

We are in the age where mobile devices have become an indispensable part of our lives. They bring applications and services anytime and anywhere. The applications and services rely, mainly, on our cellular data plans. As the cellular infrastructure is more reliable and has better coverage, it shoulders the majority of user traffic globally. Due to that it is getting overburdened and is forcing customers to go on complaining about poor performances. Though there are many alternative solutions for service providers to choose from, mobile traffic offloading is regarded as affordable and better as a short term solution due to many reasons. Less or no modification to existing infrastructure, availability of the required enabling technologies in existing mobile devices, and price worthiness of the initial investment and operational cost are among the reasons that appeal to both operators and customers.

Therefore, in this research work a feature-rich interworking architecture is proposed for the purpose of mobile traffic offloading. It is fully-integrated, hybrid and Quality of Service -aware. It integrates three different wireless technologies, namely, cellular, wireless local area network and device-to-device communication. Moreover, it has multiple links between the cellular and the other two networks: very tight and loose couplings. In addition, it supports both delay tolerant and delay sensitive traffic types. On top of that, a traffic offloading algorithm is devised that can monitor the instantaneous network conditions and decide through which specific path to en-route the packets while fulfilling the Quality of Service demand of the initiated data traffic.

The implementation is done using the NS3 simulation tool followed by performance evaluation using six key performance metrics. Packet loss ratio, end-to-end delay, jitter, packet delivery ratio, offloading ratio and handover delay are the applied metrics.

The results obtained are very much promising and encouraging. For instance, it is found that up to 90% of delay tolerant traffic and above 50% of delay sensitive traffic can be offloaded from the cellular network towards the wireless local area and device-to-device connections.

Though the results obtained may not be conclusive the various performance indicators demonstrated the huge potential of such a feature-rich interworking architecture with many possible future improvements.

Keywords: *Mobile Traffic Offloading, Network Architecture, Heterogeneous Networks, Network Performance Measurement, NS3.*

Table of Contents

CHAPTER ONE: INTRODUCTION **1**
1.1 Background **1**
1.2 Overview of Wireless Technologies **2**
1.2.1 Cellular Networks 2
1.2.2 WLAN Networks 11
1.2.3 Device-to-Device Communications 14
1.3 The Status of Cellular Networks 23
1.4 The Concept of Mobile Traffic Offloading 25
1.5 Other Alternatives to Mobile Traffic Offloading 28
1.6 Statement of the Problem **30**
1.7 Objectives **31**
1.7.1 General Objective 31
1.7.2 Specific Objectives 31
1.8 Research Methodology **32**
1.9 Scope and Contributions of the Research **33**
1.10 Organization of the Rest of the Dissertation **35**
CHAPTER TWO: INTERWORKING ARCHITECTURES FOR MOBILE TRAFFIC OFFLOADING: STATE-OF-THE-ART **36**
2.1 Introduction **36**
2.2 Major Classifications **39**
2.3.1 Level of Integration 40
2.3.2 Technologies Involved 44
2.3.3 Nature of Traffic Offloaded 46
2.3 Efforts towards Feature-rich Interworking Architecture **49**
2.3.1 Research Efforts 50
2.3.2 Standardization Efforts 71
2.3.3 Product Manufacturers 85
2.4 Current Challenges and Open Issues **91**
2.5 Summary **94**
CHAPTER THREE: PROPOSED INTERWORKING ARCHITECTURE: FULLY INTEGRATED, HYBRID AND QoS-AWARE **97**
3.1 Introduction **97**
3.2 The Proposed Interworking Architecture **97**
3.2.1 Major Components and Their Interconnections 99
3.2.2 Traffic Offloading Algorithm 102
3.3 Summary **109**
CHAPTER FOUR: IMPLEMENTATION OF THE PROPOSED OFFLOADING SYSTEM USING NS3 .... **111**
4.1 Introduction **111**

4.2 The 4G/LTE Subsystem 115
4.3 The WLAN Subsystem 118
4.4 The D2D Subsystem 122
4.5 Integrating the Three Subsystems and Hybridizing the Whole System 123
4.6 Making the Offloading System QoS-Aware 125
4.7 The Traffic Offloading Algorithm 129
4.8 Summary 133
CHAPTER FIVE: PERFORMANCE EVALUATION OF THE IMPLEMENTED OFFLOADING SYSTEM .. 134
5.1 Introduction 134
5.2 System Configurations and Simulation Parameters 134
5.3 Performance Evaluation 138
5.3.1 Packet Loss Ratio (PLR) 138
5.3.2 Packet Delivery Ratio (PDR) 140
5.3.3 End-to-End Delay 141
5.3.4 Jitter 142
5.3.5 Offloading Efficiency (OE) 143
5.3.6 Handover Delay 148
5.4 Analysis of Results 148
5.5 Summary 149
CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORKS 151
6.1 Conclusions 151
6.2 Recommendations for Future Works 153
References 155

List of Figures

Figure 1.1: A typical cellular network with frequency reuse factor of 1/4 [3]. 3
Figure 1.2: GSM Network Architecture. Adapted from [5]..5
Figure 1.3: Architectural Elements of LTE System. Adapted from [15]. 9
Figure 1.4: Application Scenarios of D2D Communications in Urban area setting. Adapted from [27]..... 22
Figure 1.5: Classification of mobile traffic offloading based on current research efforts. 27
Figure 2.1: Classification of the Features Common in Interworking Architectures. 39
Figure 2.2: LTE Network Architecture [21]. 41
Figure 2.3: Device-to-Device Communication Classification, Adapted from [36]. 44
Figure 2.4: Evolution of Cellular/Mobile Technologies (adapted from [45]). 46
Figure 2.5: Classification of the Efforts towards Integrated Interworking Architecture. 49
Figure 2.6: Hybrid Interworking Architecture [61]. 51
Figure 2.7: Algorithm for HotZones [8]. 52
Figure 2.8: Algorithm for MixZones [8]. 52
Figure 2.9: MADNet Architecture [44]. 54
Figure 2.10: Communication flow in MADNet Architecture [44]. 54
Figure 2.11: An Offloading System Supporting Uplink Traffic [15]. 54
Figure 2.12: Implementation of D2D System in 4G [84]. 60
Figure 2.13: Architectural and Protocol Enhancement of LTE and LTE - A Systems to Support D2D Communication [35]. 61
Figure 2.14: FlashLinQ Prototype Modem [88]. 61
Figure 2.15: Conceptual MOTO Architecture [90]. 63
Figure 2.16: Offloading Possibilities in MOTO [90]. 63
Figure 2.17: Very Tight Coupling at the Protocol Stack Level. 64
Figure 2.18: 3GPP System to WLAN Inter-networking Simplified Reference Model [127]. 72
Figure 2.19: Untrusted non-3GPP access..75
Figure 2.20: Trusted non-3GPP access..76
Figure 2.21: Non-Roaming Architecture within EPS using S5, S2a, S2b [29]...78
Figure 2.22: Non-Roaming Architecture within EPS using S5, S2c [29]. 78
Figure 2.23: Context aware CnE to enable seamless connectivity as described in [152]..........................86
Figure 2.24: Data offloading solution architecture by Alvarion [153]. 86
Figure 2.25: Trusted Wi-Fi access integration with the packet core network advertised in [79]. 87
Figure 2.26: Simple architecture diagram of Passpoint-enabled Wi-Fi network [162]. 89
Figure 2.27: Passpoint-enabled mobile device states [162]. 96
Figure 2.28: Summary of the Efforts towards Feature-rich Interworking Architecture from technologies involved, nature of traffic and integration perspectives by the main contributors. 96
Figure 3.1: A Generic Interworking Architecture Showing the Various Wireless Technologies Involved and the Possible Traffic Transmission paths. 100
Figure 3.2: Detailed Implementation Diagram of the Proposed Feature-rich Interworking Architecture.101
Figure 3.3: A Flowchart Presenting the Workflow of the Offloading Algorithm Accompanying the Proposed Interworking Architecture. 104

Figure 3.4: Traffic Offloading or Steering Algorithm That Exploits the Proposed Internetworking Architecture Presented Using Pseudo Code Format. ……………………………………………………………………………………………106
Figure 3.5: Pseudo Code Representation of the Algorithm That Utilizes the Very Tight Coupling between the eNB and Wi-Fi AP…….107
Figure 3.6: Pseudo Code Snippet of the Algorithm to Send/Upload Delay Tolerant Traffic to an External Netwok………......107
Figure 3.7: The Pseudo Code that Estimates the Channel Condition of the Complementary Networks Based on Their Channel Frequency……………………………………………………………………………………………………….108
Figure 4.1: Organization of the main NS3 modules. Adapted from [4]……………………………………………….112
Figure 4.2: The LTE-EPC simulation model as part of the NS3 implementation. Taken as-is from [7]………..116
Figure 4.3: Shows the screen capture of the three models - LTE-EPC system, Wi-Fi AP-based network and D2D communication………........117
Figure 4.4: LTE-EPC data plane protocol stack [4]…………………………………………………………………………....118
Figure 4.5: C++ script used to create, configure and link the LTE/4G subsystem components……….……..119
Figure 4.6: WifiNetDevice Architecture [8]……………………………………………………………………………………..120
Figure 4.7: Shows the NS3 code fragment of the WLAN subsystem implementation that consists of the four nodes described previously………………………………………………………………………………………………………120
Figure 4.8: C++ code extract showing the implementation of the D2D subsystem……………………………..122
Figure 4.9: NS3 lines of code used to implement QoS in to LTE subsystem……………………………………….127
Figure 4.10: QoS implementations of the Wi-Fi AP, the 802.11 UE devices (StaWifiMacs), and the D2D ……….127
Figure 4.11: DSCP or ToS field of IPv4 header………………………………………………………………………………..128
Figure 4.12: NS3 code fragment that used to tag and extract user traffic. …………………………………………128
Figure 4.13: Function that estimates the channel conditions of the WLAN and D2D subsystems…………...130
Figure 4.14: Portion of the OffloadTraffic() function that handles delay sensitive traffic. ………………………131
Figure 4.15: Portion of the OffloadTraffic() function that takes care of delay tolerant traffic…………….…..133
Figure 5.1: Packet loss ratio of the various traffic paths. ………………………………………………………………….139
Figure 5.2: End-to-End Delay of the D2D and WLAN paths………………………………………………………………..140
Figure 5.3: End-to-End Delay of the eNB, the tight and loose links………………………………………………….141
Figure 5.4: Delay variability of the five links……………………………………………………………………………………142
Figure 5.5: The ratio of packets delivered at the destination via the respective paths…………………………143
Figure 5.6: Offloading efficiency of those paths used for delay sensitive traffic……………………………….…145
Figure 5.7: Efficiency of the D2D, WLAN, and the loose coupling links…………………………………………….…145
Figure 5.8: The delay induced during vertical handoff…………………………………………………………………….147

List of Tables

Table 1.1: 802.11 PHY layer standards []16]……………………………………………………………………………………...13
Table 1.2: Bluetooth Device Classes Operating Specification [18] [19]……………………………………………………….15
Table 1.3: Summary of alternative solutions to mobile traffic offloading adapted from [46]…………………….....29
Table 2.1: Details of Cellular, Wi-Fi/WLAN and D2D Technologies………………………………………………………………48
Table 2.2: Standardized QCI Characteristics for 3GPP and WLAN Interworking……………………………………….…49
Table 2.3: Scenarios and their Capabilities [127]…………………………………………………………………………………..…72
Table 4.1: NS3 key concepts and abstractions with their contextual meaning and supporting examples..... 113
Table 4.2: Mapping between QoS values: QCI, DSCP and ToS. The table is populated with data from [10], [11] and [12]. 126
Table 5.1: System configurations and simulation parameters for the D2D subsystem. 135
Table 5.2: System configurations and simulation parameters for the WLAN subsystem. 136
Table 5.3: System configurations and simulation parameters for the 4G/LTE subsystem. 13
Table 5.4: Simulation parameters of the very tight and loose couplings. 137
Table 5.5: Comparison of obtained results in this work with international benchmarks …………………………..148

List of Acronyms

6LOWPAN	Low-Power Wireless Personal Area Networks for IPv6
AAA	Authentication, Authorization and Accounting
AFH	Adaptive Frequency-Hopping
A.K.A	Also Known As
AMUSE	Adaptive Bandwidth Management through User-Empowerment
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
BS	Base Station
CAGR	Compound Annual Growth Rate
CAPEX	Capital Expenditure
CDMA	Code Division Multiple Access
CEMA	Central and Eastern Europe and Middle East and Africa
CGW	Converged Gateway
CSMA	Carrier Sense Multiple Access
D2D	Device-to-Device
DOPS	Distance-Based Opportunistic Publish/Subscribe
DPI	Deep Packet Inspection
DROiD	Derivative Re-injection to Offload Data
DSCP	Differentiated Services Code Point
DSMIPv6	Dual-Stack Mobile IP
DSSS	Direct Sequence Spread Spectrum
eNB	Evolved Node B or E-UTRAN Node B
EPC	Evolved Packet Core
EPDG	Evolved Packet Data Gateway
EPS	Evolved Packet System
ETA	Ethiopian Telecommunications Agency
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FDMA	Frequency Division Multiple Access
FHSS	Frequency-Hopping Spread Spectrum
GBR	Garanteed Bit Rate

GCB	Green Content Broker
GFSK	Gaussian Frequency Shift Keying
GGSN	Gateway GPRS Support Node
HARQ	Hybrid Automatic Repeat Request
HSDPA	High Speed Downlink Packet Access
HSS	Home Subscriber Server
HTO	Heuristic Traffic Offloading
HYPE	Hybrid Opportunistic and Cellular
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IFOM	IP Flow Mobility
IMO	Intelligent Mobile Offload
IMS	IP Multimedia System
IMSI	Internationalal Mobile Subscriber Identity Module
ISM	Industrial, Scientific and Medical
ISPs	Internet Service Providers
ITU	International Telecommunications Union
KPIs	Key Performance Indicators
LAA	Licensed-Assisted Access
LAN	Local Area Network
LIPA	Local IP Access
LMA	Local Mobility Anchor
LTE	Long Term Evolution
LTE-A	Long Term Evolution - Advanced
LTE-U	LTE in the Unlicensed Spectrum
MAC	Media Access Control
MADNet	Metropolitan Advance Delivery Network
MIH	Media Independent Handover
MIMO	Multiple Input Multiple Output
MIP	Mobile IP
MME	Mobility Management Entity

MMS	Multimedia Messaging System
MMS	Multimedia Messaging System
MNOs	Mobile Network Operators
MoSoNets	Mobile Social Networks
MOTO	Mobile Opportunistic Traffic Offloading
MSC	Mobile Switching Center
MTC	Machine Type Communications
NAS	Non Access Stratum
NFC	Near Field Communication
NIC	Network Interface Card
NoA	Notice of Absence
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
O-QPSK	Offset Quadrature Phase Shift Keying
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PGW	PDN Gateway
PMIP	Proxy Mobile IP
ProSec	Proximity-based Services
PSTN	Public Switched Telephone Network
QATO	Quality-Aware Traffic Offloading
QCI	QoS Class Identifier
QoS/QoE	Quality of Service/Quality of Expericne
RNC	Radio Network Controller
RSSI	Received Signal Strength Indicator
RSVP	Resource Reservation Protocol
RTS/CTS	Ready-To-Send/Clear-To-Send
SAE	System Architecture Evolution
SAU	Support Alarm Unit

SC-FDMA	Single Carrier Frequency Division Multiple Access
SCTP	Stream Control Transmission Protocol
SDMA	Space Division Multiple Access
SGSN	Serving GPRS Support Node
SGSN	Serving GPRS Support Node
SGW	Serving Gateway
SIM	Subscriber Identity Module
SIP	Session Initiation Protocol
SIPIHO	SIP-Based Intelligent Handover
SMP	Service Management Platform
SMS	Short Messaging System
SWLANC	Smart Wireless LAN Connectivity
TDLS	Tunneled Direct Link Setup
TDMA	Time Division Multiple Access
TFTs	Traffic Flow Templates
ToS	Type of Service
TSGs	Technical Specification Groups
TTG	Tunnel Termination Gateway
TWAG	Trusted WLAN Access Gateway
TWAN	Trusted Wi-Fi Access Network
UE	User Equipment
UICC	Universal Integrated Circuit Card
UMTS	Universal Mobile Telecommunications System
UPnP	Universal Plug and Play
VHO	Vertical Handover
WAG	WLAN Access Gateway
WBSn	Wi-Fi Base Station
W-CDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WLANs	Wireless LAN
WPA2	Wi-Fi Protected Access V2

WPS	Wi-Fi Protected Setup
ZeroConf	Zero Configuration Networking

CHAPTER ONE

INTRODUCTION

1.1 Background

Since the invention of wireless communication by Guglielmo Marconi in 1897 it has evolved through the years and now has become one of the necessities of modern life style. One cannot imagine modern one's life without mobile phones, for instance, it feels like it is impossible to spend a day without it. Satellite communications, mobile phones, social media, televised transmissions, and radio broadcasts are all the fruits of wireless communication evolutions and advancements.

These and other wireless communication technologies and systems are aimed at providing some specific services. For example, the cellular network is a kind of telecommunication network designed to provide voice and data transmission for mobile users. It can provide connectivity for a larger geographic area using what is called base stations (BSs). The installation of these base stations is in such a way that each covers a regularly shaped pattern. Cellular networks are widely used in highly populate urban areas. Cellular network operators invest huge amount of money to provide various connectivity services like voice calls, short messaging system (SMS), multimedia messaging system (MMS), video calls, and Internet access. The kind of service a user gets depends on the type of service they subscribed for, the kind of device they have, and the availability and performance of the network at that particular place and time, among others.

On the other hand, wireless local area networks (WLANs) are intended to provide wireless connectivity in a small geographic area cheaply and, usually, as an extension of a wired local area network. There are devices called Wi-Fi access points (APs) that act as a connection point where mobile users or fixed users connect to access the LAN resources or connect to the Internet. What user devices need to get connected with these WLANs is a wireless network interface card (NIC) having the correct protocol configured like 802.11a/b/g/n/ac and future ax. These wireless connections are most of the time open for the public without any cost since most of the owners intend to extend their LAN service for their employees or customers without additional cost. Nowadays, urban areas throughout the globe are being overwhelmed by WLANs in parks, hotels, cafeterias, universities, and government and non-government vicinities [1] [2].

1.2 Overview of Wireless Technologies

1.2.1 Cellular Networks

Cellular networking was initially proposed by Bells Labs around 1971 as geographic services divided into smaller cells. A cell, covering a particular geographic area, is centered on a fixed-location transceiver called cell site or base station [3]. Each cell uses a different set of frequency from neighboring cells to avoid interference. This guarantees bandwidth within each cell.

The aggregation of these cells provides radio coverage over a wide geographic area. This allows the use of portable transceivers like mobile phone, also known as cellular phone, to communicate with each other and with fixed BSs, even if some of the transceivers are moving through more than one cell during transmission.

The cellular telephone systems throughout the world are based on this kind of cellular networks. They are called telecommunications service providers or internet service providers. Major ones include, but not limited to, T-Mobile, Orange, Huawei, Vodafone, MTN and locally, EthioTelecom.

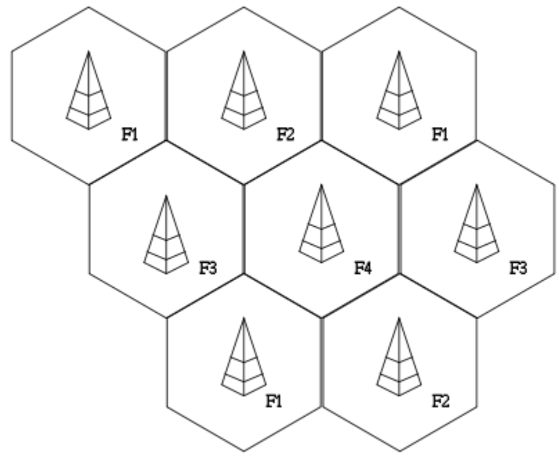


Figure 1.1: A typical cellular network with frequency reuse factor[1] *of 1/4 [3].*

They mainly provide voice and data services using cellular networks which covers most of the inhabited land of the earth. This allows cellular phones and other mobile computing devices to be connected to the public switched telephone networks and public Internet.

In a cellular network system, a land area to be supplied with radio service is divided into regular shaped cells, which can be hexagonal, square, circular or similar regular shapes. Each of these cells is assigned with multiple frequencies like F1-F4 as shown in Figure 1.1 where each of them is equipped with a radio base station. The group of frequencies will be reused in other cells, as long as the same frequencies are not reused in adjacent neighboring cells as that would cause co-channel interference.

The increased capacity in a cellular network, compared with a network with a single transmitter, comes from the mobile communication switching system developed by Amos Joel of Bell Labs [4] that permitted multiple callers in the same area to use the same frequency by switching calls made using the same frequency to the nearest available cellular tower having that frequency available and from the fact that the same radio frequency can be reused in a different area for a

[1] Frequency reuse factor is the rate at which the same frequency can be used in the network. It is 1/k where k is the number of cells which cannot use the same frequencies for transmission.

completely different transmission. If there is a single plain transmitter, only one transmission can be used on any given frequency. Unfortunately, there is inevitably some level of interference from the signal from the other cells which use the same frequency.

Frequency reuse is a key characteristic of cellular networks that enable them re-use frequencies to increase both coverage and capacity. Adjacent cells reuse frequencies. The elements that determine frequency reuse are the reuse distance and the reuse factor. The reuse distance, D is calculated as

$$D = R\sqrt{3N} \tag{1.1}$$

Where R is the cell radius and N is the number of cells per cluster. Cells vary in radius from hundreds of meters up to tens of kilometers. The boundaries of the cells can also overlap between adjacent cells and large cells can be divided into smaller cells. A cluster comprises several base stations, 3 or more, but, usually, less than 20 [3].

As a typical example, a simplified view of a cellular mobile-radio network consists of the following:

- A network of radio base stations forming the base station subsystem.
- The core circuit switched network for handling voice calls and text.
- A packet switched network for handling mobile data.
- The public switched telephone network to connect subscribers to the wider telephony network.

This network is the foundation of the Global System for Mobile communications (GSM) system network which is depicted in Figure 1.2. There are many functions that are performed by this network in order to make sure customers get the desired service including mobility management, registration, call set up, and handover.

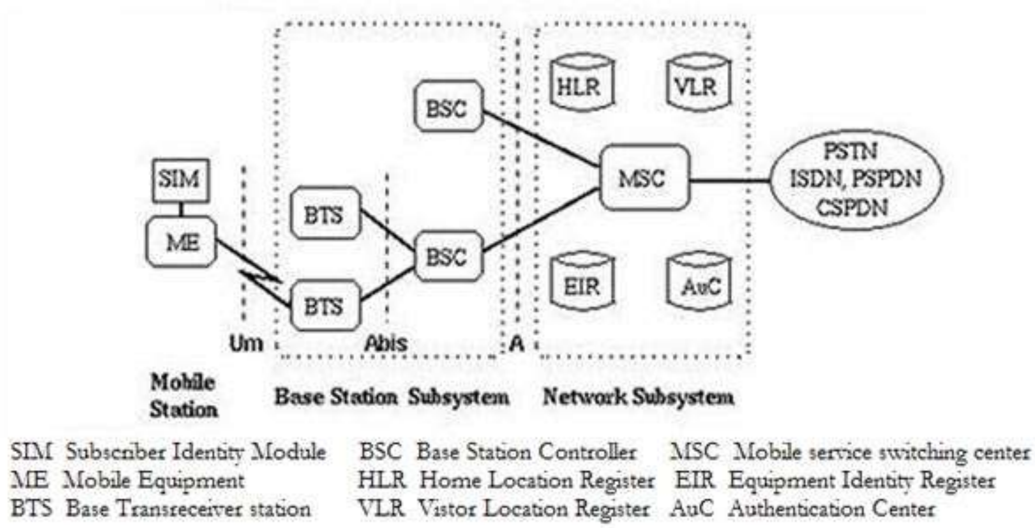


Figure 1.2: GSM Network Architecture [5].

Any phone connects to the network via a BS at a corner of the corresponding cell which in turn connects to the mobile switching center (MSC). The MSC provides a connection to the public switched telephone network (PSTN). The link from a phone to the BS is called an uplink while the other way is termed downlink.

Radio channels effectively use the transmission medium through the use of various multiplexing and access schemes: frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and space division multiple access (SDMA).

Cellular networks are observing inherently huge traffic overload. This is basically because of many reasons. First and foremost is the colossal increase of mobile cellular subscribers. That is true especially in developing world which is estimated by International Telecommunications Union (ITU) as there are 5.4 billion mobile subscriptions in the developing world which is 78% of global subscriptions [5]. The same report indicates that though mobile penetration in the developing world now reached 90.2%, still there is a huge potential for growth, particularly in Africa which has the lowest mobile penetration worldwide which is at 69.3%. On the same report similar claim was made by Portio Research that mobile subscribers worldwide reached 7.5 billion by the end of 2014 and 8.5 billion by the end of 2016. The 2017 report in [6] by Bahjat El-Darwiche et el., ascertained, besides other things, the fact that the above facts are actually happened.

The other reason for the increased traffic demand of cellular networks is the omnipresence of mobile devices like smartphones, tablets and laptops. Number of smartphone users in the US from 2010-2018 is expected to grow from 62.6 million to 220 million, and from 11.7 million to 103.6 million in Indonesia [5]. Similar issues are reflected on [6]. The availability of such devices not only increases the demand for voice and data but also for streaming video, social media, Internet gaming, among others. This further exacerbated the demand for mobile connectivity due to the deployment of 3G and 4G technologies that support such bandwidth hungry systems and applications. Many having such mobile devices and 3G subscriptions want to exchange Multimedia Messaging System (MMS) or upload or download video or image habitually, to say the least.

Yet another emerging challenge of cellular networks is the increased wireless and mobile connectivity demand. Cisco reported that global mobile data traffic reached to 820 petabytes (2^{50} Bytes) per month at the end of 2012 and 1.5 Exabyte (2^{60} Bytes) per month at the end of 2013 [7]. In the same report it is estimated to ascend to 15.9 Exabyte per month by 2018. This is an explosive mobile data traffic growth trend which is 61.2% per year. Similar interesting forecasts and analysis are made in [1].

Due to the fact that cellular networks are being overwhelmed by increasing number of subscribers being equipped with smartphones, tablets, or laptops, with 3G and 4G services the poor performance of these networks is creating upsets by many around the globe [1][8][9].

These trends are not much expected by major and prominent mobile operators throughout the globe though it potentially brings much revenue to them [9]. They rushed to their engineers for possible solutions of increasing their capacity. Expanding to new frequencies, devising new access technologies, considering possible optimization strategies and offloading traffic to complementary wireless networks are among the different practical solutions proposed and being utilized [10].

The above fact clearly shows that there is an expected increase of mobile connectivity demand from the existing cellular networks. This threatens the quality of service (QoS) for delay sensitive services like voice and video creating network overload and congestion. This further increases the number of dropped connections and ultimately rising prices for transmission.

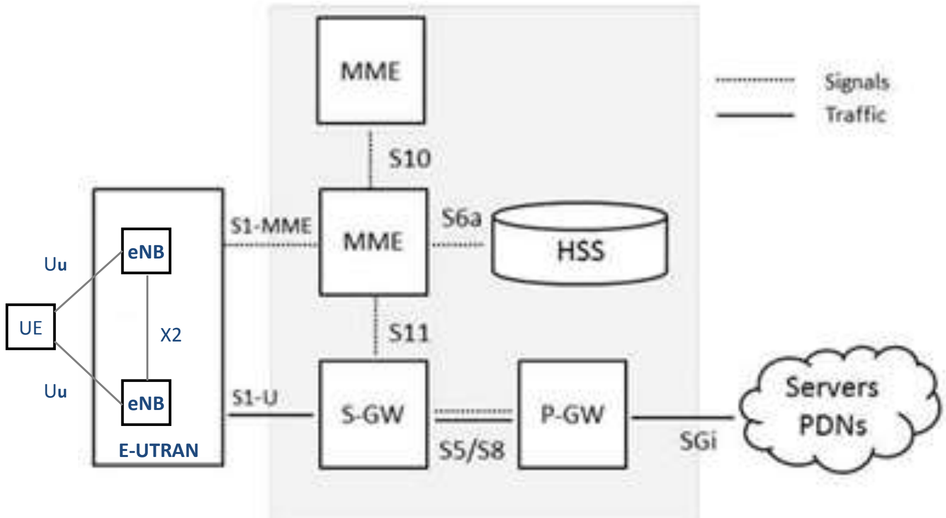
The effort to expand the wireless spectrum especially the licensed one as it is where cellular networks operate is not an immediate solution. That is because not only the process is too slow but also the available one will never solve the problem – it will never be enough [9].

Rolling out new and advanced access technologies like 4G technologies is being used by many service providers like EthioTelecom using Wideband Code Division Multiple Access (W-CDMA), and High Speed Downlink Packet Access (HSDPA) since 2009 have implemented Long Term Evolution (LTE) technology. Building denser cellular networks to increase the connectivity and higher bandwidth availability is another step taken by rather many providers worldwide which incurs a huge amount of money. Due to the observable increased availability of wireless networks, like Wi-Fi, femtocells, and Worldwide Interoperability for Microwave Access (WiMAX), Internet Service Providers (ISPs) are considering to offload their traffic through these complementary networks [9][11][12].

As 4G or LTE is the cellular network whose traffic is going to be offloaded or diverted away to other complementary wireless networks like WLAN and device-to-device (D2D) communications in the following section a brief discussion of this technology is provided.

In the latest releases of 3rd Generation Partnership Program (3GPP) specification documents [13] and [14] the overall architecture of LTE or 4G is described in detail. The architecture is composed of three major components as depicted in Figure 1.3:

- User Equipment (UE) that has a mobile terminal capable of handling all the communication functions and a Universal Integrated Circuit Card (UICC) or, as usually called, Subscriber Identity Module (SIM) or Universal SIM (USIM) that stores user-specific data like phone number, home network identity and security information.
- Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) – is the air interface or the access part of the overall network. It comprises of Evolved Node B or E-UTRAN Node B (eNB) interconnected via x2 interfaces. It handles the radio communications between the mobile and the Evolved Packet Core (EPC). The EPC comprises of the SGW, PGW, and the MME. Each eNB is a base station that controls the UEs in one or more cells. The base station that is being communicating with the UE is known as its serving eNB. The eNB and the UEs communicate via the Uu interface.



Key:

U_u	Interface between UE and eNB.
X2	Interface used by eNBs to communicate to other other.
S1-U	User plane interface between eNB and SGW.
S1-MME	Interface between eNB and MME modules.
S10	Interface between MMEs among each other.
S11	IP interface between MME and SGW.
S6a	Interface between MME and HSS modules.
S5/8	Interface between SGW and PGW.
SGi	Reference point between the PDN GW and the packet data network.

Figure 1.3: Architectural Elements of LTE System. Adapted from [15].

- Evolved Packet Core (EPC) – is the core network which is responsible for the overall control of the UE and establishment of bearers. The main logical components or nodes of the EPC are:
    - Packet Data Network (PDN) Gateway (PGW) – is responsible for IP address allocation for the UE, QoS enforcement and flow-based charging according to the rules set in the Policy and Charging Rules Function (PCRF). It performs downlink user IP packet filtering in to the different QoS-based bearers which is performed based on Traffic Flow Templates (TFTs). It also communicates with the outside world using SGi interface.
    - Serving Gateway (SGW) – serves as the local mobility anchor for the data bearers when the UE moves between eNB. All user traffic passes through this gateway. It

simply routes data between the eNB and the PGW. It holds the information about the bearers when the UE is in the idle state and temporarily buffers downlink data while the Mobility Management Entity (MME) initiates paging of the UE to reestablish the bearers.

  - MME – is the control node that processes the signaling between the UE and the EPC. It performs tasks related to bearer and connection management. The protocols running between the UE and the EPC are known as the Non Access Stratum (NAS) protocols.

- The other components of the EPC or LTE Core Network are the PCRF (not shown in Figure 1.3) and Home Subscriber Server (HSS). The former is responsible for policy control decision making and flow-based charging functionalities and the latter holds user information related to EPS-subscribed QoS profile and access restrictions for roaming. It also holds information about the PDNs to which the user can connect.

The Serving and PDN Gateways also serve another interesting purpose – they act as anchors for other network technologies during interworking.

LTE network is designed to support only packet-switched services, unlike its predecessors. It provides seamless IP connectivity between UE and the PDN. As the name indicates LTE not only encompasses the evolution of the Universal Mobile Telecommunications System (UMTS) radio access through the E-UTRAN but also it is accompanied by an evolution of the non-radio aspects under the System Architecture Evolution (SAE) which includes the Evolved Packet Core (EPC) network. Both LTE and Support Alarm Unit (SAU) constitute the Evolved Packet System (EPS).

The EPS uses the concept of EPS bearers to route IP traffic from the SGW/PGW to the UE. A bearer is an IP packet flow with a defined QoS between the gateways and the UE. The E-UTRAN and EPC together set up and release bearers as required by applications.

The LTE, also known as the E-UTRAN, is completely a new air interface system providing higher data rates (of up to 150Mbps in the downlink and 50Mbps in the uplink), lower latencies (lower than 10 milliseconds), high spectral efficiency, and optimized packet transmissions. It uses Orthogonal Frequency Division Multiple Access (OFDMA) radio access for the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) for the uplink.

1.2.2 WLAN Networks

Wireless Fidelity (Wi-Fi) is a marketing name given for local area wireless technology that allows devices to exchange data or connect to networks. The Wi-Fi Alliance defines Wi-Fi as any WLAN product that is based on the IEEE 802.11 standards.

IEEE 802.11 is a set of medium access control (MAC) and physical layer (PHY) specifications for implementing WLAN computer communication in the 0.9, 2.4, 3.6, 3.7, 5, 45 and 60 GHz frequency bands. They are created and maintained by the IEEE LAN/MAN Standards Committee – formally referred as IEEE 802. The base version of the standard was released in 1997 and has had subsequent amendments. The standard and amendments provide the basis for wireless network products using the Wi-Fi brand.

The 802.11 family consists of a series of half-duplex over-the-air modulation techniques that use the same basic protocol. 802.11-1997 was the first wireless networking standard in the family, but 802.11b was the first widely accepted one, followed by 802.11a, 802.11g, 802.11n and 802.11ac.

Other standards in the family (c, d, e, f, h, j) are service amendments and extensions or corrections to the previous specifications.

Table 1.1 summarizes the notable features of 802.11 PHY layer standards. In the Table, it is very easy to see how the capabilities of the various standards evolved through time.

From practical point of view Wi-Fi technology has many advantages as part one of the wireless technologies. It primarily allows cheaper deployment of LANs. It is also ideal to create connectivity between and extend to devices in outdoor areas and old buildings where cables cannot be used. The other attraction of WLAN is the interface that enables connectivity is built-in most mobile devices like smartphones, tablets and laptops. The Wi-Fi Alliance certifies such devices for backward compatibility and interoperability by giving the logo “Wi-Fi Certified” for such products. This makes any standard Wi-Fi device to work anywhere in the world.

One of the main drawbacks of this technology especially when compared with wired networks like Ethernet is that an intruder does not need a physical link to sniff around the network. The latest Wi-Fi Protected Access v2 (WPA2) encryption is designed to provide a secure passphrase during the establishment of connection with these devices.

Generally, Wi-Fi spectrum assignments and operational limitations are not consistent worldwide. For instance, Australia and Europe allow for an additional two channels beyond those permitted in the US for the 2.4 GHz band, which means 1 – 13 in the former and 1 – 11 in the latter, whereas Japan has even one more on top of that Australia and Europe, viz. 1 - 14.

Table 1.1: 802.11 PHY layer standards [16].

802.11 Protocol	Release Date	Frequency (GHz)	Bandwidth (MHz)	Data Rate (Mbps)	Max. MIMO Channels	Modulation Technique	Approximate Range (meters)	
							Indoor	Outdoor
802.11-1997	Jun 1997	2.4	22	1 – 2*	N/A	DSSS, FHSS	20	100
a	Sep 1999	5, 3.7	20	6 – 54*	N/A	OFDM	35	120,5000*
b	Sep 1999	2.4	22	1 – 11*	N/A	DSSS	35	140
g	Jun 2003	2.4	20	6 – 54*	N/A	OFDM, DSSS	38	140
n	Oct 2009	2.4,5	20, 40	7.2 – 135*	4	OFDM	70	250
ac	Dec 2013	5	20, 40, 80, 160	7.2 – 780*	8	OFDM	35	115
ad	Dec 2012	60	2160	Up to 6,912	N/A	OFDM	N/A	N/A

*Values are taken from the respective standard IEEE documents from which only the minimum and maximum values are considered here.

Channels 1, 6, and 11 are the only group of three non-overlapping channels in North America and the United Kingdom. In Europe and Japan using Channels 1, 5, 9, and 13 for 802.11g and 802.11n is recommended [16].

Many newer consumer devices support the latest 802.11ac standard, which uses the 5 GHz band and is capable of multi-station WLAN throughput of at least 1 gigabit per second. According to a study [16], devices with the 802.11ac specification have become common from 2015 onwards globally.

In our city, Addis Ababa, the use of IEEE 802.11b is staggering to say the least. Among twenty-four thousand Wi-Fi access points being scanned 91% is 802.11b based [17].

1.2.3 Device-to-Device Communications

D2D communication, since its inception in early 1990s, has evolved from simple wireless communication modality to complex protocols being used from simple application like TV remote control to modern applications like vehicle-to-vehicle communications [18]. D2D communication refers to a radio technology that enables nearby devices to communicate each other directly without traversing the data through another network, device or infrastructure. It is designed to use frequency in the Industrial, Scientific and Medical (ISM) spectrum.

There are many D2D technologies that exist independently and integrated with other systems. The latter category includes D2D features in LTE and LTE-A, for instance. The most common D2D technologies that exist separately are Bluetooth, Wi-Fi Direct, Near Field Communication (NFC), and ZigBee [18] [19]. These technologies are described as follows.

Bluetooth: Bluetooth is a wireless technology standard for transferring data over short distances. Theoretically, the distance ranges from 10 – 100 meters. It uses the ISM band ranging from 2402 to 2480 MHz or 2400 to 2483.5 MHz based on the frequency-hopping spread spectrum (FHSS) radio technology with the carrier modulated using Gaussian frequency shift keying (GFSK). It divides transmitted data into packets, and transmits each packet on one of 79 designated Bluetooth channels. Each channel has a bandwidth of 1 MHz. It usually performs 800 hops per second, with Adaptive Frequency-Hopping (AFH) enabled. Bluetooth low energy (BLE) uses 2 MHz spacing, which accommodates 40 channels.

Bluetooth technology has been integrated into many types of business and consumer devices, including cell phones, laptops, automobiles, printers, keyboards, mice, and headsets. This allows users to form ad hoc networks between a wide variety of devices to transfer voice and data. Bluetooth is a low-cost, low-power technology that provides a mechanism for creating small wireless networks on an ad hoc basis, known as piconets. A piconet is composed of two or more Bluetooth devices in close physical proximity that operate on the same channel using the same frequency hopping sequence. An example of a piconet is a Bluetooth-based connection between a cell phone and a headset.

Bluetooth technology was originally conceived by Ericsson in 1994. Ericsson, IBM, Intel, Nokia, and Toshiba formed the Bluetooth Special Interest Group (SIG), a not-for-profit trade association developed to drive development of Bluetooth products and serve as the governing body for Bluetooth specifications. Bluetooth is standardized within the IEEE 802.15 Working Group for Wireless Personal Area Networks that formed in early 1999 as IEEE 802.15.1-2002.3.

The range of Bluetooth basic rate/enhanced data rate (BR/EDR) devices is characterized by three classes that define power management. Table 1.2 summarizes the classes, including their power levels in mill watts (mW) and decibels referenced to one mill watt (dBm), and their operating ranges in meters (m). Most small, battery powered devices are Class 2, while Class 1 devices are typically universal serial bus (USB) adapters for desktops and laptops, as well as access points and other mains powered devices.

Table 1.2: Bluetooth Device Classes Operating Specification [18] [19].

Type	**Power**	**Max. Power Level**	**Designed Operating Range**	**Sample Devices**
Class 1	High	100 mW (20 dBm)	Up to 100 meters	USB adapters, access points
Class 2	Medium	2.5 mW (4 dBm)	Up to 10 meters	Mobile devices, Bluetooth adapters, smart card readers
Class 3	Low	1 mW (0 dBm)	Up to 1 meter	Bluetooth adapters

To allow Bluetooth devices to find and establish communication with each other, discoverable and connectable modes are specified. A device in discoverable mode periodically monitors an inquiry scan physical channel (based on a specific set of frequencies) and responds to an inquiry on that channel with its device address, local clock (counter) value, and other characteristics needed to page and subsequently connect to it. A device in connectable mode periodically monitors its page scan physical channel and responds to a page on that channel to initiate a network connection. The frequencies associated with the page scan physical channel for a device are based on its Bluetooth device address. Therefore, knowing a device's address and local clock is important for paging and subsequently connecting to the device.

Wi-Fi Direct: Wi-Fi direct is an 802.11 technology configured to work in ad-hoc mode. It enables devices to connect with each other without a wireless AP. It is a peer-to-peer connection where two devices configured in ad-hoc mode establish a direct Wi-Fi connection. During initial connection establishment the devices negotiate who should act as a coordinator. Devices can make a one-to-one connection or a group of several devices can connect simultaneously [20].

Wi-Fi Direct essentially embeds a software access point, Soft AP, into any device that must support Wi-Fi Direct communication. The soft AP provides a version of Wi-Fi Protected Setup with its push-button or PIN-based setup.

When a device enters the range of the Wi-Fi Direct host, it can connect to it, and then gather setup information using a Protected Setup-style transfer. Soft APs can be as simple or as complex as the role requires. A digital picture frame might provide only the most basic services needed to allow digital cameras to connect and upload images. A smart phone that allows data tethering might run a more complex soft AP that adds the ability to bridge to the Internet. The standard also includes WPA2 security and features to control access within corporate networks. Wi-Fi Direct-certified devices can connect one-to-one or one-to-many and not all connected products need to be Wi-Fi Direct-certified. One Wi-Fi Direct enabled device can connect to legacy Wi-Fi certified devices.

The Wi-Fi Direct certification program is developed and administered by the Wi-Fi Alliance, the industry group that owns the "Wi-Fi" trademark [21].

Wi-Fi Direct device connections can happen anywhere, anytime - even when there is not access to a Wi-Fi network. Wi-Fi Direct devices emit a signal to other devices in the area, letting them know a connection can be made. Users can view available devices and request a connection, or may receive an invitation to connect to another device. When two or more Wi-Fi Direct-certified devices connect directly, they form a Wi-Fi Direct Group using Wi-Fi Protected Setup (WPS) and the latest Wi-Fi security.

A common example is the Wi-Fi Protected Setup system included in most access points built since 2007 when the standard was introduced. WPS allows access points to be set up simply by entering a PIN or other identification into a connection screen, or in some cases, simply by pressing a button [22]. The Protected Setup system uses this information to send data to a computer, handling the information needed to complete the network setup and connect to the Internet. From the user's point of view, a single click replaces the multi-step, jargon-filled setup experience formerly required.

While the Protected Setup model works as intended, it was intended only to simplify the connection between the access point and the devices that would make use of its services, primarily accessing the Internet. It provides little help *within* a network - finding and setting up printer access from a computer for instance. To address those roles, a number of different protocols have developed, including Universal Plug and Play (UPnP), Devices Profile for Web Services (DPWS), and Zero Configuration Networking (ZeroConf). These protocols allow devices to seek out other devices within the network, query their capabilities, and provide some level of automatic setup.

Many devices are being shipped that are equipped with Wi-Fi direct technology. Laptops, mobile devices, game consoles, and televisions, portable media players, printers, cameras, scanners other common home and office devices. There are a wide range of applications of Wi-Fi direct like sharing files and printing documents, for instance.

Wi-Fi Direct-certified devices can be used for all kinds of applications - to share content, synch data, socialize, play games, play audio and video, and more - all the things you do with your Wi-Fi devices today, only easier and without worrying about finding an internet connection. Wi-Fi Direct-certified devices can form connections with nearly all the Wi-Fi Certified devices around us.

Wi-Fi Direct can provide a wireless connection to peripherals. Wireless mice, keyboards, remote controls, headsets, speakers, displays and many other functions can be implemented with Wi-Fi Direct. This has begun with Wi-Fi mouse products, and Wi-Fi Direct remote controls that were shipping circa November 2012 [22].

Near Field Communication (NFC): NFC is a short range wireless technology that can carry data between nearby devices in a secure manner. The connection establishment and data flow mechanisms are simplified when compared with other similar technologies. It can be used to make contactless transactions like Point of Sale (PoS) systems, to access digital content and to interconnect electronic devices without contact [23].

NFC uses inductive-coupling technique at 13.56 MHz frequency which is publicly usable portion of the radio spectrum.

NFC technology has evolved from a combination of contactless identification and interconnection technologies including Radio Frequency IDentification (RFID) and it allows connectivity to be achieved very easily over distances of a few centimeters, about 4 or 5 centimeters. Simply by bringing two electronic devices close together they are able to communicate and this greatly simplifies the issues of identification and security, making it far easier to exchange information. In this way, it is anticipated that NFC technology will allow the complex set-up procedures required for some longer range technologies to be avoided.

NFC has a variety of applications. These include mobile phones, PDAs, personal computers, check-out cash registers or "point-of-sale" equipment, vending machines, parking meters, and ATMs.

A further application that was proposed was that NFC connections could be used to configure the connection between two wireless devices. All that was required to configure them to operate together wirelessly would be to bring them together to effect the NFC "connection". This would initiate the set-up procedure; communication could take place over the NFC interface to configure the longer range wireless device such as Bluetooth, 802.11 or other relevant standard. Once set up, the two devices could operate over the longer range allowed by the second communication system.

Zigbee: ZigBee is a wireless communication system primarily designed for use in remote control and sensor applications where harsh radio environments and isolated locations are expected [23].

It is based on the 802.15.4 IEEE standard which defines the physical and MAC layers. It also defines the application and security layer specifications enabling interoperability between products from different manufacturers. Without any relay it can cover up to 70 meters of distance between the two communicating devices.

It operates in one of the three free bands at 2.4 GHz, 915 MHz for North America and 868 MHz for Europe. At 2.4 GHz, there are sixteen different channels available, and the maximum data rate is 250 kbps. In the 915 MHz band, there are ten channels and supporting a maximum data rate of 40 kbps, while at 868 MHz frequency, there is only one channel supporting maximum data rate of 20 kbps.

The direct sequence spread spectrum (DSSS) modulation technique is used in all cases. But, for the 868 and 915 MHz frequencies the actual form of modulation used is binary phase shift keying (BPSK) and for the 2.4 GHz, offset quadrature phase shift keying (O-QPSK) is employed.

As the technology is designed for harsh environments where interference is expected to be high the specification supports some features to guarantee reliable operations. These include a quality assessment, receiver energy detection and clear channel assessment. Carrier Sense Multiple Access (CSMA) techniques are used to determine when to transmit.

The data is transferred in packets of maximum size 128 bytes which allow a maximum payload of 104 bytes.

The standard supports 64-bit MAC addresses as well as 16-bit short addresses. The MAC addresses are used to uniquely identify every device. After a network is set up, the short addresses are used enabling over 65000 nodes to be supported.

The three different network topologies supported by ZigBee are the star, mesh and cluster tree or hybrid. As each of them has their own strengths it gives the opportunity of using then for different situations.

The major applications of ZigBee are for control and monitoring applications where relatively low levels of data throughput are needed, and with the possibility of remote, battery powered sensors where low power consumption is a key requirement. This includes but not limited to sensors, lighting controls, and security.

The above four wireless technologies are the most common candidates for D2D communications though there are others too.

D2D communication in cellular networks: D2D communication has been planned by ITU to be part of 5G [22] [25]. And, it is already specified by 3GPP to be used in LTE networks in its Rel-

12 [26] – [30]. The main focus is being on Public Safety applications and proximity-based services. This drives further development and standardization of the technologies especially Wi-Fi direct.

D2D communication in cellular networks is defined as direct communication between two mobile users without traversing the BS or core network. D2D communication is generally non-transparent to the cellular network and it can occur on the cellular spectrum (i.e., inband) or unlicensed spectrum (i.e., outband).

In a traditional cellular network, all communications must go through the BS even if both communicating parties are in range for D2D communication. This architecture suits the conventional low data rate mobile services such as voice call and text message in which users are not usually close enough for direct communication. However, mobile users in today's cellular networks use high data rate services (e.g., video sharing, gaming, proximity-aware social networking) in which they could potentially be in range for direct communications (i.e., D2D). Hence, D2D communications in such scenarios can highly increase the spectral efficiency of the network. Nevertheless, the advantages of D2D communications are not only limited to enhance spectral efficiency but also they can potentially improve throughput, energy efficiency, delay, and fairness.

In principle, exploiting direct communication between nearby mobile devices will improve spectrum utilization, overall throughput, and energy efficiency, while enabling new peer-to-peer and location-based applications and services. D2D-enabled LTE devices have the potential to become competitive for fallback public safety networks that must function when cellular networks are not available or fail.

Proximal communications represent an important set of use cases, including machine type communications (MTC) (massive as well as critical), national security and public safety situations, and vehicle-to-vehicle and intelligent transportation system applications, and also support local social networking.

The research community has already started to explore new avenues and further developments for D2D in the context of fifth-generation (5G) networks [28] - [30].

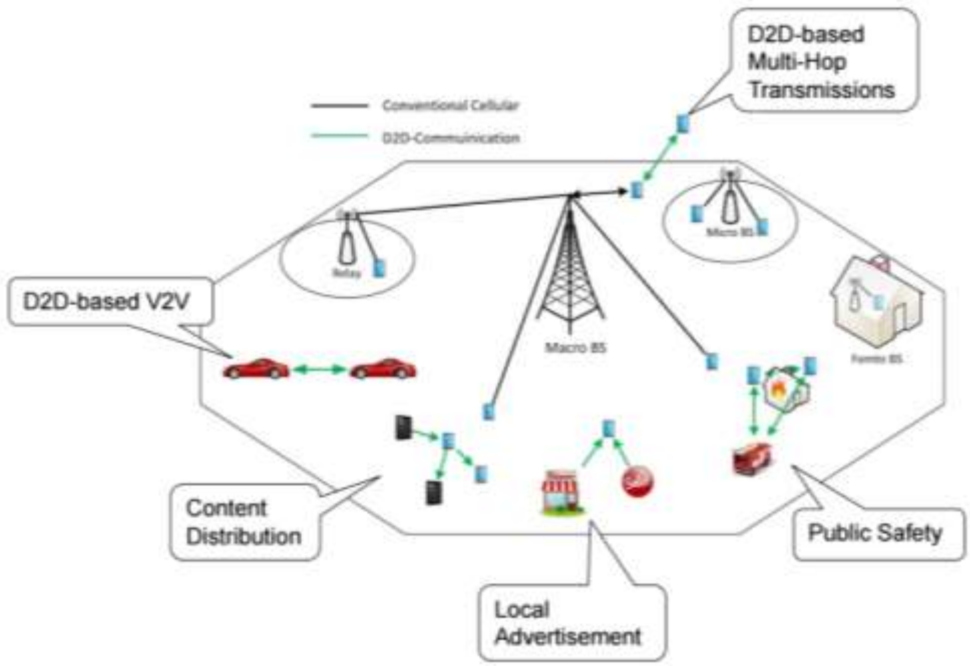


Figure 1.4: Application Scenarios of D2D Communications in Urban area setting. Adapted from [27].

As shown in Figure 1.4, cellular and D2D communications make up the wide range of applications and services for urban area residents. To make connectivity available ubiquitously it is oftentimes desired to make them interwork seamlessly. One typical purpose of seamless interworking is mobile traffic offloading.

1.3 The Status of Cellular Networks

OpenSignal estimated, in their latest report entitled Global State of Mobile Networks [31], from 95 countries sampled 93 of them had 3G or better cellular signal availability for above 50% of the time and the majority of them had availability of greater than 75% of the time. The same report found out that the global as well as country-wise speed, measured in terms of Mbps – Megabits per second, is increasing due to mainly the rollout of LTE cellular networks in many countries of the world. It also confirmed that 3G and 4G cellular networks have been and are being used for both data and voice.

Nevertheless, it is reported that many customers complained about the cellular network congestion or poor performance throughout the globe like in USA [32], India [33] [34], some European countries [35], and UK [36].The places where cellular network outage occurred were highly populated urban areas like city centers, stadiums and down towns.

So, why this paradox – while cellular network performance is highly increasing, like from 3G to 4G, and customers are yet complaining about cellular network outage? The reasons for that are multifaceted.

First, mobile hardware capability is getting improved continuously as forecasted by Gordon Moore back in 1965. It is not only their computing power but also storage and other features that are improving every day. Surprisingly, their prices are getting down making them affordable for low income individuals [37].

Second, this brings higher mobile subscription rate throughout the world as ITU reported in [38] and [39]. That is due to the fact that mobile devices has become affordable and ISPs decreased mobile subscription rates. In the report, the global mobile subscription rate has already overtaken the fixed telephone subscriptions. Since then (2012) it has grown above 20% per year and reached 4.3 billion at the end of 2017. It is also known that many of us possess multiple SIMs. That is supported by OpenSignal report [31]. It is found that though the use of multi SIM devices is rare, only 2% of the sample, in northern America it is huge in the third world countries like Nigeria, Bangladesh, Philippines, and India reaching up to 40% of the sampled users [31].

Third, the combined effect of improved computing capability of mobile devices with increasing number of mobile subscription on top of affordable price for broadband connection brings the development of mobile apps and service that consume ever increasing bandwidth. Ericson in its report [4] predicted that by 2022 the mobile data traffic in Central and Eastern Europe and Middle East and Africa (CEMA) will increase by 12 times. The same report predicted that the global total mobile data traffic is expected to rise as a compound annual growth rate (CAGR) of around 45%. Moreover, other similar reports support Ericson's prediction. Three different organizations reported that the traffic demand on cellular networks is increasing at alarming rate[1] – [4] [6].

Fourth, this is expected to be the beginning of even a much higher demand for bandwidth as these mobile devices bring services that require anytime anywhere anything connectivity. The demand for traditional communication of voice and text is being shifted to explosive data flows. Moreover, online gaming and streaming video access via the Internet is becoming a de-facto for the increasing population which is using tablets and smartphones [2][4].

So, what are the possible solutions operators wish to have, researchers contemplate and users are anticipating? Offloading cellular traffic towards other complementary wireless networks like Wi-Fi, WiMAX, microcells (like femtocells and Pico cells) and opportunistic device to device communications is one of them. Using emerging technologies like cognitive radio, network function virtualization and proactive caching are among the other category.

1.4 The Concept of Mobile Traffic Offloading

Mobile traffic offloading is the use of a complementary wireless technology to transfer data originally targeted to flow through the cellular network. D2D technologies, Wi-Fi, WiMAX, femtocell and Pico-cells are the most common complementary wireless technologies considered for this purpose [18] [19] [46].

Wi-Fi is being chosen as an immediate and temporary solution for this particular purpose by many ISPs worldwide [46]. Among the various reasons that makes Wi-Fi as the choice for offloading includes but not limited to:

- It is the last-100-meter solution for infrastructure-based networks and they are becoming ubiquitous due to its affordability, ease of configurability and use.
- WLANs are available at work, hotels, cafeterias, and even at home. All kind of organizations and homes are having their own Wi-Fi. Wi-Fi availability especially in urban areas is expected to grow further at alarming rate which opens a great opportunity for cheaper offloading of cellular traffic [47].
- It is well-known that cellular connectivity degrades within buildings and some specific areas due to fading, shadowing and other reasons. By using wireless routers in cafeterias and hotels, for instance, it is possible to extend the coverage and accessibility of mobile services and apps.
- Wi-Fi is one of the technologies embedded inside almost all smartphones, tablets and laptops. This makes offloading through Wi-Fi networks to be economically and technically feasible and preferable solution.
- Beside all these, the nature of the two wireless technologies, cellular and Wi-Fi, is another attraction for many telecom service providers. On the one hand cellular networks are better for long range communications but with high cost, lesser bandwidth and higher latency whereas Wi-Fi networks are better suited for short range communications with better bandwidth, much less cost and lower latency. Moreover, the spectrum used by cellular

networks is very expensive and strictly regulated unlike Wi-Fi networks where it uses the unlicensed spectrum giving the opportunity to use the technology with less total cost of ownership.

- Last but not least, for ISPs using Wi-Fi APs is more appealing since designing, operating, deploying, maintaining and managing such networks is very much less costly than doing the same for cellular base stations. More specifically, it shifts the central and costly administration away to decentralized and distributed management style.

And therefore, studying the nuts and bolts of mobile data offloading of cellular traffic over Wi-Fi networks has become a hot research issue currently [34][37][48] - [51][108][187].

The wide spread research efforts in mobile traffic offloading techniques by the industry and academia is summarized in Figure 1.5 as surveyed in [50].

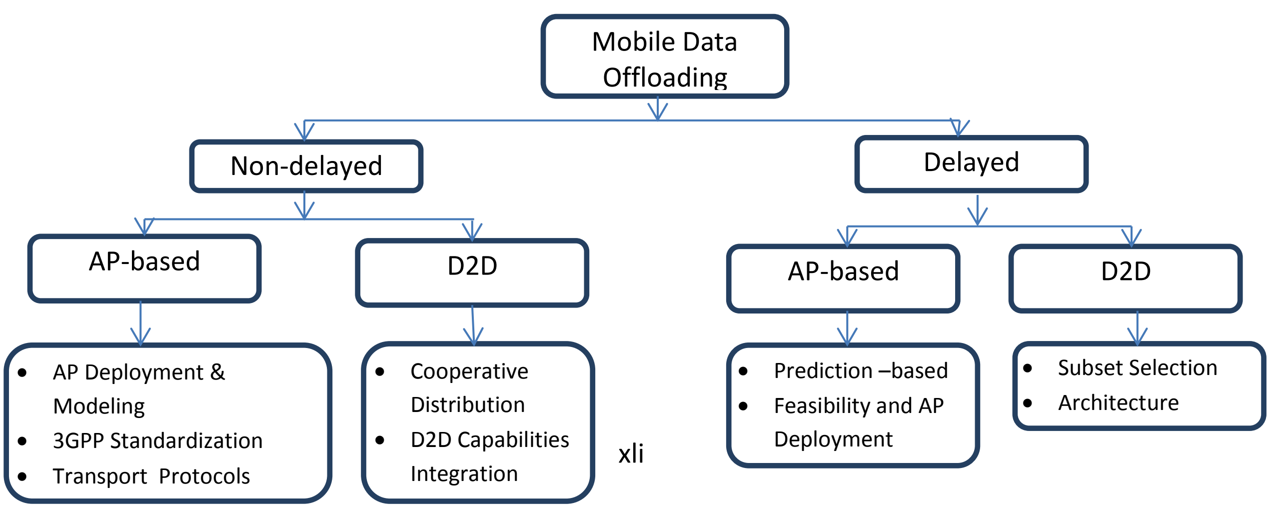


Figure 1.5: Classification of mobile traffic offloading based on current research efforts [50][124].

Many research papers pointed out that there are many research challenges that should be addressed to make mobile data offloading a reality. The major research challenges are described in detail in [5][7][49]-[52][76] and [81]. The first and foremost is lack of feature-rich internetworking architecture between the wireless technologies involved. An architecture that considers the degree of cohesion between the two as either loosely or tightly coupled is considered as one direction. Moreover, the choice of an offloading technique as either D2D or AP-based is another challenge. The choice of offloading infrastructure can be combined with the handled traffic nature based on the QoS requirements of the initiated traffic which could be either delayed or non-delayed is yet another challenge.

As pointed out in the aforementioned research challenges what is needed is a feature-rich mobile traffic offloading system that interworks as much wireless technologies as possible, provides multiple possibilities of traffic forwarding capabilities, and handling both delay sensitive and delay tolerant traffic types. This need to be investigated in further details identifying the major contributors together with the main features they propose or introduce. All these are presented to the minutiae in the next chapter as part of the comprehensive review of related works.

1.5 Other Alternatives to Mobile Traffic Offloading

To alleviate the problem of cellular network capacity there are many alternative solutions proposed to offload traffic away from it. Adding new small-sized base stations like femtocells, multicasting or broadcasting data inside the cell, integrating cellular networks with cognitive radio systems, and proactive pushing of popular content to devices are outstanding mechanisms. For the sake of brevity these alternative solutions are summarized in Table 1.3 showing their peculiar features, merits and demerits.

As it can be seen from the above table there are many alternatives for mobile traffic offloading. They are either not matured lacking the confidence of operators to invest on them or too expensive that the financial return is not guaranteed in the long term.

New macro-cell or small-cell deployment, technology upgrading, cognitive radio integration, and proactive caching are among the common suggestions [46][53]. New macro cell site establishment is taken by operators as the-last-resort since its capital expenditure (CAPEX) and operational expenditure (OPEX) are as high as $ 60,000 per year per macro cell [54]. Alternatively, the use of small cells and femtocells is rather a cheaper one [46]. However, missing the majority of the cellular bandwidth users of the macro cell and interference management issues are new challenges operators face in this arena [8] [46] [87].

Utilizing new spectrum is also frustrating as it is licensed, very much expensive and not an immediate solution [7] [46] [87]. Incorporating software defined and cognitive radio features in the core network is still expensive and entails some kind of modification to the existing infrastructure [55]. Proactive caching is storing popular content near the mobile devices that are expected to request. The performance of the prediction techniques is the crucial factor in the utilization and realization of this choice [5][30].

Table 1.3: Summary of alternative solutions to mobile traffic offloading adapted from [5][30][46].

№	Alternative Solution	Peculiar Feature	Pros	Cons
1	Erecting additional macro-cells	• Establishing new macro-cell sites where coverage is poor	• Covering large geographic area totally eradicates the problem	• Too expensive for operators
2	Adding small-sized cells	• Building more small-sized BTSs to scale up the RAN • Most popular by ISPs	• Reduces the size of macro-cells • Increases available bandwidth	• Building additional BTSs • CAPEX and OPEX increases without

			• Minimizes transmission power of BTSs	addressing the majority of mobile consumers
3	Multicasting/broadcasting	• Sharing a radio link by many nearby users within a cell for same data requests.	• High resource saving	• Require modifications of the cellular network. • Variability in data rates for users at different proximity to the transmitter. • Diminishes overall performance.
4	Integration with cognitive radio	• Exploits unused spectrum opportunistically.	• Enhance spectrum efficiency and network capacity	• Long-term solution as not yet matured.
5	Proactive caching	• Minimizes congestion at the RAN by localizing popular data nearer to users.	• Reduces traffic, delay, and load to the server. • Used together with next request prediction and pre-fetching techniques.	• Not yet available. • Deserves maturity.

And yet, another immediate and cheaper alternative is mobile traffic offloading through complementary networks like WLANs and D2D technologies, among others. It is the technique of using public or private Wi-Fi AP-based or small distance technologies like Bluetooth to direct traffic initially intended for the cellular infrastructure as much as and whenever possible.

1.6 Statement of the Problem

Currently and for the foreseeable future, there is an increasing pattern of computing capability of mobile devices, their usage and ownership, and mobile network connectivity and subscription. The abundance of highly capable mobile devices and betterment of mobile connectivity facilitated the development and usage of bandwidth hungry apps and services. This is exacerbating the performance of the already-saturated cellular network infrastructure in urban areas. Even there are complaints in many cities around the world regarding cellular network performance degradation and service interruption [17] [18].

It seems the right away solution for such demands to go for a new macro cell in the areas where the demand is aroused. However, erecting a new macro cell is expensive for operators in terms of OPEX and CAPEX as explained in [5][30][46] and [54]. Moreover, as analyzed in [54], around 80% of the total revenue of telecom services is generated from only 4% of mobile subscribers which strategically discourages service providers from establishing new macro cells.

So, operators look for other alternative solutions to fulfill the demand. As summarized in table 1.3 each of the alternative solutions have associated risks prohibiting operators to use them.

And yet, according to [5] [7] [46] [51] and [54], mobile traffic offloading is a better alternative based on many practical reasons. As it is well summarized in Figure 2.28 there are many efforts by different stakeholders to device an interworking architecture for mobile traffic offloading. However, none brought a single architecture equipped with features aspired by both operators and users.

The challenges to realize mobile traffic offloading are well described in [5] [27] and [28]. The outstanding one is lack of interworking architecture that has as many features as possible enabling it to divert as much cellular traffic as possible towards complementary networks.

Therefore, in this research work a feature-rich internetworking architecture is proposed for offloading cellular traffic over Wi-Fi networks and D2D opportunistic encounters. The proposed architecture has multiple data flow paths where there are multiple linking between the networks enabling different paths for delay tolerant and delay sensitive traffic types. It also has traffic steering algorithm that monitors instantaneous network conditions in order to maximize the efficiency of offloading the initiated traffic away from cellular systems seamlessly.

1.7 Objectives

1.7.1 General Objective

The aim of this research work is to propose and evaluate a feature-rich interworking architecture for cellular traffic offloading over WLAN and D2D networks.

1.7.2 Specific Objectives

More specifically this research undertaking is aimed at:

- Studying the availability, capacity and performance of cellular and Wi-Fi networks.
- Performing a comprehensive review of literatures and related works to identify the state-of-the-art, concepts and theories behind mobile traffic offloading, interworking architectures and techniques.
- Designing and propose a feature-rich (QoS-aware and with multiple coupling techniques) interworking architecture exploiting many possible features of cellular, WLAN and D2D systems.
- Implementing and evaluate the proposed interworking architecture and traffic steering algorithm using the NS3 network simulation tool.

1.8 Research Methodology

As a research methodology, this work applied and utilized a myriad of approaches based on their appropriateness to the specific undertaking. The major methods used include:

- Data collection and analysis of Wi-Fi APs and cellular networks to have the glimpse of their availability and capability.
- Performing rigorous literature review and comprehensive review of related works – here around 200 literatures and related works are reviewed that cover the knowledge area from its genesis to the-state-of-the-art. This spans from basic concepts and theories in interworking architectures and mobile traffic offloading to the recent research, standardization and production and marketing efforts.

- Extending the effort exerted on reviewing related works by identifying and enumerating grey areas.
- Designing a feature-rich interworking architecture and corresponding traffic steering algorithm for the purpose of mobile traffic offloading has been done.
- Implementation of the proposed interworking architecture and the traffic offloading algorithm has been done using the event-driven discrete network simulator known as NS3 – Network Simulator 3.
- Using discrete-event technique performs simulation study of the overall offloading system by using practical oriented and simulation parameters.
- Finally, performance evaluation has been made to showcase the potential of the proposed interworking architecture for mobile traffic offloading purposes.

1.9 Scope and Contributions of the Research

Expanding cellular networks is not an easy investment that could be done right away for a particular traffic demand. The various efforts in the review of related works section showed that it is technically feasible and economically profitable to use existing Wi-Fi APs and D2D technologies to offload as much cellular network traffic as possible over these unlicensed and cheaper networks especially in urban areas.

In light of this, the contributions of this research can be enumerated as follows:

- For future endeavors in the area, a comprehensive survey of mobile traffic offloading interworking architectures and techniques has been made showing the state-of-the-art from various possible dimensions including outstanding open challenges.

- A thorough investigation of both Wi-Fi APs and cellular networks has been done pioneering and showcasing their availability, capacity and performance in the city of Addis Ababa. This work can be extended to analyze the existing wireless infrastructure of the city for mobile traffic offloading, among others.
- A conceptual interworking architecture that possesses many features that can be further implemented and evaluated using specific technologies and tools.
- Implementations and modifications of the NS3 simulation tool incorporating new features and functionalities. For example, the introduction of the GetPropagationLossModel() function in the YansWifiChannel class and counting the packets dropped by WifiMac::MacTxDropEvent by designing a new function named WifiPktDropCount().
- In the proposed interworking architecture, the linking between the Wi-Fi AP and the eNB is encouraging way of making the cellular network control a WLAN and instantaneous network conditions thereof.
- The use of two independent variables, at different stages, to determine through which path a data packet better be forwarded is another insight pursued. The two variables used are the instantaneous received signal strength of the WLAN and D2D paths and the QoS requirement of packets. Figure 3.7 on page 106 and Figure 4.13 on page 131 describes in detail how the instantaneous received signal strength is calculated and implemented whereas Table 4.2 and Figures 4.9 to 4.12 details the descriptions and implementations of the QoS aspects of both the cellular, WLAN and D2D subsystems.
- The integration of three different technologies in terms of geographical coverage as WAN, LAN and PAN representing the LTE, Wi-Fi APs and the D2D, respectively.

- Possibility of total avoidance (> 90%) using the eNB to transfer a delay tolerant traffic.

As a limitation, at the implementation level, this research work is limited to or considers only:

- The LTE cellular network,
- 802.11b–based WLANs,
- 802.11b–based D2D communication,
- Data traffic, and
- A network of only two user nodes, dubbed as user equipment or UE, where one initiates data traffic while the other sinks it.

1.10 Organization of the Rest of the Dissertation

The rest of this dissertation report is organized in to five chapters. The next one, chapter two, gives a very wide and deep revision of related works in the area of mobile traffic offloading architectures, frameworks and systems. Chapter three presents the proposed feature-rich internetworking architecture for the purpose of mobile traffic offloading. The implementation of the proposed architecture is presented in chapter four which is followed by chapter five where the performance evaluation of the simulated internetworking offloading system is made using key performance metrics. Chapter 6 culminates the overall work by providing a summarized conclusion and pointing out future directions.

CHAPTER TWO

INTERWORKING ARCHITECTURES FOR MOBILE TRAFFIC OFFLOADING: STATE-OF-THE-ART

2.1 Introduction

As described in the previous chapter due to the advent of mobile devices like laptops, tablets and smartphones, the traffic demand on cellular networks is increasing at alarming rate [1] - [4]. This is expected to be the beginning of even a much higher demand for bandwidth as these mobile devices bring services that require anytime anywhere connectivity. The demand for traditional communication of voice and text is being shifted to explosive data flows. Moreover, online gaming and streaming video access via the Internet is becoming a de-facto for the increasing population which is using tablets and smartphones [2][4]. It is even observed that customers complained at some operators on the poor mobile network performances and congestions [17][18].

To alleviate this problem, there are many efforts being exerted by researchers, product manufacturers, and standardization bodies worldwide. Again as pointed out previously in subsections 1.2 and 1.3 and Table 1.3 among the various alternatives mobile traffic offloading is cheaper in terms of total cost of ownership, simplicity of introducing in terms of the modifications needed on user devices and operator infrastructure, and technically feasible due to its ISM usage, especially for Wi-Fi and D2D cases. The performance improvements harvested are also very much attractive as indicated in [13] - [16] [19] [20].

To realize mobile traffic offloading effectively, one of the open issues that need to be worked out is to come up with an architecture that exploits the possible technologies and techniques – an feature-rich interworking architecture [5]. This primarily relieves operators from the repeated effort and cost of introducing offloading solutions into their network infrastructure.

There are various architectures and techniques proposed to offload cellular traffic which consider different kinds of radio access technologies and traffic natures. These efforts can be put in to three major categories based on *level of integration*[2]as loosely, tightly and, recently, very tightly coupled; *technologies involved* as, especially in the unlicensed spectrum, D2D technologies and Wi-Fi; and finally, the *nature of traffic considered* during the offloading process as delay tolerant and delay sensitive. Here, in the category of technologies involved only wireless LAN and Personal Area Network (PAN) technologies using the ISM band are considered.

The efforts of mobile traffic offloading are technically diverse, edgy to realize, and unrecognizably spread out. For instance, there is much research effort by academia and industry to conceptualize a reliable and cheap solution which is followed by manufacturers' eagerness to market that product, usually, paralleled with the standardization and conformance efforts by professional associations and standardization bodies. Therefore, to have a better picture of what has been done in the area of interworking offloading architectures and techniques, this survey work comprehensively investigates and summarizes these efforts proposed by academia and the industry, standardized by standardization bodies, and advertised by product manufacturers and marketers.

The main contributions of this survey work are multifold:

- Generally:
  - To classify and summarize the various efforts in the area starting from its inception to the state-of-the-art,

[2] In this work, *coupling* and *integration* in the context of interworking cellular and complementary wireless networks is used interchangeably.

    - To suggest a fully integrated hybrid internetworking architecture for future implementation, and
    - To show some outstanding current open issues and realization challenges.
- Specifically
    - For the research community
        - To show what has been and being done in the various research dimensions, identifying the trend in the area,
        - To point out what the-state-of-the-art is in the area, and which directions are open and worth pursuing,
    - For standardization bodies
        - To show how much their work is being taken by manufacturers,
    - For product manufacturers
        - To show what kind of product is needed by the industry, and
        - To show which trend is worth supporting and need to plan ahead of time.

The rest of the chapter is organized as follows. Section 2.2 presents the classification of the main features a feature-rich interworking architecture is expected to have. Section 2.3 summarizes the various efforts and contributions made by stakeholders for the realization of such architecture. That is classified under three categories of efforts made by researchers, standardization bodies and product manufacturers. Finally, section 2.4 shows some outstanding realization issues and challenges and section 2.5 concludes the chapter.

2.2 Major Classifications

From a typical mobile offloading system users expect cheap and ubiquitous connectivity alternatives. And, operators wish to have a solution simpler to integrate with existing infrastructure, with least possible modifications especially in the core network, easier to manage and maintain, and better revenue generation from their investments and services. To have a complete solution of mobile traffic offloading that can satisfy the main beneficiaries, users and operators, there are a myriad of efforts around the globe. Based on these and other sources, in this work, the major features expected of a fully integrated internetworking system are categorized in to three main classes. They are, 1) **Level of Integration** - based on the point of linking between the different networks, 2) **Technologies Involved** - based on the type of wireless technologies integrated together, and 3) **Nature of Traffic** -based on the kind of traffic the system supports. The detail of this classification is visualized in Figure 2.1 and discussed in the following subsections.

Figure 2.1: Classification of the Features Common in Interworking Architectures.

2.3.1 Level of Integration

Here the level of integration of the complementary wireless technologies with the cellular network is discussed. Particularly, this work considers the integration of the three technologies – cellular, Wi-Fi and D2D. In this regard, an attempt has been made to consider all technically plausible possibilities.

It is possible to make integration between cellular and Wi-Fi, between cellular and D2D, and among the three technologies. The level of integration between cellular and WLAN[3] is usually classified into three types, based on the coupling point of the two networks. For example, for LTE networks, the integration at the access network, the core network or at the eNB defines the loose, tight and very tight coupling, respectively [21]. Figure 2.2 depicts, in detail, the LTE network architecture and where this linking is made.

Loose Coupling: apparently, the two networks, cellular 3G/4G and WLAN, are completely separate and independent [21][22]. They have separate logical addresses for their interfaces. All connections and communications are managed by each network without the knowledge and consideration of the other. In 3GPP 3G and 4G cellular networks the loose coupling point is before reaching the gateways, Serving GPRS Support Node (SGSN) for 3G and SGW for 4G. The WLAN uses interconnecting equipment like WLAN Access Gateway (WAG) for 3G and Trusted WLAN Access Gateway (TWAG) for 4G, to support authentication and billing for roaming services. Both gateways use the Authentication, Authorization and Accounting (AAA) server to connect with the Internet without any direct link to the 3G/4G system. Hence, the WLAN traffic goes directly, or via other IP networks, to the Internet without passing through the core network. However, the 3G/4G subscriber databases are extensively used to provide mobile users with direct data access to the Internet. At the gateways, Mobile IP (MIP) is used for mobility management across the networks and to communicate with the 3G/4G AAA server [21][23].

The good side of this approach is that there is no need for any modification of the cellular infrastructure. It can be simply integrated by developing or modifying the operating software platform that can handle the authentication, billing and mobility issues [21]. One typical challenge,

[3] The terms Wi-Fi, to mean Wi-Fi AP, and WLAN are used interchangeably in the context of interworking architectures since in WLANs the contact point for cellular networks is the Wi-Fi AP.

here, is the efficient way of managing the mobility of mobile users moving between the networks and the higher handover delays [22]. For this, there are many suggestions and efforts [24][25][26][27].

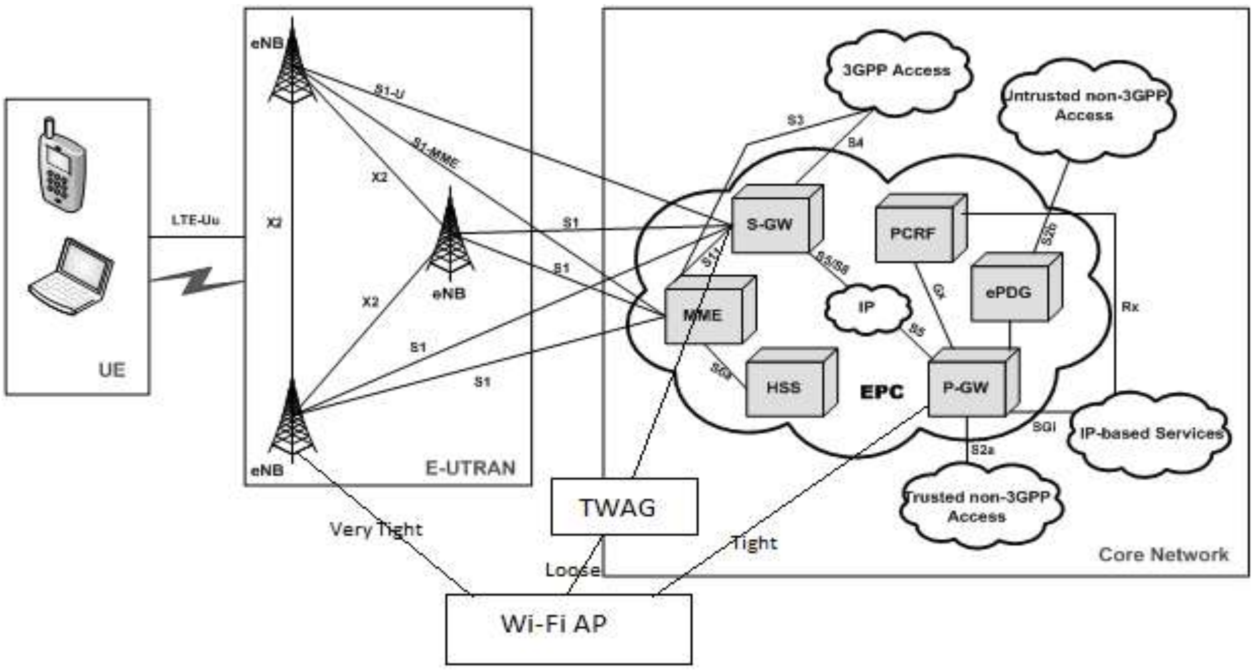


Figure 2.2: LTE Network Architecture [21].

UE – User Equipment
eNB – Evolved Node B
E-UTRAN – Evolved UMTS Terrestrial Radio Access Network
MME – Mobility Management Entity
S-GW – Serving Gateway
P-GW – PDN (Packet Data Network) Gateway

HSS – Home Subscriber Server
EPC – Evolved Packet Core
PCRF – Policy and Charging Rules Function
ePDG – Evolved Packet Data Gateway
3GPP – 3rd Generation Partnership Program

Tight Coupling: The two networks are integrated in such a way that the WLAN network appears to be a regular radio access network to the cellular network [28]. The Wi-Fi AP is directly connected to the core network at layer 2 of the IP stack, specifically, either at the core network level (SGSN - Serving GPRS Support Node or GGSN - Gateway GPRS Support Node for 3G and S-GW or P-GW for 4G) or access level (e.g. RNC – Radio Network Controller for 3G and S-GW for 4G). This enables the cellular network to treat the WLAN network as its own RAN. During communication, there is only one IP address and hence one connection for the two interfaces – Wi-Fi and cellular.

3GPP classified such networks as trusted and untrusted non-3GPP accesses [29]. A trusted non-3GPP network can be directly connected to the core network whereas the untrusted one requires an evolved packet data gateway (ePDG) as intermediate equipment that provides security mechanisms.

For tight coupling of the two networks one need to modify their services, protocols, and interfaces in the existing infrastructure which is challenging for operators to deploy such solutions. On the other hand, mobility management is not an issue here as it is handled automatically by the core network. Moreover, services like billing and security are fully taken care of by the cellular network. This enables inter-technology seamless mobility and session continuity reducing the handoff delay experienced in loose coupling [30][31]. The typical challenges of this approach are the increased burden on the core network due to the added WLAN traffic and the need for the WLAN to be operated by the cellular operator as it is directly connected to the core network.

Very Tight Coupling: It's the mechanism by which a Wi-Fi AP is connected directly with the eNB that covers it. The Wi-Fi AP is managed as part of the eNB subsystem. For 3G and WLAN systems it is created as in [32] using a new interface between Wi-Fi AP and the RNC of the UTRAN. Between LTE and Wi-Fi it is recently proposed by Xavier Langrange [33]. Since connection is made at the lower layer than tight coupling, namely the Packet Data Convergence Protocol (PDCP)[33][34], vertical handover is automatically handled between the two technologies. Moreover, it gives the opportunity for the Wi-Fi AP to reuse the security features of the cellular system like in the tight coupling case.

Similar to the tight coupling case, billing, security, mobility and session issues are fully handled by the cellular network and this adds more traffic to the core network. Besides, full implementation requires the modification of some protocol components alongside to the addition of new equipment.

The interworking interfaces or components of a WLAN with 4G networks can be visualized from the depiction given in Figure 2.2 which is adapted from [21].

The aforementioned kinds of integrations apply, mainly, for cellular and Wi-Fi technologies. When it comes to the interconnection between cellular and D2D technologies the work by Arash Asadi et al. in [118] gives a very good discussion. The classification is initially based on either the cellular spectrum or the ISM band is used for the D2D communication. The former is referred as inband and the latter as outband. Again, when a single licensed spectrum is shared by both technologies it is called underlay inband and when different licensed frequencies are used it is called overlay inband D2D. Yet, outband D2D communication occurs when the D2D uses the unlicensed ISM spectrum. This kind of transmission can be controlled by either the eNB, known as controlled, or by the user, known as autonomous. It is here, in the outband D2D communication, that the real challenge of power management and interworking architectural issues arise [35].

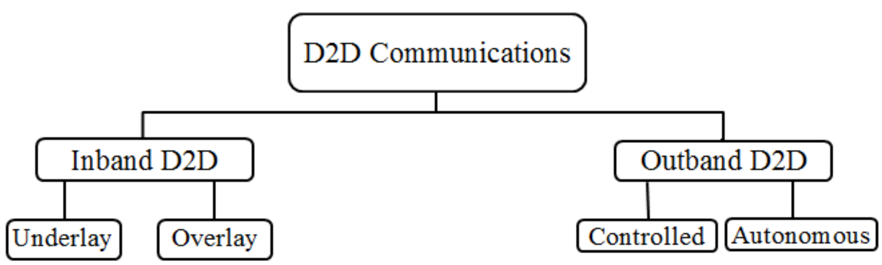


Figure 2.3: Device-to-Device Communication Classification, Adapted from [36].

2.3.2 Technologies Involved

The technologies that are coupled with 3GPP cellular networks as part of the interworking architecture are briefed here. Though, there are works that attempt to integrate Ethernet [37], WiMAX [23][38] and other networks with 3GPP-baseed 3G and 4G mobile networks, nevertheless, only WLAN and D2D technologies are considered in this work. Apparently, between these three different wireless technologies there is a considerable amount of work to interwork as depicted in the matrix presented in Figure 2.28. First, an attempt has been made to briefly see the three technologies involved and turn to their integration specifics. More discussions can be obtained in chapter one subsection 1.2.

D2D Technologies: D2D technologies facilitate direct communication between nearby devices enabled with such technologies as Wi-Fi in ad-hoc mode, NFC, Bluetooth, and ZigBee. For the sake of brevity, only D2D technologies that are related to cellular networks are considered here. Basically, mobile devices are already equipped with Wi-Fi, Bluetooth and NFC technologies; therefore, these technologies as depicted in Figure 2.3 can be exploited for opportunistic communication among mobile devices as they get in close proximity. The attractiveness of these technologies, in offloading cellular traffic, is their low transmission power and use of the unlicensed ISM band as described in Table 2.1. There are many works exploiting Wi-Fi, in ad-hoc mode, and Bluetooth to offload cellular traffic opportunistically [39] - [40].

Wi-Fi/WLAN Technology: Wi-Fi Alliance defines Wi-Fi in [41] as *"any WLAN product based on the IEEE 802.11 standards."* Wi-Fi APs or devices have two modes of operation; infrastructure and ad-hoc. In infrastructure mode a Wi-Fi AP acts as last-mile connection for the wired network extending it and adding mobility features. In this mode, a mobile device has to authenticate and perform a handshake to get access to the Intranet or Internet via the AP.

In ad-hoc mode, Wi-Fi devices can communicate directly among each other. This mode has been exploited in many popular multiplayer gaming consoles and consumer electronics [36]. Wi-Fi alliance, since 2010, worked towards a specification called Wi-Fi Direct that is initially intended for file transfers and multimedia sharing [42]. There is also Tunneled Direct Link Setup (TDLS) that enables two devices on the same Wi-Fi network to communicate directly with AP involvement [42].

Wi-Fi or WLAN, in both infrastructure and ad-hoc modes, is exploited to offload cellular traffic where the former is referred in this work (and in many others), as AP-based offloading and the latter as direct D2D communication [36][43][44]. In both modes, the devices involved in the communication use the frequency bands specified in Table 2.1, in their respective columns.

Cellular Technologies: Being aware of other standards and specifications of cellular network technologies by 3GPP2, this work focuses on those proposed by 3GPP. The picture depicted in Figure 2.4 shows the evolutionary progression of cellular network technologies. In a nutshell, the progression of 3GPP mobile networks can summarized as follows: GSM, GPRS, EDGE, HSPA [HSUPA and HSDPA], HSPA+, LTE and LTE-A [45]. As a matter of fact, these should not be considered as distinct technologies or eras, but chronological introduction of a new, on top of the existing, commercial cellular infrastructure. For instance, GPRS adds, in to the circuit-switched GSM, a packet-switched capability rather than fully replacing it.

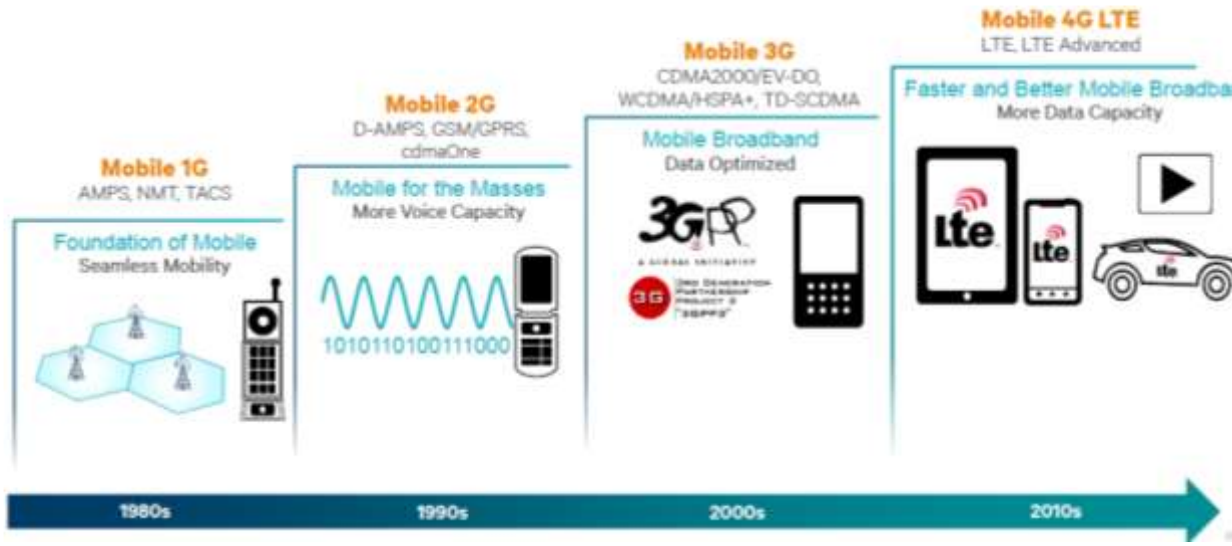


Figure 2.4: Evolution of Cellular/Mobile Technologies (adapted from [45]).

2.3.3 Nature of Traffic Offloaded

Here the nature of traffic offloaded in terms of delay guarantees, is deliberated. One of the issues during offloading cellular traffic is guaranteeing the minimal delay requirements of applications or services being taken off the cellular network. Though, initially, IP networks didn't think of Quality of Service issues, various enhancement works had been started by ITU-T, IEEE, IETF and 3GPP. In its E.800 document [46] ITU-T defined Quality of telecommunication services where it specified concepts, models, objectives and dependability plans for End-to-End (E2E) QoS requirements. IEEE also made amendments, among others, to the existing 802.11 standards by introducing QoS at the MAC layer [47]. The IETF set many RFCs related to QoS requirements and specifications to be considered by implementers. Among them is RFC2205 – Resource ReSerVation Protocol (RSVP)[25] –defines a resource reservation setup protocol for integrated services via the Internet. This is followed by the likes of RFC1633 Integrated Services in the Internet Architecture [48] – overview of a proposed extension to the Internet architecture and protocols to provide integrated services to support real-time and non-real-time services, alike. Then follows RFC2474 [49] and RFC2475 [50] that define the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers and an architecture for implementing scalable service differentiation in the Internet, respectively. Moreover, RFC2212 explains the requirements for network elements that support guaranteed services [51].

Based on these recommendations, specifications and requirements, 3GPP defined and enforced QoS policies and technical specifications for 3GPP technologies and WLAN internetworking. Those 3GPP technical reports and specifications are TR 23.838 [52] outlining QoS policy aspects, ETSI TS 123.107 [53] describing the E2E QoS concept and architecture, and ETSI TS 23.203 [54] standardizing QoS class identifier (QCI) characteristics for 3GPP and WLAN internetworking. Table 2.2 summarizes the traffic classification and QoS requirements to be enforced in during such internetworking scenarios.

There are considerable efforts with regard to offloading delay tolerant traffic via APs [15][54] and to D2D [17][39][44]. Even, some claimed mobile traffic offloading is better suited for such traffics but not for delay sensitive ones [56][39]. On the contrary, many others show the possibility of offloading delay sensitive traffic too [55][57]. And yet, some other investigated how to handle both kinds of traffic exploiting all opportunities to minimize the cellular infrastructure. The work of Minsoo Lee et al. [55] showed how to differentiate the user traffic followed by the identification of the delay requirements of the traffic flow, deciding whether to use the WLAN or the cellular network to transmit the traffic is a typical work.

Table 2.1: Details of Cellular, Wi-Fi/WLAN and D2D Technologies.

Technologies / Features	Cellular (LTE/HSPA+/HSxPA/EDGE/GSM/GPRS)	Wi-Fi/WLAN/ (802.11b/g/n/ac)	D2D (Bluetooth/Wi-Fi Direct)
Distance/coverage	LTE: 100 km HSPA+: 200 km HSxPA: 200 km EDGE: 26 km GSM: 26 km GPRS: 26 km	30-300 m	~10 m
Data Rate (Theoretical, Max) **(UL/DL)**	LTE: (50-75/100,150,300) HSPA+: (168/22) HSPA: (0.384/14.4) EDGE: (0.5/1.6) GSM: (0.4/0.4) GPRS: (0.04/0.08)	11/54/135/780 Mbps	3 Mbps
Achievable Throughput (DL/UL)	LTE: (51.11±16.9/17.9±5.1) HSPA+: (4.5±1.3/1.3±0.2) HSPA: (0.384/14.4) EDGE: (0.237/0.059) GSM: (0.034/0.021) GPRS: (0.014/0.01)	5/20/74/200 Mbps	~2.1 Mbps
Radio Technology	LTE: OFDMA/MIMO/SC-FDMA HSPA+: CDMA/FDD MIMO HSPA: W-CDMA/CDMA-FDD/CDMA-FDD-MIMO EDGE: TDMA/FDD GSM: TDMA/FDD GPRS: TDMA/FDD	DSSS/OFDM/ MIMO-OFDM (both n & ac)	Additive FHSS
Max. Power Allowed	LTE:100 mW @ 200 Mbps HSPA+: 44.2 mW@1990 Mbps HSxPA: 250 mW EDGE: 2000 mW/1000 mW GSM: 2000 mW@850/900, 1000 mW@1800/1900 MHz GPRS: 500 mW	100/100/200/250 mW	2.5 mW (4 dBm)
Frequency Used	LTE: 700, 800, 900, 1800, 2600 MHz HSPA+: 2100-2200 MHz/1885-2025 MHz HSPA: 2100-2200 MHz/1885-2025 MHz EDGE, GSM, GPRS: 850, 900/1800, 1900 MHz	2.4/2.4/(2.4/5)/5 GHz	2.4/(2.4/5) GHz
(UL/DL)* represents uplink and downlink, respectively. NA – Not Available			

Table 2.2: Standardized QCI Characteristics for 3GPP and WLAN Interworking.

<table>
<tr><th>QCI</th><th>Resource Type</th><th>Priority</th><th>Packet Delay</th><th>Packet Error Loss Rate</th><th>Example Services</th></tr>
<tr><td>1</td><td rowspan="4">GBR (guaranteed bitrate)</td><td>2</td><td>100 ms</td><td>10^{-2}</td><td>Conversational voice (e.g. VoIP)</td></tr>
<tr><td>2</td><td>4</td><td>150 ms</td><td rowspan="2">10^{-3}</td><td>Real-time video (e.g. Video call)</td></tr>
<tr><td>3</td><td>3</td><td>50 ms</td><td>Real-time gaming</td></tr>
<tr><td>4</td><td>5</td><td>300 ms</td><td rowspan="3">10^{-6}</td><td>Buffered video</td></tr>
<tr><td>5</td><td rowspan="5">Non-GBR</td><td>1</td><td>100 ms</td><td>IMS signaling</td></tr>
<tr><td>6</td><td>6</td><td>300 ms</td><td>Buffered video streaming, TCP-based services (e.g. email, chat, ftp)</td></tr>
<tr><td>7</td><td>7</td><td>100 ms</td><td>10^{-3}</td><td>Voice, Real-time video, Interactive gaming</td></tr>
<tr><td>8</td><td>8</td><td rowspan="2">300 ms</td><td rowspan="2">10^{-6}</td><td rowspan="2">Buffered video streaming, TCP-based services (e.g. email, chat, ftp)</td></tr>
<tr><td>9</td><td>9</td></tr>
</table>

2.3 Efforts towards Feature-rich Interworking Architecture

The efforts towards integrated interworking architecture are technically diverse, conceptually delicate and practically pealing. In this section, an attempt has been made to classify the efforts made from three major dimensions. This is pictorially depicted in Figure 2.5.

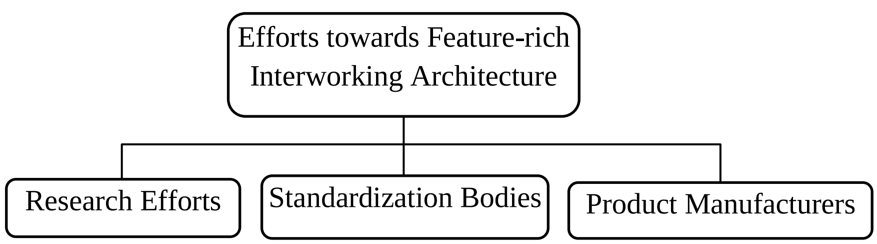


Figure 2.5: Classification of the Efforts towards Integrated Interworking Architecture.

Under research efforts the hard works made by academia, research institutes/centers, and the industry are considered. Under product manufacturers those organizations who design, implement, manufacture, market and sell products like network hardware, operating software systems for wireless networking devices, Wi-Fi APs or routers, network controllers, mobile devices or phones or equipment, and similar systems are considered. Typically, these includes, but not limited to, Cisco Systems, Qualcomm, and Aruba Networks. When it comes to standardization efforts, those professional and governmental institutes, organizations, associations and consortia that are responsible to set working and operating conditions, performance limitations, certifications, conformance tests and those engaged in similar activities are examined. IEEE, IETF, ITU-T, 3GPP, and Wi-Fi Alliance are some of them.

Henceforth, the effort of researchers in academia and the industry, technical specifications and recommendations of standardization bodies and advertisements of product manufacturers and marketers are summarized. The advertisements of product manufacturers and marketers are gathered from the white papers they disseminate, though the academic and research credibility of such sources is still very much questionable. However, it will give us a glimpse of what features is being realized into products that are conformant with the research and standardization efforts.

2.3.1 Research Efforts

Globally, there is a considerable work by researchers in academia and industry in devising an interworking architecture among cellular, WLAN and D2D technologies since the idea is presented for the first time back in 2002 [58]. In this section, the research efforts made, mainly, to integrate these networks are investigated by critically examining the specific contributions. For instance, internetworking involves features and activities like mobility management and session continuity, vertical handoff techniques, neighbor discovery, content injection and distribution, movement prediction and target set selection which researchers propose, optimize, analyze, implement and evaluate. Moreover, some propose models or architectures or frameworks that handle some set of interworking features.

Interworking Models/Architectures/Frameworks

The pioneering work of Shiao-Li Tsao et al. proposed three different internetworking strategies between UMTS and WLAN [59]. As early as 2002 they tried to show both loose coupling, using mobile IP, and tight coupling, using the gateway approach, of integrations. The system architecture is shown in Figure 2.6. They also depicted some specific deployment scenarios. The following year, in 2003, M. Buddhikot et al. presented a very interesting internetworking system to interwork WLAN and CDMA2000 in loose coupling [60]. They developed software based on WLAN gateway and client application for the proposed offloading system. They have used a test bed to evaluate the performance of the overall system. A similar work is conducted by Jee-yound Song et al. in 2003 [61] as shown in Figure 2.7. The interesting idea proposed was a hybrid coupling scheme between UMTS and WLAN by combining tight and loose coupling techniques. The former used for real-time and the latter for non-real-time traffic. This is decided based on the QoS required of the traffic to be transmitted.

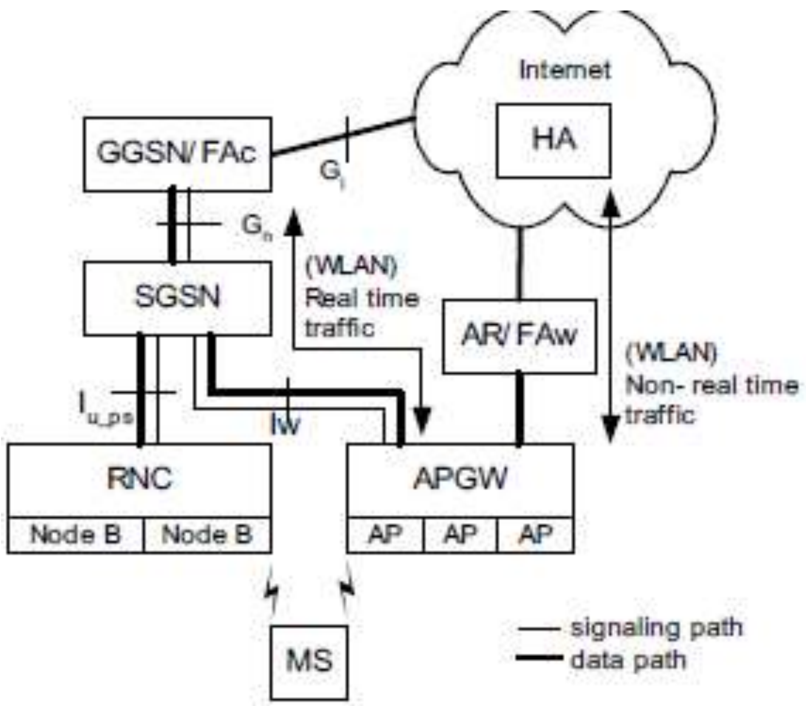


Figure 2.6: Hybrid Interworking Architecture [61].

Procedure 1 Serving user's requests in a network with HotZones.

```
if (S_r^u(t) ≠ ∅) then
  if (c ∈ H) then
    Turn on WiFi interface;
    Try to serve all r ∈ S_r^u(t) via WiFi;
    //a success with probability p
  else
    Get τ_H^u = time before u enters a cell ∈ H;
    for all r ∈ S_r^u(t) do
      if (τ_r expires in ≤ τ_H^u) then
        Serve r via 3G;
      else
        Do nothing;
      end if
    end for
  end if
end if
```

Figure 2.7: Algorithm for HotZones [8].

Procedure 2 Serving user's requests in a network with MixZones.

```
if (c ∈ M) then
  Turn on WiFi interfaces in set U_c(t);
  Opportunistic WiFi transfers among users;
else
  for all users u in cell c do
    Get τ_M^u = time before u enters a cell ∈ M;
    for all r ∈ S_r^u(t) do
      if (τ_r expires in ≤ τ_M^u) then
        Serve r via 3G;
      else
        Do nothing;
      end if
    end for
  end for
end if
```

Figure 2.8: Algorithm for MixZones [8].

In 2004, Apostolis K. Salkintzis in [62] discussed and analyzed three different kinds of architectures and corresponding message flow diagrams based on the requirements of the most common internetworking scenarios put forth in [63] 3GPP Technical Report. What makes the work unique is the detailed depiction and discussion of the architectures and their interfaces between UMTS and WLAN, progressing from loose to a tighter coupling in the process, adding new features in to the internetworking system. The researcher, together with others, extended the effort in testing one of the architectures proposed in the previous work for seamless continuity of real-time video sessions based on two different approaches, contention-based and contention-free where simulation results being in favor of the former [64].

In 2006, George Lampropoulos et al. studied and proposed an offloading mechanism where each connection of a UE, unlike previous works, is served by different radio access technologies based on a tight coupling of UMTS and WLAN systems getting some improvements in connection and handover blocking probabilities [65]. In the next year, Yu Zhou et al. designed and implemented a dual-mode UE in OPNET, targeting to achieve load balancing between WLAN and UMTS. The utility-based access control algorithm, doing vertical handover based on the utility values, perform better than without it in terms of throughput and load balancing [66].

In the following years, extensive works on different offloading techniques that exploit emerging enabling mechanisms (like LIPA – Local IP Access, IMS – IP Multimedia System, and SIP – Session Initiation Protocol) emerged. Nikodin Ristanovic et al. in 2010 [8] designed two different algorithms, MixZones, based on opportunistic transfers and HotZones supported by Wi-Fi APs, for delay-tolerant offloading of large, socially recommended content from 3G networks. The two algorithms are presented in Figures 2.8 and 2.9, respectively. The solutions in addition to offloading a significant amount of 3G traffic, up to 50% for delay tolerant, bring down user battery consumption. Continuing the effort, a more complete offloading system, named Wiffler, is designed and implemented by Aruna Balasubramanian et al. in 2010 [55]. Wiffler exploits two innovative ideas - leveraging delay tolerance nature of some traffic and fast switching to overcome the poor availability and performance of Wi-Fi. For delay tolerant applications, Wiffler uses a simple model of the environment to predict Wi-Fi connectivity based on which it postpones transfers to offload more data on Wi-Fi whenever it available. However, this is done only if delaying reduces 3G usage and the transfer can be completed within the application's tolerance threshold. For applications that are extremely sensitive to delay, like VoIP, it quickly switches to 3G if Wi-Fi fails to successfully transmit the packet within the applications QoS limit. Using vehicular testbed they found traffic reduction of 45% for a delay tolerance of 60 seconds. In terms of completeness, another similar effort is made by John L. Tomici in [67] by proposing a Converged Gateway (CGW) based hybrid conceptual network architecture that can support Home IP, Femto, Wi-Fi and D2D networks. It supports radio-interference, RAN and CN -level offloading mechanisms via LIPA and other enabling techniques. Though, the idea is noble, there is no any evaluation of the proposed system, merely conceptual discussions.

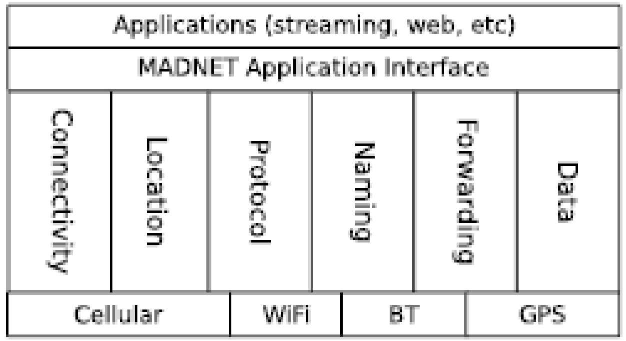


Figure 2.9: MADNet Architecture [44].

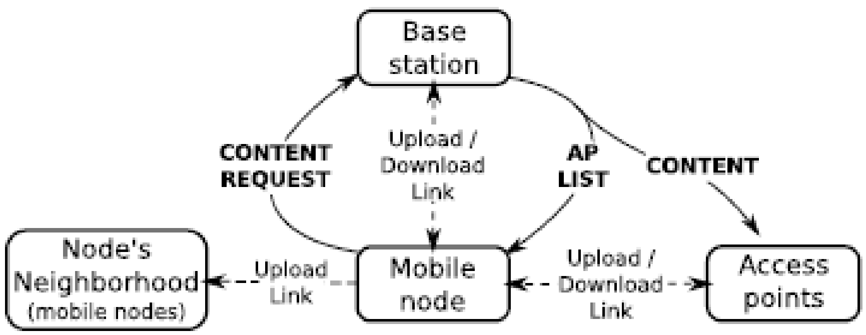


Figure 2.10: Communication flow in MADNet Architecture [44].

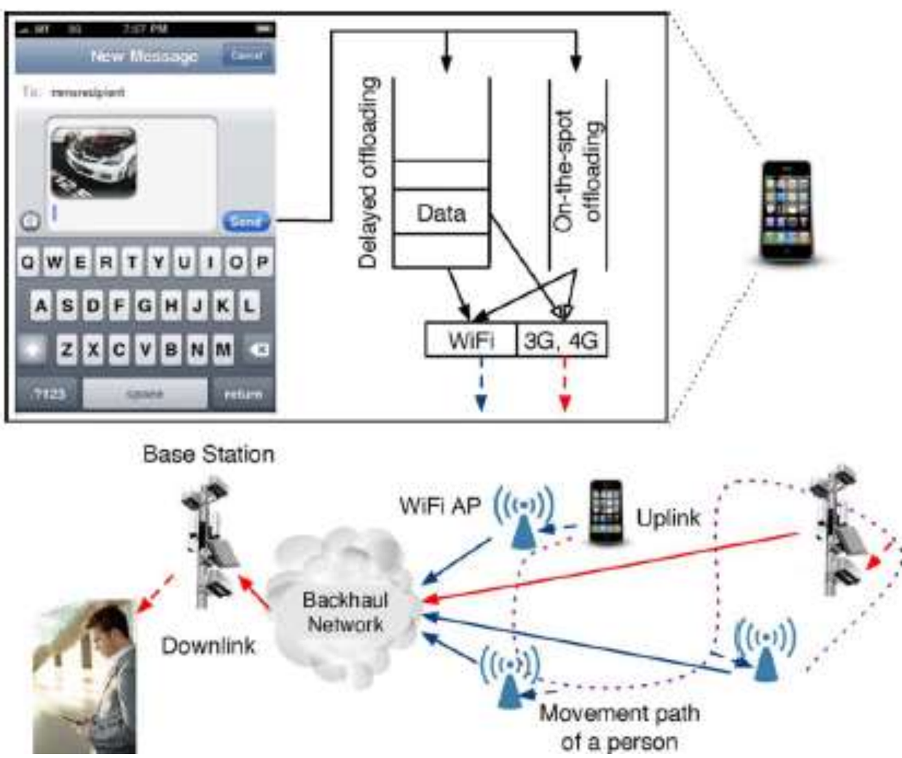


Figure 2.11: An Offloading System Supporting Uplink Traffic [15].

Another complete offloading system for 3G systems is proposed and evaluated by Savior Dimatteo et al. in [44], Aaron Yi Ding et al. in [68] and [68]. As shown in Figures 2.9, 2.10 and 2.11 they came up with Metropolitan Advance Delivery Network (MADNet) integrated architecture that consists of 3G, Wi-Fi and mobile-to-mobile Pocket Switched Networks. They used two separate queues to manage the traffic from mobile devices, one for delayed tolerant and another for on-the-spot or delay sensitive ones. They also developed an algorithms based on the message flows depicted in Figure 2.11. The architecture is evaluated for performance using real mobility traces and for energy usage using energy-aware algorithm giving, *surprisingly* as they put it, 65% traffic reduction using on-the-spot offloading, and better end-to-end data rates during night times than working hours. Safdar Rizvi proposed [30] two tight coupling strategies - interconnecting WLAN with SGSN and Gateway GPRS Support Node (GGSN). Using OPNET modeler results show that the latter outperformed the former in terms of various applications and measurement parameters mainly due to, in the former case, the Wi-Fi AP requires additional processing for successful communication than the latter. On the contrary, G. Vijayalakshmy recently proposed a loosely coupled WiMAX-UMTS-WLAN convergence architecture using IMS-SIP protocol. The QoS performance metrics on different traffic delays, as delay sensitive (for Voice over IP, Video streaming, and Web browsing) and background traffic (like email and FTP), are evaluated using OPNET and show better results when compared with a system without the IMS-SIP protocol [23].

Yet another compete offloading system are proposed, implemented and tested by Shahriar Nirjon et al. [69], Youngbin Im et al. [57], and Vinicius F.S. Mota et al. [70]. In the first case they design MultiNets software-based system that handles seamless switch-over between 3G and Wi-Fi in real-time. The system has three operating modes; Energy saving, offload, and performance modes. It is implemented on Android devices as a prototype implementation. Results show that on average

79.82% of data traffic is offloaded by automatically switching to Wi-Fi whenever it is available. This saves, on average, 21.14 KJ of energy per day. In the second case they propose cost-aware Wi-Fi offloading system, dubbed AMUSE - Adaptive bandwidth Management through USer-Empowerment, based on user's throughput-delay tradeoffs and 3G budget constraints. Microsoft Windows 7 tablets implementation and evaluation shows improved performance for both heavy and light traffic users. In the third case, they propose, design and test a multi-criteria decision-making framework called OpenLite that handles 3G traffic to opportunistic mobile-to-mobile networks. Results show that up to 6% traffic can be offloaded when there is no delay tolerance and went up to 36% for delay tolerance of 20 minutes. Patrick Baier et al. devised an opportunistic traffic offloading system that uses movement predictions of mobile users named TOMP – Traffic Opportunistic using Mobile Prediction [71]. Based on the movement prediction of mobile user the system distributes messages to the target set. The system is tested over HSDPA and Bluetooth devices.

There are extensive analytic works on specific offloading features and issues too. Yong Li et al. investigated the multiple mobile data offloading problem where multiple kinds of content with different delay tolerances and sizes by formulating it as a Sub-modular Function Maximization problem and designing three algorithms to proximately solve it [39]. The analysis involves 3G networks based on delay tolerant offloading over Bluetooth enabled devices in realistic network assumptions. A more interesting work is done by Sarabjot Singh et al. in [72] who analyze the effects of offloading in M-RAT K-tier wireless heterogeneous networks. The model consists of M different RATs, each deploying up to K different tiers of APs (s), where each tier differs in transmit power, path loss exponent, deployment density and bandwidth. Each class of s is modeled as an independent Poisson point process (PPP), with mobile user locations modeled as another independent PPP, all channels further consisting of i.i.d. Rayleigh fading. They made intensive mathematical modeling and analysis of offloading behaviors showing, among others, the optimum fraction of traffic offloaded to maximize SINR coverage is not, in general, the same as the one that maximizes rate coverage. Further analytical study of offloading has been made by Fidan Mehmeti et al. in [56] using 2D Markov chains based queuing model specifically for delayed offloading, by Yoora Kim et al. in [73] using embedded Markov process, again, for delayed offloading system, by Andrey Krendzel et al. in [74] using a generic, independent of the underlying offloading technology, analytical approach for studying the impact of offloading techniques on the networks of MNOs (Mobile Network Operators) by modeling user's activity and offloading times as strictly alternating independent ON/OFF processes, and Patrice Raveneau et al. in [75] studied how offloading between 3G and Wi-Fi works in a large urban settings. What is unique about this study is the use of a real world cellular and Wi-Fi deployments and data gathered thereof.

The works summarized up to this point are mainly focused on efforts towards offloading system of 3G traffic over Wi-Fi APs. Due to the advent of the 3GPP 4G technology (termed as LTE – Long Term Evolution), works kick off to integrate the same both with Wi-Fi APs and D2D technologies. Here, the reader should be reminded that unlike its predecessor 3GPP technologies, LTE already considered the issue of offloading both on WLAN and D2D systems from its inception [76][77]. Therefore, here follows research efforts attempting to interwork LTE with WLAN (Wi-Fi s) or D2D technologies or both.

Fidan Mehmeti et al. studied, using queuing analytic model, the performance improvements achievable by Wi-Fi-based data offloading as a function of Wi-Fi availability and performance, user mobility and traffic load, and the coverage ratio and respective rates of different cellular technologies like EDGE, HSPA and LTE [78]. Though the work is only for on-the-spot offloading scenario, it considered various performance metrics like offloading gain, transmission delay, and availability ratio. On the other hand, Jonathan Ling provided a detailed theoretical discussion of LTE-WLAN integration – loose, tight and hybrid coupling [79]. As it has been done for UMTS and WLAN in [61], here the hybrid coupling exploits both the Wi-Fi and cellular radio interfaces for simultaneous communication.

Rather few attempts have been made to interwork LTE with WLAN/Wi-Fi when compared with D2D technologies. Souheir EIDO et al. proposed different offloading scenarios from Femto and macro cellular networks using the SIPTO approach using loose coupling over WLANs [80]. On a similar work, Liang Hu et al. investigated the downlink performance of Wi-Fi and Femto cells as indoor offloading solutions for LTE networks [81]. Results show that out-band Femto and Wi-Fi's give similar and better results than out-band Femto cells deployment due to the sever interference between the Femto and macro cell in the latter. Mehdi Bennis et al. proposed a fully distributed and dynamic traffic offloading framework between LTE and Wi-Fi using cross-system learning as traffic steering policy [82]. This tight coupling seamlessly steer traffic between cellular and Wi-Fi RATs depending on the traffic type, users' QoS requirements, network load, and interference levels. A very recent work by Yingzhe Li et al. validated a stochastic geometry framework for analyzing the coexistence of overlaid Wi-Fi and LTE networks in the unlicensed band showing positive results of LTE in the Unlicensed spectrum (LTE-U) and Licensed-Assisted Access (LAA)[83].

Alexander Pyattaev et al. studied the performance of Wi-Fi Direct overlaid with LTE in network-assisted mode in urban environments [42]. Extending their work in [84], as the implementation system diagram is shown in Figure 2.12, they made a full-featured network-assisted deployment with all the necessary support to provide real-time D2D connectivity over the emerging Wi-Fi-Direct technology as shown in Figure 2.13. Sergey Andreev et al. proposed and evaluated a conceptual network-assisted D2D cellular offloading system [85]. The system is based on Wi-Fi direct over

LTE in-band underlay mode. The uniqueness of this work is the use of statistical geometry (statistical modeling of spatial relationships) to capture the topological randomness and system dynamism of D2D users in a home-made advanced system level simulator (SLS).

Balaji Raghothaman et al. described a comprehensive system architecture that enables mobile-to-mobile communications to enhance the capacity of the cellular network which is based on the LTE Release-10 [86]. Neighbor discovery mechanisms, the resource partitioning for the mobile-to-mobile link and changes to the design of the HARQ (Hybrid Automatic Repeat Request) entity structure and timing are discussed very well. Wenjie Hu and Guohong Cao proposed a Quality-Aware Traffic Offloading (QATO) framework over Wi-Fi direct [87]. The android implementation and photo uploading plication system shows reduction of 38% of battery power in downloading and 70% in uploading, and reduction of delay by 45% in downloading and 88% in uploading. Another comprehensive work done by Xingqin Lin develops, as they call it, *a baseline hybrid network model* for cellular and Wi-Fi ad-hoc nodes [43]. The model uses Poisson point processes to model the random and unpredictable locations of mobile users. Several important multicast D2D metrics including coverage probability, mean number of covered receivers per multicast session, and multicast throughput are analytically characterized under the proposed modal. D2D mode selection which enables D2D nodes switch between D2D and cellular modes is incorporated into the hybrid network model. Extending the model, they added two spectrum sharing models, overlay and underlay. Further extending the baseline model, they added a multi antenna transmission to study the relationship between massive MIMO and underlay D2D networking. The spectral efficiency of such a system is studied under both perfect and imperfect channel state information (CSI) assumptions. When compared to the case without D2D, a loss in cellular spectral efficiency was observed due to D2D underlay. With perfect CSI, it is suggested that the spectrum loss can be completely overcame if scaling is used appropriately. Whereas, with imperfect CSI a new asymptotic underlay contamination effect was observed.

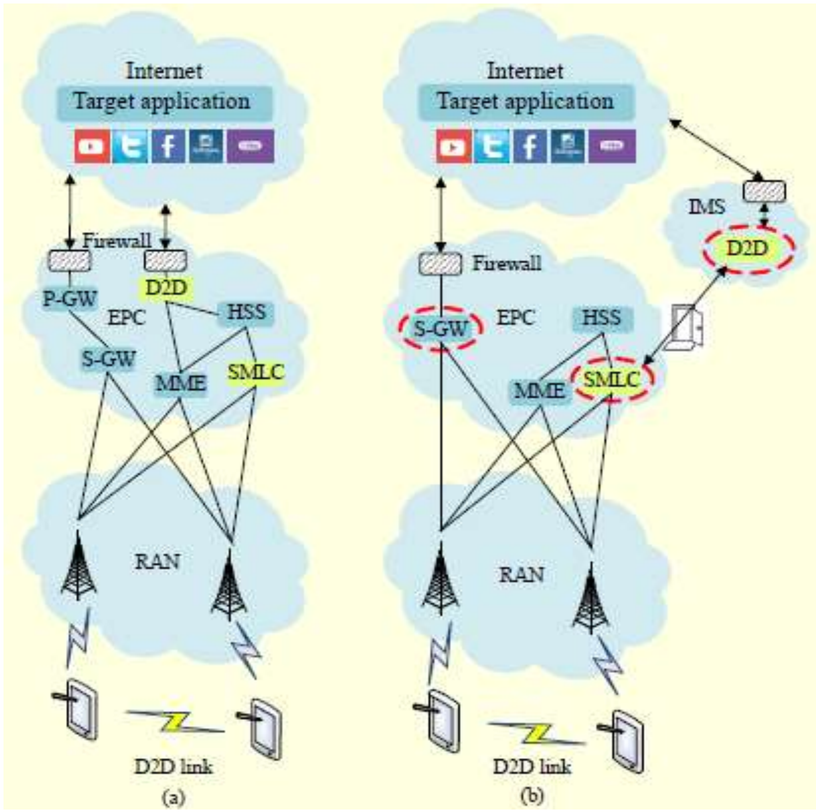


Figure 2.12: Implementation of D2D System in 4G [84].

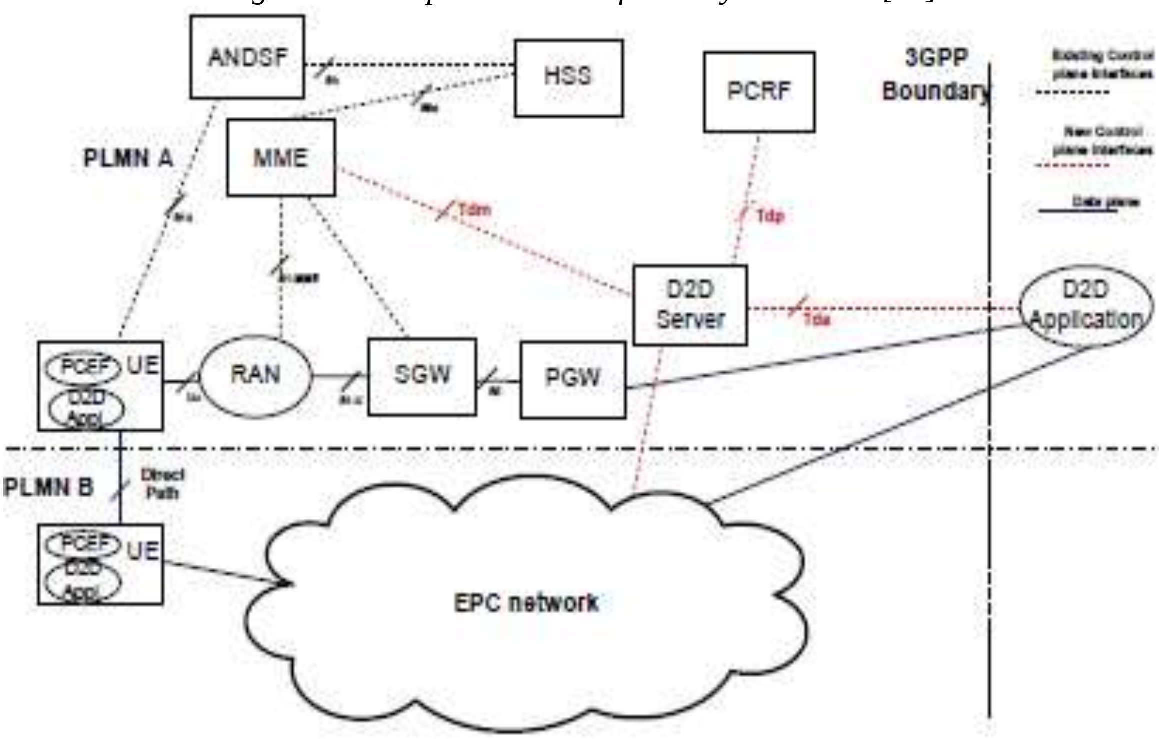


Figure 2.13: Architectural and Protocol Enhancement of LTE and LTE - A Systems to Support D2D Communication [35].

Figure 2.14: FlashLinQ Prototype Modem [88].

Wei Peng et al. proposed a model that exploits opportunistic proximity via Near Field Communication (NFC) or Wi-Fi Direct between smartphones [89]. Unlike other similar works they investigated “who (will download the content through the cellular link),” and “when” to locally estimate (their own and their acquaintances) topological importance on the opportunistic proximity link. Xinzhou Wu in [88] proposes FlashLinQ a synchronous peer-to-peer wireless PHY/MAC network architecture that leverages the fine-grained parallel channel access offered by OFDM, and incorporates an analog energy-level based signaling scheme that enables SIR based distributed scheduling. The new system shows significant spectral efficiency gains over the CSMA/CA system using RTS/CTS boosting the importance of licensed spectrum scenarios. FlashLinQ is complete in the sense that it contains timing and frequency synchronization, peer discovery, link management, and channel-aware distributed power management, data-rate and link scheduling. Its implementation, using DSP/FPGA platform, is shown in Figure 2.14.

Balaji Raghothaman et al. proposed architectural and protocol enhancement for LTE-based D2D communication as shown in Figure 2.15 [35]. Whereas, Zehua Wang et al. in [90] and Arash Asadi in [91] proposed LTE-based D2D offloading system without any fundamental change on the existing system. The former proposes a scheme to reuse the downlink licensed spectrum of the cellular network for D2D in the in-band scenario and the latter proposes a scheduling protocol that assists offloading between LTE and Wi-Fi direct.

Lorenzo Valerio et al. propose completely new approach of adaptive offloading solution based on the Reinforcement Learning (RL) framework [92]. In this system, the controller of the dissemination process, trained by the RL selects a proper number of content replicas to be injected in the opportunistic network to guarantee the timely delivery of contents to all interested users. The system when compared with Actor–Critic and Q-Learning algorithms automatically learn a very efficient strategy to reduce the traffic on the cellular network offloading up to 97% over Wi-Fi in ad-hoc mode.

A project partially funded by EU designed, implemented and evaluated a complete offloading system named MOTO[4] – Mobile Opportunistic Traffic Offloading as shown in Figure 2.15 and 2.16 [93][94]. They conceptualized offloading architecture that can exploit Wi-Fi transmissions and D2D opportunities, has core MOTO services (like content diffusion, AAA, and network status information), protocols (to communicate with external infrastructure operators, terminals, and external content providers) and APIs (to link core services with different RAT and with the content provider) for both the application and infrastructure. The platform can be integrated either inside or outside operators' network. This is a typical full-fledged offloading architecture which after implementation and testing can provide real world offloading benefits.

[4] MOTO is a €4.8 billion project under the FP7-ICT with project reference no. 317959.

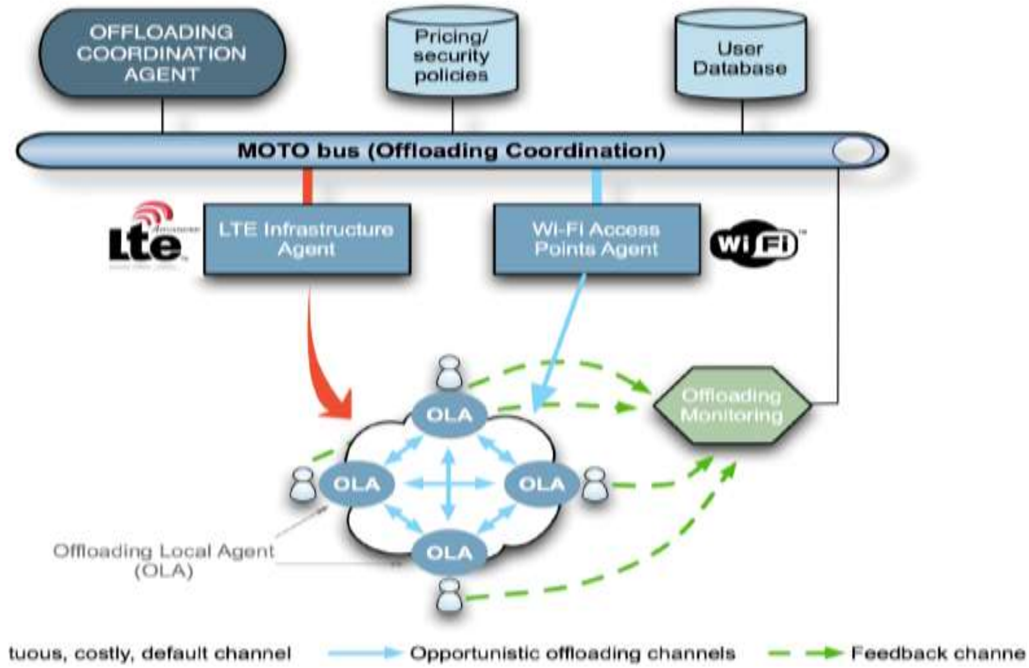


Figure 2.15: Conceptual MOTO Architecture [90].

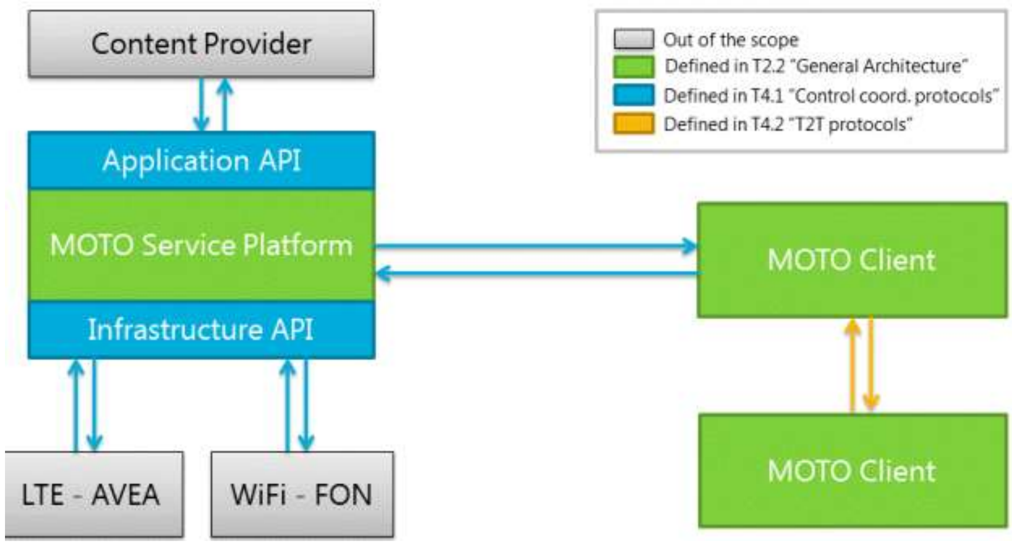


Figure 2.16: MOTO Specification Overview [91].

Very recently Xavier Lagrange came up with a new way of coupling between 4G and Wi-Fi networks referred to as Very Tight Coupling [33]. The idea is, simply, to connect the Wi-Fi AP to the eNB currently covering it. As shown in Figure 18 the Packet Data Convergence Protocol (PDCP) is the common layer between the two networks. All control functions like security, mobility, and session management are done by the LTE using Wi-Fi for data transmission only. It allows the UE to quickly set up layer two connections with Wi-Fi APs allowing it to offload traffic via the covered Wi-Fi network.

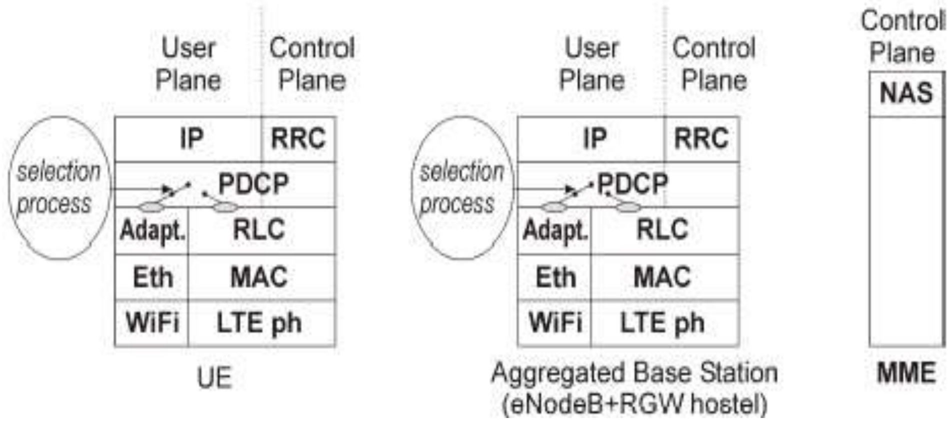


Figure 2.17: Very Tight Coupling at the Protocol Stack Level.

Another innovative new approach was made by Miguel Angel Patino Gonzalez et al. in pointing out the possibility of adapting Multipath TCP for mobile traffic offloading, specifically, in terms of the radio access [95]. Implications and requirements for a network operating with MPTCP-enabled devices with various architectural elements with associated costs for users and operators are presented.

The effort of internetworking 3GPP emerging cellular technology after LTE, known as LTE-A was made by Salam Doumiati et al. in [40] who investigated the required enhancements to the LTE-A network in order to integrate D2D communication at the network architecture level including cloud services into the offloading arena specifically for the ProSec based opportunistic environments.

Vertical Handoff Mechanism, Mobility Management and Session Continuity

George Lampropoulos et al. performed an in depth investigation on handover mechanisms for the internetworking of Wi-Fi and 3G systems [96]. The various works are summarized based on the point of system integration, peculiar features are presented, and the strength and weaknesses are discussed in detail. A much similar effort has been made by Hyeyeon Kwon et al. in performing a comparative performance analysis of five popular Virtical Handover (VHO) algorithms between 3G and WLAN networks [97]. They showed that what specific working environment each perform better. M. Lott proposed a hybrid (of location-based, velocity-based and prior knowledge-based) VHO mechanism based on critical investigation of various recent works of VHO triggers between 3G and WLAN systems [98]. They contemplated their proposed system will have less resource requirements and signaling overhead.

There are works done to evaluate the recommendations of both Institute of Electrical and Electronics Engineers (IEEE) and Internet Engineering Task Force (IETF) for the purpose of vertical handover and mobility management. Massimo Bernaschi et al. used the MIPv6 protocol in loose coupling between WLAN and 3G networks [79]. They evaluated the handoff performance at the network and transport layers getting a specific downside of the network level handoff due to the mismatch of RTT between the two networks incurring a slow-start of the handoff process. A similar work was conducted by Xavier Perez Costa and Hannes Hartenstein who evaluated the performance of MIPv6 via simulation using various metrics involving various traffic types [99]. Ashutosh Dutta et al. used a test bed to evaluate their implementation of the IEEE 802.21, also known as . Media Independent Handover (MIH), showing that 802.21-assisted handover helps perform the secured proactive handover with reduced delay and packet loss to a level that is acceptable for interactive VoIP and streaming traffic [100]. More detailed discussion of the IEEE 802.21 specification is presented in subsection B on page 17 of this per. On a similar effort, Thikrait Al Mosawi et al. proposed a solution for vertical handover of a multimedia session with the triggers being provided by MIH for efficient terminal mobility management in a SIP environment, which they named SIP-based Intelligent Handover (SIPIHO)[101]. The peculiarity of this handover process is that it can be either network-assisted or mobile node controlled. Uniquely so, Li Ma, et al., used Stream Control

Transmission Protocol (SCTP) to handle the VHO between UMTS and WLAN challenging works based on SIP, that require a server system, and MIP, which needs home and foreign agents, that their solution is network-independent [102].

Continuing the effort Minsoo Lee et al. in [103] proposed and implemented a location based services (LBS) broker for both loose and tight coupling cases of 3G and WLAN. The LBS Broker includes location-aware authenticator for fast secure roaming using the concepts of the direction of the user and pre-warming zone. Nagendra Prasad Mandru et al. in [104] proposed a simplified VHO algorithm based on received signal strength values and available radio resources, to enable transparent roaming across WLAN and UMTS.

Wi-Fi AP Deployment, Neighbor Discovery, Movement Prediction and Target Set Selection

Eyuphan Bulut and Boleslaw K. Szymanski in [105] and [106] proposed and evaluated, via real world traffic traces, a Wi-Fi deployment algorithm based on the density of user data request frequencies. The results show that, using the same number of APs, better offloading ratio is obtained by using Integer Linear Programming (55%) than greedy algorithm (45%).

Bo Han et al. investigated the target-set selection problem for information delivery in 3G and D2D based Mobile Social Networks (MoSoNets) using opportunistic D2D to facilitate information dissemination within the delay threshold using three algorithms – greedy, heuristic, and random. Performance evaluation of a prototype implementation, named Opp-Off on Nokia N900 phones, and trace-driven simulation shows that the heuristic algorithm can offload up to 73.66% [107].

Zhigang Li et al. presented the idea of the predictability of temporal and spatial characteristics of human mobility which can be exploited for efficient opportunistic communication[108]. Based on regularity of human mobility [109] they proposed the NodeRank algorithm which uses the encounter characteristics between nodes to choose the target-nodes and disseminate information. This algorithm combines the contact time and inter-contact times as evaluation metrics. The results are promising in terms of offloading efficiency and network redundant copies.

Qiaoyang Ye developed a low-complexity distributed algorithm of user association to achieve load balancing and improve cell edge performance between small cells and D2D. They further extended their study form single base station to multiple-antenna Multiple Input Multiple Output (MIMO) system to investigate further the resource allocation between LTE and D2D networks [110].

Implementation and Performance Evaluation (Analytical, Simulation, Testbed, Prototyping, etc)

Daniel Camps-Mur et al. provided an experimental evaluation of Wi-Fi Direct technology from the real world usage perspective [111]. To be more specific, they quantified the expected delays when Wi-Fi Direct devices discover (discovery delay) each other and establish a connection (formation delay). For the former, they found 3 seconds as the maximum value and less than 5, 8 and 9 seconds for Autonomous, Persistent and Standard group formation cases, respectively. They also evaluated the performance of Notice of Absence (NoA) power saving protocol which shows up to 66% power reduction, especially, when used together with adaptive policy without significant performance degradation. In a much similar endeavor, Sahar Hoteit evaluated the capacity and energy saving gains of offloading LTE over Wi-Fi Passpoint hotspots (also known as . HS2.0 – Hot Spot 2.0) [112][113]. It's obtained that up to 15% capacity and 13% energy saving gains are obtained when compared with those without Passpoint APs. Moreover, installing Passpoint hotspots in the outer annulus of the macro cell coverage provides the maximum capacity gain. An important point to mention here is that HS2.0-enabled device can recognize 802.11u [114] broadcasts and proceed to authentication automatically.

In another work, Icaro da Silva et al. demonstrated a Wi-Fi and 3GPP common performance monitoring system based on some Key Performance Indicators (KPIs) like throughput [115]. Whereas, Marty Glapa and Amit Mukhopadhyay presented a similar model to assess and optimize Wi-Fi offloading potential in a network based on applications and user behaviors in the process showing techniques of creating traffic density maps and identifying high traffic areas[116].

Content Injection and Distribution

Filippo Rebecchi et al. proposed, designed, and evaluated a re-injection strategy, called DROiD - Derivative Re-injection to Offload Data, to finely control the distribution of contents in a hybrid mobile network consisting of HSPA and D2D-enabled devices using Wi-Fi or Bluetooth [117][118]. The idea is, unlike existing techniques, bind re-injection to statically defined objective functions, to dynamically adapt to the current network topology. Performance assessment of the strategy using a vehicular dataset to disseminate contents under different tolerances of delay reveals significant savings in the infrastructure load that is above 50%. In a similar effort, Vincenzo Sciancalepore formulated the dissemination of content, injected through the cellular interface, in an opportunistic network with varying node mobility [119]. The system they designed based on HYbrid oPportunistic and cEllular (HYPE) technique minimizes the load of the cellular network while meeting the constraint in terms of delay guarantees. HYPE consists of two main building blocks: (i) the Content Server, and (ii) the Mobile applications. The Content Server runs inside the network infrastructure, while the Mobile applications run in mobile devices that are equipped with cellular connectivity, as well as able to directly communicate with each other via short range connections (e.g., via WLAN or Bluetooth). The proposed approach substantially outperforms previous approaches, like heuristics-based ones.

Tao Han et al. proposed a mobile traffic offloading scheme/algorithm called content brokerage [120]. In this scheme, a new network node called the green content broker (GCB) is introduced to arrange the content delivery between the content requester and the content owner. The specialty of the GCB is its use of green energy, like solar energy, to reduce the power consumption of mobile networks. Evaluation of a heuristic traffic offloading (HTO) algorithm to proximate the optimal solution, using the amount of energy and bandwidth as NP problem, provides reduced energy consumption and improved QoS, mainly; by retrieving contents from their neighboring peers instead of remote BSs. Another Distance-based Opportunistic Publish/Subscribe (DOPS) content dissemination model was proposed by Xiaofeng Lu et al. very recently that is composed of three layers: application layer, decision-making layer and network layer [121]. When a user wants new content, he/she subscribes on a subscribing server. A user decides whether to send content based on the distance measurement between him/her-self and the recipient.

User Satisfaction and Incentive Mechanism

Xuejun Zhuo et al. investigated the tradeoff between the amount of traffic being offloaded and users' satisfaction [122][123]. They proposed an incentive mechanism, WinCoupon algorithm, to motivate users to leverage their delay tolerance for 3G traffic offloading. Users with high delay tolerance and large offloading potential will be given higher priority to minimize the incentive cost for a given offloading target. They've used a reverse auction mechanism to let users proactively express their delay tolerance via bids. The delay tolerant network based case study illustrates how to predict offloading potentials using stochastic analysis.

The works of Lin Gao et al. [124] and Stefano Paris et al. [20] show, through two totally different approaches, the interactions between cellular and Wi-Fi operators and the economic viability of their cooperation. In the former case, they've used the Nash Bargaining Theory as a tool to model and analyze the interaction between Mobile Network Operators (MNOs) and AP Operators (APOs). They used a one-to-many bargaining game and explored the bargaining solution (game equilibrium) under two bargaining protocols – sequential bargaining (then MNO bargains with APOs sequentially) and concurrent bargaining (the MNO bargains with all APOs concurrently). Results shed light on the economic aspects and the possible outcomes of the MNO/APOs interactions, which can be used as a roadmap for designing policies for data offloading solutions. In the latter case, a reverse auction algorithm is used to exploit third parties for offloading. The conditions put forth were, first, to offload the maximum amount of data according to available capacity of the APs, second, to foster participation of AP owners (individual rationality), and third, preventing market manipulation (incentive compatibility). Following, they used three alternative greedy algorithms to help solve the problem as a combinatorial auction, and a payment rule was designed to guarantee both, individual rationality, and truthfulness for realistic scenarios in which only part of the data traffic can be offloaded. Numerical results demonstrate that the proposed schemes captured the economical and networking essence of the problem very well.

2.3.2 Standardization Efforts

In this subsection, the standardization efforts of mobile traffic offloading processes by internationally or regionally recognized standardization bodies, like ITU, IETF, IEEE, 3GPP, 3GPP2 and Wi-Fi Alliance, will be revised. For the sake of brevity, the 3GPP standards on and related to mobile traffic offloading can be categorized in to three stages of maturity or evolution as *Stage 1* –

Feasibility Studies, *Stage 2 - System Architectures, Descriptions and Procedures*, and *Stage 3 - Technology Enhancements and Extensions*.

Stage 1 - Feasibility Studies

Feasibility study of internetworking 3GPP systems with WLAN has been done by 3GPP Technical Specification Groups (TSGs), in [63], [125] and [126] as early as in 2002. Mainly, these documents identify and describe first, scenarios for the internetworking, second, the internetworking service requirements and third, guidelines for standardization of the internetworking system. The documents, after briefing on a generic and simplified 3GPP system to WLAN internetworking reference model to be used as a reference, as shown in Figure 2.18, went on describing various progressive scenarios. In total, six different scenarios are defined and discussed in detail with respect to their service and operational capabilities as depicted in Table 2.3.

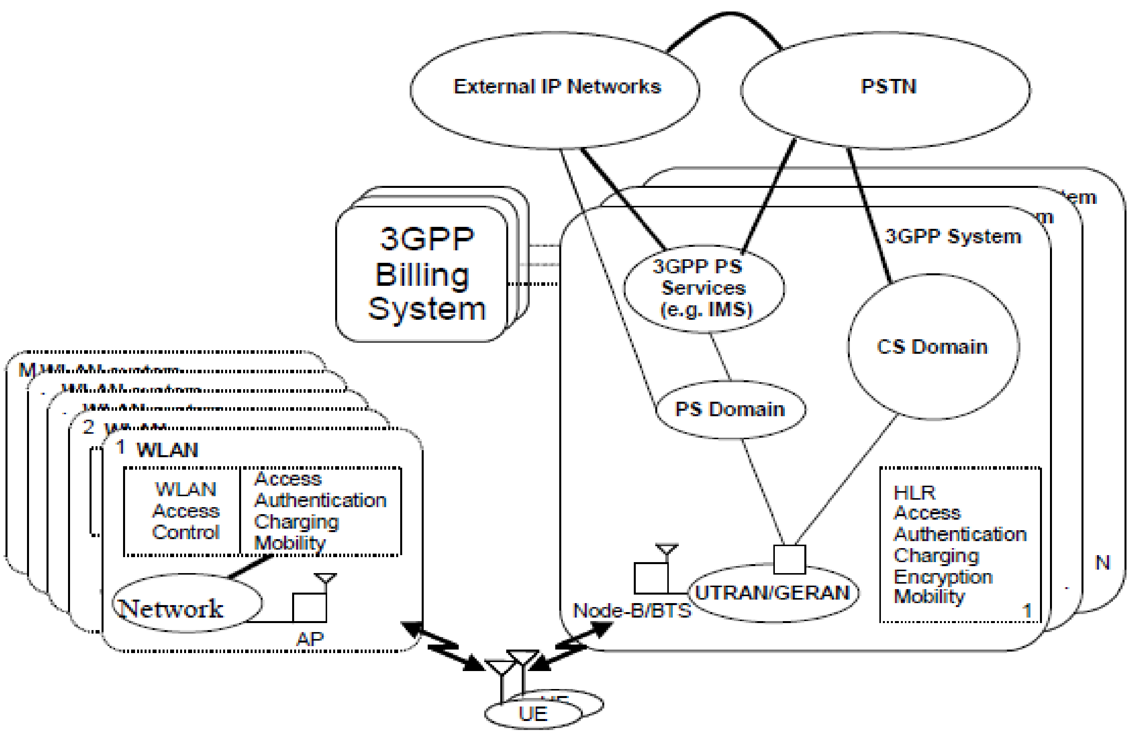


Figure 2.18: 3GPP System to WLAN Inter-networking Simplified Reference Model [127].

Table 2.3: Scenarios and their Capabilities [127].

Scenarios	Scenario 1:	Scenario 2: 3GPP	Scenario 3:	Scenario 4:	Scenario	Scenario 6:

Service & Operational capabilities	Common Billing and Customer Care	system based access control and charging	Access to 3GPP system PS based services	Service continuity	5: Seamless services	Access to 3GPP system CS based services
Common billing	x	x	x	x	x	x
Common customer care	x	x	x	x	x	x
3GPP system based access control		x	x	x	x	x
3GPP system based access changing		x	x	x	x	x
Access to 3GPP system PS based services from WLAN			x	x	x	x
Service Continuity				x	x	x
Seamless Service Continuity					x	x
Access to 3GPP system based services with seamless mobility						x

The documents further discussed other essential considerations that need to be made during deployment of the internetworking system including, but not limited to, ownership of the WLAN, operations like charging and billing, internetworking trust, various service capabilities as use cases, and service requirements for each of the six scenarios. Other feasibility study that is related to offloading is 3GPP TR 23.771 [128]. Specifically, it studies the mechanisms needed to support IMS Emergency sessions over WLAN, both in the Trusted-S2a and in the untrusted-S2b cases or interfaces (please refer to Figure 2.23 and 2.24 for the interfaces). The support of emergency session is only studied for UEs which, first, have valid credentials to access EPC over WLAN and, second, are authorized to connect to EPC over WLAN in the location where they initiate an emergency session. In another phase, as a continuity of the above study, it investigated cases like users who are not authenticated or has no USIM, roaming cases, support of session continuity at inter-access mobility, and Trusted Wi-Fi Access Network (TWAN) accesses.

3GPP in TR 23.937 studies and defines the appropriate solutions for supporting mobility and roaming between 3GPP-WLAN internetworking system so that ongoing 3GPP packet switched based services can be maintained with minimal impact on the end-user's perceived quality on the services at a change of the access network [129]. The solutions should impose minimum changes to the 3GPP packet switched core network and the terminals as well as the WLAN access.

Stage 2 - System Architectures, Descriptions and Procedures

As early as in 2002 3GPP worked on the functional and architectural definition of 3GPP system to WLAN internetworking [63]. The work describes high level requirements and principles of the internetworking mechanism like access control requirements, authentication methods and charging requirements. It also discussed possible internetworking architectures for the cases of roaming and non-roaming scenarios, described the reference points, protocols, procedures (for instance, for authentication, authorization and subscriber profile update). Though the document declared it is as generic (can be used for any 3GPP networks), the era and some discussions are based on the GSM technology. Here one has to remember that mobile 3G technologies like WCDMA and CDMA2000 were launched in June 2001 showing the much early work of 3GPP towards internetworking.

3GPP in its latest release TS 23.234 gave detailed description of 3GPP to WLAN internetworking system [38]. In this document 3GPP reaffirmed the intent of the internetworking idea by saying *"The intent of 3GPP–WLAN Inter-networking is to extend 3GPP services and functionality to the WLAN access environment providing bearer services allowing a 3GPP subscriber to use a WLAN to access 3GPP PS based services."* The description can also ply to other IP based access networks like LAN, WiMAX and Frame Relay. Extending the functional and architectural definition made in [63], this document defines 3GPP-based system architectures and procedures to do the following:

- Provide AAA services to the 3GPP-WLAN internetworking system based on subscription.

- Provide access to the locally connected IP network, e.g. the Internet, if allowed by subscription.
- Provide WLAN UEs with IP bearer capability to the operator's network and PS services, if allowed by subscription.
- Provide WLAN UEs with IP bearer capability to access IMS emergency calls for both UICC and UICC-less cases.

This document is further enhanced by the details in [29] and [130], among others. The main architectural and functional suggestions can be summarized as follows. A 3GPP System can be interworked with WLAN in any one of two access types; *trusted* and *untrusted* non-3GPP accesses. Non-3GPP access includes any packet switched network like Wi-Fi, WiMAX, LAN and CDMA. 3GPP systems range from GSM/GPRS to LTE-A, where here emphasis is given for 3G and 4G networks.

Untrusted non-3GPP access for Wi-Fi: For the first time in 3GPP documents, the issue of *internetworking trust* is described and defined in 2002 [126]. At that time, Wi-Fi security features were not well developed and Wi-Fi was regarded as unsecure. Untrusted access includes any Wi-Fi access that the mobile operator has no control over, like public Wi-Fi, home Wi-Fi, and corporate Wi-Fi. It also includes Wi-Fi accesses that don't provide any adequate security mechanisms like authentication and encryption.

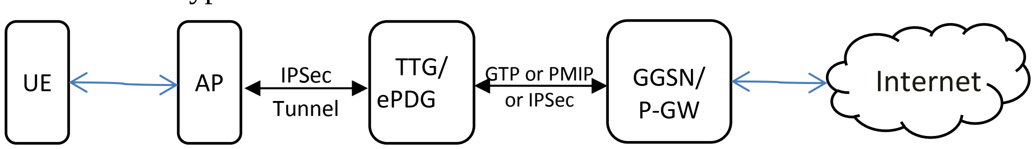


Figure 2.19: Untrusted non-3GPP access.

The untrusted model, as shown in Figure 2.19, requires no change to the Wi-Fi Radio Access Network (RAN) but has an impact on the device side which requires an IPSec client in the device as defined in [131]. In 4G networks the device is connected directly to the Evolved Packet Data Gateway (ePDG) in the EPC through a secure IPSec tunnel. The ePDG is connected to the P-GW where each user session is transported through a secure tunnel via GTP [132] or PMIP [26]. A similar concept is also used in non-EPC 3G networks where the device is connected to a Tunnel Termination Gateway (TTG) through a secure IPSec tunnel. The TTG is in turn connected to the GGSN via GTP [130] [100]. Because the communication is secured end-to-end between the device and EPC, this option can be used with any Wi-Fi network.

Trusted non-3GPP access for Wi-Fi: Trusted non-3GPP Wi-Fi access was first introduced with the LTE standard in 2008 [133]. Trusted access is often assumed to be an operator-built Wi-Fi access with encryption in the Wi-Fi RAN and a secure authentication method. However, it is always up to the home operator to decide what is to be considered trusted. In practice the Wi-Fi access network must support the following features to be considered as trusted; 1) 802.1x-based authentication which in turn requires encryption of the RAN, 2) 3GPP-based network access using EAP method for authentication, and 3) IPv4 and/or IPv6 support.

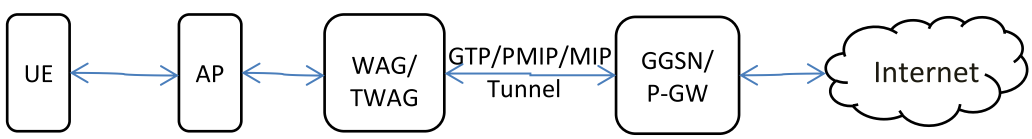


Figure 2.20: Trusted non-3GPP access.

In a trusted access case, as shown in Figure 2.20, the UE is connected through a Trusted Wireless Access Gateway (TWAG) in the Wi-Fi core. The TWAG is in turn connected directly with the Packet Gateway (P-GW) in the Evolved Packet Core (EPC) through a secure tunnel GTP [130][134], MIP [135] or PMIP [26]. IETF in RFC 5555 defined Mobile IPv6 Support for Dual Stack Hosts and Routers [24]. This document specifies extending the Mobile IPv6 and Network Mobility specifications to allow the registration of IPv4 addresses and prefixes. A similar concept is also used in non-EPC 3G networks where a Wireless Access Gateway (WAG) is connected with the GGSN through a secure GTP tunnel [130][134]. Parameters in the subscriber profile are needed in order to setup the GTP tunnel. This will normally require knowledge about the user's International Mobile Subscriber Identity Module (IMSI). Therefore, trusted 3GPP Wi-Fi access is not possible for devices without SIM cards. The following two figures (Figure 2.21 and 2.22) are taken from [29] that show the non-roaming (single operator) scenario within the EPS using two possible set of interfaces.

Mobility Management Protocols by Internet Engineering Task Force (IETF) [24], [26] and [135] defined mobility management protocols in its Request For Comments (RFCs) documents. The basic essence of the matter is that the basic IP protocols did not address mobility of the end-devices and these are handled in a series of Mobile-IP standards. These can be grouped into two main categories: client-based and network-based solutions. Client-based solutions require some special functionality in the client device, and make use of a mobility agent in the network, whereas, network-based solutions rely on the network for both agent and client functionalities, thus making the mobile device agnostic to these mobility functions and therefore simpler to implement. One of the main goals of any of these mobility protocols is to provide seamless mobility as the device moves from network to network. This is essentially achieved by preserving the IP address of the mobile device via the concept of Home IP Address (which stays invariant) and associated Care-Of IP Address (which changes due to mobility). The main client-based approach used to provide seamless mobility is based on the Mobile IP (MIP) protocol [24], which lately has been extended into the Dual-Stack Mobile IP (DSMIPv6) architecture [24]. The main network-based approaches are based on the Proxy Mobile IP (PMIP) protocol [RFC 5213], also an extension of the MIP protocol.

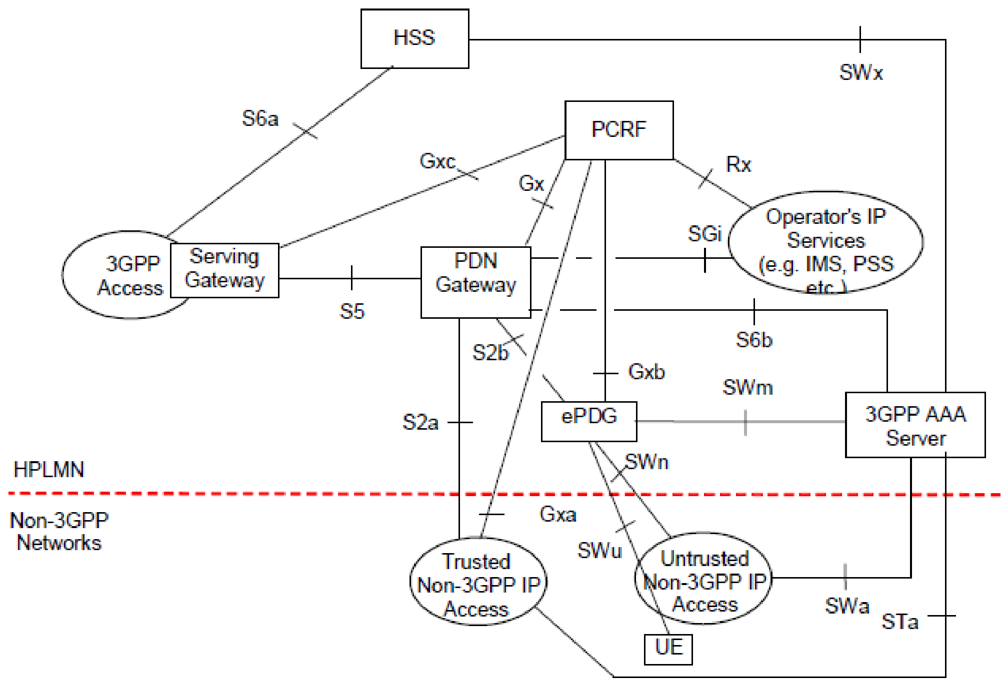


Figure 2.21: Non-Roaming Architecture within EPS using S5, S2a, S2b [29].

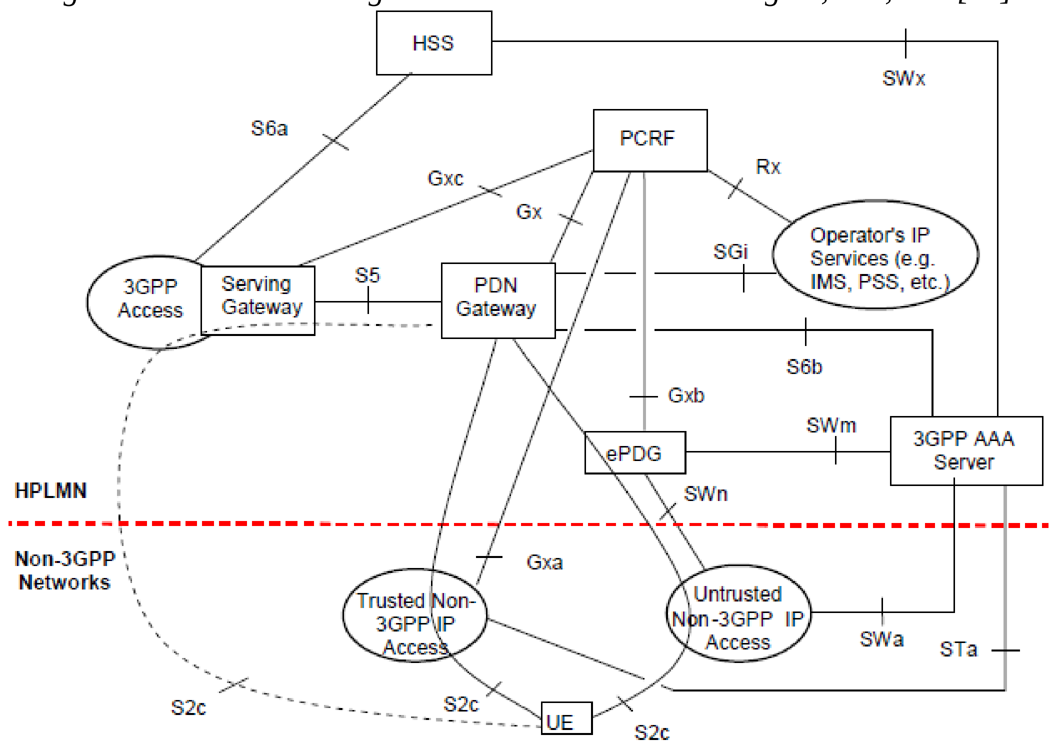


Figure 2.22: Non-Roaming Architecture within EPS using S5, S2c [29].

<u>Components</u>:

AAA – Accounting, Authentication and Authorization
ePDG – evolved Packet Data Gateway
HSS – Home Subscriber Server
PCRF – Policy and Charging Rules Function
PDN-Gateway (also known as . P-GW) – Packet Data Network Gateway
S-GW – Serving Gateway

UE – User Equipment e.g. mobile device

Interfaces:

Gx – It provides transfer of (QoS) policy and charging rules from PCRF to Policy and Charging Enforcement Function (PCEF) in the PDN GW.

Gxa – It provides transfer of (QoS) policy information from PCRF to the Trusted Non-3GPP accesses.

Gxb – This interface is not specified within this Release of the specification.

Gxc – It provides transfer of (QoS) policy information from PCRF to the Serving Gateway.

S2a – It provides the user plane with related control and mobility support between trusted non 3GPP IP access and the Gateway.

S2b – It provides the user plane with related control and mobility support between ePDG and the Gateway.

S2c – It provides the user plane with related control and mobility support between UE and the Gateway. This reference point is implemented over trusted and/or untrusted non-3GPP Access and/or 3GPP access.

S5 – It provides user plane tunneling and tunnel management between Serving GW and PDN GW. It is used for Serving GW relocation due to UE mobility and in case the Serving GW needs to connect to a non-collocated PDN GW for the required PDN connectivity.

S6a – This interface is defined between MME and HSS for authentication and authorization. It is defined in TS 23.401 [4].

S6b – It is the reference point between PDN Gateway and 3GPP AAA server/proxy for mobility related authentication if needed. This reference point may also be used to retrieve and request storage of mobility parameters. This reference point may also be used to retrieve static QoS profile for a UE for non-3GPP access in case dynamic PCC is not supported.

SGi – It is the reference point between the PDN Gateway and the packet data network. Packet data network may be an operator external public or private packet data network or an intra-operator packet data network, e.g. for provision of IMS services. This reference point corresponds to Gi for 3GPP accesses.

STa – It connects the Trusted non-3GPP IP Access with the 3GPP AAA Server/Proxy and transports access authentication, authorization, mobility parameters and charging-related information in a secure manner.

SWa – It connects the Untrusted non-3GPP IP Access with the 3GPP AAA Server/Proxy and transports access authentication, authorization and charging-related information in a secure manner.

SWm – This reference point is located between 3GPP AAA Server/Proxy and ePDG and is used for AAA signaling (transport of mobility parameters, tunnel authentication and authorization data). This reference point also includes the MAG-AAA interface functionality, RFC 5779 [43] and Mobile IPv6 NAS-AAA interface functionality, RFC 5447 [44].

SWn – This is the reference point between the Untrusted Non-3GPP IP Access and the ePDG. Traffic on this interface for a UE-initiated tunnel has to be forced towards ePDG.

SWu – This is the reference point between the UE and the ePDG and supports handling of IPSec tunnels. The functionality of SWu includes UE-initiated tunnel establishment, user data packet transmission within the IPSec tunnel and tear down of the tunnel and support for fast update of IPSec tunnels during handover between two untrusted non-3GPP IP accesses.

SWx – This reference point is located between 3GPP AAA Server and HSS and is used for transport of authentication, subscription and PDN connection related data.

Rx – This is the reference point between the (plication Function) AF and the PCRF's component plication Manager (AM). Enables the AF subscription to notifications on IP-CAN bearer level events (e.g. signaling path status of AF session) in the IP-CAN.

The above architecture for the trusted and untrusted 3GPP access network for WLAN is worked out by 3GPP. The beauty of the recommendation is that 3GPP used predefined standards from many other standardization bodies like IEEE, ITU, and IETF. This fact is depicted on Figures 2.19 and 2.20, for trusted and untrusted 3GPP accesses. Here, a little deliberation is made on the major enabling standards from the aforementioned standardization bodies.

MIP supports uninterrupted routing of IP-packets to and from a mobile device and provides session continuity by means of a Home Agent (HA), which is an entity located at the Home Network of the mobile device (also referred to as a Mobile Node – MN) that anchors the permanent IP address assigned to the mobile device, known as Home Address (HoA). The HA keeps the device's HoA when the MN has moved from the home network, and redirects traffic to the device's current location. The HA is informed of the current location by the MN, using a temporary IP address or Care-of Address (CoA) that the MN acquires from the visited network. A bi-directional IP-tunnel between the MN and the HA is then used to redirect traffic between these nodes. MIP, defined originally for IPv4 devices and networks, was subsequently extended to MIPv6 to be applicable to IPv6 devices and networks. DSMIPv6 is a further extension of MIPv6, where the basic mobility functionality is extended to support dual stack IPv4/IPv6 devices and networks. Accordingly, DSMIPv6 extensions are also defined to register IPv4 addresses and transport of both IPv4 and IPv6 packets over the tunnel between the HoA and the visited network. These extensions enable the mobile device to roam between IPv4 and IPv6 access networks seamlessly, and are considered crucial as IPv4 networks and devices gradually evolve to IPv6.

PMIP and its IPv6 extension, PMIPv6, are examples of Network-based IP mobility solutions, which manage the mobility of the mobile device entirely to the network. In this way, the device is not required to perform any signaling or updates, as it changes of its point-of-attachment (i.e. visited network) due to its mobility. Hence, these changes become transparent to the mobile terminal's IP protocol stack, resulting in simpler device solutions than those based on baseline MIP.

The PMIP-enabled mobile IP network architecture consists of a central entity, called Local Mobility Anchor (LMA), and a number of Mobile Access Gateways (MAGs), which together define a mobility domain. The LMA plays the role of a local HA (as in DSMIP networks) and anchors the

IP prefixes used by the MNs. MAGs reside close to the mobile node, usually in the Access Routers (which in turn are either collocated with the APs or directly connected to them).

Detection of movement of mobile devices as well as implementing associated signaling is done by the MAGs. Typically, the MAG detects mobility through standard terminal operations, such as router and neighbor discovery or by means of link-layer support, without any mobility specific support from the device. Bi-directional tunnels between the LMA and the MAGs are set up, so that the mobile device is able to keep the originally assigned IP address within the mobility domain despite any location changes. Since the LMA is aware of the actual location of the mobile device, any packets addressed to the device are tunneled to the appropriate MAG, relieving the mobile device of the need to manage the IP packet routing due to its own mobility.

A typical application of MIP, PMIP, DSMIP, IPSec and GTP can be seen from Figures 2.20 and 2.21 above where they are implemented via the interfaces S2a and S2b for the first three and SWn and Wu for IPSec and S2a/Gn' for GTP.

Stage 3 - Technology Enhancements and Extensions: There are works by 3GPP that add some new features to facilitate and support offloading as extensions and enhancements. As a matter of fact, non-3GPP access network [29] is an extension of 3GPP 3G and 4G pure cellular architectures to support non-3GPP accesses like WiMAX, WLAN and LAN. Among the various extensions to 3GPP systems that are related to WLAN offloading are ANDSF mobility object [136], IP flow mobility and seamless WLAN offload [137][138], Location Services architecture for 3GPP WLAN internetworking systems [139], and Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) [133] are the most outstanding ones.

In addition to enhancing existing features there are works to optimize the performance of current systems [140]. In TR 23.890, 3GPP described optimization of WLAN access during mobility between 3GPP RATs. It identifies potential inefficiencies and suggests new solutions to address them. It also describes further extensions to ANDSF in order to enable policy differentiation for 3GPP RATs and WLAN to enable a UE to distinguish preferences of WLAN with respect to specific 3GPP RATs upon acquiring connectivity or upon mobility.

ETSI TS 123 402 v12.9.0 in [130] describes UMTS/LTE Architecture Enhancements for Non-3GPP Accesses and Stage 2 Service Description for providing IP connectivity using non-3GPP accesses to the Evolved 3GPP PS Domain where the protocols between its Core Network elements are IETF-based. ITU-T Recommendation I.130 [128] details a three-stage method for characterization of telecommunication services, and ITU-T Recommendation Q.65 [150] defines stage 2 of the method. The specification covers both roaming and non-roaming scenarios and covers all aspects, including mobility between 3GPP and non 3GPP accesses, policy control and charging, and authentication, related to the usage of non-3GPP accesses. TS 23.401 [141] covers architecture aspects common to the Evolved 3GPP Packet Switched domain. It also defines detailed considerations for Non-seamless WLAN offload.

Other standards, of 3GPP or other standardization bodies, related to offloading are presented as follows. 3GPP TS 123.303 define Proximity-based Services (ProSec) [142] that describes direct communication paths between two or more ProSec-enabled UEs via E-UTRAN or WLAN. Some architectural enhancements to the 3GPP system to support Proximity-based Services are suggested in [143]. Other 3GPP system specifications and enhancements can be obtained from the official website [77].

3GPP is a collaboration of telecommunications associations called Organizational Partners who are grouped together to make a globally applicable 3G mobile phone system specification based on GSM within the scope of IMT-2000 of ITU [77]. The participating members are mainly from Europe, USA, Japan, China, Korea and India. Though latter the scope is extended to the development and maintenance of GSM, GPRS, EDGE, UMTS (like HSPUA, HSPDA, HSPA), and LTE. Its work encompasses Radio, Core Network and Service architectures. Whereas, 3GPP2, though a similar consortia of companies trying to standardize mobile phone system specifications, is based on ITU's IMT-2000 project for CDMA2000 based on its predecessor cdmaOne. Here, the participating associations are the same as for 3GPP excluding Europe. Some works that support WLAN offloading of cellular traffic are the 3GPP2 A.S0025-0 defining Interoperability Specification for cdma200 and M2M communications [132] and the 3GPP - WLAN Inter-networking specification 3GPP2 S.R0087-0 describing the requirements of the internetworking between the two networks extending 3GPP2 packet data services and capabilities to the WLAN environment [144].

IEEE 802.21, also known as . Media Independent Handover services [145][146], is a standard that enables seamless handover, not only, between homogeneous networks but also among heterogeneous networks. It provides data flow among IEEE 802.3, 802.11, 802.15, 802.16, 3GPP and 3GPP2 networks via different handover mechanisms. Another effort by IEEE is the development of Cellular Traffic Offloading Use cases for 802.11ah [147]. In this work the offloading requirements for future 802.11ah technology is made to ensure the solution is viable technically and economically. The IEEE Std 802.11u, named WLAN MAC PHY Specification – Amendment 9 – Inter-networking with External Networks [148], defines enhancements to the IEEE 802.11 MAC to support WLAN internetworking with external networks like 3GPP networks.

ITU-T Q.3202.1 under its *Series Q: Switching and Signaling recommendation* document specified authentication protocols based on E-AKA for internetworking among 3GPP, WiMAX, and WLAN for next generation networks [149]. This recommendation proposes to ply the E-AKA protocol to non-3GPP network devices not equipped with UICC for internetworking among the three networks mentioned before. 3GPP, in its standard document 3GPP TS 33.234, already defined similar protocol for internetworking between 3GPP, WiMAX and WLAN networks. But the WiMAX or WLAN device require an external UICC reader to ply the E-AKA. This ITU-T recommendation removes this setback.

When it comes to IETF, in its RFC 4555, for instance, describes IKEv2 Mobility and Multihoming Protocol (MOBIKE) that allow IP addresses associated with IKEv2 and tunnel mode IPSec security associations to change. That enables the mobile client to keep the connection with the gateway active while moving from one address to another [150].

Wi-Fi Certified Passpoint launched in 2012 as an industry-wide solution to streamline network access in hotspots and eliminate the need for users to find and authenticate a network each time they connect. In Wi-Fi networks that do not support Passpoint, users must search for and choose a network, request the connection to the AP each time, and in many cases, must re-enter their authentication credentials. Passpoint automates that entire process, enabling a seamless connection between hotspot networks and mobile devices with the highest WPA2 security support. Passpoint is enabling a more cellular-like experience when connecting to Wi-Fi networks. Passpoint was developed in Wi-Fi Alliance through partnerships between mobile device manufacturers, network equipment vendors, and operators.

2.3.3 Product Manufacturers

Hyperboles of product manufacturers and marketers are examined, in this section, to see the extent of the realization of research outcomes and conformity with standardization bodies' technical specifications and recommendations.

Ruckus Wireless Inc. advertised their implementation of a set of PMIP based interfaces, such as S2a and S2b, in their product Ruckus SmartCell Gateway (SCG) and the PDSN/PGW to enable the seamless integration and handover between access technologies, mainly via 3G and WLAN [151]. Ruckus SCG enables, as they claimed, handover and roaming between WLAN and 3G networks. The whole implementation in embedded in to Ruckus Access Point. Other products that support some feature of offloading include Ruckus carrier scale wireless controller and gateway that supports 3GPP TTG/PDG.

Qualcomm Inc. advertised a 3G/4G and Wi-Fi offload framework as connectivity engine (CnE), as shown in Figure 2.21, to manage inter-system radio connections and applications [152]. Specifically, it's declared that their CnE has many attractive features like a mechanism to provide operator's policy to the device in a dynamic fashion, algorithms in the device to detect characteristics of unplanned Wi-Fi networks, and a mechanism to allow seamless handovers between 3G/4G and Wi-Fi networks. CnE, optionally, uses DSMIPv6 and IP Flow Mobility (IFOM) to ensure a smooth handover (session continuity) and mobility that works in a make-before-break manner.

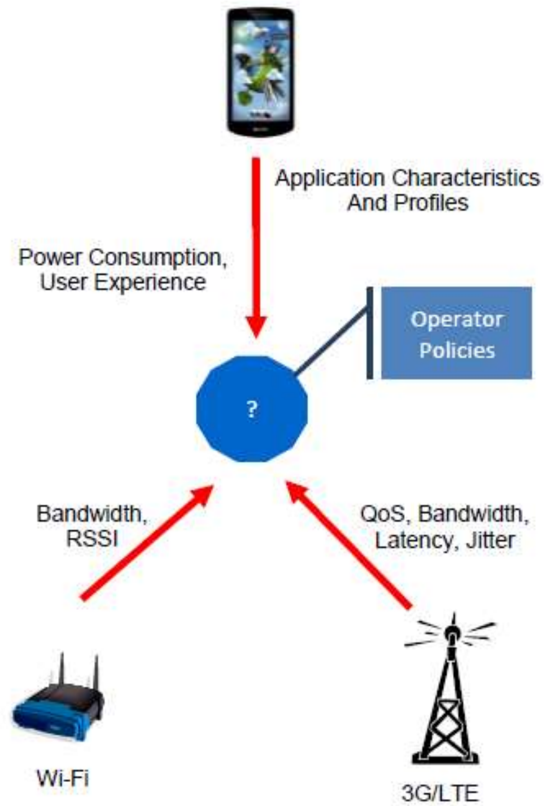


Figure 2.23: Context aware CnE to enable seamless connectivity as described in [152].

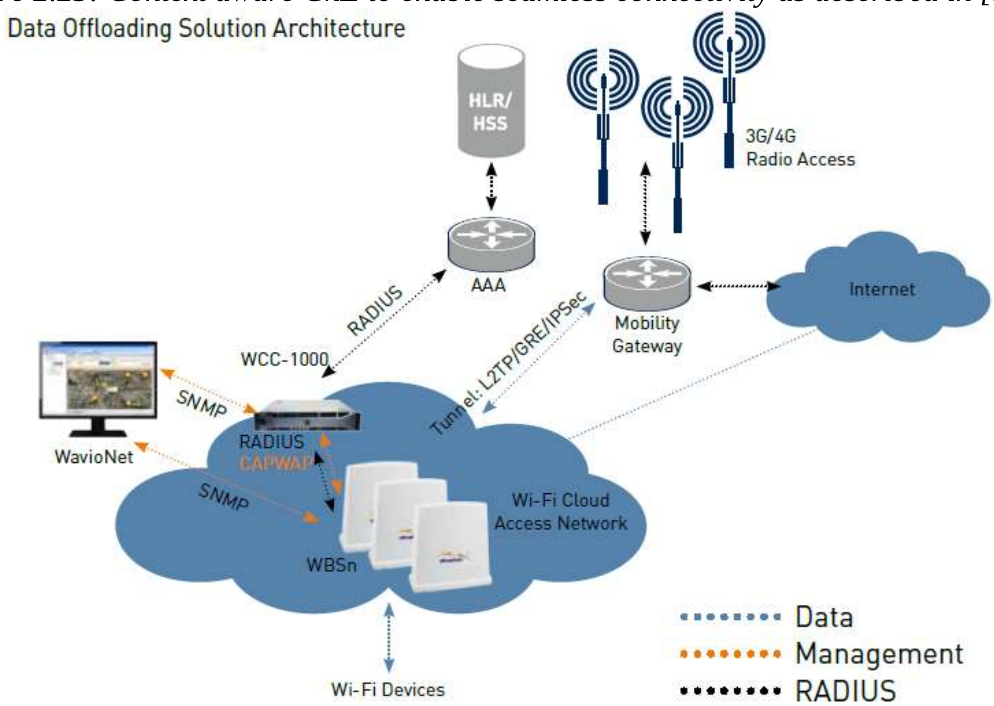


Figure 2.24: Data offloading solution architecture by Alvarion [153].

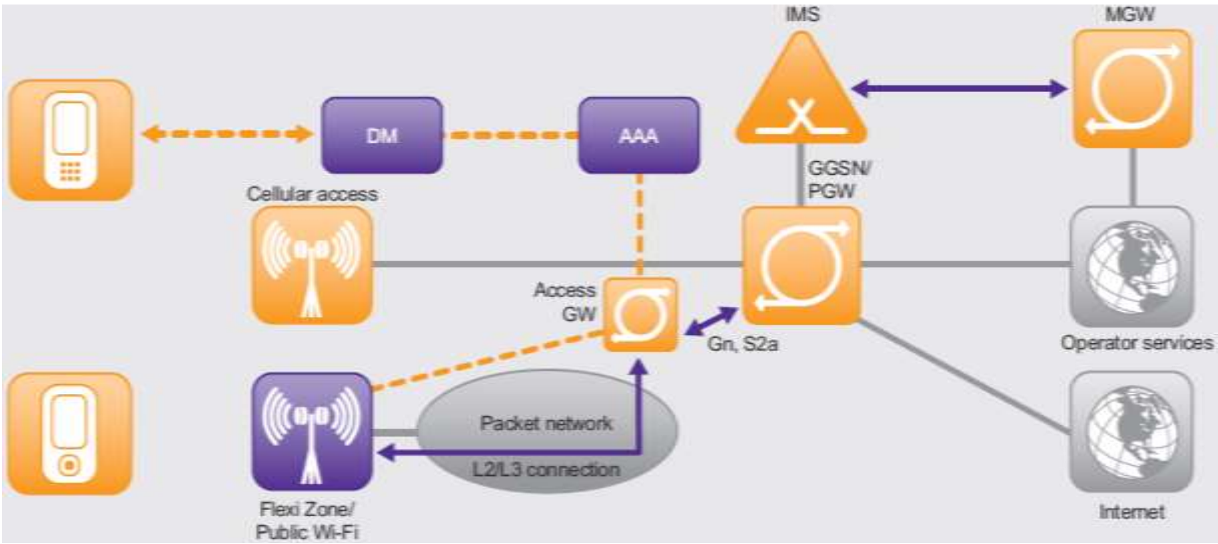


Figure 2.25: Trusted Wi-Fi access integration with the packet core network advertised in [79].

Alvarion Ltd. is Israeli based company claiming a mobile data offloading solution for 3G/4G networks [153]. Their solution, dubbed as Wi-Fi Base Station (WBSn), is based on the IEEE 802.11n standard that supports 2.4 and 5 GHz radios, 3x3 MIMO and three spatial data streams. The overall architecture is shown in Figure 2.22. WavioNet and AlvariStar are two network element management solutions with full FCS[5] element management support, providing a single point of management for controlling and configuring WLANs. Another solution specified is the WCC-1000 which is Alvarion's Wi-Fi Cloud Controller. Each WCC-1000 described as cable of managing tens of thousands of users. Multiple WCC-1000 units may be used, with redundantly, for a true scalable, pay-as you-grow network of millions of Wi-Fi users. Other attractive features include, RADIUS based authentication proxy, base station registration, mobility support, and resource management.

Cisco Systems Inc. enumerated products that can be used in HetNets for various purposes [154]. Among these is Cisco ASR 1000 series, also known as . Aggregation Services Routers and Cisco Intelligent Services Gateway (ISG), which are products claimed to offer flexible policy and

[5]FAPCS (an acronym for *fault, configuration, accounting, performance, and security*) is a model/framework for network management as defined by ITU-T (available in http://www.itu.int/rec/T-REC-X.701-199708-I and http://www.itu.int/rec/T-REC-M.3010)

authentication of Wi-Fi subscribers and full set of the functions required from the WAG. Another one is the Cisco ASR 5000 Series having features that can support packet core functions and, with the Small Cell Gateway, be able to provide functions relevant to Wi-Fi/3G/4G handover capabilities such as P-GW, TTG, ePDG, and PCEF. A similar set of products are announced by Motorola [155]. They are categorized in to three key areas: indoor, outdoor, and large area. Their offloading features range from basic data offload to the more complex integrated data or data and voice offload using the Generic Access Network Controller (GANC) or the I-WLAN integrations. These features are implemented and available in AP series (like 300, 650, 6511, 5131, 5181, 7131 and 7181), the RFS WLAN controller series (6000 and 7000), the backhaul and management products with the PTP series (100, 200, 300, 500, 600 and 800), and the PMP series (100, 320 and 430). Moreover, it is stipulated that the AirDefense, LANPlanner, Wireless Manager and BroadbandPlanner are the products that provide the various features for the successful offloading deployment and management.

Other products that provide specific feature of offloading includes the Qosmos Mobile Data Offloading solutions that implement differentiated flow steering via a reliable application detection technology relying on their own DPI (Deep Packet Inspection) engine to implement better offloading strategies on total flow visibility [156]. The DPI Engine is embedded in WLAN Controllers named ixEngine built in to x86 based server with Intel DPDK (Data Plane Development Kit) data plane for high speed packet processing. Similarly, Netkrom in [157] and Aptilo Networks in [88] advertised a niche of products for 3G and WLAN offloading. The former describe a product that can act as a bridge between the core network and the access network like the TDM Microwave and the Wireless Ethernet Bridge. Whereas, the later provide a Service Management Platform (SMP) that manages traffic offload to Wi-Fi further supporting local breakout offloading for the whole 3G or 4G network including the mobile core.

Apart from the aforementioned products providing a narrow and very specific offloading features there are manufacturers that announce products with full-fledged offloading functionality. Aruba Networks in [158] offers, as they claim, *a complete edge-to-core solution* for offloading 3G/4G networks. In the mobile core it has Aruba Mobility Controller, Aruba AirWave Network Management, and Aruba ClearPass. In the access network it has Aruba Instant Access Points, Aruba Mobility Controller, and Aruba Campus Access Points. It also advertised the support of third-party applications providing business intelligence, advertised services, and content security via, what they call it, Aruba Activate Zero Touch Provisioning. Aruba Networks Inc. is marketing another set of products that can support different offloading features as described in [159]. In the same token, Nokia Siemens Networks claimed having such comprehensive offloading solution, dubbed as, Smart Wireless LAN Connectivity (SWLANC) between Wi-Fi and 3GPP networks, specifically HSPA and LTE, making the former a seamless extension of the later [79] as depicted in Figure 2.23. It is an overlay to existing mobile and Wi-Fi networks allowing operators to exploit existing core and radio network infrastructures. Their Flexi Zone APs support simultaneous HSPA, LTE and Wi-Fi s declaring it as *the first small cells implementation* using the concept of Liquid Radio baseband pooling. Swedish based company, named Clavister, is yet another manufacturer who announces an intelligent and secure mobile data offloading solution nicknamed Clavister Intelligent Mobile Offload (IMO). Its specialty is the ability to directly execute on COTS x86 platforms or embedded in hypervisors such as VMware or KVM supporting Intel Virtualization Technology [160].

Wi-Fi Alliance also announced, with detailed technical specification as described in [161], Wi-Fi certified Passpoint, also known as HotSpot 2.0 (HS2.0), with the following key features: automated network discovery and selection, seamless network access and roaming between hotspots, immediate account provisioning, WPA2-Enterprise security, and easy implementation of operator's policies. These services are rendezvous between the Hotspot 2.0 entities - APs, mobile devices, hotspot operators and service providers [162]. The architecture of Passpoint-enabled WLAN and mobile device states are depicted in Figure 2.6 and 2.27.

A very interesting, rather unique, product is announced by Anritsu in [163]. The product is offloading test solution named MD8475A 3GPP BTS Simulator. It is a test environment for services like Access Network Discovery and Selection Function (ANDSF), ePDG, PCEF, and PDN-GW. It's claimed that the product can provide three different kinds of test cases: functional test cases for the authentication functions, connectivity test cases between 3GPP and Wi-Fi networks, and composite test cases to test the validity of various services provided on 3GPP and Wi-Fi networks concurrently.

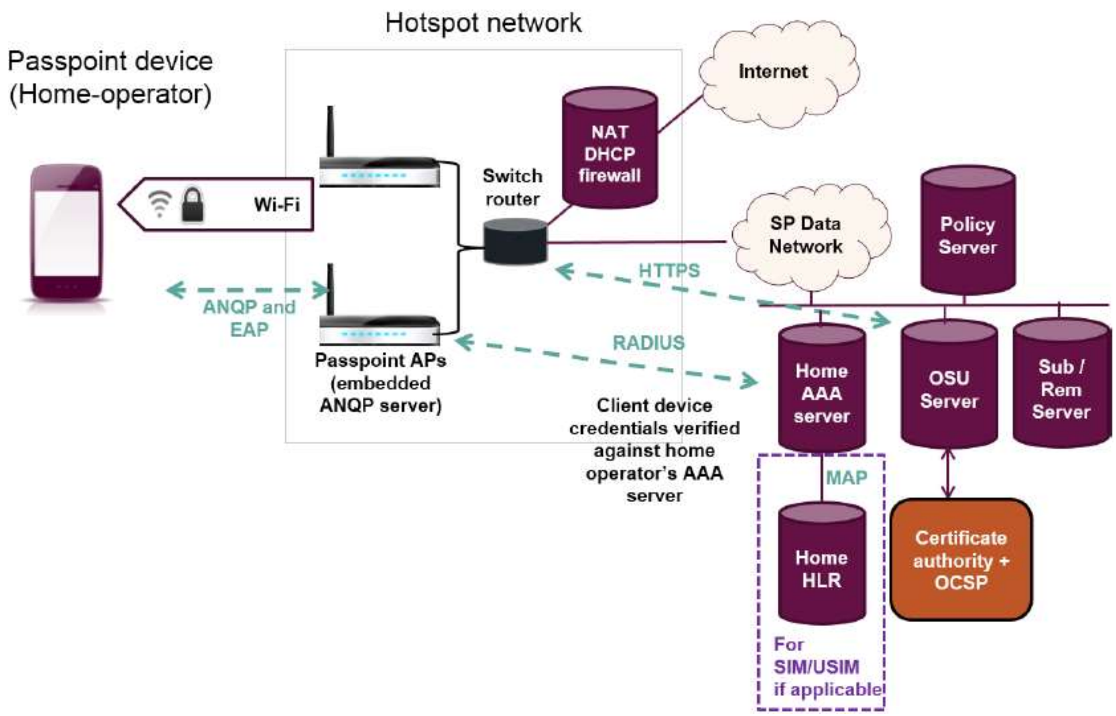


Figure 2.26: Simple architecture diagram of Passpoint-enabled Wi-Fi network [162].

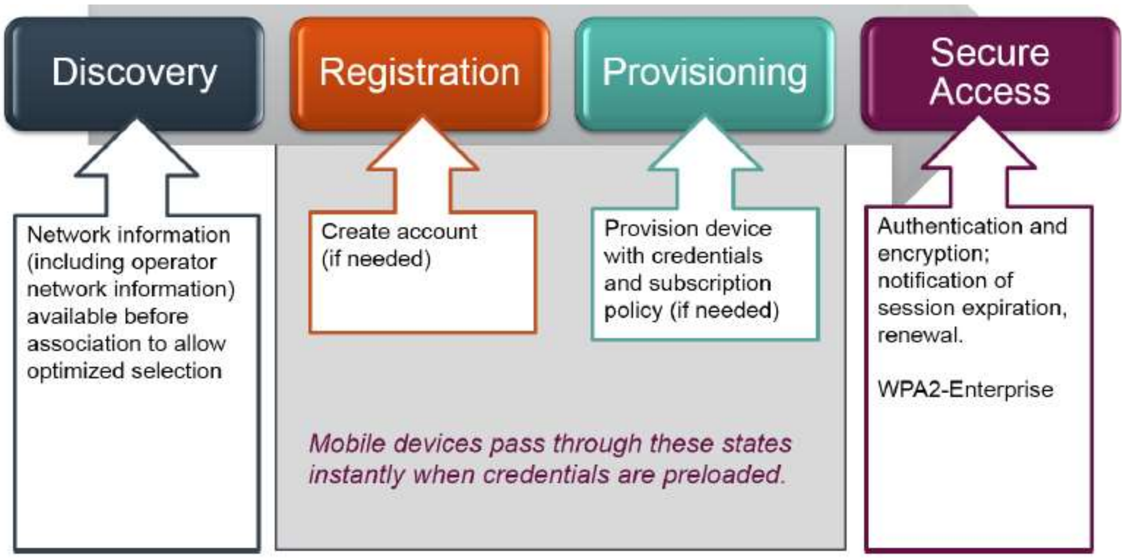


Figure 2.27: Passpoint-enabled mobile device states [162].

2.4 Current Challenges and Open Issues

To realize such a feature-rich internetworking architecture there are many challenges to be investigated and worked out. These challenges and open issues are summarized into five major categories as summarized below.

Interoperability, Deployment and Mobility Related: A unified fully integrated internetworking architecture supporting the various wireless access technologies readily available in mobile devices is what users and operators are looking for. There is little work in this regard. Such architecture needs to perform advanced traffic differentiation based on QoS requirements using optimized and improved packet inspection techniques. There are many techniques proposed to seamlessly handle vertical handover between cellular and WiMAX/WLAN. As there is no single solution that fits all which mechanism best suites what kind of scenario deserves critical investigation. The interoperability of cellular and WLAN regarding automatic user authentication, admission control, and guaranteed QoS is one area that deserves deep investigations. Moreover, what such techniques are required to steer traffic among WiMAX/WLAN and D2D technologies is another challenge to be worked out. A similar issue must be investigated between cellular and D2D technologies.

Furthermore, user-controlled versus network-controlled WLAN offloading and network-assisted versus autonomous D2D offloading together with the inband and outband scenarios are not yet fully understood. These separate efforts should come together to bring an efficient and viable solution for mobile traffic offloading. These issues need to be worked out for improved user experience and operator acceptability.

The mobility of users or nodes based on human behavior and movement patterns has to be investigated especially for opportunistic communications. Smarter connection management techniques and usage agnostic offloading process that facilitate improved session continuity and better inter-technology management is required for the existing and emerging mobile devices and services. With regard to Wi-Fi deployment, besides the traditional cite surveying and deployment planning, the question of at which locations Wi-Fi APs should be placed so as to maximize coverage and return of investment is critical for both the user and the operator, respectively.

Security and Privacy Related: It is known that wireless communication systems are more vulnerable for security and privacy threats than wired systems. Therefore, the security and integrity of the overall internetworking system on top of the individual users and devices deserve investigation. Distributed trust, as described in [93], is one proposal especially between user devices since one user is *'strange'* to the other. The high dynamicity of users together with the ad hoc nature of communications makes the issue of security and privacy more complex.

D2D communication entails a different and unmatched security and privacy vulnerability and threat to the overall system. Though there are some works to handle such challenges as in [5] and [92] further real world scenario-based investigation is needed.

Performance Related: The various overheads introduced between two UEs negatively affect the user experience and performance of the overall offloading system. The inter-technology switching, the various database accesses and processing (during handoff and roaming, for instance), and core network delays contribute to the overall system delay. This is apparently critical for real-time traffic.

Another critical issue is the power consumption due to multiple interface usage, constantly scanning for connection probes and packet processing. There are some works like designing energy-aware algorithm to be used during the offloading process as in [8] and [13]. Interested readers can refer to [164] which is a very concise survey on the issue. Therefore, how to minimize and optimize the battery power consumption of interfaces and protocols is required for mobile devices that are battery powered.

The performance of D2D technologies, like Bluetooth and Wi-Fi Direct, is another challenge. As devices are distributed and connections are ad hoc the association and re-association mechanism and the associated overhead, interference with other unlicensed devices, and as the connection is sometimes short-lived the efficiency of data transmission mechanisms deserve further investigation. Channelization of lower layer protocols, clustering mechanisms and scheduling algorithms are needed to minimize intermittent link failures and energy wastage.

Applications and Services Related: To use Wi-Fi, especially in ad hoc mode, in real-time applications in dense environments, bring challenges due to interferences, mobility, radio resource management, and propagation characteristics, among others. The variability of the traffic, complex user demands and preferences, incentive mechanisms especially for delayed transmissions, and battery and storage constraints are continuously challenged as the need increases. May be it is time to evolve from traditional Peer-to-Peer communications to distributed D2D transmissions via distributed offloading mechanisms?

Techno-economic Related: The benefit of using Wi-Fi or femtocell APs for cellular traffic offloading is analyzed by Lin Gao et al. in [7] in terms of macro-cell operators collaboration with third party owners for win-win scenario. Such analysis should be extended to D2D, WiMAX and other wireless operators in a comprehensive manner to harvest the benefits of heterogeneous networks fully. Questions like how operators generate revenue, how to incentivize owners and mobile users (for delayed offloading, for instance), devising fair price models, and how to minimize the Total Cost of Ownership (TCO) of carrier grade APs has to be answered diligently.

2.5 Summary

Globally, there is considerable work by researchers in academia and industry towards devising a feature-rich internetworking architecture among cellular, WLAN and D2D technologies since the inception of the idea back in 2002. The main efforts are, due to various reasons, to address some specific feature or aspect of the overall offloading process like VHO, session continuity, neighbor discovery and target set selection, content distribution and movement prediction, user satisfaction and incentivizing. There is also considerable work in devising an integrated architecture or framework that exploits WLANs and D2D technologies at various degrees of investigation.

Therefore, in this laborious review of related works the various efforts of researchers, standardization bodies and product manufacturers are summarized from level of integration, technologies involved and nature of traffic considered. The overall review of related works is summarized using a four-dimensional table as shown in figure 2.26.

However, to the best of our knowledge, there is no single effort towards combining various attractive features, for instance, which is ***fully integrated*** (exploiting various wireless technologies), ***hybrid*** (support different level of integrations between cellular and WLANs) and ***QoS-aware*** (capable of en-routing packets based on their delay or QoS requirements). And, that is where this research work is heading!

	Integrated Technologies												
	3G/-based		3G/D2D		3G/-based/D2D		4G/-based		4G/D2D		4G/-based/D2D		
Academic and Research Efforts	[105][57][68][56][124][13][44][55][73][78][106][69][30][64][123][23][145][103][22][165][60][98][62][104][97][83][166][6][167][99][101][115][66][65][61][102]	[37]	[39][107][122][20][120][71][70][110][121]		[8][168]		[112][81][113][75][6][79][21][80][82][60][83][78][74][169][72]	[34][169][6][27]	[118][117][86][119][42][43][87][92][84][91][90][40][108][89][88][121][170][35]	[85]	[93][94][67]		**Delayed**
	[55]*		[70]* [110]*		[168]*		[78]*				[93]* [94]*		**Real-Time**
Standardization Efforts	[161][125][130][38][63][51]110][24][135][26][136][150][148][171]						[157]* [120]* [126]* [32]* [172][173][133][174][149] [62]* [50]* [22]* [131]* [24]* [132]* [144]*		[98] [102]				**Delayed**
													Real-Time
White Papers	[175][152][153][154][156][158][176][160][155][157][177][162]						[166]* [149]* [150]* [152]* [154]* [163] [167]* [151]* [168]* [158]*	[159]					**Delayed**
													Real-Time

Loosely Coupled	**Tightly Coupled**	**Autonomous**	**Network-Assisted**	**Loosely Coupled**	**Tightly Coupled**	**Loosely Coupled**	**Tightly Coupled**	**Autonomous**	**Network-Assisted**	**Loosely Coupled**	**Tightly Coupled**
Integration Levels											

* indicates a work cited again having additional feature.

Figure 2.28: Summary of the Efforts towards Feature-rich Interworking Architecture from technologies involved, nature of traffic and integration perspectives by the main contributors.

CHAPTER THREE

THE PROPOSED INTERWORKING ARCHITECTURE: FULLY INTEGRATED, HYBRID AND QoS-AWARE

3.1 Introduction

As indicated in the previous chapter, one of the aspirations of the researchers, ISPs and standardization bodies is to have feature-rich interworking architecture for mobile traffic offloading. A ***“full-fledged”*** or ***“fully integrated”*** architecture that interconnects as many wireless technologies as possible (if not wired ones though not the scope of this work), supports both real-time and delay tolerant traffic types and en-route s traffic via multiple coupling capabilities is a big stride forward in this research endeavor.

Therefore, in this work, a ***fully integrated***, ***hybrid*** and ***QoS-aware*** interworking mobile traffic offloading architecture is proposed. It is ***fully integrated*** as it interworks 4G/LTE cellular system with WLAN and D2D technologies. It is ***hybrid*** because it provides both loose and very tight coupling between the 4G network and the WLAN. And, it is ***QoS-aware*** since it supports both delay sensitive and delay tolerant traffic types based on the QoS requirement of the initiated traffic.

In this chapter, the features of the proposed interworking architecture, its major components, their interconnections and interactions, and the algorithm that steer the overall offloading system are presented in detail.

3.2 The Proposed Interworking Architecture

The primary purpose of the proposed mobile traffic offloading interworking architecture is to minimize the burden of cellular networks by diverting as much traffic as possible that is initially targeted to it towards WLAN and D2D systems, as they are readily available in today's mobile devices. As discussed so far in this work, there are many efforts that contribute something towards fulfilling the expectations of, mainly, operators and users in devising a feature-rich offloading system. Such an interworking architecture is a system that has, to the minimum, the following major offloading features:

1. Interwork different wireless technologies: D2D communications (mainly Bluetooth or Wi-Fi Direct) and WLAN with cellular systems.
2. Handles both delay sensitive and delay tolerant transmissions based on the QoS requirements of the applications or services.
3. Some combination of the coupling techniques, like loose and tight or loose and very tight or tight and very tight, between the cellular and WLAN systems.
4. Operator/network-controlled or -assisted D2D that works as an opportunistic access for both the WLAN and cellular systems.

On top of these major features it has to support various offloading possibilities and other crucial features as described in the various works reviewed in the previous chapter. Mobility management, AAA services, security features, and other network management features are vital for the realization of such a system. Mobility management considers, mainly, handover mechanisms within and between the various wireless technologies comprising the offloading solution. The AAA services provide, among others, the needed accounting, charging, and billing system for the operator and users. Security issues are very important because they handle the privacy of users and the integrity of the overall system. Other issues like incentive mechanisms and cooperation strategies can be integrated into the core components or put as part of the network management module. The network management component is responsible for issues related to load balancing, congestion control, QoS/QoE, and security issues.

In this work, these features are assumed to be part of and provided by the cellular network backhaul as it is part of a complex, organized and regulated system. For instance, here in Ethiopia, EthioTelecom is the only national company providing telecommunication services since its inception back in the late 1800's. The ISP provides both Internet and telephone services via its landline, microwave radio, and satellite systems. It is regulated by the Ethiopian Telecommunications Agency (ETA) which is a member of the International Telecommunications Union (ITU). It has GSM (2G), UMTS (3G) and LTE (4G) cellular systems throughout the country. So, all the support and management systems and features needed as part of such interworking offloading system are exploited from the EthioTelecom cellular infrastructure.

To this end, the details of the major components, their interconnections and interactions of the proposed interworking architecture for mobile traffic offloading are described as follows.

3.2.1 Major Components and Their Interconnections

As described in Figure 3.1, the proposed internetworking architecture comprises of three different and independent wireless communication subsystems; cellular network, WLAN, and D2D communication.

The cellular network is based on the 4G or LTE which consists of two user equipment that desire to communicate. The 4G network, as usually the case is, connected to the Internet directly, allowing users to connect to it, for instance, via data plan.

The WLAN is Wi-Fi AP-based network where two users equipment are connected to the AP to share data or access resources from an external network. As the traditional configuration of such networks dictate, it is connected to an external network like an Ethernet-based or broadband network that is ultimately extended to the Internet. This interconnection of the WLAN to an external network represents a loose coupling between the WLAN and the cellular systems. There is no any vertical handover procedure between the two.

The third one is the direct device to device communication that connects two user mobile devices without any coordinator or intermediary. Though, as described in section 1.2.3 of this report there are many candidate wireless technologies Wi-Fi in ad hoc mode of configuration is chosen for this particular implementation. That is due to, mainly, its recent popularity in mobile devices and long distance coverage as it is based on 802.11 technologies when compared to other D2D systems like Bluetooth.

Figure 3.1 shows the interconnections of the three wireless technologies and their interaction mechanisms in a fully integrated architecture. It also contains a loose coupling between the 4G cellular system and the WLAN as, eventually; both are connected to the Internet. It also depicts a very tight linking between the eNB of the 4G system and the AP of the WLAN system.

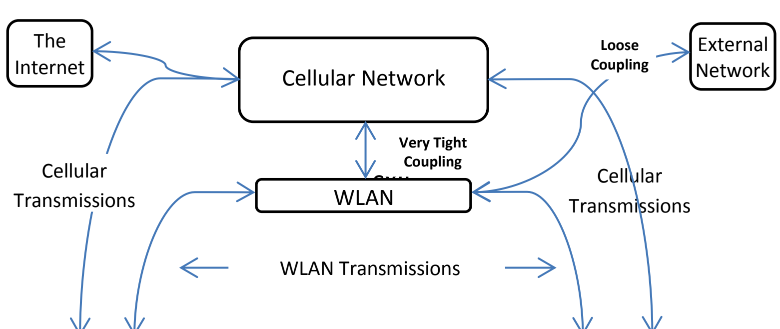


Figure 3.1: A Generic Interworking Architecture Showing the Various Wireless Technologies Involved and the Possible Traffic Transmission paths.

The eNB of the 4G system and the Wi-Fi AP of the WLAN are interconnected representing a very tight coupling between the two networks. Here, both connections can be used simultaneously in such a way that the WLAN uses the advanced security features of the eNB stack in the tight coupling scenario whereas it sends delay tolerant traffic directly to an external server in the loose coupling case. This makes the attachment and transmission faster as long as the AP and the UE are attached to the same eNB which is the assumption in this work. This is primarily used during, for instance, diverting 4G traffic as soon as the WLAN condition improves and back to the 4G whenever the WLAN network condition deteriorates. This is specifically useful for delay sensitive traffic.

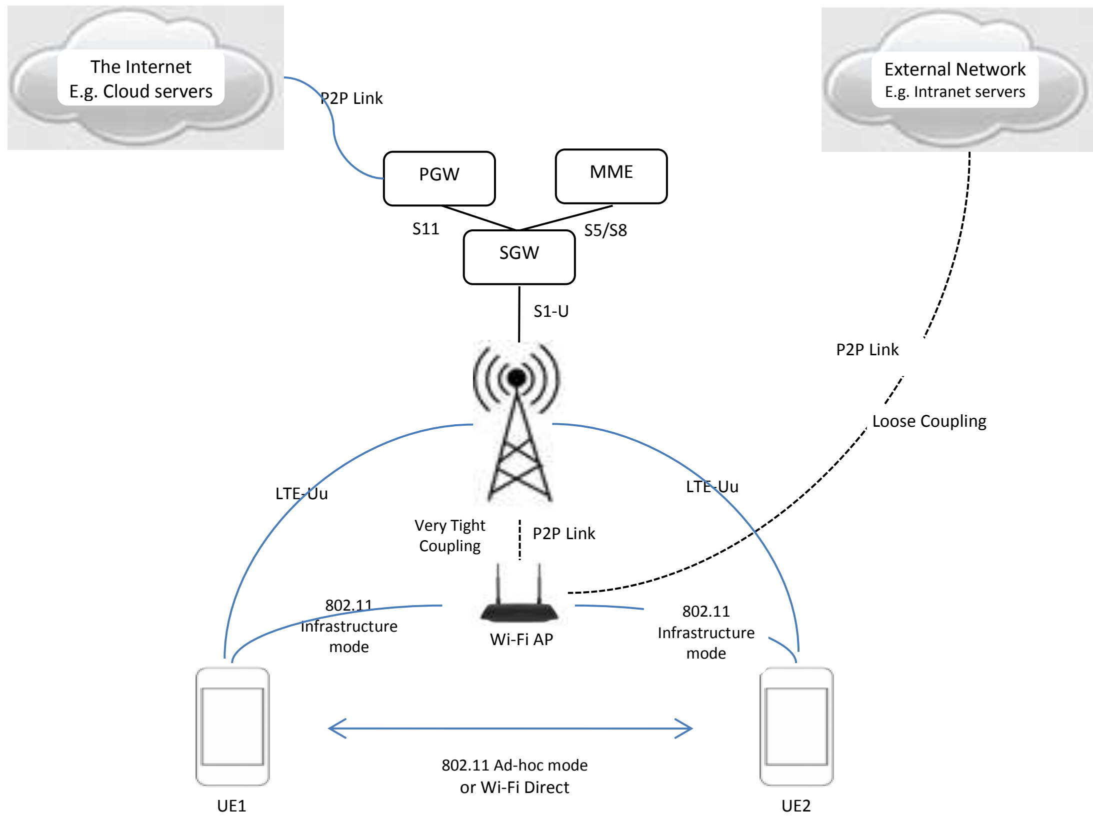


Figure 3.2: Detailed Implementation Diagram of the Proposed Feature-rich Interworking Architecture.

In a more detailed diagram, Figure 3.2, the significance of the interconnections can be explained as follows. First, any traffic initiated by any user equipment (like mobile phone, tablet, laptop, or smart watch) attempts to utilize any direct device to device communication in the proximity. The system that manages such a transaction may cache popular content and keep the profile of nearby devices. If such an opportunity is not available or the QoS demand of the initiated traffic is higher or tighter then it tries to use the WLAN system to forward the traffic. Based on the availability and potential of the WLAN system, for instance, depending on the received signal strength of the destination UE, it may defer from using it going directly towards the cellular transmission. Apparently, this is called on-loading. The software system that manages such an integrated and hybrid interworking system needs to keep many profiles of nearby users, network capabilities, trust levels, popular data, among others, to exploit the full potential of such an architecture.

The hybrid nature of the interworking architecture can be exploited by continuously monitoring the instantaneous network conditions or parameters of the WLAN and D2D systems.

As described by many research works[33][31][5] and [62], there is no such fully integrated and hybrid interworking system having crucial features as envisioned by users and operators.

And yet, such an interworking architecture is useless as long as it is not supported by a software system, like an algorithm, that monitors the initiation of user traffic, extracts QoS requirements of the initiated traffic and other details, gathers the instantaneous network conditions, and decides through which path to en-route the traffic. This is not erstwhile activity but a continual and iterative one.

As a result, the next subsection describes the traffic steering or offloading algorithm that exploits the features described above in an attempt to offload as much cellular traffic as possible via the direct D2D or the WLAN systems.

3.2.2 Traffic Offloading Algorithm

The proposed traffic steering or offloading algorithm is presented using two representations – flowchart and pseudo code. The flowchart presents the overall offloading process exploiting the various interconnections of the architecture whereas the pseudo codes are divided in to three snippets for ease of comprehension.

Figure 3.3 depicts the flowchart of the offloading algorithm. It is triggered whenever a UE initiates a 4G/LTE cellular traffic destined for the other UE. The algorithm, to make a decision on which particular pathway to forward the 4G traffic, expects five values, viz., RSSI of the D2D and

WLAN channels, QoS value which is either delay sensitive or delay tolerant, the number of packets to be transferred, and the source and destination addresses.

Then, the algorithm executes as follows:

1. Assume UE1 initiates a 4G cellular traffic, and hence, the offloading algorithm is triggered with the five initial values supplied appropriately: RSSI, QoS of the initiated traffic, number of packets, the source (UE1) and destination (UE2) IP addresses.
2. Then, it immediately checks whether there is actually packets to be transferred as a matter of early checkup.
3. Then, it goes on checking the delay requirement of the initiated traffic.
4. If the traffic is a time critical one, like voice, it just checks whether it is possible to use either the D2D or the WLAN system to forward the traffic. That is done by cross checking the instantaneous network conditions of the two networks, first the D2D then the WLAN. From many literatures like [178], [16] and [179] the 802.11 networks can be used to transfer real-time traffic as long as their RSSI is better than -67 dBm.

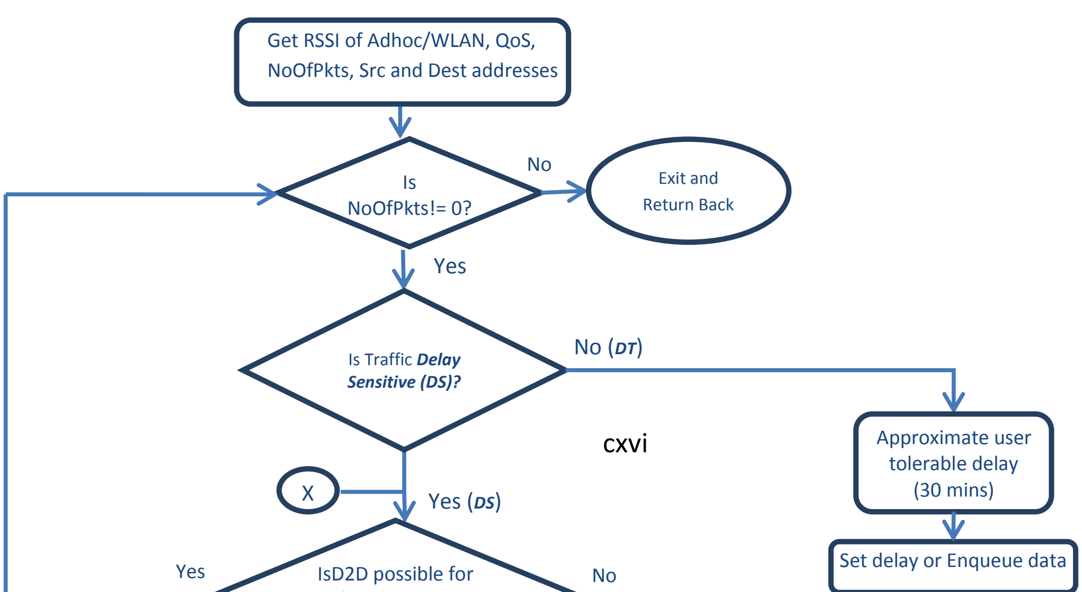


Figure 3.3: A Flowchart Presenting the Workflow of the Offloading Algorithm Accompanying the Proposed Interworking Architecture.

5. Step 4 continues to be executed as long as all the packets are transferred. This is done actively monitoring the conditions of the WLAN.
    i. If the condition of the network deteriorated beyond the delay threshold then the traffic, which is delay sensitive, is sent to UE2 via the eNB of the 4G network.
    ii. As and when the condition of the WLAN improves beyond the minimal requirement of the traffic, transmission diverted towards using it. This is done since the AP is

connected with the eNB covering it. This is dubbed as ***very tight coupling*** between the two networks.

6. If the traffic is **delay tolerant** then the algorithm sets a 30 minutes delay putting it in a queue. This value can be configured depending on various factors like the nature of the delay insensitive data and other network conditions. For example, appropriate values can be chosen depending on the time-of-the-day of industrial and residential areas. This value is chosen based on the discussion made on [15].
7. When the delay expires it attempts to check whether a direct D2D transmission or a Wi-Fi AP is available within the coverage of the devices.
    i. It either of the two networks are available (with minimal RSSI of -96 dBm) then it starts sending packets. This continues as long as all the packets are transferred. The value -96 dBm is selected because it is the minimal sensitivity of the 802.11 devices in the NS3 Wi-Fi module.
    ii. After the 30 minutes delay of the traffic if there is no coverage of both complementary networks then the traffic is sent to an external network like a cloud server or the Internet so that whenever UE2 gets access it will download the data. This is known as ***loose coupling*** between the WLAN and the 4G cellular system.
8. Finally, after all the packets are transferred via one of the various possible traffic flow paths the algorithm terminates gracefully.

```
                UE1 initiates cellular traffic destined to UE2
OffloadTraffic (RSSI, QoS, NoOfPkt, Src, Dest)
If (NoOfPkt > 0)
    If(QoS) //delay sensitive traffic
        Is RSSI of D2D > -67 dBm
                Send a packet directly to UE2
                Decrement NoOfPkt
        Is RSSI of WLAN > -67 dBm
                Send a packet via Wi-Fi AP to UE2
                Decrement NoOfPkt
        Send traffic to UE2 via the eNB//very tight coupling is exploited here
   Else   //delay tolerant traffic
   Set delay_counter to 30 //30 mins
   Do until offload_attempt is zero // offload_attempt = 3
     Do until delay_counter is zero
        Is RSSI of D2D > -96 dBm
                Send a packet directly to UE2
```

Figure 3.4: Traffic Offloading or Steering Algorithm That Exploits the Proposed Internetworking Architecture Presented Using Pseudo Code Format.

This traffic offloading algorithm is further represented using four pseudo codes. That is done, to make it easier to see, among other functionalities, when and how the loose and very tight linking between the WLAN and the 4G system are performed. Figure 3.4, 3.5, 3.6 and 3.7 presents the whole offloading algorithm in pseudo code representation.

Figure 3.4 depicts the snippet that steers delay sensitive traffic between the Wi-Fi AP and the eNB depending on the instantaneous condition of the WLAN and D2D systems. To minimize the traffic flow via the 4G system the algorithm monitors whether the D2D or the WLAN network conditions are improving by instantaneously checking via their RSSI values. Whenever it is improved beyond the delay requirements of the traffic the algorithm diverts it to these networks. It shows, in bold and italics, when very tight and loose couplings are triggered.

Figure 3.5 shows the algorithm that steers traffic between the eNB and the Wi-Fi AP, exploiting the very tight link between them, re-iterates transferring data via the eNB and the Wi-Fi AP as long as all the packets are transferred.

```
Send Traffic to UE2 via the eNB – Very Tight Coupling
SendTraffic_eNB(NoOfPkt, src, dest)
 Do
    Send traffic via eNB to UE2
    Decrement NoOfPkt
    If(NoOfPkt > 0)
        Do
            If RSSI of D2D channel > -67 dBm //instantaneous RSSI
               Send a packet directly to UE2
               Decrement NoOfPkt
```

Figure 3.5: Pseudo Code Representation of the Algorithm That Utilizes the Very Tight Coupling between the eNB and Wi-Fi AP.

```
Send Traffic to External Network via the Loose Coupling
SendTraffic_LC (NoOfPkt, src, dest)
 Do until NoOfPkt = 0
    Send/upload traffic to an external network based server
    Decrement NoOfPkt
 When NoOfPkt = 0
         End Transmission
```

Figure 3.6: Pseudo Code Snippet of the Algorithm to Send/Upload Delay Tolerant Traffic to an External Network.

Figure 3.6, in its part, presents the loose coupling scenario where a delay tolerant traffic is sent to an external network, so that, latter, when the receiving end got access, it can download the data. The algorithm simply sends or uploads all the data to the external network via the Wi-Fi AP. This external network could be a local server on the intranet or a cloud server in the Internet.

$$PL(d > d_0) = PL(d_0) + 10 \times n \times log_{10}\left(\frac{d}{d_0}\right) + \gamma \tag{3.1}$$

$$PL(d_0) = 20 \times log10\left(\frac{4\pi f d_0}{c}\right) \tag{3.2}$$

$PL(d > d_0)$ – transmitted power of the transmitter node.
$PL(d_0)$ – received power at the receiver node.
f – the center frequency of the channel. f = 2422MHz for the WLAN and f = 2412 MHz for the D2D.
d_0– reference distance, usually, taken as 1 m.
d – distance between the transmitter and receiver nodes.
c – the speed of light.
n – path loss exponent, n = 3 is taken.
γ–a zero-mean Gaussian distributed random variable with standard deviation σ. Here, assumed to be 0.
$PL(d > d_0)$ and $PL(d_0)$ are measured in terms of dB.

```
Estimate the Received Signal Strength
EstChanCond (freq)
Initialize Pt = 17.02 dBm, pi = 3.14, c = 3.0e+8, d0 = 1 m, and n = 3.
From the mobility model get the position/distance of the dest node which is d.
  Calculate path loss at d0 as PL = 20*log10((4*pi*freq*d0)/c) //Using Friis Free Space Path Loss Model
  Then calculate path loss at d as PL (d > d0) = PL + 10*n*log10(d/d0) //Using Log-distance path loss model
  Then Pr = PL (d > d0) + Pt
Return Pr
```

Figure 3.7: The Pseudo Code that Estimates the Channel Condition of the Complementary Networks Based on Their Channel Frequency.

Last, but not least, Figure 3.7 shows the pseudo code that does instantaneous channel estimation or calculation for the D2D and WLAN subsystems. It estimates the received signal power, Pr, using the Log Distance Path Loss Model. The model which is shown in Eqn 3.1 is ideal for this scenario since it is a generic model that can be used to predict the propagation loss for various environments including urban areas, indoors and outdoors [8]. To calculate the path loss at the reference distance, which is 1 meter, the Friis free space path loss model [9], which is shown in Eqn 3.2, is used. The D2D and WLAN subsystems are differentiated based on the assigned channel frequency during configuration of the networks.

3.3 Summary

In this chapter, the proposed interworking architecture and the accompanying traffic offloading algorithm are presented in detail. They bring forth a full-fledged traffic offloading system whose main features are enumerated below.

- It provides an ***integration of three different wireless communication technologies***: cellular, 802.11 based WLAN and D2D networks. This is why the proposed interworking architecture is dubbed as fully integrated as these technologies span from few meters to thousands of kilometers.

- It provides ***two different ways or places of linking between the cellular network and the WLAN***, viz., very tight and loose coupling. This makes the proposed interworking architecture a ***hybrid*** one.
- In loose coupling the Wi-Fi AP is linked with an external network like local server or the Internet so that without much interaction with the cellular network it can send ***delay tolerant*** traffic away from it.
- In very tight coupling the accompanying algorithm steers delay sensitive traffic between the eNB and the Wi-Fi AP depending on the instantaneous network condition of the latter. This enables ***simultaneous use*** of both devices, for example, one for the delay sensitive traffic and the other for the delay tolerant one.
- These two coupling techniques give the glimpse of ***network assisted and independent*** functioning of the WLAN with the cellular network.
- The traffic offloading algorithm make decisions, primarily, based on the instantaneous network conditions of the WLAN and the D2D channels. This makes the overall offloading system ***QoS-aware***.
- All in all, the internetworking architecture has five different pathways to push traffic from one UE to another. The direct D2D, the Wi-Fi AP, the eNB, the loose and very tight links make up the possibilities.
- It does not only offload traffic away from the cellular system but also on-loads back to it whenever the WLAN and D2D network condition deteriorates. That applies only for delay sensitive traffic.

CHAPTER FOUR

IMPLEMENTATION OF THE PROPOSED OFFLOADING SYSTEM USING NS3

4.1 Introduction

The proposed conceptual interworking architecture for mobile traffic offloading presented in the previous chapter is implemented using the NS3 (Network Simulator 3) tool. The detailed implementation of the proposed architecture and the accompanying traffic offloading algorithm is presented and discussed in this chapter. For the sake of completeness, however, it would be appropriate to have a little discourse about NS3.

NS3 is a discrete event-driven network simulator. It is an open-source platform that is freely available, via GNU[6] GPLv2[7] license, for research and educational purposes. It has well validated and documented core module, more realistic models that capture the entire simulation workflow, and many protocol implementations being used by network researchers worldwide. It also has enthusiastic maintainers and researchers who are available 24/7 online in the various platforms like user groups and mailing lists on best effort basis [1]. Some researchers went further in showing how capable and better NS3 is when compared with other similar network simulators in [2] and [3], among others. More details about NS3 can be found from [4].

Here, how NS3 is organized as a software object is briefly discussed. The concept of ***modules*** and ***models*** are central to its organization. NS3 is composed of independent modules that are built (or compiled) as a separate software library. User written simulation programs or scripts can link the modules depending on the specific implementation needed to conduct the simulation at hand. In

[6]GNU stands for a circular acronym that represents Gnu is Not Unix.
[7] GPLv2 is one of the latest open source licenses that stand for General Public License version 2.

other words, these modules are the libraries of NS3 that capture and represent the various networking technologies (like LTE, Wi-Fi, WiMAX, and 6LoWPAN), protocols (like IPv4, IPv6, LTE user and control planes, 802.11a/b/g/n, and CSMA), and algorithms (like routing, mobility, and propagation). Figure 4.1 depicts how the major modules are arranged and interrelated.

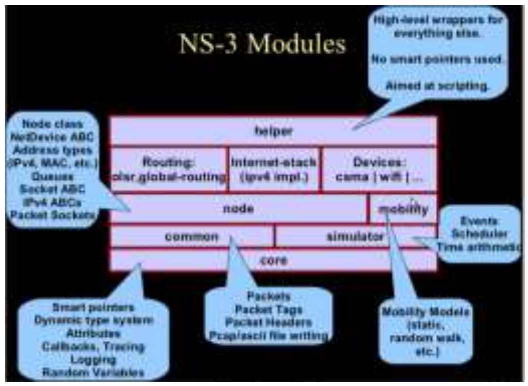


Figure 4.1: Organization of the main NS3 modules. Adapted from [182].

Models, on the other hand, are abstractions of real-world objects, protocols, devices, and so on. So, a module usually consists of multiple models like, for example, the ***internet*** module contains models for both Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). And generally, however, NS3 models do not span multiple software modules.

Anyone who wishes to use NS3 needs to work at the command line and with C++ and/or Python program development languages. Though it primarily used on Linux system it is possible to use on FreeBSD, and Windows (via Cygwin and Windows Visual Studio).

Above and beyond the aforementioned discussions, before attempting to use NS3 to simulate a networking system one has to understand a few key core concepts and abstractions thereof. They are node, application, channel, net device, and topology helpers, among others. Some of these key concepts and abstractions are presented in Figure 4.1 and Table 4.1. The later describes NS3 key concepts and abstractions with contextual meanings in NS3 supported with typical examples.

In Table 4.1 the column "generic meaning" represents the meaning of the concept or abstraction in the broader networking and Internet arena while the next column describes what the same term means in NS3. The last column itemizes typical examples. It is very simple to see that NS3 is based on Object Oriented paradigm.

Table 4.1: NS3 key concepts and abstractions with their contextual meaning and supporting examples [182].

№	Concept or Abstraction	Generic Meaning	Meaning and representation in NS3	Example
1	Node	Computing device, host, end system.	Basic computing device modeled using ***class Node***.	NodeContainer ueWiFi; Ptr<Node> ueLte;
2	Application	A kind of software that does something specific for the user. Always lives on top of system software like operating systems.	A user program that generates some activity to be simulated. Modeled using class Application. It has many specializations.	Application UdpEchoServerApplication Application UdpEchoClientApplication
3	Channel	Communicationmedia between two devices.	Basic communication sub-network between nodes, represented by class Channel, has many specializations	PointToPointChannel p2p;
4	Net device	Peripheral card, like NIC (network interface card) or device controller.	Representation of both a simulated network device and its driver software. "Installed" on a node. A node can have multiple net devices. Represented by class NetDevice.	NetDeviceContainer PointToPointNetDevice; Ptr<NetDevice> PointToPointNetDevice;
5	Topology helper	NS3 object creation and configuration facilitators	Simplify interconnecting and configuring Nodes, NetDevices, and Channels. Available for many constructs like NetDevices and Channel.	PointToPointHelper pointToPoint;

Before using these key concepts and abstractions for the implementation and simulation of the proposed internetworking architecture one needs to enumerate the general steps followed in writing a simulation program using NS3. The process of performing computer network simulations using NS3 typically involves the following five steps:

1. Define the networking scenario to be implemented and simulated.

2. Write a simulation program that recreates the desired scenario topology or architecture. This is done by accessing the appropriate NS-3 model libraries and APIs.
3. Specify configuration parameters of the objects that are being used for the simulation. This can be done using input files or hardcoding within the simulation program.
4. Configure the desired output to be produced by the simulator.
5. Run the simulation.

Off course, one may need to manipulate or post-process the output using tools like python to get the behavior of the networked system, after all, that is the primary purpose of doing a simulation.

NS3 having all these attractive features and capabilities, for this specific work, it is chosen mainly because it has the needed modules and library utilities to simulate LTE[8]-based networks and 802.11-based WLANs with both infrastructure and ad-hoc modes of operations. In NS3, the 4G component is dubbed as **LTE module** and the 802.11-based one is known as **wifi module**. Both are equipped with the necessary features required to capture the basic functionalities and behaviors of the proposed internetworking architecture and the accompanying traffic offloading algorithm.

The prototype implementation of the overall offloading system, including both the architecture and the offloading algorithm, consists of four different components; the LTE/4G subsystem, the WLAN subsystem, the D2D communication subsystem and finally the traffic offloading or steering algorithm.

Therefore, this chapter presents the implementation of the three subsystems of the internetworking architecture, in the sequence described above, followed by the traffic offloading algorithm using NS3.

---

[8]The LTE module and set of libraries are developed by the LENA (LTE/EPC Network simulAtor) project at CTTC – Center Technogic de Telecomunicacions de Catalunya [http://networks.cttc.es/mobile-networks/software-tools/lena/].

4.2 The 4G/LTE Subsystem

As described in chapter one, LTE cellular system has three major components: the UE, eNB and CN (SGW, PGW, MME, HSS, and PRCF, among others). From these components the NS3 LTE module modeled and implemented UEs, eNBs, LTE (with and without EPC support) and their interfaces, protocols and entities as recommended by 3GPP in [5] and [6]. The LTE component with EPC support has, though in a single node, the functionalities of the SGW, PGW and MME. All the 3GPP LTE radio protocol stack is implemented entirely within the UE, eNB and PGW nodes. Figure 4.2 shows the LTE-EPC simulation model implemented in NS3 which can be found at ***ns-3.26/src/module/lte*** after installing NS3.

Consequently, the LTE/4G subsystem of the proposed internetworking architecture is implemented using the features present in the aforementioned NS3 module. As shown in Figure 4.3 the LTE/4G subsystem has two UEs, one eNB, one PGW and one remote host that represent the Internet. That is really adequate enough to model the features and study the behavior of LTE/4G systems in the proposed internetworking system.

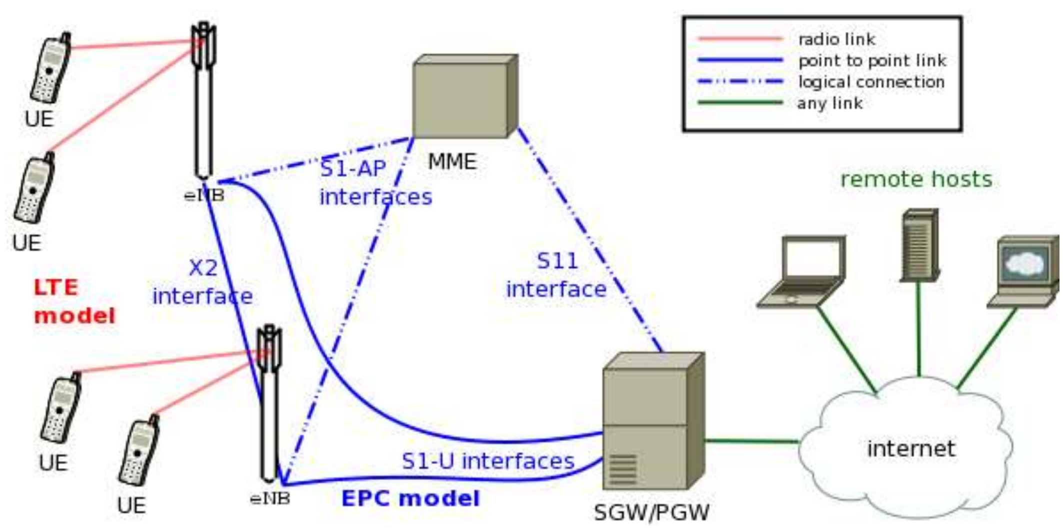


Figure 4.2: The LTE-EPC simulation model as part of the NS3 implementation [7].

The UE represents the mobile user device like smart phone or tablet that is equipped with all the components needed to initiate, store, and transfer 4G traffic transmissions like USIM and LTE radio interfaces. The eNB represents a 4G base station that has implemented the main functionalities of the E-UTRAN. Last, but not least, the PGW node has all the features of the SGW, MME and the PGW of LTE systems.

This implementation is shown in Figure 4.3. In the figure the LTE/4G subsystem comprises of the two user devices, marked as UE0 and UE1, the eNB that is put at the xy coordinate of (75.0,75.0), the EPC node which is labeled as **pgw/sgw/mme** painted with reddish color, and the Internet host labeled as **pgwrh**. These five nodes model a full-fledged LTE/4G system that facilitates the communication between the two UEs and a remote host that enable the UEs to traverse through the Internet and reach out the world. The **eNB** and the **pgw/sgw/mme** represent the LTE-EPC simulation model of NS3 platform presented earlier in Figure 4.2. It contains the LTE radio protocol stack residing entirely within the UEs and the eNB nodes. It also contains the EPC component which includes core network interfaces, protocols and entities. These entities and protocols reside within the **pgw/sgw/mme** node and partially within the **eNB** node.

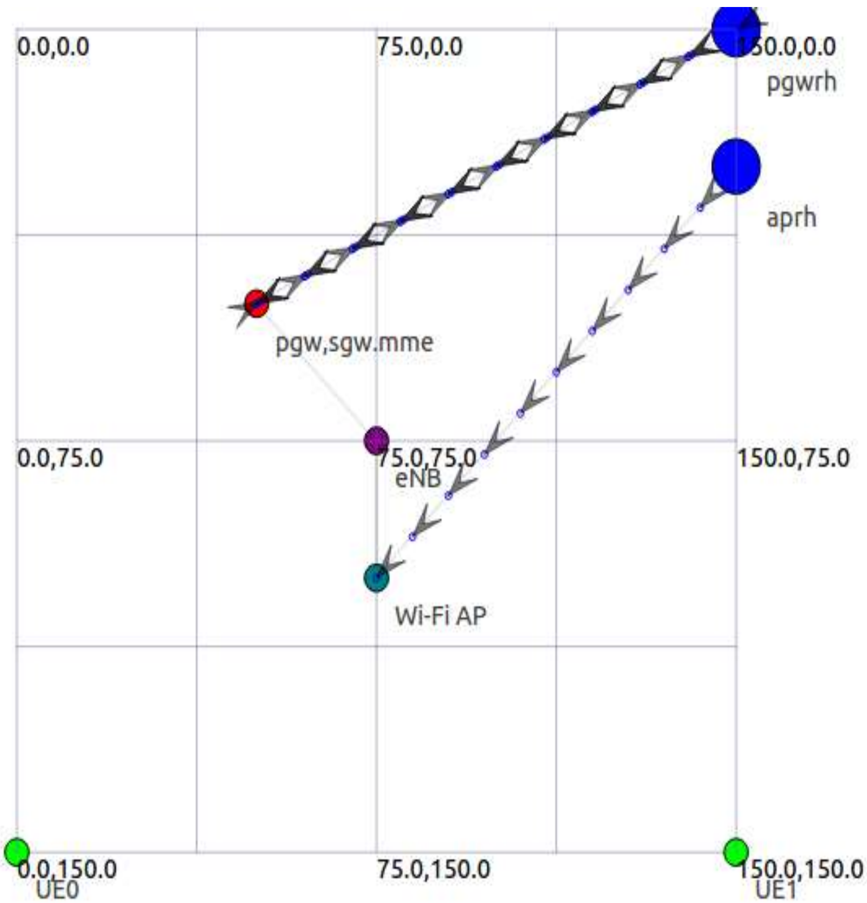


Figure 4.3: Shows the screen capture of the three models - LTE-EPC system, Wi-Fi AP-based network and D2D communication.

Figure 4.4 shows the LTE-EPC data plane protocol stack that clearly shows the entities and protocols involved in the full-fledged LTE/4G subsystem implemented in this work. The UE has the full TCP/IP protocol stack modified to support LTE traffic by including the PDCP and RLC layers between the IP and MAC components. This equips it to communicate with the eNB establishing the LTE Radio link. The eNB also contains another stack called the S1-U protocol stack or GTP/UDP/IP tunneling stack where it connects with the SGW/PGW/MME components; these are integrated in the NS3 implementation as a single node. The **pgw/sgw/mme** node has GTP/UDP/IP stack to forward and receive IP-based traffic between the Internet and the eNB. In NS3 this is referred as the

EpcSgwPgwApplication which is a specialization of the **Application** class. Simply, this is what is implemented in Figure 4.3 as LTE/4G subsystem.

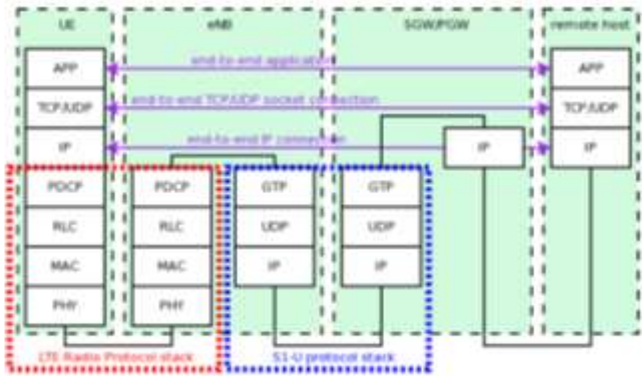


Figure 4.4: LTE-EPC data plane protocol stack [182].

The LTE/4G subsystem is created using the code snippet shown in the Figure 4.5. Line 5 - 9 creates the remote host node, labeled as pgwrh, which is used as a device in the Internet sinking traffic away from the PGW node. Line 10 creates one eNB and two UE nodes. Lines 11 and 12 install the LTE protocol stack on the eNB and UEs, respectively. And line 13 links the UEs with the eNB while line 14 shorts the **pgw** and **pgwrh** nodes. Using the code at line 1 alone, the LTE access network is created, but only when lines 2 to 4 are added it becomes a full-fledged EPC system allowing IPv4 networking over LTE.

Though the statement at line 3 implicitly creates PGW node, it is recommended by the LTE module designers to use the statement at line 4 to explicitly create it. That line enables the inclusion of a **pgw/sgw/mme** node to the simulation scenario. Lines 8, 9 and 14 install the IPv4 stack on the remote host and then create a point-to-point link between the **pgw/sgw/mme** node and the remote host.

4.3 The WLAN Subsystem

The WLAN or the Wi-Fi AP-based subsystem consists of two UEs, 802.11b-based Wi-Fi AP that facilitates communication between the UEs and a remote host that allows the AP to escape traffic to the Internet via any wired infrastructure. This is implemented in Figure 4.3 comprising the four nodes – the two UEs, the Wi-Fi AP and the remote host labeled as **UE0**, **UE1**, **Wi-Fi AP**, and **aprh,** respectively.

```
1. Ptr<LteHelper> lteHelper = CreateObject<LteHelper> ();
2. Ptr<PointToPointEpcHelper>  epcHelper =
   CreateObject<PointToPointEpcHelper> ();
3. lteHelper->SetEpcHelper (epcHelper);
4. Ptr<Node> pgw = epcHelper->GetPgwNode ();
5. NodeContainer remoteHostContainer;
6. remoteHostContainer.Create (2);//creates two remote hosts
7. Ptr<Node> pgwrh = remoteHostContainer.Get (0);//obtains one of the
   remote hosts
8. InternetStackHelper internet;
9. internet.Install (pgwrh);
10. NodeContainer ueNodes;
    NodeContainer enbNode;
    enbNode.Create (numberOfNodes);// numberOfNodes=1
    ueNodes.Create (numberOfUeNodes);// numberOfUeNodes=2
11. NetDeviceContainer enbLteDevs = lteHelper->InstallEnbDevice (eNB);
12. NetDeviceContainer ueLteDevs = lteHelper->InstallUeDevice (ueNodes);
13. lteHelper->Attach (ueLteDevs, enbLteDevs.Get (0));
14. NetDeviceContainer pgwrhDev = p2ph.Install (pgw, pgwrh);
```

Figure 4.5:C++ script used to create, configure and link the LTE/4G subsystem components.

To implement this scenario in NS3 the Wi-Fi module is exploited. The various constructs of the model provide MAC-level implementations and packet-level abstractions of PHY-level details for 802.11a/b/g/n/ac specifications. The **WifiNetDevice** architecture is depicted in Figure 4.6 detailing the MAC High, MAC Low, **WifiPhy** and **WifiChannel** layers and corresponding message flows that implement the MAC and Physical equivalents of 802.11 technologies.

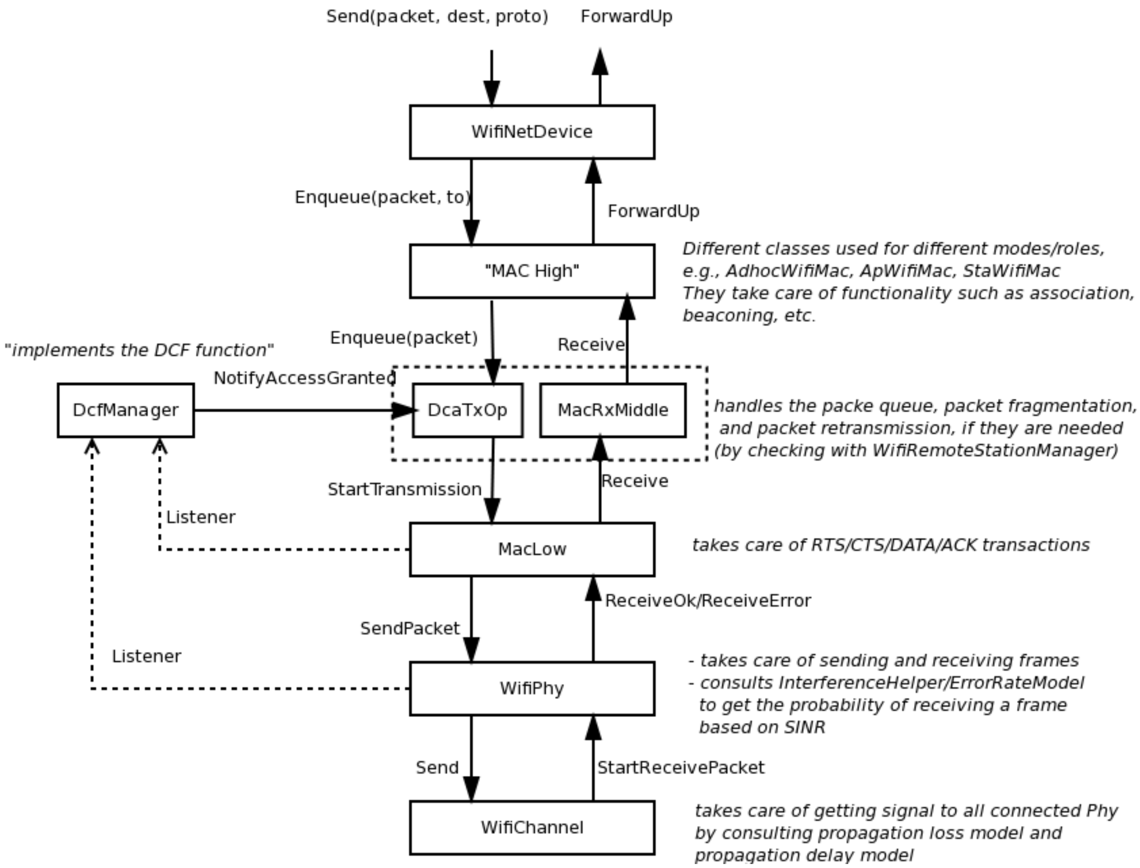


Figure 4.6: WifiNetDevice Architecture [8].

```
1. NodeContainer apNode;
   apNode.Create(numberOfNodes);// numberOfNodes = 1
   Ptr<Node> ap = apNode.Get (0);
2. WifiHelper wifi;
3. wifi.SetStandard (WIFI_PHY_STANDARD_80211b);
4. YansWifiPhyHelper wifiPhy =  YansWifiPhyHelper::Default ();
5. YansWifiChannelHelper wifiChannel;
   wifiChannel.SetPropagationDelay ("ns3::ConstantSpeedPropagationDelayModel");
   wifiChannel.AddPropagationLoss("ns3::LogDistancePropagationLossModel",
                                  "Exponent",DoubleValue (3.0));
6. wifiPhy.SetChannel (wifiChannel.Create ()); //by default, channel 1 AP and STA
7. QosWifiMacHelper wifiMac = QosWifiMacHelper::Default();
8. wifiMac.SetType ("ns3::StaWifiMac","QosSupported",BooleanValue(true));
9. NetDeviceContainer staDevs = wifi.Install (wifiPhy, wifiMac, ueNodes);
   NetDeviceContainer devices = staDevs;
10. wifiMac.SetType ("ns3::ApWifiMac","QosSupported", BooleanValue (true));
11. NetDeviceContainer apDevice = wifi.Install (wifiPhy, wifiMac, apNode);
12. devices.Add (apDevice);
13. internet.Install(ap);
14. NetDeviceContainer aprhDev = p2ph.Install (ap, aprh);
```

Figure 4.7: Shows the NS3 code fragment of the WLAN subsystem implementation that consists of the four nodes described previously.

In Figure 4.7 the statements under line 1 create Wi-Fi AP node and converts its instance object `apNode` to a smart pointer name `ap` for future manipulation of the node. The two UEs created for the LTE/4G subsystem are reused in this subsystem too. This, rationally, deserves some more explanation which is postponed for the right place. A node in the NS3 abstraction of a networking device can hold more than one Net Devices. A closer look at line 9 where **staDevs** are created as a new Net Device with preconfigured **wifiPhy** and **wifiMac** values and installed on **ueNodes**. **ueNodes** are the UE nodes created at line 10 of Figure 4.5.

The `wifiHelper` at line 2, the `YansWifiPhyHelper` at line 4 and the `YansWifiChannelHelper` at line 5 are used to configure the 802.11 specific technology, physical layer, and channels, respectively. In more detail, line 5 is used to configure the **wifiChannel** propagation delay and propagation loss models, line 6 is used to create the Wi-Fi channel, line 9 is used to create the Net Devices that is going to be installed on the **ueNodes**, lines 11 - 13 are used to install the Wi-Fi AP configurations on the `apNode`, and line 14 is used to interconnect the **Wi-Fi AP** and the **aprh**. The type of channel created at line 6 is the default one which is channel 1 with center frequency of 2142 MHz and 20 MHz width. The statement at line 12 is more interesting because it interconnects the two `staDevs`, the Net Devices configured with appropriate parameters of `wifiPhy`, `wifiMac` and `wifiChannel`, on the `apNode,` creating a WLAN consisting of the Wi-Fi AP and the two UEs.

Finally, the statements at lines 13 and 14 install IPv4 stack and creates a point-to-point link between the Wi-Fi AP and remote host, `aprh`, respectively.

One issue that might deserve a short explanation is why the 802.11b standard is chosen. As per the work of Asrat M. et el. in [18] it is found that more than 90% of Wi-Fi APs deployed in cities like Addis Ababa are of 802.11b.

4.4 The D2D Subsystem

In NS3, 802.11-based Net Devices can be configured to work either in infrastructure or ad-hoc mode of operation. Apparently, the MAC-High model has three Wi-Fi topological alternatives for the programmer to choose from: Access Point (AP) as ***ns::ApWifiMac***, non-AP Station (STA) as ***ns3::StaWifiMac*** and STA in Independent Basic Service Set (IBSS – or ad hoc network) as ***ns3::AdhocWifiMac***. In the previous subsection, the WLAN subsystem contains one AP in infrastructure mode, as ***ns::ApWifiMac*** and two STA nodes in non-AP Station mode, ***ns3::StaWifiMac***. Here in the D2D subsystem, the UEs are 802.11b Net Devices configured to work in ad hoc mode. To avoid inter-channel interference they are assigned with a different channel that is not identical with the WLAN subsystem.

In Figure 4.3, on page 116, the D2D subsystem is represented by the two UEs, **UE0** and **UE1**. These are the same nodes used in both of the previous subsystems. Just another Net Device is created, configured and installed on them. This makes the two UE nodes to be equipped with three different Net Devices representing three different wireless technologies. Figure 4.8 presents the NS3 code snippet of the D2D subsystem implementation.

```
1. WifiHelper wifi;
2. wifi.SetStandard (WIFI_PHY_STANDARD_80211b);

3. YansWifiChannelHelper adhocChannel;

4. adhocChannel.SetPropagationDelay
   ("ns3::ConstantSpeedPropagationDelayModel");
5. adhocChannel.AddPropagationLoss
   ("ns3::LogDistancePropagationLossModel","Exponent",DoubleValue (3.0));
6. YansWifiPhyHelper adhocPhy =  YansWifiPhyHelper::Default ();
   adhocPhy.Set("ChannelNumber", UintegerValue (6));
   adhocPhy.SetChannel(adhocChannel.Create());
7. wifiMac.SetType ("ns3::AdhocWifiMac", "QosSupported", BooleanValue
   (true));
8. NetDeviceContainer adhocDevs = wifi.Install (adhocPhy, wifiMac,
   ueNodes);
```

Figure 4.8: C++ code extract showing the implementation of the D2D subsystem.

In Figure 4.8 lines 1 and 2 are reused from the WLAN subsystem. Lines 3, 4 and 5 creates the ad hoc Wi-Fi channel, configures the propagation delay and loss models. The propagation loss model defined here is the **LogDistancePropagationLossModel**. The three statements under line number 6 creates the PHY layer of the Net Device and configures it to use channel 6 with center frequency 2422 MHz and 20 MHz width. Line 7 specifically sets the MAC type of the Net Device as ad hoc. Line 8, practically, installs the `adhocPhy`, `wifiMac`, and `adhocChannel` onthe `ueNodes`. The `ueNodes` are being reused for the third time. In many literatures and standardization body documents, this kind of D2D system is referred as Wi-Fi Direct as described in chapter one, section 1.2.3.

4.5 Integrating the Three Subsystems and Hybridizing the Whole System

Here an attempt has been made to show and discuss how the subsystems are interconnected to realize the integrated feature of the offloading system. That is followed by how the overall offloading system is made hybrid by describing the way it handles both real-time and delay tolerant traffic types.

The explanations in the preceding three subsections of this chapter are about how the three different wireless communication subsystems are implemented in NS3. They need to be linked and interlinked together to realize the proposed offloading system.

Technically, the D2D subsystem works independently. That is to mean it is not controlled either by the Wi-Fi AP or the eNB. It is an opportunistic proximity-based communication mechanism. When one of the devices become within the coverage of the other, it initiates data transmission which stays until the signal quality didn’t drop below a certain limit. For delay tolerant case signal strength of -96 dBm is the threshold to initiate transmission. Whenever the signal strength goes above -67 dBm it can be used for real-time transmissions as long as there is something to transfer.

When it comes to the WLAN and LTE/4G subsystems, the Wi-Fi AP and the eNB are interlinked using a point-to-point link as

```
NetDeviceContainer apenbDev = p2ph2.Install (ap, eNB);
```

where `p2ph2` is a *`PointToPointHelper`* instance object configured with appropriate parameters. This link between the eNB and the Wi-Fi AP will be exploited when, for instance, while UE0's real-time traffic is being sent via the eNB towards UE1and instantaneously the WLAN signal strength goes above -67 dBm where the next bunch of packet is then offloaded towards the Wi-Fi AP to be delivered to UE1. And, while a real-time traffic is being transferred via the Wi-Fi AP and immediately the signal strength lowers below -67 dBm then the next group of packets is on-loaded via the eNB. This back and forth continues as long as there are packets to be send and the condition of the network changes. This linking of the Wi-Fi AP with the nearest eNB is what is referred to as very tight coupling.

The Wi-Fi AP, besides its link with the eNB, has a connection with the Internet via an external network. This external network represents, for example, a wire-line intranet with gateway to the Internet. Having this data path allows the WLAN subsystem to send delay tolerant traffic that is delayed beyond a certain limit, like 30 minutes in this implementation, towards the Internet through the external network [171]. This delay is introduced before a single packet is transmitted. The final destination can be a cloud system, for instance, so that whenever the other device has got access to the Internet it will download the data. This is done without the involvement of the LTE/4G subsystem. However, the Wi-Fi AP can be configured and allowed to use the LTE/4G security mechanism as it better than the former in terms of maturity and management. This is what is called loose coupling between the two subsystems. The following line of code shows the implementation to achieve this feature.

```
NetDeviceContainer aprhDev = p2ph.Install (ap, aprh);
```

4.6 Making the Offloading System QoS-Aware

On top of the above features that make it fully integrated and hybridized, the implemented interworking architecture has, through the offloading algorithm, the capability of providing priority based on the QoS requirement of the traffic initiated by the UEs'. In this subsection, how the QoS awareness is introduced in the overall offloading system is discussed.

First, the LTE subsystem is configured to be aware of type of traffic. QCI is defined by 3GPP in its standard document [10] for LTE networks to make sure that the QoS requirements of different traffic types get the adequate bearer resources. It defines specific packet forwarding behavior like packet rate and packet delay budget for each data flow called, by 3GPP, service data flow (SDF). Therefore, each kind of SDF has separate bearer and associated QCI. For instance, voice calls and video gaming are real-time by their nature whereas SMS and chat are delay tolerant. These two groups of traffic types will have different bearer that interconnects the source and destination with different resources as defined by the respective QCI and other parameters like Guaranteed Bit Rate (GBR), Maximum Bit Rate (MBR) and Allocation and Retention Priority (ARP) which are network parameters pre-configured by the network operator.

Table 4.2: Mapping between QoS values: QCI, DSCP and ToS. The table is populated with data from [10], [11] and [12].

QCI	Resource Type	Priority	Packet Delay	Packet Loss	DSCP/Dec/Hex	ToS (Dec/Hex)	Service Example(s)
1	GBR_CONV_VOICE	2	100 ms	10^-2	EF/46/2E	184/B8	Conversational Voice
2	GBR_CONV_VIDEO	4	150 ms	10^-3	EF/46/2E	184/B8	Conversational Video
3	GBR_GAMING	3	50 ms	10^-3	EF/46/2E	184/B8	Real Time Gaming
4	GBR_NON_CONV_VIDEO	5	300 ms	10^-6	AF41/34/22	136/88	MM Streaming
5	NGBR_IMS	1	100 ms	10^-6	AF31/26/1A	104/68	IMS Signaling
6	NGBR_VIDEO_TCP_OPERATOR	6	300 ms	10^-6	AF31/26/1A	104/68	TCP, FTP, apps High Priority
7	NGBR_VOICE_VIDEO_GAMING	7	100 ms	10^-3	AF21/18/12	72/48	VoIP, video, apps non-GBR
8	NGBR_VIDEO_TCP_PREMIUM	8	300 ms	10^-6	AF11/10/0A	40/28	TCP, FTP, apps Reg. Priority
9	NGBR_VIDEO_TCP_DEFAULT	9	300 ms	10^-6	BE/0/00	0/00	TCP, FTP, apps Low Priority

This feature is implemented using NS3 in its Eps-Bearer class with specific values and parameters as tabulated in the first five columns of Table 4.2. Hence, in this work, this implementation of NS3 is exploited using the lines of code presented in Figure 4.9.

In the code snippet there are two bearers created: the first at line 10 and 11 and the second at line 18 and 19 with the highest and lowest bearer resources as *GBR_CONV_VOICE* and *NGBR_VIDEO_TCP_DEFAULT* (see Table 4.2).The former is used for real-time traffic whereas the latter is used for delay tolerant transmissions. The LTE network, in NS3, does not require any additional component or code to route user initiated traffic towards these two bearers. It automatically detects the QoS requirements of the packets and redirects through the right bearer as long as the user traffic is labeled or tagged with the appropriate QoS values.

The WLAN and the D2D subsystems are made to be aware of QoS via the lines of code shown in Figure 4.10.

```
for (uint32_t u = 0; u < ueNodes.GetN (); ++u)
{
Ptr<NetDevice> ueDevice = ueLteDevs.Get (u);
GbrQosInformation qos;
qos.gbrDl = 1024;
qos.gbrUl = 1024;
qos.mbrDl = qos.gbrDl;
qos.mbrUl = qos.gbrUl;

enum EpsBearer::Qci q1 = EpsBearer::GBR_CONV_VOICE;
EpsBearer bearer1 (q1, qos);
bearer1.arp.priorityLevel = 2;
bearer1.arp.preemptionCapability = true;
bearer1.arp.preemptionVulnerability = true;
lteHelper->ActivateDedicatedEpsBearer (ueDevice, bearer1, EpcTft::Default ());

enum EpsBearer::Qci q2 = EpsBearer::NGBR_VIDEO_TCP_DEFAULT;
EpsBearer bearer2 (q2, qos);
bearer2.arp.priorityLevel = 9;
bearer2.arp.preemptionCapability = true;
bearer2.arp.preemptionVulnerability = true;
lteHelper->ActivateDedicatedEpsBearer (ueDevice, bearer2, EpcTft::Default ());
}
```

Figure 4.9: NS3 lines of code used to implement QoS in to LTE subsystem.

```
1.QosWifiMacHelper wifiMac = QosWifiMacHelper::Default();
2.wifiMac.SetType ("ns3::ApWifiMac",
                          "QosSupported", BooleanValue (true));
3.wifiMac.SetType ("ns3::StaWifiMac",
                          "QosSupported", BooleanValue (true));
4.wifiMac.SetType ("ns3::AdhocWifiMac",
                          "QosSupported", BooleanValue (true));
```

Figure 4.10: QoS implementations of the Wi-Fi AP, the 802.11 UE devices (StaWifiMacs), and the D2D.

Lines 2, 3 and 4 configure the Wi-Fi AP, the Wi-Fi station devices on the two UEs, and the Wi-Fi devices for D2D communication to be aware of QoS transmissions, respectively. Line 1 creates QoS enabled Wi-Fi MAC object using the `QosWifiMacHelper` instance.

After enabling the three Net Devices on each UE nodes to be aware of QoS traffic, the next step is to initiate traffic that is tagged with QoS bits based on the DSCP (or ToS) field of the IPv4 header. The type of service, or ToS, is based on or derived from the six-bit Differentiate Services Code Point (DSCP) field of the IPv4 header. RFC 2474 [13] reserved the first six bits of DS field (or IPv4 ToS) for the DSCP, and RFC 3168 [14] reserved the last two bits for Explicit Congestion Notification, as shown in Figure 4.11. The DSCP name, decimal and hex values with corresponding ToS decimal and hex values are given in the last two columns of Figure 4.11. The conversion from six bits DSCP to eight bits ToS is done by appending two zeros at the end of the latter. Then, rearranging them in to nibbles and representing them in to hexadecimal or converting them in to decimal can be followed depending on the need.

0	1	2	3	4	5	6	7
DSCP						ECN	

Figure 4.11: DSCP or ToS field of IPv4 header.

Therefore, based on the values set in the DSCP or ToS bits, by the operator of the network or the user who initiated the traffic, a packet is prioritized and hence requests a route for a low-delay and high-throughput or vice versa.

In this work, the user is assumed to tag or label the packets of the initiated traffic depending on the nature of the application. The traffic assumed is data traffic which can be either delay sensitive or delay tolerant. Packet tagging, at the source node, and extraction, at the destination node, is performed using the NS3 code snippet presented in Figure 4.12.

```
1. std::vector<uint8_t> tosValues = {0, 184};

2. InetSocketAddress localAddr (Ipv4Address ("7.0.0.2"),dlPort);
3. localAddr.SetTos(tosValues[rnd]);//rnd is either 0 or 1

4. InetSocketAddress remoteAddr (Ipv4Address ("7.0.0.3"),dlPort);
5. tosPktRcvd = remoteAddr.GetTos();
```

Figure 4.12: NS3 code fragment that used to tag and extract user traffic.

As can be seen from the NS3 statements of Figure 4.12 the `InetSocketAddress` objects are used to set at line 2 and 4, using the Socket base class function `SetTos()`and `GetTos()`, at lines 3 and 5, respectively, to tag and extract ToS values from the IPv4 header of the user traffic. Line 1 is a vector or array of two values that are used to define, in decimal, the two traffic types as 00 to mean best effort (BE) or a transmission with lowest priority and 184 to represent expedited forwarding (EF) or highest priority traffic like conversational voice (see Table 4.2).

Now, after enabling the devices to be aware of QoS traffic, tagging the initiated traffic and detecting the same of inbound packets with QoS bits, the next step is devising a mechanism to make decisions based these values, which path the packet should follow without compromising its QoS requirements. That is where the traffic offloading or steering algorithm comes in to picture.

4.7 The Traffic Offloading Algorithm

As has been said earlier, having an interworking architecture whose components are put in place systematically and interconnected cleverly may not do much unless and otherwise supported with a software component that makes decisions on how to route and re-route packets diligently based on the instantaneous conditions of the overall network. Therefore, the proposed offloading algorithm presented and explained using the flowchart, in Figure 3.3, and a sequence of pseudo codes, in Figure 3.4 – 3.6, are implemented using C++ in NS3 to steer traffic via the various routes ingrained in the offloading architecture.

The offloading algorithm is implemented using the following two program portions with signatures `void OffloadTraffic(double rxPwr, uint8_t qos, uint32_t pktsz, Ptr<Socket> srcsocket, Ptr<Socket> destsocket)` and `double EstChanCond(double f).` Here, an attempt has been made first to see how the second function estimates the channel conditions of the complementary wireless networks and then how the former is used to steer traffic based on instantaneous network conditions.

The channel estimation function takes the frequency of the channel, 2412 MHz for the WLAN and 2422 MHz for the D2D and then it calculates the signal strength of the signal at the receiver. As explained previously in chapter 3 using Figure 3.7 and equations 3.1 and 3.2 the Log-distance propagation path loss model is used in this estimation.

```
double EstChanCond(double f) //f = 2422 for WLAN & f = 2412 for D2D in MHz)
{
double Pr = 0.0;
double Pt = 17.02;//dBm
double pi = 3.14;//unit-less
double c = 3.0e+8;//m/s - speed of light
double n = 3;//path loss exponent for urban area
double d0 = 1;//meters -- reference distance
for (NodeContainer::Iterator j = ueNodes.Begin(); j != ueNodes.End (); ++j)
{
```

Figure 4.13: Function that estimates the channel conditions of the WLAN and D2D subsystems.

Figure 4.13 presents the implementation of the same in NS3. The center frequency of the WLAN and D2D networks is passed as a parameter during the function invocation. In the same figure, to estimate the signal strength at a given distance d, first the distance itself should be obtained. For that the mobility model used by the mobile node is used with the appropriate member functions as can be seen at lines 11 to 14. The function, `EstChanCond(double f)`, finally returns the estimated received signal power which is extensively used by the offloading function to decide through which specific path to push the user traffic.

The offloading functionality is implemented using a function named `OffloadTraffic()`, equipped with the appropriate parameters before its invocation. Inside the function there are two major compartments; the first part takes care of delay sensitive traffic and the second handles delay tolerant one. They are presented in Figure 4.14 and Figure 4.15, respectively.

The first and the sixth lines checks whether the D2D and the WLAN subsystems can be used to transfer a delay sensitive traffic or not, respectively. That is done by getting the signal strength of the respective channels by calling the `EstChanCond()` function. Then, if the D2D subsystem is in good quality, i.e. its signal strength is above -67 dBm, then the transmission is sent directly to UE1 using the code at lines 3 and 4. The same is done for the case of the WLAN using the line of codes from 6 to 9. But, before going for the WLAN the D2D connection is checked. It is when the D2D is not usable for delay sensitive traffic the WLAN is checked, in the first place.

The interesting part of the code is seen when both the D2D and the WLAN subsystems fail to be useful for delay sensitive traffic. Since the focus is on delay sensitive traffic when both the complementary networks cannot satisfy the QoS requirements of the initiated traffic it is sent to the destination via the eNB of the LTE network, which is done using the lines 16 and 17. If this data traffic is delayed the QoS or QoE of the user degrades. Sending the traffic back to the eNB is referred as on-loading. This definitely doesn't help the cellular network and desired to be minimized.

```
1.  if((EstChanCond(adhocFreq)) > -67)
2.  {
3.    int rmtadhocnode = sourceAdhoc->Connect(remoteAdhoc);
4.    int sentPkt = sourceAdhoc->SendTo(Create<Packet> (NumPkt),0,remoteAdhoc);
5.  }
6.  else if((EstChanCond(apFreq)) > -67)
7.  {
8.    int rmtadhocnode = recvSinkSta0->Connect(localSta0);
9.    int sentPkt = recvSinkSta0->SendTo(Create<Packet> (NumPkt),0,localSta0);
10. }
11. else //sending traffic via eNB, exploiting very tight coupling
12. {
13.   int tempPkts = 1024;
14.   while(tempPkts > 0)
15.   {
16.    int rmtadhocnode = sourceUeLte0->Connect(remoteUeLte0);
17.    int sentPkt = sourceUeLte0->SendTo(Create<Packet> (tempPkts),0,localUeLte1);
18.    while((EstChanCond(apFreq)) > -67)//then offloading again!
19.    {
20.     rmtadhocnode = sourceAPeNBdl->Connect(localAPeNBdl);
21.     sentPkt = sourceAPeNBdl->SendTo(Create<Packet> (tempPkts),0,localAPeNBdl);

22.     NumPkt = NumPkt - tempPkts;
23.     if(NumPkt < 1024) tempPkts = NumPkt;
24.    }
25.   }
26. }
```

Figure 4.14: Portion of the OffloadTraffic() function that handles delay sensitive traffic.

There is still more interesting feature of the architecture and its implementation. Line 18 is used to check the WLAN system whether its channel condition is improved or not after sending a group of packets. This is done by the eNB. The eNB and the Wi-Fi AP are interconnected, which is dubbed as ***very tight coupling***. That means the former can control or monitor the latter. Then if it is found that the WLAN channel condition is improved the next bunch of packets are offloaded via the Wi-Fi AP instead of the eNB. Otherwise, transmission continues using the eNB. This back and forth continues until all the packets are transmitted exhaustively.

When the traffic is delay tolerant the second part of the offloading function is initiated. As presented in Figure 4.15, lines 1 and 6 checks whether the D2D and the WLAN subsystems can be used to transmit traffic, in that sequence. That is done, again, by calling the channel estimation function. When their channel condition is adequate enough that means it is better than -94 dBm, then they can be used to send the traffic directly, using the D2D connection or via the Wi-Fi AP. The value -94 dBm is chosen because it is the signal sensitivity of most mobile devices' Wi-Fi hardware according to [15].

One may ask that what if the two complementary networks are not available at all? Then the initiated traffic is sent to an external system like a cloud server through the Wi-Fi AP which is done using lines 13 and 14. It is the reality that the main purpose of Wi-Fi APs is to extend the wired intranet to mobile users where cables are not convenient or cost effective. The intranet, in turn, is connected to the Internet via landline connectivity like DSL or fiber optic. This direct interconnection of the Wi-Fi AP with the external network is what is known as ***loose coupling***. A delay tolerant transmission can be sent to such a system until the destination is ready and whenever it is ready it can fetch the data from the external network.

```
1. if((EstChanCond(adhocFreq)) > -94)
2. {
3. int rmtadhocnode = sourceAdhoc->Connect(remoteAdhoc);
4. int sentPkt = sourceAdhoc->SendTo(Create<Packet>
   (NumPkt),0,remoteAdhoc);
5. }
6. else if((EstChanCond(apFreq)) > -94)
7. {
8. int rmtadhocnode = recvSinkSta0->Connect(localSta0);
9. int sentPkt = recvSinkSta0->SendTo(Create<Packet> (NumPkt),0,localSta0);
10. }
11. else
12. {
```

Figure 4.15: Portion of the OffloadTraffic() function that takes care of delay tolerant traffic.

4.8 Summary

In this chapter, the implementation of the interworking architecture, which is composed of LTE, WLAN and D2D components, and the traffic steering or offloading algorithm is made and explained in detail using the NS3 network simulator. The integration of the three wireless technologies, the hybridization of the architecture using very tight and loose coupling, and the realization of QoS-awareness of the three networks are all described in detail.

In addition, how the offloading algorithm makes decisions based on the instantaneous network conditions and QoS requirements of the packets is shown. Moreover, how the algorithm exploits the hybrid nature of the architecture for delay sensitive and delay tolerant traffic types is discussed.

All in all, this chapter shows how the full integration, hybrid-ness, and QoS-aware-ness feature of the interworking architecture is implemented using NS3.

Next, the performance evaluation of the overall offloading system is done using key performance indicators.

CHAPTER FIVE

PERFORMANCE EVALUATION OF THE IMPLEMENTED OFFLOADING SYSTEM

5.1 Introduction

In the previous two subsequent chapters it is presented that a feature-rich interworking architecture is conceptually proposed using generic components, technologically mapped using specific wireless technology building blocks and contextually implemented using NS3. Obviously, the next step is evaluating the implemented system using key performance indicators that are appropriate to show the behavior of the overall system.

Therefore, in this chapter, the performance of the implemented interworking architecture and offloading algorithm are evaluated using the following key performance indicators: Packet Loss Ratio, Packet Delivery Ratio, End-to-End Delay, Jitter, Offloading Efficiency and Handover Delay. Before going for these evaluations, though, the basic system configurations and simulation parameters are briefly presented.

5.2 System Configurations and Simulation Parameters

In this evaluation, oftentimes, the default system configurations are used unless specified otherwise. That means, in the following tables, the system configurations and simulation parameters specified are used to indicate those features specifically modified from the default parameters and system values of the respective NS3 objects.

5.2.1 Wi-Fi Ad Hoc Network

In the following table, Table 5.1, the parameters used in the D2D subsystem, viz., the Wi-Fi ad- hoc network, are presented. More specifically, those values used in the NetDevice of the two UEs which is implemented by using the **`adhocDevs`** are tabulated.

The standard chosen, as described earlier, is the 802.11b which is the one widely deployed in Addis Ababa based on [17]. From the 14 channels available with each of them have 22 MHz bandwidth, the sixth channel is assigned for the D2D subsystem. This channel has center frequency of 2422 MHz.

The other peculiar configuration to note here is the mobility model used which is **ConstantVelocity** with speed of 1.5 m/s or 5.4 km/s. This is a typical walking speed for mobile users according to [16].

Table 5.1: System configurations and simulation parameters for the D2D subsystem.

Parameter	Value
Wi-Fi Physical Standard	WIFI_PHY_STANDARD_80211b
wifiMac Type	"ns3::AdhocWifiMac"
Bandwidth	22 MHz
Channel Number	6
Channel Frequency	2422 MHz
MAC Type	QosWifiMac
PropagationDelayModel	ConstantSpeed, speed "2.99792e+08"
PropagationLossModel	LogDistance with exponent = 3, ReferenceDistance "1", ReferenceLoss "46.6777"
Mobility Model	ConstantVelocity, v = 1.5 m/sec
EnergyDetectionThreshold	-94 dBm
TxPowerEnd	16.0206 mW

5.2.2 WLAN or Wi-Fi AP-based Network

Table 5.2 presents the simulation parameters used in the WLAN subsystem. It is implemented by using the Net Devices on the two UEs and the Wi-Fi AP as **`staDevs`** and **`apDevice`,** respectively. Much of the configuration values are identical with that of the D2D subsystem except the channel frequency assigned. Here the channel with center frequency of 2412 MHz is chosen to avoid interference with the D2D system.

It is already explained that both the D2D and the WLAN subsystems are implemented using the 802.11b technology in ad hoc and infrastructure mode of operations, respectively.

Table 5.2: System configurations and simulation parameters for the WLAN subsystem.

Parameter	**Value**
Wi-Fi Physical Standard	WIFI_PHY_STANDARD_80211b
wifiMac Type	"ns3::ApWifiMac"
wifiMac Type	"ns3::StaWifiMac"
Bandwidth	22 MHz
Channel Number	1
Frequency	2412 MHz
MAC type	QosWifiMac
PropagationLossModel	LogDistance with exponent = 3, ReferenceDistance "1", ReferenceLoss "46.6777"
PropagationDelayModel	ConstantSpeed, speed "2.99792e+08"
Mobility Model	ConstantVelocity, v = 1.5 m/sec
EnergyDetectionThreshold	-94 dBm
TxPowerEnd	16.0206 mW

5.2.3 LTE or 4G Network

The LTE subsystem, which is modeled by using the NetDevice on the two UEs as **ueLteDevs,** configurations and simulation parameters are given here. Table 5.3 presents all the values that are default except the two bearers created for delay sensitive and delay tolerant data transmission. As described in chapter 4, the two bearers are configured with QCI values of GBR_CONV_VOICE with ARP priority level of 1 and NGBR_VIDEO_TCP_DEFAULT with ARP priority level of 9, respectively. The two configuration values are chosen simply because they are the highest and the lowest values in the QCI enumerator of NS3.

Table 5.3: System configurations and simulation parameters for the 4G/LTE subsystem.

Parameter	Value
LteEnbPhy::TxPower	30 dBm
LteUePhy::TxPower	10 dBm
LteEnbNetDevice::UlBandwidth	25 MHz
LteEnbNetDevice::DlBandwidth	25 MHz
LteHelper::PathlossModel	ns3::FriisPropagationLossModel
PointToPointEpcHelper::S1uLinkDataRate	"10000000000bps"
RadioBearerStatsCalculator::EpochDuration	"+250000000.0ns
RadioEnvironmentMapHelper::	NoisePower "1.423e-13
LteEnbNetDevice/Mtu	"30000" Bytes
LteUePhy/TxPower	"10" dBm
EpsBearer::Qci q1 = EpsBearer arp.priorityLevel	GBR_CONV_VOICE QCI=1
EpsBearer::Qci q2 = EpsBearer arp.priorityLevel	NGBR_VIDEO_TCP_DEFAULT QCI=9

5.2.4 The Very Tight and Loose Coupling Channels

In this subsection, the very tight and loose coupling links' simulation parameters are presented using Table 5.4. The very tight coupling data path, which interconnects the eNB (the **`enbNode`**) and the Wi-Fi AP (the **`apNode`**), is created using the **`apenbDev`** NetDevice attached on the two nodes. Whereas, the loose coupling that shorts the Wi-Fi AP with the external network or the Internet is created using the **`aprhDev`** NetDevice.

Table 5.4: Simulation parameters of the very tight and loose couplings.

NetDevices	Parameter	Value
`apenbDev`	DataRate	50 Mb/s
	Mtu	3000 Bytes
	Delay	20 ms
`aprhDev`	DataRate	100 Gb/s
	Mtu	1500 Bytes
	Delay	10 ms

Table 5.4 shows, in its top three rows, the **`apenbDev`** and, in its bottom three rows, the **`aprhDev`** parameters, respectively. In the former case the maximum performance is determined by the eNB as it is the lowest in terms of data rate and delay for the maximum data size. In the latter case the

maximum performance could be assumed to be that of the Internet as it involves the process of uploading data to an external computer via the Internet.

Using all these system configurations and simulation parameters the total simulation time is 100 seconds. This time is fixed based on the LTE and WLAN performance analysis made in [183] – [185]. In these papers the LTE average call setup time is found to range from 5 – 10 seconds and with average E2E delay of 2 to 2.6 msecs. This gives a total data transmission session of around 80 seconds for LTE traffic. In the case of WLAN, specifically in 802.11b networks, the Ready-To-Send/Clear-To-Send (RTS/CTS) and acknowledgment (Ack) transmissions take around 2 msecs. After which data transmission commences. Therefore, for the actual data transmission both networks have from 90 to 95 seconds for a single simulation run. This is repeated 150 times to get rich simulation output data in text formats. From these text files python scripts are used to extract specific values of the aforementioned performance metrics that are saved in to another file for further processing. The analysis and presentation of these data is done using MATLAB and SIMULINK software version R2016a.

5.3 Performance Evaluation Metrics

To evaluate the overall offloading system, i.e., both the interworking architecture and the offloading algorithm, six key performance indicators or metrics are chosen. Subsequently, they are briefly defined and used to show the performance of the overall offloading system.

5.3.1 Packet Loss Ratio (PLR)

Packet loss ratio can simply be defined as the ratio of the difference between the total packets sent and total packets received to the total packets sent between the source and destination. It can also be put as the percentage of packets lost with respect to packets sent. Mathematically, this can be put as follows:

$$PLR = \frac{Total\ Packets\ Sent - Total\ Packets\ Recieved}{Total\ Packets\ Sent} \quad (5.1)$$

Eqn. 5.1 is used to calculate the PLR for each of the five traffic flow paths. Data traffic is assumed to be initiated by UE1 destined to UE2.

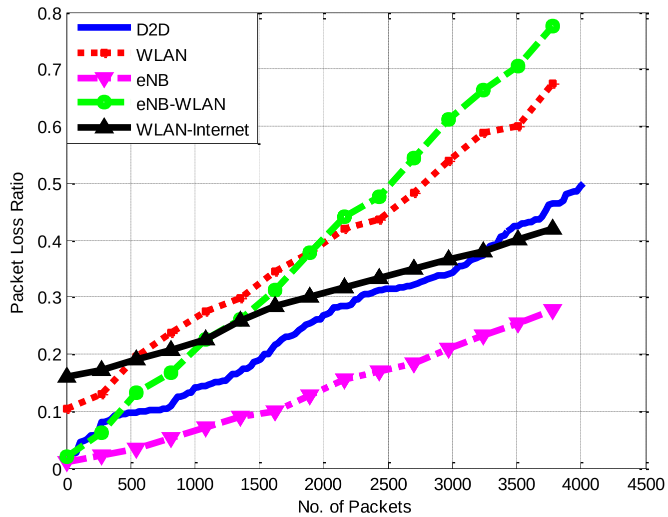


Figure 5.1: Packet loss ratio of the five different traffic paths.

As can be seen in Figure 5.1 the traffic through the WLAN faced the highest packet loss when compared to the D2D link and the LTE. The higher PLR of WLAN could be accounted for wireless channel conditions as there are much possible interference, the medium access control (MAC) retry limits and the used physical layer (PHY) modes of 802.11b technology. The LTE network is the better one among them as it is expected to be more reliable as it is primarily designed for the purpose of voice communication. When we see the tight (eNB to Wi-Fi AP) and loose (Wi-Fi AP to the Internet) links the former suffered as high as 70% loss when compared to the latter whose maximum loss is around 24%. This is expected as the tight coupling is, apparently, a vertical handover between the LTE's eNB and the WLAN's AP. Moreover, at smaller packet sizes PLR is tolerable in all the paths but at larger packet sizes it starts increasing especially for the tight coupling link.

5.3.2 Packet Delivery Ratio

Packet delivery ratio (PDR) is the ratio of number of packets received at the destination to the number of packets sent by the source. Here it is calculated by dividing the total packets successfully received by the destination via a particular path to the total packets sent by the source via that specific path.

$$PDR = \frac{Total\ Packets\ Recieved}{Total\ Packets\ Sent}\ x\ 100 \tag{5.2}$$

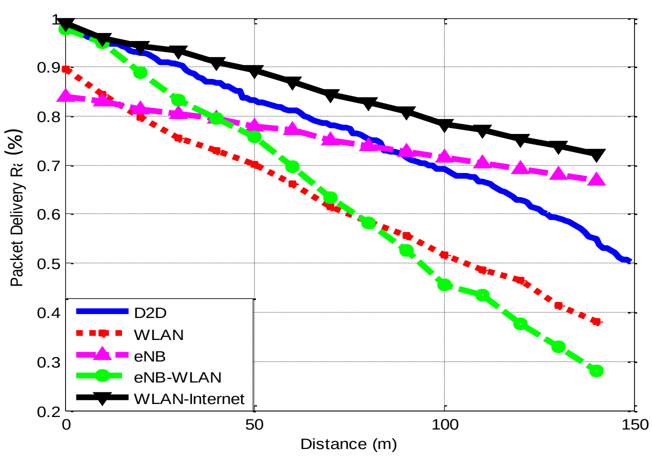


Figure 5.2: The ratio of packets delivered at the destination via the respective paths.

The above figure, Figure 5.2, shows the PDR of the five data paths. The WLAN path with almost linear decreasing as distance increases has a better performance. The D2D also has almost a linear decline in PDR values with distance which has exhibited a lower performance when compared with the WLAN path. The LTE has little PDR variation with distance when compared to the others as it is not much affected by distance variations. The loose coupling shows a better PDR than the tight coupling which has the highest negative slope when compared with all other paths.

5.3.3 End-to-End Delay

End-to-end delay, also known as one-way delay, measures the time taken for a packet to be transmitted across a network from source to destination. Mathematically, this can be calculated by subtracting the time stamp of the packets arrival time at the destination from the packets outbound time at the source. Then that result is averaged over the entire packets as shown in eqn. 5.3 for each of the five paths. The result is plotted in Figures 5.3 and 5.4.

$$E2E\ Delay = \frac{\sum_i^n (t_{ai} - t_{si})}{n} \tag{5.3}$$

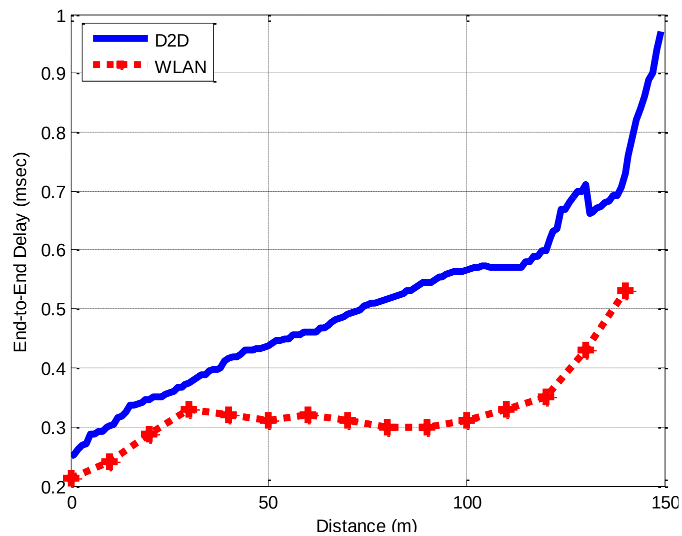


Figure 5.3: End-to-End Delay of the D2D and WLAN paths.

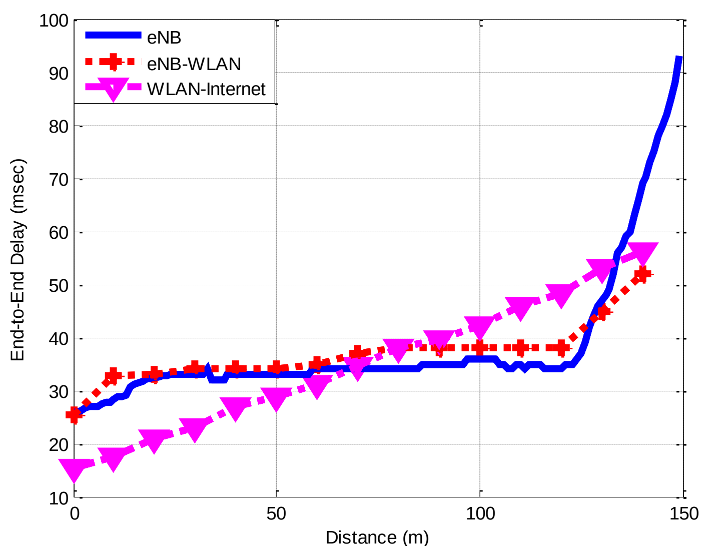


Figure 5.4: End-to-End Delay of the eNB, the tight and loose links.

The end-to-end delay exhibited by the WLAN and the D2D which are based on the 802.11b technologies is much better than the LTE, the tight and loose coupling paths. The WLAN has the lowest delay of packets followed by the D2D. The LTE faced the worst delay at higher distances indicating that is better to use it for delay sensitive traffic at short distances and for delay tolerant traffic at longer distances. Among the two coupling links, viz., the tight and loose coupling, the former met the worst especially at smaller distances. The latter shows almost a linear increase in the end-to-end delay as distance between the nodes increases. The cause of the higher delay in eNB-WLAN link is due to additional vertical handover delay introduced. These are all are depicted using Figures 5.3 and 5.4.

5.3.4 Jitter

Jitter, or packet delay variation or variability, is the difference in end-to-end delay between packets in the data transmission. For each data path, jitter is calculated by the NS3 built-in packet probes and classifiers defined under the Flow Monitor module. It classifies and calculates, for each unicast transmission, the sum of all end-to-end delay jitter values for all received packets of the flow. The difference of successive packets' E2E delay gives the jitter. Then it is averaged over the total number of packets transmitted from UE1 to that particular destination as depicted in eqn. 5.4. The result is plotted in Figure 5.5.

$$Jitter = \frac{\sum_{i=0}^{n}(t_{ai+1} - t_{si+1}) - (t_{ai} - t_{si})}{n} \tag{5.4}$$

As can be seen in Figure 5.5 the D2D path has the lowest delay variation whereas the loose coupling link has the worst. The D2D link has lowest jitter due to the fact that it doesn't face much interference and hindrance during data transmission as it's done in short distances and directly between the communicating devices, while the loose coupling link faced the worst as it traverses through the Internet to reach a remote destination.

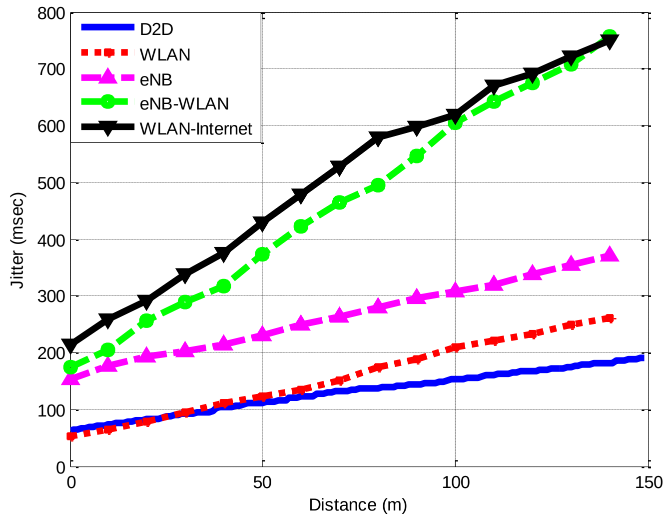


Figure 5.5: Delay variability of the five links.

5.3.5 Offloading Efficiency (OE)

Here, the two different kinds of traffic types, delay tolerant and delay sensitive, deserve separate treatment. That is because they may follow different transmission paths in the architecture based on the decision made by the traffic offloading algorithm. In the interworking architecture presented in Chapter 3, Figure 3.1, there are multiple paths for data traffic initiated by the UE in the left (UE0) to reach the other UE found in the right (UE1), directly or indirectly. From these data transmission paths the delay tolerant traffic may follow:

DT-Path 1: The direct D2D path – from one of the UE to the other directly,

DT-Path 2: The path via the Wi-Fi AP or WLAN towards the other UE, and

DT-Path3: The path via the Wi-Fi AP or WLAN towards the external network from which the other UE fetches the data later.

These paths are selected based on their availability which, apparently, means if their RSSI value is better than -94 dBm. On the other hand, the delay sensitive traffic may follow one of the following paths:

DS-Path 1: The direct D2D path – from one of the UE to the other directly,

DS-Path 2: The path via the Wi-Fi AP or WLAN to the other UE,

DS-Path 3: The path via the eNB towards the other UE – data flow from one UE to another via the eNB,

DS-Path 4: The path via the eNB and down to the WLAN or Wi-Fi AP towards the other UE.

One of these paths will be selected, in the sequence they appear. How they are used or selected is discussed in Chapter 3 under section 3.2.2.

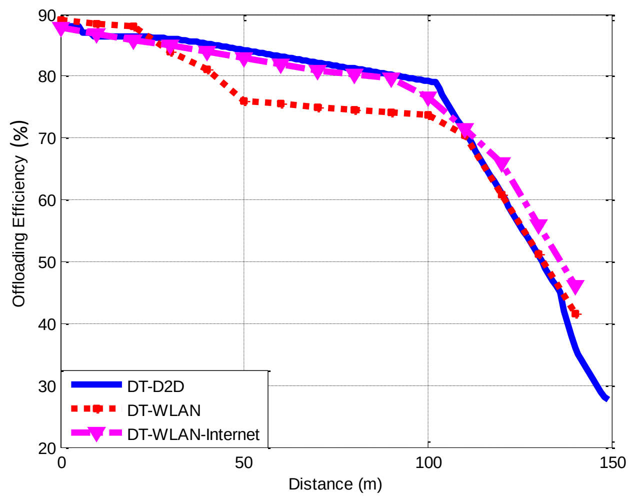


Figure 5.6: Offloading efficiency of those paths used for delay tolerant (DT) traffic.

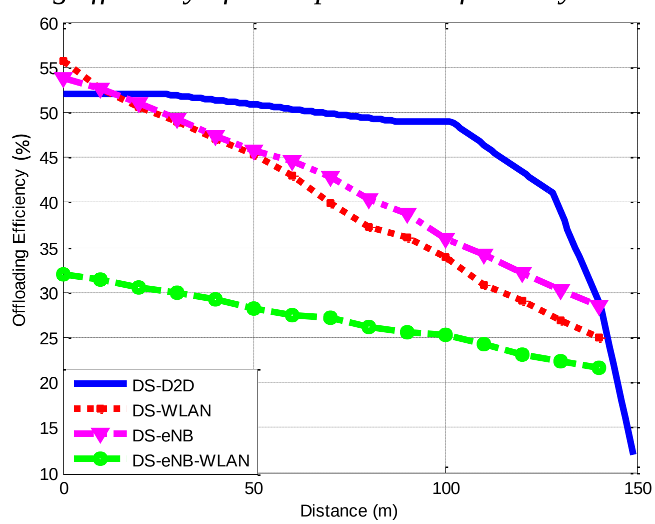


Figure 5.7: Efficiency of the D2D, WLAN and the loose coupling links – delay sensitive

(DS) paths.

To this end, in this interworking architecture, data traffic is said to be offloaded when it uses one of these paths to en-route the initiated data traffic from one of the UE0 to UE1. Thus, offloading efficiency can be calculated as the ratio of the total traffic initiated from the source UE sent via that particular data transmission path and delivered to the source UE to the total traffic initiated by source to the destination. For example, the total offloading efficiency for delay tolerant (DT) traffic will be

$$OE_{DT} = \frac{total\ traffic\ delivered\ via\ DT{-}Path1,\ \ DT{-}Path2\ and\ DT{-}Path3}{total\ DT\ traffic\ initiated} \qquad (5.5)$$

This formula is the aggregation of the offloading efficiency of the D2D, the WLAN and the loose coupling paths. It is also possible to calculate the individual offloading efficiencies and sum up at the end.

In the same token, the total offloading efficiency of delay sensitive (DS) traffic will be

$$OE_{DS} = \frac{total\ traffic\ delivered\ via\ DS{-}Path1, DS{-}Path2, DS{-}Path3, and\ DS{-}Path4}{total\ DS\ traffic\ initiated} \qquad (5.6)$$

Similarly, the offloading efficiency of individual paths can be calculated separately and summed up to get the total efficiency. It is technically the percentage of data traffic initiated from UE0 towards UE1 that is targeted to pass through the eNB but used that particular path instead.

As shown in Figure 5.6, up to 88%, 85% and 87% of delay tolerant traffic can be offloaded using the loose coupling, the WLAN and the D2D paths, respectively. It is observed that, as it might be expected, better offloading is done when the UEs are closer to the complementary networks.

In the case of delay sensitive traffic depicted using Figure 5.7 the four paths (viz., the D2D, the WLAN, the eNB, and the tight coupling) offloaded around 54%, 55%, 53% and 35% of the data traffic during the simulation runs, respectively. That is a considerable amount of traffic. The total amount of traffic initiated by UE0 passed through the eNB that reached UE1, which is around 54%, is the direct consequence of this fact. That is almost a reduction of half of the expected traffic.

The tight coupling link introduced between the eNB of the LTE and the Wi-Fi AP of the WLAN offloaded, on average, above one-third (35%) of the initiated total traffic. This, if further fine-tuned using a more robust algorithm, it can take away more cellular traffic towards the WLAN.

5.3.6 Handover Delay

The handover delay is the measure of the time delay induced during the data transmission changes from the LTE to the WLAN network. This occurs when delay sensitive traffic is being transmitted via the eNB because both the D2D and the WLAN channels were poor, i.e. worse than -67 dBm. However, as the traffic steering algorithm is continuously monitoring the WLAN by estimating its instantaneous RSSI value, whenever this value is better than -67 dBm, the next bunch of packets are sent via the Wi-Fi AP towards the other UE.

Consequently, the handover delay is the time taken as the eNB stops sending data, triggers the P2P channel between itself and the Wi-Fi AP and the first packet arrives at the latter.

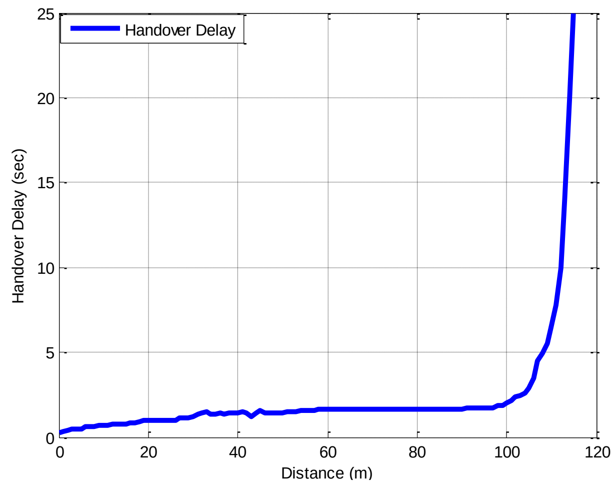


Figure 5.8: The delay induced during vertical handoff.

The handover delay between the eNB and the Wi-Fi AP is shown in Figure 5.8. It increases with distance almost exponentially. It is found that the vertical handover suffered a lot after the distance between the two UEs reached around 110 meters. That is because the RSSI of the 802.11b technology jumps to -74 dBm and gets worsening just after the user goes away the 110 meter mark. That is mainly due to the transmitted signal gets attenuated due to various signal interferences and hindrances.

5.4 Analysis of Results

Table 5.5 summarizes the results obtained in this work for the six performance evaluation metrics with respect to international benchmarks. The benchmarks are collected from various research outputs like [188] - [193] and standardization bodies like ITU [194]. The results are mainly for the tight coupling link that shorts the eNB and the Wi-Fi AP. That is because the other links are standard links that exist even if the there is no interworking between the various wireless technologies.

Table 5.5: Comparison of obtained results in this work with international benchmarks.

№	Performance Metric	Obtained Results		Benchmarks	
		GBR	NGBR	GBR	NGBR
1	PLR	~ 4 %	~ 12 %	2.5 %	5 %
2	PDR	~ 7 %	~ 25 %	3 %	16 %
3	E2E Delay	30 – 50 msec	75 msec	~ 150 msec	~ 150 msec
4	Jitter	~ 14 msec	~ 300 msec	40 msec	~ 400 msec
5	Offloading Efficiency	~50 % for DS	~90% for DT	~38% for DS	~ 78% for DT
6	Handover Delay	200-1200 msec		400–1000 msec	

As the results show the E2E delay, jitter and offloading efficiencies are found to be much better than the benchmarks. This provides the minimal requirements for data transmission applications like email, SMS and file transfers. That is mainly because such kind of traffic is bursty by its nature having tolerable loss and delivery ratios without losing the content of the data. More specifically, the offloading efficiency is very good enabling the LTE network to have a much lesser burden of delay tolerant traffic (90% offloaded). This gives an opportunity for the cellular network to be used for delay sensitive applications that are real-time by their nature like gaming, video calls and voice.

When it comes to the delay sensitive traffic the proposed system can offload around 50% of the traffic away from the cellular network. That is an increase of around 12% from the maximum achieved since today. This will be very attractive for cellular network operators as it very much lightens the burden on their core network in long term if further techniques are investigated like utilizing high end Wi-Fi APs than the ones currently used. Moreover, if latest 802.11 technologies like 802.11a/e/h are used a better performance in terms of wider coverage and lesser power usage can be obtained. This demands further investigation to be conclusive.

The handover delay experienced by data packets in the proposed system almost overlaps with the benchmarks with slight extension in both ends which deserve further investigation to fine tune the extreme cases. And yet, based on the results obtained in this work the proposed system can be used for data transmission of both delay sensitive and delay tolerant traffic types without losing their QoS requirements.

For the proposed offloading system to be used for real-time traffic like video calls and gaming further investigations are required to study and device mechanisms to lower the average PLR and PDR below the benchmarks.

5.5 Summary

This chapter evaluated the performance of the implemented offloading system using six key performance indicators or metrics. Both the interworking architecture, via the various data flow paths, and the traffic steering algorithm are evaluated. There are few deviations from the expected performance in some of the paths like the WLAN poorer throughput and the higher handover delays. On the contrary, there are promising and encouraging results obtained like the offloading efficiency obtained for both delay sensitive and delay tolerant data flows. It is found that it is possible to offloading up to 88% of the cellular traffic as long as the data traffic has low QoS requirements. Moreover, for delay sensitive traffic it is possible to offload above 50% of the traffic relieving the LTE from above half of the data traffic it is supposed to shoulder.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORKS

6.1 Conclusions

There are many efforts by researchers and standardization bodies towards devising a fully interworking architecture that fulfills the expectations of users and operators between cellular, WLAN and D2D technologies, among others. The majority of the efforts focus on addressing some specific feature or aspect of the overall offloading process. There is less work in devising a feature-rich interworking architecture that integrates wireless technologies creating multiple data paths meeting the QoS demand of applications.

To this end, there is little work that can be considered as a fully integrated, hybrid and QoS-aware interworking architecture addressing both delay tolerant and delay sensitive traffic exploiting the various technologies at different levels of integrations. On the basis of which, this work proposed a conceptual interworking architecture that is fully integrated in a sense that it interconnects D2D and WLAN systems with 4G/LTE networks. It is also hybrid as it has multiple levels of integrations between the three wireless technologies, viz., very tight and loose couplings.

The conceptual interworking architecture is implemented using NS3 and it is thoroughly tested. As the network architecture alone cannot do much without a software element that monitors the status of the data paths and decides through which specific path to en-route traffic, a traffic steering algorithm is formulated and implemented. The algorithm estimates the D2D and WLAN instantaneous conditions via their RSSI values and decides whether to en-route traffic towards them or not. More specifically and briefly, if the either of the networks has a connection to the other UE with RSSI value better than and equal to -67 dBm then it can be used to offload delay sensitive traffic. Whereas, their RSSI value is better than -94 dBm then they can be used to send delay tolerant traffic. Otherwise, they are regarded as unavailable.

The performance evaluation of both the interworking architecture and the traffic steering algorithm is done using six key performance indicators, viz., packet loss ratio, end-to-end delay, packet delivery ratio, jitter, offloading efficiency and handover delay. Much of the results obtained are promising and encouraging for further efforts. The major findings and contributions can be enumerated as follows:

- A review of the state-of-the-art is very much comprehensive that researchers in the area can get what had been done, what is being done and is going to be done from three different contributors: academic and industry researchers, standardization bodies and product manufacturers.
- A conceptual feature-rich interworking architecture that is fully integrated (having three different technologies), hybrid (supporting both very tight and loose coupling) and QoS-aware (handling both delay sensitive and delay tolerant traffic) is proposed.
- The generic network architecture is contextually implemented to interwork LTE, WLAN and D2D systems using the NS3 simulation tool.
- A traffic steering algorithm that monitors the instantaneous network conditions of the WLAN and decides whether to use it or not, among others, is designed and implemented.
- Much of the results obtained are promising and encouraging. It is found that around 90% of delay tolerant and more than 50% delay sensitive traffic can be offloaded from the LTE network.

- This proposal and implementation can be taken as a starting point for many further efforts in this research direction as pointed out in the next section.

6.2 Recommendations for Future Works

As this research is one part of the next generation ubiquitous and heterogeneous all-in-one networks there are many directions to pursue and push forward too. The major future works that can emanate from this work are summarized below.

One straight forward direction could be using the same offloading system for real-time applications and services like voice, video conferencing and online gaming and critically evaluate its performances.

In this work while the eNB monitors the instantaneous network conditions of the WLAN it uses the RSSI value alone. Using a multi-valued parameter to decide upon would be considered a better alternative as there are other factors that affect the wireless radio like power and SINR. This might be more interesting when both the LTE and WLAN/D2D networks are available and users or applications/services are allowed to choose through which specific path to forward their data.

Using cognitive and software defined radio to support or streamline the traffic steering algorithm is another facet that could be suggested. It is also possible to test the overall idea in the context of IPv6 as it is unarguably the future of the Internet.

As the capabilities of mobile devices continue improving it would be fascinating to think of implementing the features of the steering algorithm using multithreaded capabilities. For instance, in the steering algorithm while data is being transmitted using one thread the other thread may monitor the conditions of the other networks in parallel. Moreover, using the various data transmission paths or links simultaneously would be another interesting effort.

Last but not least, using real-time local traffic data to study and analyze the performance of the existing local wireless networks would be an innovative endeavor. Here, two approaches can be seen extending this work. One is performing a rigorous analysis and evaluation of the local wireless networks, cellular and WLAN, using local real-time traffic data on top of the proposed offloading system see the possibilities of devising a mobile traffic offloading strategy. Another one is just a thorough study and analysis of the existing networks to have a glimpse of their real-time performances spatiotemporally. These not only showcase the existing performance bottlenecks but also pinpoint potential future improvement strategies.

References

[1]. UMTS Forum, “Mobile traffic forecasts 2010-2020,” Jan, 2011.

[2]. R. Wood, “Wireless network traffic worldwide : forecasts and analysis 2013-2018,” Oct, 2013.

[3]. Portio, “Portio Research Mobile Factbook 2013,” Feb, 2013.

[4]. Ericsson, “Ericsson Mobility Report,” Jun, 2016.

[5]. F. Rebecchi, M. De Amorim, V. Conan, and et el., “Data Offloading Techniques in Cellular Networks : A Survey,” IEEE Commun. Surv. Tutorials, pp. 1–25, 2014.

[6]. Bahjat El-Darwiche, Pierre Peladeau, Christine Rupp and Florian Groene, 2017 Telecommunications Trends - Aspiring to Digital Simplicity and Clarity in Strategic Identity, strategy& report, pwc publishing, 2017.

[7]. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016-2021, March 2017. Document ID:1454457600805266. Avaialble: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.

[8]. J. Okwuibe, M. Liyanage and M. Ylianttila, "Provider Assisted Wi-Fi Offloading Leveraging on SDN," 22th European Wireless Conf., 2016, pp. 1-6.

[9]. L. Gao, G. Iosifidis, J. Huang and L. Tassiulas, "Economics of mobile data offloading," 2013 Proc. IEEE INFOCOM, Turin, 2013, pp. 3303-3308.

[10]. N. Ristanovic, J.-Y. Le Boudec, A. Chaintreau, and V. Erramilli, “Energy Efficient Offloading of 3G Networks,” 2010.

[11]. Y. Liu, “An SDN Platform for Traffic Offloading",” Department of Computer Science, University of Helsinki, Helsinki, Finaland, Apr, 2015.

[12]. S. Jose, “Site Survey Guide for Deploying Cisco 7920 IP Phones,” April, 2005.

[13]. J. G. Andrews, “Seven Ways that HetNets Are a Cellular Paradigm Shift,” IEEE Comm. Mag., pp. 136–144.

[14]. E. Hossain, Heterogeneous Wireless Access Networks: Architectures and Protocols. Winnipeg: Springer, 2008.

[15]. A. Y. Ding et al., "Enabling energy-aware collaborative mobile data offloading for smartphones," IEEE Intl Conf. on Sensing, Comm. and Networking (SECON), New Orleans, LA, 2013, pp. 487-495.

[16]. Amazon, “Cellular-Wi-Fi Integration,” Jun, 2012.

[17]. K. Lee, J. Lee, and Y. Yi, “Mobile Data Offloading : How Much Can WiFi Deliver ?,” ACM Conex., vol. 21,

no. 2, pp. 536–550, 2010.

[18]. V. F. S. Mota, D. F. Macedo, Y. Ghamri-Doudane, and J. M. S. Nogueira, “On the feasibility of WiFi offloading in urban areas: The Paris case study,” in IFIP Wireless Days, 2013.

[19]. Asrat Mulatu, J. Casademont Serra, and Yalemzewd Negash “Availability and Capacity of Wi-Fi s in Addis Ababa,” KICS- 2016 Korea Ethiopia ICT Int. Conf., AAU-IT PhD Program, Addis Ababa, pp. 113-119, May 2016.

[20]. K. Fukuda and K. Nagami, “A Measurement of Mobile Traffic Offloading.” LNCS, vol. 7799, 2013, pp 73-82.

[21]. S. Paris, F. Martignon, I. Filippini, and L. Chen, “An Efficient Auction-based Mechanism for Mobile Data Offloading,” vol. 1233, pp. 1–14, 2014.

[22]. O. I. Oshin, C. U. Ndujiuba, and F. E. Idachaba, “3GPP LTE : The Relevance of Interworking,” 2015.

[23]. W. Song, W. Zhuang, and A. Saleh, “Interworking of 3G cellular networks and wireless LANs,” 1998.

[24]. G. Vijayalakshmy, “Loosely Coupled Heterogeneous Networks Convergence using IMS-SIP-AAA,” vol. 85, no. 16, pp. 29–45, 2014.

[25]. H. Soliman, “Mobile IPv6 Support for Dual Stack Hosts and Routers,” RFC-5555.

[26]. R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, “Resource ReSerVation Protocol (RSVP),” RFC-2205, 1997.

[27]. Q. Wu, “Proxy Mobile IPv6 (PMIPv6) Localized Routing,” Sep 2012.

[28]. S.-J. Wu, “A New Integrated Mobile Architecture for Heterogeneous Wireless Networks,” 6th Int’l Conf. IIHMSP, pp. 159–162, 2010.

[29]. Y. Khadraoui, X. Lagrange and A. Gravey, "A Survey of Available Features for Mobile Traffic Offload," 20th European Wireless Conference, Barcelona, Spain, 2014, pp. 1-4.

[30]. 3GPP, “ETSI TS 123 402 v13.5.0 Architecture Enhancements for non-3GPP Accesses,” 2016.

[31]. OpenSignal, Inc., "Global State of Mobile Networks," February 2017.

[32]. Users and T-mobile support, “t-mobile users complaint on lte extremely slow in San Fransisco,” March 2017. Available: https://support.t-mobile.com/thread/138003.

[33]. Try to act against telcos on network congestion rule breach, “mobile user complain due to network congestion”, The Indian Express, September 2016. Available:http://indianexpress.com/article/technology/tech-news-technology/trai-to-act-against-telcos-on-network-congestion-rule-breach-3028690/.

[34]. National Consumer Complaint Board Portal, Airtel Network Problem, Available:https://www.complaintboard.in/complaints-reviews/.

[35]. Zlata Rodionova, The 10 most common customer service fails by mobile phone operators, Oct 2015, Avaialble:https://www.independent.co.uk/news/business/news/.

[36]. Katie Morley, I complained about poor mobile signal and was told 'move to China', The Telegraph, Dec 2014. Available:https://www.telegraph.co.uk/finance/personalfinance/household-bills/11291123/.

[37]. S. Rizvi, A. Aziz, N. M. Saad, N. Armi, and M. Z. Yusoff, “Tight Coupling Internetworking Between UMTS and WLAN : Challenges, Design Architectures and Simulation Analysis,” Int’l Comp. Networks, vol. 3, no. 2, pp. 116–134, 2011.

[38]. Y. Khadraoui, X. Lagrange and A. Gravey, "Very tight coupling between LTE and WiFi: From theory to practice," 2016 Wireless Days (WD), Toulouse, 2016, pp. 1-3.

[39]. S. Velentzas and T. Daguiklas, “4G Cellular/WLAN Interworking,” Motorola, ATKE, 2002.

[40]. X. Lagrange, T. Bretagne, and D. Irisa, “Very Tight Coupling between LTE and Wi-Fi for Advanced Offloading Procedures,” IEEE WCNCW'14, Apr 2014. DOI: 10.1109/WCNCW.2014.6934865.

[41]. Y. Khadraoui, X. Lagrange and A. Gravey, "Very tight coupling between LTE and WiFi: a Practical Analysis," CoRes 2016, May 2016, Bayonne, France. <hal-01306438>

[42]. B. Raghothaman, E. Deng, R. Pragada, G. Sternberg, T. Deng and K. Vanganuru, "Architecture and protocols for LTE-based device to device communication," 2013 Intl Conf. on Computing, Networking and Commun. (ICNC), San Diego, CA, 2013, pp. 895-899.

[43]. A. Asadi, Q. Wang, and V. Mancuso, “A Survey on Device-to-Device Communication in Cellular Networks,” 2014.

[44]. Y. Lin, W. Chen, S. Tsai, and Y. Lin, “Design and Implementation of An Offloading Technology for 3.5G Networks,” 2010.

[45]. 3GPP, “TS 23.234 v12.0.0 System to WLAN Interworking System Description,” 2014.

[46]. Y. Li, G. Su, P. Hui, D. Jin, L. Su, and L. Zeng, “Multiple Mobile Data Offloading Through Delay Tolerant Networks,” 2010.

[47]. S. Doumiati, H. Artail, and D. M. Gutierrez-estevez, “A framework for LTE-A Proximity-based Device-to-Device Service Registration and Discovery,” vol. 34, pp. 87–94, 2014.

[48]. Vangie Beal. (2008, April 24th). Webopedia Definition of Wi-Fi. Avaiable: http://www.webopedia.com/TERM/W/Wi_Fi.html.

[49]. A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "3GPP LTE Traffic Offloading onto WiFi Direct," pp. 135–140, 2013.

[50]. X. Lin, "Integrated Cellular and Device-to-Device Networks," PhD Dissertation, Dept of Elec. & Comp. Engg, Univ. of Texas, 2014.

[51]. S. Dimatteo, P. Hui, B. Han, and V. O. K. Li, "Cellular traffic offloading through WiFi networks," in Proc. of 8th IEEE Intl Conf. on Mobile Ad-hoc and Sensor Systems, MASS 2011, 2011, pp. 192–201.

[52]. Qualcomm, "The Evolution of Mobile Technologies," Jun., 2014.

[53]. ITU-T, "E.800 Series E: Overall Network Operation, Telephone Service, Sevice Operation and Human Factors" Aug, 2008.

[54]. IEEE, "Std 80.11e-2005 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8," Nov, 2005.

[55]. R. Braden, Clark D., and S. Shenker, "Integrated Services in the Internet Architecture : an Overview," RFC-1633, 1994.

[56]. K. Nichols, S. Blake, F. Baker, and D. Black, "rfc2474 - Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," 2010.

[57]. S. Blake, D. Black, M. Carlson, Z. Wang, E. Davies, and W. Weiss, "An Architecture of Differentiated Services," RFC-2475, 1998.

[58]. IETF, "Specification of Guaranteed Quality of Service," RFC-2212.

[59].3GPP, "QoS and Policy Aspects of 3GPP-WLAN Interworking," 3GPP TR 23.836 v1.0.0, December, 2005.

[60]. 3GPP-ETSI, "Quality of Service (QoS) concept and architecture," ETSI TS 123 107, Jun., 2016.

[61]. 3GPP-ETSI, "Policy and charging control architecture," ETSI TS 123.203, Sep., 2014.

[62]. A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in Proc. 8th Int'l Conf. MobiSys, 2008, pp. 209–222

[63]. L. Templiers, "Optimization of Delayed Mobile Data Offloading," 2013.

[64]. Y. Im, C. Joe-Wong, S. Ha, S. Sen, T. T. Kwon, and M. Chiang, "AMUSE : Empowering Users for Cost-aware Offloading with Throughput-Delay Tradeoffs," 2012.

[65]. Govt. of India Ministry of CIT Radio Division, “Mobile Data Offload,” 2008.

[66]. Shiao-Li Tsao and Chia-Ching Lin, “Design and evaluation of UMTS-WLAN Interworking Strategies,” in Proc. VTC 2002-Fall, Vol. 2, pp. 777-781, 2002.

[67]. M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, “Integration of 802.11 and Third-Generation Wireless Data Networks,” in IEEE Infocom, 2003.

[68]. J. Song, S. Lee, and D. Cho, “Hybrid Coupling Scheme for UMTS and Wireless LAN Interworking,” pp. 2247–2251, 2003.

[69]. A. K. Salkintzis, "Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks," in IEEE Wireless Comm., vol. 11, no. 3, pp. 50-61, June 2004.

[70]. 3GPP, “TR 22.934 WLAN-UMTS Interworking Rel-6,” 2008.

[71]. A. K. Salkintzis, G. Dimitriadis, D. Skyrianoglou, and N. Passas, “Seamless Continuity of Real-Time Video across UMTS and WLAN Networks: Challenges and Performance Evaluation,” 2010.

[72]. G. Lampropoulos, N. Passas, A. Kaloxylos, and L. Merakos, “A flexible UMTS / WLAN architecture for improved network performance,” pp. 889–906, 2007.

[73]. Y. Zhou, Y. Rong, H. Choi, J. Kim, J. Sohn, and H. I. Choi, “A Dual-mode Mobile Station Module for WLAN/UMTS Interworking Systems,” 2010.

[74]. J. L. Tomici and P. Chitrapu, "Multi-RAT traffic offloading solutions for the bandwidth crunch problem," IEEE Long Island Systems, Applications and Technology Conf., Farmingdale, NY, 2011, pp. 1-6.

[75]. S. T. A.Y. Ding, P. Hui, M. Kojo, “Enabling Energy-Aware Mobile Data Offloading for Smartphones through Vertical Collaboration,” in CoNEXT Studnet’12, 2012, pp. 27–28.

[76]. A. L. Singh and J. A. Stankovic, “MultiNets: A System for Real-Time Switching between Multiple Network Interfaces on Mobile Devices,” vol. 9, no. 39, 2012.

[77]. F. S. Mota, D. F. Macedo, and Y. Ghamri-doudane, “Managing the Decision-Making Process for Opportunistic Mobile Data Offloading,” 2014.

[78]. P. Baier, D. Frank, and K. Rothermel, “TOMP : Opportunistic Traffic Offloading Using Movement Predictions,” in 37th IEEE Conf. Local Computer Networks(LCN), 2012, pp. 1–9.

[79]. S. Singh, and H. S. Dhillon, “Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights,” 2013.

[80]. Y. Kim, “An Analytical Framework to Characterize the Efficiency and Delay in a Mobile Data Offloading System,” 2015.

[81]. A. Krendzel and M. Portoles-comeras, “Modeling Network Traffic in Mobile Networks Implementing Offloading,” 2011.

[82]. P. Raveneau, R. Stanica, M. Fiore, S. Uppoor, and M. Cunche, “Urban-scale Cellular Offloading through Wi-Fi Access Points : a Measurement-based Case Study,” 2015.

[83]. F. Salvatore, “Long Term Evolution (LTE): from an architectural point of view,” 2010.

[84]. 3GPP. (2016, Mar 7th). 3GPP Specifications Website. Available: http://www.3gpp.org/specifications/.

[85]. F. Mehmeti and T. Spyropoulos, "Performance Analysis of Mobile Data Offloading in Heterogeneous Networks," in IEEE Trans. on Mobile Computing, vol. 16, no. 2, pp. 482-497, Feb., 2017.

[86]. J. Ling, S. Kanugovi, S. Vasudevan and A. K. Pramod, "Enhanced capacity and coverage by Wi-Fi LTE integration," in IEEE Comm. Mag., vol. 53, no. 3, pp. 165-171, Mar 2015.

[87]. S. Eido and A. Gravey, “How much LTE traffic can be offloaded ?,” EUNICE’14, 2014.

[88]. H. Cong, L. Hu, C. Coletti, and et el., “Macro Network Realistic Indoor Wi-Fi and Femto Deployment Study as the Offloading Solution to LTE Macro Networks,” Proc. IEEE VTS Veh. Tech. Conf., 2012, pp. 1–6.

[89]. M. Bennis, M. Simsek, A. Czylwik, W. Saad, S. Valentin and M. Debbah, "When cellular meets WiFi in wireless small cell networks," in IEEE Comm. Mag., vol. 51, no. 6, pp. 44-50, June 2013.

[90]. Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan and J. C. Zhang, "Modeling and Analyzing the Coexistence of Wi-Fi and LTE in Unlicensed Spectrum," in IEEE Transac. on Wireless Comm., vol. 15, no. 9, pp. 6310-6326, Sept. 2016.

[91]. A. Pyattaev, J. Hosek, K. Johnsson, R. Krkos, M. Gerasimenko, and P. Masek, “Trial Implementation of Live D2D Technology,” vol. 37, no. 5, pp. 877–887, 2015.

[92]. S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina and Y. Koucheryavy, "Cellular traffic offloading onto network-assisted device-to-device connections," in IEEE Comm. Mag., vol. 52, no. 4, pp. 20-31, April 2014.

[93]. B. Raghothaman, G. Sternberg, S. Kaur, R. Pragada, T. Deng and K. Vanganuru, "System Architecture for a Cellular Network with Cooperative Mobile Relay," IEEE Vehicular Technology Conference (VTC Fall), San Francisco, CA, 2011, pp. 1-5.

[94]. W. Hu and G. Cao, "Quality-Aware Traffic Offloading in Wireless Networks," in IEEE MobiHoc'14, 2014,

pp.277-286.

[95]. X. Wu et al., "FlashLinQ: A Synchronous Distributed Scheduler for Peer-to-Peer Ad Hoc Networks," in IEEE/ACM Transac. on Networking, vol. 21, no. 4, pp. 1215-1228, Aug. 2013.

[96]. W. Peng, F. Li, X. Zou and J. Wu, "The Virtue of Patience: Offloading Topical Cellular Content through Opportunistic Links," 2013 IEEE 10th Intl Conf. on Mobile Ad-Hoc and Sensor Systems, Hangzhou, 2013, pp. 402-410.

[97]. Z. Wang and V. W. S. Wong, "A novel D2D data offloading scheme for LTE networks," 2015 IEEE Intl Conf. on Comm. (ICC), London, 2015, pp. 3107-3112.

[98]. A. Asadi and V. Mancuso, "WiFi Direct and LTE D2D in action," 2013 IFIP Wireless Days (WD), Valencia, 2013, pp. 1-8.

[99]. L. Valerio, R. Bruno, A. Passarella, L. Valerio, R. Bruno, and A. Passarella, “Cellular traffic offloading via Opportunistic Networking with Reinforcement Learning,” September, 2015.

[100]. EU FP7 Project, “Mobile Opportunistic Traffic Offloading - Release A.” 2011.

[101]. EU FP7 Project, “Mobile Opportunistic Traffic Offloading - Release B.” 2013.

[102]. M. A. Patiño González, T. Higashino and M. Okada, "Radio access considerations for data offloading with multipath TCP in cellular/WiFi networks," The Intl Conf. on Information Networking 2013 (ICOIN), Bangkok, 2013, pp. 680-685.

[103]. G. Lampropoulos, N. ikos Passas, L. Merakos, and A. Kaloxylos, “Handover Management Architectures in Integrated WLAN/Cellular Networks,” IEEE Commun. Surv. Tutorials, vol. 7, no. 4, pp. 30–44, 2005.

[104]. Hyeyeon Kwon, Mi-Jeong Yang, A. S. Park and S. Venkatesan, "Handover prediction strategy for 3G-WLAN overlay networks," NOMS 2008, Salvador, Bahia, 2008, pp. 819-822.

[105]. M. Lott, M. Siebert, S. Bonjour, D. Von Hugo, and M. Weckerle, “Interworking of WLAN and 3G systems,” IEEE Proc. Commun., vol. 151, no. 5, pp. 507–513.

[106]. H. Hartenstein, “A simulation study on the performance of Mobile IPv6 in a WLAN-based cellular network,” Elsevier - Computer Networks, vol. 40, pp. 191–204, 2002.

[107]. A. Dutta, S. Das, D. Famolari, et al., “Seamless Handover across Heterogeneous Networks - An IEEE 802.21 Centric Approach,” Apr., 2005.

[108]. T. A. Mosawi, D. Wisely and H. Aghvami, "A Novel Micro Mobility Solution Based on Media Independent

Handover and SIP," IEEE Vehicular Technology Conference, Montreal, Que., 2006, pp. 1-5.

[109]. T. R. Li Ma, Fei Yu, and Victor C.M. Leung, “A NEW METHOD TO SUPPORT UMTS /WLAN VERTICAL HANDOVER USING SCTP,” IEEE Comm. Mob. Resour. Mangmt, 2010, pp. 44–51.

[110]. M. Lee, G. Kim, S. Park, S. Jun, J. Nah, and O. Song, “Efficient 3G / WLAN Interworking Techniques for Seamless Roaming Services with Location-Aware Authentication,” 2010, pp. 370–381.

[111]. N. P. Mandru, D. P. Bavirisetti, and S. Khara, “An Efficient Vertical Handoff Technique for Two-Tier Heterogeneous Networks,” Int’l P2P Netw. Trends Technol., vol. 2, no. 2, pp. 50–55, 2012.

[112]. E. Bulut and B. K. Szymanski, “WiFi Access Point Deployment for Efficient Mobile Data Offloading,” PINGEN '12, vol. 17, no. 1, 2016, pp. 45-50.

[113]. E. Bulut and B. K. Szymanski, “Efficient Mobile Data Offloading Using WiFi Access Points,” in ACM SIGMOBILE, vol. 17(1), Jan 2013.

[114]. B. Han, P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao and A. Srinivasan, "Mobile Data Offloading through Opportunistic Communications and Social Participation," in IEEE Trans. on Mobile Computing, vol. 11, no. 5, pp. 821-834, May 2012.

[115]. Z. Li, Y. Shi, S. Chen, and J. Zhao, “Cellular Traffic Offloading through Opportunistic Communications Based on Human Mobility,” KSII Transac. on Internet and Inform. Systems,vol. 9, no. 3, pp. 872–885, 2015.

[116]. A.-L. Barabasi, “The origin of bursts and heavy tails in human dynamics,” Nature, vol. 435, pp. 207–211.

[117]. Q. Ye, “Small Cell and D2D Offloading in Heterogeneous Cellular Networks,” PhD Dissertation, University of Texas, 2015.

[118]. D. Camps-Mur, A. Garcia-Saavedra and P. Serrano, "Device-to-device communications with Wi-Fi Direct: overview and experimentation," in IEEE Wireless Comm., vol. 20, no. 3, pp. 96-104, June 2013.

[119]. S. Hoteit, S. Secci, S. Wietholter, A. Wolisz, C. Ziemlicki, and Zbigniew Smoreda, “Quantifying the Achievable Cellular Traffic Offloading Gain with Passpoint Hotspots Categories and Subject Descriptors,” WiMobCity’14.

[120]. S. Hoteit, S. Secci, G. Pujolle, A. Wolisz, and C. Ziemlicki, “Mobile Data Traffic Offloading over Passpoint Hotspots,” Comput. Networks, pp. 76–93, 2015.

[121]. IEEE Computer Society, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 9 : Interworking with External Networks.” 2012.

[122]. I. da Silva, Y. Wang and F. Mestanov, "A common performance monitoring system for Wi-Fi and 3GPP networks," IFIP/IEEE Intl Symposium on Integrated Network Management (IM 2013), Ghent, 2013, pp. 1078-1079.

[123]. M. Glapaand A. Mukhopadhyay, "Optimizing Wireless Networking of Wi-Fi and LTE," Oct., 2011.

[124]. F. Rebecchi, M. Dias de Amorim and V. Conan, "DROid: Adapting to individual mobility pays off in mobile data offloading," IFIP Networking Conf., Trondheim, 2014, pp. 1-9.

[125]. F. Rebecchi, M. D. De Amorim, and V. Conan, "Adaptive Mobile Traffic Offloading," ALGOTEL 2014, pp. 1–4, 2014.

[126]. V. Sciancalepore, D. Giustiniano, A. Banchs, and A. Picu, "Offloading Cellular Traffic through Opportunistic Communications : Analysis and Optimization," 2015.

[127]. T. Han and N. Ansari, "Offloading Mobile Traffic via Green Content Broker," in IEEE Internet of Things Journal, vol. 1, no. 2, pp. 161-170, April 2014.

[128]. X. Lu, P. Lio, and P. Hui, "Distance-Based Opportunistic Mobile Data Offloading," Sensors, pp. 1–11, 2016.

[129]. X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-Coupon : An Incentive Framework for 3G Traffic Offloading," pp. 206–215, 2011.

[130]. X. Zhuo, W. Gao, G. Cao, and Y. Dai, "Win-Coupon : An Incentive Framework for 3G Traffic Offloading," pp. 206–215, 2011.

[131]. L. Gao, G. Iosifidis, J. Huang, L. Tassiulas and D. Li, "Bargaining-Based Mobile Data Offloading," in IEEE Journal on Selected Areas in Comm., vol. 32, no. 6, pp. 1114-1125, Jun 2014.

[132]. 3GPP, "3GPP TR 37.834 V12.0.0," 2013.

[133]. 3GPP, "ETSI TR 122 934 V6.2.0," 2003.

[134]. 3GPP, "TR37.834 Study on Wireless Local Area Network (WLAN) - 3GPP radio interworking," Release 12, 2013.

[135]. 3GPP, "3GPP TR23.771 Study on Sytem Impacts of IMS Emergency Sessions over WLAN," Jun, 2015.

[136]. 3GPP, "TR23.937 Feasibility Study on Mobility between 3GPP-WLAN Interworking and 3GPP Systems," Aug, 2008.

[137]. 3GPP, "ETSI TS 123 401 v13.5.0 GPRS Enhancements for E-UTRAN Access," 2016.

[138]. S. Kent and K. Seo, "RFC:4301 Security Architecture for the Internet Protocol," Dec, 2005.

[139]. 3GPP2, “Interoperability Specification (IOS) for cdma2000 Machine to Machine (M2M) Communications,” Aug, 2012.

[140]. 3GPP, “Local IP Access and Selected IP Traffic Offlaod (LIPA-SIPTO),”TR 23.829, Oct., 2011.

[141]. 3GPP, “std 23852-c00 GTP SaMOG 3GPP.” Aug 2013.

[142]. D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6,” Jun 2004.

[143]. 3GPP, “ETSI TS 124 312 v13.1.0 ANDSF MO.” 2016.

[144]. 3GPP, “ETSI TS 123 261 v13.0.0.” 2016.

[145]. 3GPP, “TR 23.861 v13.0.0 Network based IP Flow Mobility.” 2008.

[146]. 3GPP, “TR 23.837,” Release 7, 2006.

[147]. 3GPP, “TR 23.890,” Release 12, 2013.

[148]. 3GPP, “ETSI TS 124 234 v12.2.0 3GPP System to WLAN Interworking WLAN UE to Network Protocols Stage 3,” 2015.

[149]. 3GPP, “ETSI TS 123 303 v12.2.0 Proximity-based Services (ProSe) Stage 2,” 2014.

[150]. 3GPP, “TR 23.703 v2.0.0 Study on Architecture Enhancements to Support Proximity-based Services (ProSe).” 2008.

[151]. 3GPP2, “3GPP2 – WLAN Interworking,” 2004.

[152]. A. De Oliva, T. Melia, A. Banchs, I. Soto, and A. Vidal, “IEEE 802.21 (Media Independent Handover services) Overview,” vol. 21, 2012.

[153]. IEEE, “IEEE Standard for Local and metropolitan area networks-Media Independent Handover Services,” Jan, 2009.

[154]. J. Pirskanen, T. Koskela, and L. Cariou, “Considerations on Cellular Traffic Offloading Use case in 802.11ah,” Jun, 2011.doc.: IEEE 802.11-11/1011-00

[155]. IEEE Computer Society, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 9: Interworking with External Networks.” 2012.

[156]. ITU-T, “ITU-T Q.3202.1 Authentication Protocols based on EAP-AKA for Interworking among 3GPP, WiMAX, and WLAN in NGN,” 2008.

[157]. P. Eronen, “RFC:4555 IKEv2 Mobility and Multihoming Protocol (MOBIKE),” RFC-4555, 2006.

[158]. Ruckus Wireless Inc., “Interworking Wi-Fi and Mobile Networks.” 2013.

[159]. Qualcomm Inc., “A 3G/LTE Wi-Fi Offload Framework : Connectivity Engine (CnE) to Manage Inter-System Radio Connections and Applications.” 2015

[160]. Alvarion Ltd., “Mobile Data Offloading Solution for 3G/LTE Networks.” 2012.

[161]. Cisco Systems, “Architecture for Mobile Data Offload over Wi-Fi Access Networks,” 2012.

[162]. Motorola Inc., “WiFi Data Offload,” 2011.

[163]. QOSMOS Inc., “Mobile Data Offloading Use Case.” 2014.

[164]. Netkrom, “Mobile Data Offloading Solution,” 2013.

[165]. Aruba Networks, “Carrier-class public Wi-Fi for 3G/4G offload.” 2012.

[166]. Aruba Networks Inc., “HybridControl Architecture for Service Providers.” 2013.

[167]. Clavister, “Intelligent Mobile Offload Intelligent and secure mobile data offloading,” 2017.

[168]. Wi-Fi Alliance, “Hotspot 2.0 (Release 2) Technical Specification,” 2015.

[169]. Wi-Fi Alliance, “Wi- Fi CERTIFIED Passpoint: An essential and strategic solution for service provider Wi-Fi deployments,” Oct 2014.

[170]. Anritsu, “Understanding WLAN offload in cellular networks.” Iss. 2, Aug 2013.

[171]. X. Chen, J. Wu, Y. Cai, H. Zhang and T. Chen, "Energy-Efficiency Oriented Traffic Offloading in Wireless Networks: A Brief Survey and a Learning Approach for Heterogeneous Cellular Networks," in IEEE Journal on Selected Areas in Comm., vol. 33, no. 4, pp. 627-640, April 2015.

[172]. A. P. Massimo Bernaschi, Filippo Cacace, Giulio Iannello, Stefano Za, “Seamless Interworking of WLANs and Cellular Networks: Architecture and Performance Issues in a Mobile IPv6 Scenario,” IEEE Wireless. Commun., 2012, pp. 73–80.

[173]. A. Aijaz, N. Uddin, O. Holland, and A. H. Aghvami, “On Practical Aspects of Mobile Data Offloading to Wi-Fi Networks,” Aug, 2015.

[174]. Shiao-Li Tsao and Chia-Ching Lin, "Design and evaluation of UMTS-WLAN interworking strategies," Proc. IEEE 56th Vehicular Technology Conf., vol.2., pp. 777-781, 2002.

[175]. K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, “Mobile data offloading: How much can wifi deliver?,” IEEE/ACM Trans. Netw., vol. 21, no. 2, pp. 536–550, 2013.

[176]. M. Amani, T. Mahmoodi, M. Tatipamula and H. Aghvami, "Programmable policies for data offloading in LTE network," IEEE Intl Conf. on Comm.(ICC), Sydney, 2014, pp. 3154-3159.

[177]. D. Camps-Mur, A. Garcia-Saavedra and P. Serrano, "Device-to-device communications with Wi-Fi Direct: overview and experimentation," in IEEE Wireless Comm., vol. 20, no. 3, pp. 96-104, June 2013.

[178]. 3GPP, “TS 24.302,” Release 9, 2010.

[179]. 3GPP, “Overview of 3GPP Release 8 V0.3.3.” 2012.

[180]. 3GPP, “Overview of 3GPP Release 10 V0.2.1,” 2014.

[181]. 3GPP, “TR 23.852,” Release 12, 2013.

[182]. Ruckus Wireless, “Interworking Wi-Fi and Mobile Networks,” USA, 2013.

[183]. Nokia-Siemens Networks, “Wi-Fi integration with cellular networks enhances the customer experience,” Finland, 2012.

[184]. Aptilo Inc., “Mobile Data Offloading,” 2015.

[185]. G. Valadon, F. Le Goff, and C. Berger, “Daily Walks In Paris : A Practical Analysis of Wi-Fi Access Points,” Proc. CoNEXT'07, pp. 5–6, Dec 2007.

[186]. nsnam.org, NS3 Community, https://www.nsnam.org/overview/community/, 2011-2015.

[187]. Youness Jouihri and Zouhair Guennoun, Circuit Switch Fallback Improvement and the Smarter Way for Implementation, 2013 AASRI Conf. on Intelligent Systems and Control, AASRI Procedia 4, Elsevier B.V, pp. 9-17, 2013

[188]. Mipro, Latency in Numbers - LTE Systems, Available:http://www.ericsson.com/hr/about/events/archieve/2007/mipro_2007/mipro_1137.pdf, 2007.

[189]. Guo Enzo, Lin James and Zhang Yang, The Performance Analysis of LTE Network, Communication Networks-ENSC 427 Project Report, SFU - Canada, 2014.

[190]. Azana Hafizah Mohd Aman, Aisha-Hassan A. Hashim, Huda Adibah Mohd Ramli and Shayla Islam, Pakcet Loss and Packet Delivery Evaluation Using Network Simulator for Multicast Enabled Network Mobility Management, Int'l Journal of Future Generation Commun. and Networking, Vol. 10, No. 4, pp. 41-50, 2017. ISSN: 2233-7857.

[191]. Chiapin Wang and Yu-Chun Cheng, Peformance Analysis of Packet Loss Rate and Packet Time Delay of IEEE 802.11 WLANs, 2nd Int'l Symposium on Computer, Communication, Control and Automation (3CA-2013), Atlantis Press, pp. 373-377, 2013.

[192]. Jumana Bukhari and Nadine Akkari, QoS based Approach for LTE-WiFi Handover, 2016 7th Int'l Conf. on

Computer Science and Information Technology, 2016. ISBN: 978-1-4673-8913-6.

[193]. Nbil Mesabahi and Hamza Dahmouni, Delay and Jitter Analysis in LTE Networks, IEEE, 2016. ISBN: 978-1-5090-3837-4/16.

[194]. ITU-T G.1028, End-to-End Quality of Service for Voice over 4G Mobile Networks, April 2016. ID: 11.1002/1000/12748.