Incorporating Context Awareness in Cellular Networks to Enhance System Performance and User Mobility Support

Integration von Kontextbewusstsein in Mobilfunknetze zur Verbesserung der Systemleistung und Unterstützung der Benutzermobilität

von

Nandish P. Kuruvatti, M.Sc

geboren in Davanagere, Indien

Vom Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
genehmigte Dissertation

D 386

Tag der mündlichen Prüfung:	26 August 2020
Dekan des Fachbereichs:	Prof. Dr.-Ing. Ralph Urbansky
Prufüngsvorsitzender:	Prof. Dr.-Ing. Stefan Götz
1.Berichterstatter:	Prof. Dr.-Ing. Hans D. Schotten
2.Berichterstatter:	Prof. Dr.-Ing. habil. Madhukar Pandit

Own Publications

1. Kuruvatti, N.P., Klein, A., Schneider, J., Schotten, H.D., Exploiting Diurnal User Mobility for Predicting Cell Transitions, *International Workshop on Broadband Wireless Access at IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, USA, 2013.

2. Kuruvatti, N.P., Molano, J., Schotten, H.D., Mobility Context Awareness to Improve Quality of Experience in Traffic Dense Cellular Networks, *International Conference on Telecommunications (ICT)*, Limassol, Cyprus, 2017.

3. Kuruvatti, N.P., Schotten, H.D., Framework to Support Mobility Context Awareness in Cellular Networks, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Nanjing, China, 2016.

4. Kuruvatti, N.P., Zhou, W. and Schotten, H.D., Mobility Prediction of Diurnal Users for Enabling Context Aware Resource Allocation, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Nanjing, China, 2016.

5. Kuruvatti, N.P., Molano, J., Schotten, H.D., Monitoring Vehicular User Mobility to Predict Traffic Status and Manage Radio Resources, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Sydney, Australia, 2017.

6. Kuruvatti, N.P., Klein, A., Schotten, H.D., Prediction of Dynamic Crowd Formation in Cellular Networks for Activating Small Cells, *International Workshop on 5G Architecture at IEEE-Vehicular Technology Conference (VTC-Spring)*, Glasgow, Scotland, 2015.

7. Kuruvatti, N.P., Schotten, H.D., Post-Resource Sharing Power Allocation in Cellular Networks to Coexist with D2D Underlay, *International Conference on Networks of the Future (NoF)*, Buzios, Brazil, 2016.

8. Kuruvatti, N.P., Klein, A., Lianghai, J., Sattiraju, R. and etal., Robustness of Location Based D2D Resource Allocation Against Positioning Errors, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Glasgow, Scotland, 2015.

9. Kuruvatti, N.P., Mutabazi, H.M., Schotten, H.D., Exploiting Mobility Context Awareness in Cellular Networks for Assisting Vehicular Use Cases, *IEEE-Vehicular Technology Conference (VTC-Fall)*, Chicago, USA, 2018.

10. Kuruvatti, N.P., Schotten, H.D., Mobility Context Awareness in Heterogeneous Networks to Enhance Multipath Communications, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Kuala Lumpur, Malaysia, 2019.

11. Kuruvatti, N.P., Hernandez, R., Schotten, H.D.,Interference Aware Power Management in D2D Underlay Cellular Networks, *IEEE-AFRICON conference*, Accra, Ghana, 2019.

12. Kuruvatti, N.P., Mallikarjun, B.M., Kusumapani, S.C., Schotten, H.D., Mobility Awareness in Cellular Networks to Support Service Continuity in Vehicular Users, *International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2020.

13. Klein, A., Kuruvatti, N.P., Schneider, J., Schotten, H.D., Fuzzy Q-Learning for Mobility Robustness Optimization in Wireless Networks, *International Workshop on Emerging Technologies for LTE-Advanced and Beyond 4G at IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, USA, 2013.

14. Chakraborty, P., Kuruvatti, N.P., Schotten, H.D., A Novel Serious Game Engineering Based Interactive Visualization and Evaluation Platform For Cellular Technologies, *International Symposium on Networks, Computers and Communications (ISNCC)*, Marrakech, Morocco, 2017.

15. Sattiraju, R., Klein, A., Lianghai, J., Kuruvatti, N.P. and et al., Virtual Cell Sectoring for Enhancing Resource Allocation and Reuse in Network Controlled D2D Communication, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Glasgow, Scotland, 2015.

16. Ambekar, A., Kuruvatti, N.P., Schotten, H.D., Improved Method of Secret Key Generation Based on Variations in Wireless Channel, *International Conference on Systems, Signals and Image Processing (ISSIP)*, Vienna, Austria, April 2012.

17. Lianghai, J., Klein, A., Kuruvatti, N.P., Schotten, H.D., System Capacity Optimization Algorithm for D2D Underlay Operation, *International Workshop on 5G Technologies at IEEE International Conference on Communications (ICC)*, Sydney, Australia, 2014.

18. Lianghai, J., Klein, A., Kuruvatti, N.P., Sattiraju, R, Schotten, H.D., Dynamic Context-aware Optimization of D2D Communications, *International Workshop on 5G Mobile systems at IEEE-Vehicular Technology Conference (VTC-Spring)*, Seoul, Republic of Korea, 2014.

19. Kilinic, C., Monserrat, J.F., Filippou, M.C., Kuruvatti, N.P. and et al., New Radio 5G User Plane Design Alternatives, *International Workshop on 5G RAN Design at IEEE GLOBECOM*, Washinton DC, USA, 2016.

20. Kilinic, C., Ericson, M., Rugeland, P., Zaidi, A., Aydin, O., Kuruvatti, N.P. and et al., 5G Multi-RAT integration evaluations using a common PDCP layer, *IEEE-Vehicular Technology Conference (VTC-Spring)*, Sydney, Australia, 2017.

21. Pateromichelakis, E., Gebert, J., Mach, T., Belschner, J., Guo, W., Kuruvatti, N.P. and et al., Service-Tailored User-Plane Design Framework and Architecture Considerations in 5G Radio Access Networks, *IEEE Open Access Journal*, 2017

22. Sacristan, D., Herranz, C., Monserrat, J.F., Szczygiel, A., Kuruvatti, N.P., Roger, D.G. and et al., 5G Visualization: The METIS-II Project Approach, *Mobile Information Systems Journal*, 2018.

Acknowledgements

I would like to thank Prof. Dr-Ing. Hans D. Schotten for being my PhD supervisor and for giving me an opportunity to work in his research group. I am grateful to him for all the research opportunities that were provided through exceptional projects.

I would like to express my sincere gratitude towards Prof. Dr-Ing. habil. Madhukar Pandit for being my second supervisor. I thank him from the bottom of my heart for all his constructive feedback towards my PhD. Words cannot express how grateful I am for his kindness, patience and goodwill towards me.

I would like to thank my colleagues at WICON as well as collaborators from the METIS and 5GNetmobil projects for their constant support and motivation. A big thank you to Andreas Klein and Abhijit Ambekar for their mentorship during the early days of my research. A heartfelt thanks to Lianghai Ji for his generosity and support during various phases of my PhD. It was a privilege. I would like to thank my friends in Germany for being a family away from home.

Last but not the least, I would like to thank my family; my father, my mother and my aunt. Without their blessings none of this would have been possible. Thanks to my sister Sowmya and my brother-in-law Parveen for all the support. Thanks to my best friends Harsha B.S and Naveen K.R for always being there. And a bow to my wife Vinanthi for her love and support!

Abstract

In today's world, mobile communication has become one of the most widely used technologies corroborated by growing number of mobile subscriptions and extensive usage of mobile multimedia services. It is a key challenge for the network operators to accommodate such large number of users and high traffic volume. Further, several day-to-day scenarios such as public transportation, public events etc., are now characterized with high mobile data usage. A large number of users avail cellular services in such situations posing high load to the respective base stations. This results in increased number of dropped connections, blocking of new access attempts and blocking of handovers (HO). The users in such system will thus be subjected to poor Quality of Experience (QoE). Beforehand knowledge of the changing data traffic dynamics associated with such practical situations will assist in designing radio resource management schemes aiming to ease the forthcoming congestion situations. The key hypothesis of this thesis is that consideration and utilization of additional context information regarding user, network and his environment is valuable in designing such smart Radio Resource Management (RRM) schemes.

Methods are developed to predict the user cell transitions, considering the fact that mobility of the users is not purely random but rather direction oriented. This is particularly used in case of traffic dense moving network or group of users moving jointly in the same vehicle (e.g., bus, train, etc.) to predict the propagation of high load situation among cells well in advance. This enables a proactive triggering of load balancing (LB) in cells anticipating the arrival of high load situation and accommodating the incoming user group or moving network. The evaluated KPIs such as blocked access attempts, dropped connections and blocked HO are reduced by around 25%, 42% and 49%, respectively.

Further, everyday scenario of dynamic crowd formation is considered as another potential case of high load situation. In real world scenarios such as open air festivals, shopping malls, stadiums or public events, several mobile users gather to form a crowd. This poses high load situation to the respective serving base station at the site of crowd formation, thereby leading to congestion. As a consequence, mobile users are subjected to poor QoE due to high dropping and blocking rates. A framework to predict crowd formation in a cell is developed based on coalition of user cell transition prediction,

cluster detection and trajectory prediction. This framework is suitably used to prompt context aware load balancing mechanism and activate a small cell at the probable site of crowd formation. Simulations show that proactive LB reduces the dropping of users (23%), blocking of users (10%) and blocked HO (15%). In addition, activation of a Small Cell (SC) at the site of frequent crowd formation leads to further reductions in dropping of users (60%), blocking of users (56%) and blocked HO (59%).

Similar to the framework for crowd formation prediction, a concept is developed for predicting vehicular traffic jams. Many vehicular users avail broadband cellular services on a daily basis while traveling. The density of such vehicular users change dynamically in a cell and at certain sites (e.g. signal lights), traffic jams arise frequently leading to a high load situation at respective serving base station. A traffic prediction algorithm is developed from cellular network perspective as a coalition strategy consisting of schemes to predict user cell transition, vehicular cluster/moving network detection, user velocity monitoring etc. The traffic status indication provided by the algorithm is then used to trigger LB and activate/deactivate a small cell suitably. The evaluated KPIs such as blocked access attempts, dropped connections and blocked HO are reduced by approximately 10%, 18% and 18% , respectively due to LB. In addition, switching ON of SC reduces blocked access attempts, dropped connections and blocked HO by circa 42%, 82% and 81%, respectively.

Amidst increasing number of connected devices and traffic volume, another key issue for today's network is to provide uniform service quality despite high mobility. Further, urban scenarios are often characterized by coverage holes which hinder service continuity. A context aware resource allocation scheme is proposed which uses enhanced mobility prediction to facilitate service continuity. Mobility prediction takes into account additional information about the user's origin and possible destination to predict next road segment. If a coverage hole is anticipated in upcoming road, then additional resources are allocated to respective user and data is buffered suitably. The buffered data is used when the user is in a coverage hole to improve service continuity. Simulation shows improvement in throughput (in coverage hole) by circa 80% and service interruption is reduced by around 90%, for a non-real-time streaming service.

Additionally, investigation of context aware procedures is carried out with a focus on user mobility, to find commonalities among different procedures and a general framework is proposed to support mobility context awareness. The new information and interfaces which are required from various entities (e.g., vehicular infrastructure) are discussed as well.

Device-to-Device (D2D) communications commonly refer to the technology that enables direct communication between devices, hence relieving the base station from traffic routing. Thus, D2D communication is a feasible solution in crowded situations, where users in proximity requesting to com-

municate with one another could be granted D2D links for communication, thereby easing the traffic load to serving base station. D2D links can potentially reuse the radio resources from cellular users (known as D2D underlay) leading to better spectral utilization. However, the mutual interference can hinder system performance. For instance, if D2D links are reusing cellular uplink resources then D2D transmissions cause interference to cellular uplink at base station. Whereas, cellular transmissions cause interference to D2D receivers. To cope up with such issues, location aware resource allocation schemes are proposed for D2D communication. The key aim of such RA scheme is to reuse resources with minimal interference. The RA scheme based on virtual sectoring of a cell leads to approximately 15% more established links and 25% more capacity with respect to a random resource allocation.

D2D transmissions cause significant interference to cellular links with which they reuse physical resource blocks, thereby hindering cellular performance. Regulating D2D transmissions to mitigate the aforementioned problem would mean sub-optimal exploitation of D2D communications. As a solution, post-resource allocation power control at cellular users is proposed. Three schemes namely interference aware power control, blind power control and threshold based power control are discussed. Simulation results show reductions in dropping of cellular users due to interference from D2D transmissions, improvement in throughput at base station (uplink) while not hindering the D2D performance.

The concepts and framework developed for exploiting context information represent main contributions of the author to the European FP7 project METIS [MP12] and Horizon 2020 project METIS-II [Pro15] in evaluating the performance of future wireless technologies and the fifth generation of wireless communication systems (5G). In essence, the proposed concepts and presented results incite and validate the consideration of context awareness as a key enabler for intelligent and autonomous management of network. Additionally, the research activities conducted in the scope of this thesis are very relevant for future 5G wireless systems and ongoing discussions in 5G Infrastructure Public Private Partnership (5G-PPP) [GP15a].

Keywords - context awareness, radio resource management, cellular networks, next cell prediction, route prediction, congestion, hotspot, dynamic crowd, vehicular traffic jam, load balancing, small cell activation, D2D resource allocation, resource reuse, power control

Zusammenfassung

In der heutigen Welt ist die mobile Kommunikation zu einer der am weitesten verbreiteten Technologien geworden, was durch die wachsende Anzahl von Mobilfunk-Abonnements und die umfassende Nutzung mobiler Multimedia-Dienste bestätigt wird. Für die Netzbetreiber ist es eine zentrale Herausforderung, eine so große Anzahl von Nutzern und ein hohes Datenaufkommen zu bewältigen. Darüber hinaus tritt in vielen alltäglichen Szenarien wie z.B. imöffentlichen Nahverkehr, bei öffentliche Veranstaltungen, usw. eine hohe mobile Datennutzung auf. In solchen Situationen, in denen die jeweiligen Basisstationen stark belastet sind, nutzen eine Vielzahl von Nutzern Mobilfunkdienste. Dies führt zu einer erhöhten Anzahl von Verbindungsabbrüchen, zur Blockierung neuer Zugriffsversuche und zur Blockierung von Handovern (HO). Die Benutzer in einem solchen System erfahren somit eine schlechte Servicequalität (Quality of Experience (QoE)). Die a-priori Kenntnis der sich ändernden Datenverkehrsdynamik, die mit solchen praktischen Situationen verbunden ist, kann bei der Entwicklung von Programmen,die darauf abzielen, die bevorstehende überlastungssituation zu verringern, das Management von Funkressourcen zu unterstützen. Die Haupthypothese dieser Arbeit ist, dass die Berücksichtigung und Nutzung zusätzlicher Kontextinformationen über Benutzer, Netzwerk und Umgebung bei der Entwicklung solcher intelligenten Radio Resource Management (RRM)-Algorithmen wertvoll ist.

Es werden Methoden zur Vorhersage von Benutzerzellenübergängen entwickelt, wobei berücksichtigt wird, dass die Mobilität der Benutzer nicht zufällig ist, sondern richtungsorientiert ist. Dies wird insbesondere bei dichtem Datenverkehr oder bei einer Gruppe von Benutzern, die sich gemeinsam im selben Fahrzeug bewegen (z.B. Bus, Bahn, etc.), verwendet, um die Ausbreitung der Hochlastsituation zwischen den Zellen frühzeitig vorherzusagen. Dies ermöglicht eine proaktive Auslösung des Load Balancing (LB) in Zellen, vor dem Eintreffen einer Hochlastsituation. Um die eingehende Benutzergruppe oder das moving network zu berücksichtigen. Die ausgewerteten (Key Performance Indicators) KPIs wie Blocked Access attempts, Dropped user Connections und Blocked HO attempts, werden um rund 25%, 42% bzw. 49% reduziert.

Darüber hinaus wird das Alltagsszenario der dynamischen Massenbildung als weiterer potenzieller Fall einer Hochlastsituation betrachtet. In realen Szenarien wie z.B. Open-Air-Festivals, Einkaufszentren, Stadien oder öf-

fentlichen Veranstaltungen versammeln sich viele Mobilfunknutzer zu einer Menschenmenge. Dies stellt eine hohe Belastungssituation für die jeweilige bedienende Basisstation am Ort der Massenbildung dar und führt zu Congestion. Infolgedessen sind viele Nutzer aufgrund hoher Fall- und Blockierraten einer schlechten QoE ausgesetzt. Ein Framework zur Vorhersage der Massenbildung in einer Zelle wird auf Grundlage einer Kombination aus Benutzerzellen übergangsvorhersage, Cluster-Erkennung und Trajektorienvorhersage entwickelt. Dieses Framework wird geeignet eingesetzt, um kontextbezogene Lastausgleichsmechanismen anzustoßen und Small Cell (SC) an einem wahrscheinlichen Ort der Massenbildung zu aktivieren. Simulationen zeigen, dass das proaktive LB das Dropped user Connections (23%), Blocked Access attempts (10%) und Blocked HO attempts(15%) reduziert. Darüber hinaus führt die Aktivierung der SC an der Stellen der häufigen Menschenmengen entsthen zu einer weiteren Verringerung des Dropped user Connections (60%), Blocked Access attempts (56%) und Blocked HO attempts (59%).

Ähnlich wie das Framework zur Vorhersage von Massenbildung wird ein Konzept zur Vorhersage von Staus entwickelt. Viele Fahrzeugnutzer nutzen während der Fahrt täglich Breitbanddienste. Die Dichte solcher Fahrzeugnutzer ändert sich dynamisch in einer Zelle und an bestimmten Stellen (z.B. Ampeln), Es entstehen häufig Staus, die zu einer hohen Belastungssituation an der jeweiligen bedienenden Basisstation führen. Ein Verkehrsvorhersagealgorithmus wird aus der Perspektive des Mobilfunknetzes als Kombinationstrategie entwickelt, die aus der Vorhersage des übergangs zwischen Benutzerzellen, der Erkennung von Fahrzeugclustern und -bewegungen, der überwachung der Benutzergeschwindigkeit und vielen mehr besteht. Die vom Algorithmus bereitgestellte Verkehrsstatusanzeige wird dann verwendet, um LB auszulösen und die SC entsprechend ein- und auszuschalten. Die ausgewerteten KPIs wie z.B. Blocked Access attempts, Dropped user Connections und Blocked HO attempts werden um ca. 10%, 18% und 18% reduziert. Darüber hinaus reduziert das Einschalten von SC die Blocked Access attempts, Dropped user Connections und Blocked HO attempts ca. 42%, 82% bzw. 81%.

Bei zunehmender Anzahl von verbundenen Geräten und steigendem Datenverkehrsaufkommen ist es für heutige Netze auch relevant, trotz hoher Mobilität eine einheitliche Servicequalität zu gewährleisten. Darüber hinaus sind urbane Szenarien oft durch Versorgungslücken gekennzeichnet, die die Kontinuität des Dienstes behindern. Es wird ein kontextbezogenes Ressourcenzuweisungsschema vorgeschlagen, das eine verbesserte Mobilitätsvorhersage nutzt, um die Kontinuität der Dienste zu erleichtern. Bei der Mobilitätsvorhersage werden zusätzliche Informationen über den Ursprung und das mögliche Ziel des Benutzers berücksichtigt, um das nächste Straßensegment vorherzusagen. Ist mit einem Deckungsloch auf der anstehenden Strecke zu rechnen, werden dem jeweiligen Nutzer zusätzliche Ressourcen zugewiesen

und die Daten entsprechend zwischengespeichert. Die gepufferten Daten werden während der Abdeckung verwendet, um die Servicekontinuität zu verbessern. Die Simulation zeigt eine Verbesserung des Durchsatzes (in Funklöchern) um circa 80% und die Betriebsunterbrechung wird um ca. 90% reduziert für nicht-Echtzeitanwendungen, wie Streaming-Services.

Darüber hinaus wird eine Untersuchung kontextbezogener Verfahren mit Schwerpunkt auf der Benutzermobilität durchgeführt, um Gemeinsamkeiten zwischen verschiedenen Verfahren zu finden und es wird ein allgemeiner Rahmen zur Unterstützung der Sensibilisierung für den Mobilitätskontext vorgeschlagen. Die neuen Informationen und Schnittstellen, die von verschiedenen Stellen (z.B. Fahrzeuginfrastruktur) benötigt werden, werden ebenfalls diskutiert.

Mithilfe der direkte Kommunikation zwischen zwei Endnutzern (Device-to-Device (D2D) Kommunikation) kann die Datenverkehrslast im zellularen Netz gesenkt werden. Dies wird ermöglicht, in dem Nutzer, welche sich in lokaler Nähe zueinander befinden, eine gegenseitige Funkverbindung aufbauen und Daten auszutauschen, statt eine zellulare Verbindung dazu zu nutzen. D2D-Verbindungen können die Funkressourcen von Mobilfunknutzern potenziell wiederverwenden, was zu einer besseren spektralen Nutzung führt (sogenanntes D2D-Underlay). Die dabei auftretenden Interferenz können jedoch die gegenseitige Beeinflussung kann jedoch die Systemleistung beeinträchtigen. Dies ist beispielsweise der Fall wenn für D2D-Verbindungen zellulare Uplink Resourcen wiederverwendet werden. Im Gegenzug können zellulare Nutzer ebenso D2D Nutzer stören. Zur Vermeidung dieser Einflüsse, werden standortbezogene Ressourcenallokationsprogramme für die D2D-Kommunikation vorgeschlagen. Das Hauptziel eines solchen RA-Schema ist die Wiederverwendung von Ressourcen bei minimaler Beeinträchtigung. Das RA-Schema, das auf virtuellen Sektorisierung einer Zelle basiert, führt zu ca. 15% mehr Aufgebauten Verbindungen und zu einer um 25% erhöhten Zellkapazität in Vergleich zur zufälligen Ressourcenzuteilung.

D2D-übertragungen verursachen signifikante Störungen der Mobilfunkverbindungen, mit denen sie sich physische Ressourcenblöcke teilen. Die Regulierung von D2D-übertragungen zur Minderung des oben genannten Problems würde eine suboptimale Nutzung der D2D-Kommunikation bedeuten. Als Lösung wird nach der Ressourcenzuweisung eine Regulung der Sendeleistung (Power Control) bei Mobilfunknutzern vorgeschlagen. Drei Schemata, nämlich Interference Aware Power Control, Blind Power Control und Threshold based Power Control, werden diskutiert. Simulationsergebnisse zeigen eine Verringerung der Verbindungsabbrüche von Mobilfunknutzern aufgrund von Störungen durch D2D übertragungen und eine Verbesserung des Durchsatzes an der Basisstation (Uplink), ohne die D2D-Leistung zu beeinträchtigen.

Die Konzepte und Rahmenbedingungen für die Nutzung von Kontextinformationen stellen die wichtigsten Beiträge des Autors zum europäischen FP7-

Projekt METIS[MP12] und zum Horizon-2020-Projekt METIS-II[Pro15] bei der Bewertung der Leistung zukünftiger drahtloser Technologien und der fünften Generation von drahtlosen Kommunikationssystemen (5G) dar. Die vorgeschlagenen und validierten Konzepte zeigen dabei auf, wie die Berücksichtigung von Kontextinformationen als Schlüsselfaktor für ein intelligentes und autonomes Netzwerkmanagement dienen kann. Darüber hinaus sind die Forschungsaktivitäten, die im Rahmen dieser Arbeit durchgeführt wurden, relevant für das zukünftige 5G Mobilfunksystem und laufende Diskussionen in der 5G Infrastruktur Public Private Partnership (5G-PPP)[GP15a].

Contents

List of Figures

List of Tables

List of Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
4G	Fourth Generation
5G	Fifth Generation
3GPP2	3rd Generation Partnership Project 2
3GPP	3rd Generation Partnership Project
AMC	Adaptive Modulation and Coding
AMPS	Advanced Mobile Phone System
AM	Amplitude Modulation
AOA	Angle of Arrival
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
BS	Base Station
C-CAST	Context Casting
CDMA	Code Division Multiple Access
CFI	Crowd Formation Indicator
CIO	Cell Individual Offset
CQI	Channel Quality Indicator
CSI	Channel State Information
C-UE	Cellular User Equipment

D2D	Device-to-Device
DL	Downlink
D-UE	D2D User Equipment
EDGE	Enhanced Data Rates for GSM Evolution
EMA	Exponential Moving Average
eNB	Evolved Node B
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FM	Frequency Modulation
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile communications
HARQ	Hybrid Automatic Repeat Request
HOM	Handover Margin
HO	Handover
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
HYS	Hysteresis
IEEE	Institute of Electrical and Electronics Engineers
IMT-2000	International Mobile Telecommunications-2000
IMTS	Improved Mobile Telephone Service
IoT	Internet of Things
ISDN	Integrated Services Digital Network

ITU-R	International Telecommunication Union-Radio Communications
KPI	Key Performance Indicators
L2S	Link-to-System
LB	Load Balancing
LOS	Line of Sight
LTE-A	Long Term Evolution-Advanced
LTE	Long Term Evolution
M2M	Machine-to-Machine
MAC	Media Access Control
MIMO	Multiple Input Multiple Output
MLB	Mobility Load Balancing
MMS	Multimedia Messaging Service
MNO	Mobile Network Operator
MRO	Mobility Robustness Optimization
MTC	Machine Type Communication
MTS	Mobile Telephone Service
NCL	Neighboring Cells List
NGMN	Next Generation Mobile Networks
NLOS	Non Line of Sight
NMT	Nordic Mobile Telephone
NR	New Radio
OFDMA	Orthogonal Frequency Division Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OPEX	Operational Expenditures
OSS	Operations Support System
PC	Power Control

PL	Pathloss
PRB	Physical Resource Block
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RA	Resource Allocation
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RWP	Random Way Point
s2s	Site-to-Site
SC	Small Cell
SINR	Signal-to-Interference-plus-Noise Ratio
SISO	Single Input Single Output
SMS	Short Message Service
SON	Self Organizing Networks
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TDOA	Time Difference of Arrival
TSI	Traffic Status Indicator
TTI	Transmission Time Interval
TTT	Time-To-Trigger
UE	User Equipment
UL	Uplink

UMTS	Universal Mobile Telecommunication System
V2X	Vehicle-to-everything
WAP	Wireless Application Protocol
WCDMA	Wide-band Code Division Multiple Access
uMTC	Ultra-reliable Machine-Type Communication
xMBB	Extreme Mobile Broadband

Chapter 1

Introduction

Mobile communications is one of the most ubiquitously availed technologies in contemporary world, with its application horizon extending from user centric hand-held devices to industrial communication, traffic safety and health. With 7.6 billion subscriptions worldwide in year 2017 [CJC17], mobile communications can be considered as one of the key technologies of 21st century. However, the advent of mobile communications has its origins dating back as far as the 19th century. In 1865, scientist James Clerk Maxwell demonstrated theoretically the existence of Electromagnetic waves in his publication *A Dynamical Theory of the Electromagnetic Field* [Max65] and this theory was experimentally validated by Heinrich Hertz (1886) [Hei05]. Following these events, several scientists contributed to the development of wireless telegraphy with their early experimentation e.g., Nikola Tesla (1891), J.C. Bose (1894), Oliver Lodge (1894), Alexander Popov (1895) and so on. Benefiting from these fundamental works, the Hertzian wave telegraphy practical enough for communication was invented and patented by Guglielmo Marconi (1897), kicking off the commercialization of practical radio systems. Since then, the radio communication has paved its way into human lives, with its immense development and usage during World Wars I and II for information broadcast and military applications.

Ship-to-shore communications was among the first applications of mobile telephony (1919) using AM technology at 4.2 and 8.7 MHz [Sch05]. Further, the police department started making use of radio communication (1928) using small rugged radios installed in their automobiles at 35 MHz. Over the following years, a series of tests were conducted at different frequencies, transmission paths, distances etc., to study the effects of signal reflection, refraction, diffraction and so on. These tests helped in understanding propagation effects in a wireless channel and led to several other advancements in

mobile telephony e.g., usage of FM for reliable radio transmission over AM, operation at higher frequency ranges and more user channels. In 1946, the Federal Communications Commission, FCC, in the USA licensed operation of the first land-mobile telephone system. The Mobile Telephone Service (MTS) provided by Bell labs used FM transmission in the 150 MHz frequency band with channels or carrier frequencies spaced 120 kHz apart. However, since the receivers were inefficient in discriminating well between adjacent channels, neighboring cities were restricted only to use alternate channels spaced 120 kHz apart. A separation of 80 kilometers was required between systems. The system could accommodate forty channels or simultaneous calls, with calls placed through a manual operator and the calls being half-duplex in nature, i.e only one side of the connection being able to communicate at any given time in the call.

Following the parallel advancements in semiconductor technology, new semiconductor devices with low cost and power requirements were made available, improving the circuitry in phones. By 1965, this helped the Bell labs introduce Improved Mobile Telephone Service (IMTS) with distinctive features such as, reduced FM channel spacing of 30 KHz, automatic dialing, scan to select an idle channel not assigned to any user and full duplex nature of calls, i.e both parties in a connection were able to communicate at any given time in a call. However, the system was able to accommodate only 800-1000 users in a given area of 30-50 kilometers, whereas waiting list was being as long as 25,000 users [Sch05]. Thus, the limited spectrum availability for this service posed a crucial challenge for the system, to provide IMTS service to its subscribers with low blocking probability.

In order to cope up with the limited spectral resources and yet accommodate more users, engineers at Bell labs revisited a proposal made by them in 1947, to introduce a cellular geographical structure. In cellular principle, the given service region is divided into smaller contiguous geographical areas referred to as *cells* and total available spectrum (frequency channels) is divided among these cells. The same channels can be reused among cells, which are far-enough apart from one another to cause any significant interference. Even though the number of channels assigned per cell is smaller, the frequency reuse at a short-enough distance assures accommodation of more users in total service area. In such a system, however, a problem will arise when a mobile user travels from one cell to another. Their ongoing calls should be assigned a new frequency channel in respective cell the user enters. This procedure of re-assigning channels is called handover.

Before the invention of cellular radio concept and advancements in handover concepts, e.g., three-sided trunk circuit [Joe70], cellular telephone switching plan [FN73] etc., the mobile telephony was limited to bulky phones installed in vehicles as a part of PSTN. The vehicles were required to be in coverage area of a radio transmitter or base station throughout the call without a possibility of handover. Nevertheless, with subsequent advancements in cellular concepts, the first hand-held mobile phones were introduced by Motorola in 1973, eventually rolling out the generations of mobile communications that followed.

The first generation (1G) of mobile communications was based on analog transmission with a lot of research done in formulating models for entities that determined performance of analog telephone systems, for e.g., pathloss, doppler spectra, fading statistics etc. The Nippon Telephone and Telegraph (NTT) deployed the first commercial analog cellular systems in Tokyo, Japan in 1979. However, Nordic Mobile Telephone (NMT) system built in Nordic countries in 1981 was the first analog system with large coverage and standard handover procedures, which combined several cells in a large area into a single network [Mol11]. Following this, several countries developed their own standards. The standard implemented in the USA was called Advanced Mobile Phone System (AMPS). These systems used analog FM transmission and covered two 25 MHz wide bands (824-849 MHz in uplink, 869-894 MHz in downlink). The systems used frequency division multiple access (FDMA), with 25 MHz bandwidth in each direction of transmission split into 30 KHz wide channels [Sch05]. Although, cellular based AMPS systems showed improvements in mobile capacity upon their introduction, soon the additional issues arising from growing capacity demand, security and spectral efficiency led to investigations for further alternatives.

The second generation (2G) of mobile communications manifested transition from analog to digital signaling. The key research done in area of digital communications such as spectral efficient modulation schemes, impact of channel variations on digital signals, channel coding and multiple access schemes etc., prompted this transition [Mol11]. Two major standards were established namely, Global System for Mobile communications (GSM) in Europe (1991) and Interim Standard (IS-95) developed in the USA (1995). The former used time division multiple access (TDMA) whereas, the latter would use code division multiple access (CDMA). In IS-95, the 25 MHz-wide bands allocated for mobile service were split into CDMA channels 1.25 MHz wide, whereas GSM used TDMA in 25 MHz bands of 890-915 MHz for uplink

and 935-960 MHz for downlink communications [Sch05]. Several TDMA or CDMA based 2G standards were developed by different countries and countries like Japan and USA allowed a wide variety of standards work in the same geographical regions, allowing consumers to choose between different technical standards [Mol11]. In general, 2G systems provided international roaming, authentication, encryption on the wireless link, automatic location services, efficient inter-operation with ISDN systems, and a comparatively high audio quality. In addition, a short message service (SMS) with up to 160 alphanumeric characters and data services at 9.6 kbit/s were offered [Sch03]. The packet switched data services were introduced along with conventional 2G circuit switched domain, giving rise to the 2.5G standard, General Packet Radio Service (GPRS). This enabled various additional services such as Multimedia Messaging Service (MMS), instant messaging and other internet applications through wireless application protocol (WAP). Further, addition of higher order modulation and coding schemes allowed GSM to evolve into Enhanced Data Rates for GSM Evolution (EDGE) and its respective packet data counterpart Enhanced General Packet Radio Service (EGPRS) [OMM16].

The third generation (3G) of mobile communications was envisioned as a system that satisfies International Mobile Telecommunications-2000 (IMT-2000) specifications developed by the International Telecommunication Union-Radio Communications (ITU-R) [OMM16]. The key requirements were in terms of high data rates (up to 2 Mbps), bandwidth on demand, delay and quality requirements, high spectral efficiency, backward compatibility etc [HT05]. Universal Mobile Telecommunication System (UMTS) developed by the European Telecommunications Standards Institute (ETSI) was one of the first major 3G mobile communication system that met IMT-2000 specifications (2001). It adopted Wide-band Code Division Multiple Access (WCDMA) as its air interface and the system was designed for multimedia communication enabling transfer of high quality images and video as part of person-to-person communication. In addition, access to services on public and private networks was enhanced by higher data rates and flexible communication capabilities [HT05]. New specifications were developed in the framework of the 3rd Generation Partnership Project (3GPP) for evolution of 3G to 3.5G (2006), named as High Speed Packet Access (HSPA). It included High Speed Downlink Packet Access (HSDPA) added in 3GPP release 5 and High Speed Uplink Packet Access (HSUPA) added in 3GPP release 6. These standards were able to offer packet data rate up to 14.6 Mbps in

downlink and up to 5.76 Mbps in uplink. With the introduction of Multiple Input Multiple Output (MIMO) antennas, these standards evolved further to handle higher data rates [OMM16]. These 3GPP standards maintained backward compatibility while new features were being added. Thus, without affecting existing terminals in market, HSPA further evolved into HSPA+ (2008) incorporating carrier aggregation for higher peak data rates.

The fourth generation (4G) of mobile communications was the second UMTS evolution called Long Term Evolution (LTE) [OMM16]. This standard introduced a new air interface based on Orthogonal Frequency Division Multiple Access (OFDMA) and a new architecture and Core Network (CN) called the System Architecture Evolution/Evolved Packet Core (SAE/EPC). Significant capacity improvements were observed and major cost reductions were possible due to transition of cellular networks away from circuit switched functionality [OMM16]. LTE was designed not to be backward compatible with UMTS and LTE anticipating higher spectrum block allocations than UMTS. LTE was devised to support frequency carriers from 1.4 MHz to 20 MHz in width and operate with component frequency carriers that are very flexible in arrangement [OMM16]. The LTE Release 8 (2007) system had peak data rates of 326 Mbps, reduced latency (20 ms) and improved spectral efficiency. However, LTE Release 8 is often considered as a precursor to 4G technology, since it did not comply with IMT-Advanced requirements (a successor to IMT-2000 specifications), which was used to define 4G. Nevertheless, 3GPP LTE Release 10 and IEEE 802.16 m (deployed as WiMAX) fulfilled IMT-Advanced requirements and technically became the first 4G systems. WiMAX encountered problems in gaining popularity and was superseded by LTE, which incorporated several new features such as carrier aggregation and higher order MIMO in its release 10. The use of 8X8 and 4X4 MIMO in downlink and uplink respectively, along with carrier aggregation upto 100 MHz of total bandwidth facilitated peak data rates of 3 Gbps (downlink) and 1.5 Gbps (uplink) [OMM16]. The LTE Release 11 added new frequency bands, implemented coordinated multipoint transmission and reception (CoMP), relaying and interference cancellation. The LTE Release 12 (2015) enhanced support for heterogeneous networks, added higher order MIMO and enabled aggregation between Time Division Duplexing (TDD) and Frequency Division Duplexing (FDD). In addition, new solutions (LTE-M and Narrow-Band IoT) tailored to support massive Machine Type Communication (MTC) were introduced in LTE Releases 12 and 13. Extreme broadband data rates were also targeted by introducing carrier aggregation

up to 32 carriers in Release 13. Standardization of LTE by 3GPP is continuing and is anticipated to proceed to Release 13 and beyond [OMM16]. The evolution of mobile communication generations is depicted in figure 1.1, with respective services they offer.

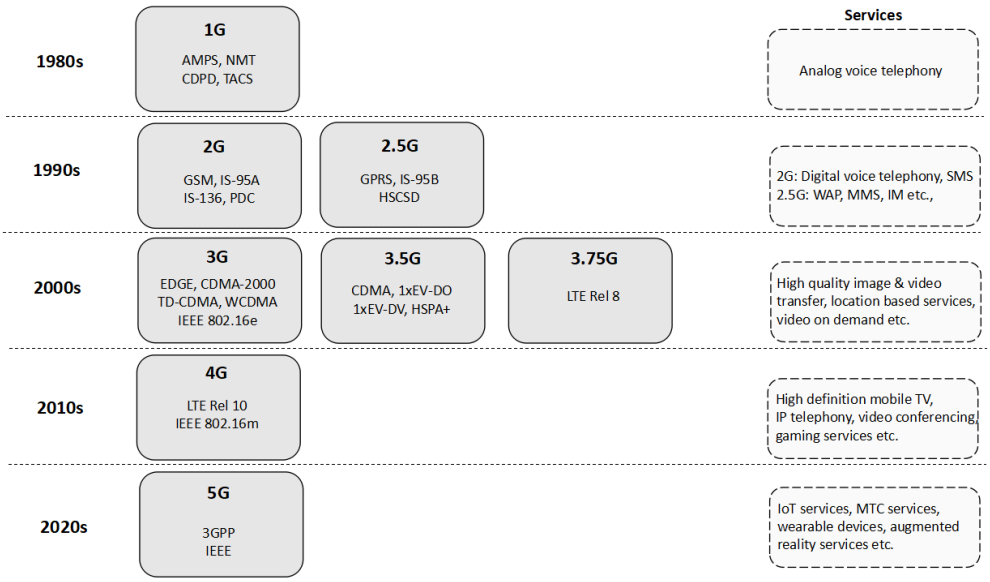

Figure 1.1: Evolution of Mobile Communication Generations

From analog systems in 1G to digital smart-phones of today, the evolution of mobile communication has come a long way and is still ongoing. The fifth generation of mobile communication (5G) is expected to be rolled out into the market by the year 2020. 5G is anticipated not only to enhance data rates in comparison to its predecessors but also support features such as ubiquitous things communicating, service in high mobility, service in crowd, fast connectivity and real-time and reliable connections [MP13a]. Further, the 5G is foreseen as a driver for industrial and societal changes, with 5G systems being designed to empower vertical industries, e.g., factories of the future, automotive, energy, e-health, media and entertainment [GP15b]. 5G services can be broadly classified as [OMM16]:

- Extreme Mobile Broadband (xMBB) - extreme high data-rate and low-latency communications are provided, with a better coverage. A uniform experience over the coverage area is provided with graceful performance degradation as the number of users increases.
- Massive Machine-Type Communication (mMTC) - wireless connectivity for tens of billions of network-enabled devices is provided with scalable connectivity for increasing number of devices and an efficient trans-

mission of small payloads. Wide area coverage and deep penetration are prioritized over data rates.

- Ultra-reliable Machine-Type Communication (uMTC) - ultra-reliable low-latency communication links are offered for network services with extreme requirements on availability, latency and reliability, e.g. V2X communication and industrial manufacturing applications. Reliability and low latency are prioritized over data rates.

Thus, 5G promises to bring novel network and service capabilities. The user experience continuity in non-trivial situations such as high mobility (e.g. in trains), sparse or densely populated areas etc., is ensured. In addition, mission critical services requiring high reliability, low latency and global coverage will be supported by the 5G platform. Further, 5G is going to be a key enabler for the Internet of Things by offering infrastructure to connect massive number of sensors, rendering devices and actuators with strict constraints of energy and transmission.

However, 5G has its fair share of challenges. The number of mobile subscriptions has been ever increasing with total number of mobile subscribers worldwide being 7.6 billion by year 2017[CJC17]. With increasing demand for mobile broadband services and roll out of machine type communications, the key challenges in 5G systems are to support, 1000 times higher data volume per area, 10-100 times more number of connected devices, 10-100 times higher data rate, 10 times longer battery life (mMTC devices), 5 times reduced End-to-End (E2E) latency. Several research and pre-standardization activities are going on in order to meet these challenges and facilitate standardization of 5G systems [GP15a] [GP17].

Figure 1.2 summarizes the key 5G scenarios, which are foreseen to become a reality in near future [OMM16] [MP13a]. Further, the vital characteristics a 5G system targets to possess are also depicted in figure 1.2. A brief summary of these features is as below:

- Fast connectivity - Enables users to experience instantaneous connectivity with negligible delay. The applications being used (work or infotainment) will have "flash" like behavior, with response being perceived as instantaneous after a single click.
  Use cases: Virtual reality office, dense urban information society etc.
- Real-time and reliable connections - Unlike the legacy systems with human centric communications, 5G envisions certain machine type com-

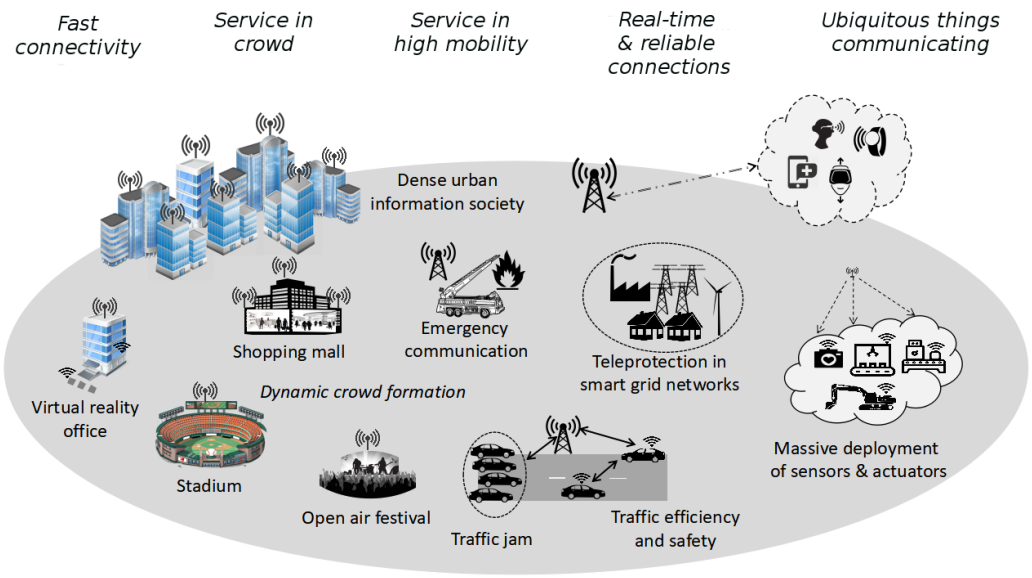

Figure 1.2: Foreseen scenarios in 5G

munications with real-time constraints. This characteristic of 5G ensures high reliability and low latency for mission critical communications.
Use cases: Traffic efficiency and safety, emergency communication, smart grids etc.

- Ubiquitous things communicating - In addition to the human centric communication, 5G enables connection of various machine type devices to the network. The range of these devices can vary from simple sensors and actuators to advanced service critical equipments.
  Use cases: Massive deployment of sensors and actuators, industrial and medical devices etc.
- Service in crowd - Contemporary mobile systems are designed to provide acceptable broadband experience to users when being sparsely distributed in public areas such as parks, bus-stops etc. However, amidst a crowd of other users, a user's experience of mobile and wireless internet access is deteriorated. Thus, this attribute of 5G is devised to ensure a better Quality of Experience (QoE) despite crowded situations.
  Use cases: Stadium, shopping mall, traffic jam, open-air festival etc.
- Service in high mobility - In today's world several users demand to avail mobile broadband services while traveling at high speeds. However, due to high mobility of such users, the mobile systems of today fail to provide a good quality of experience. This feature of 5G ensures that

a better QoE is provided to the user irrespective of his mobility e.g. pedestrian walk in city, commuter traveling in public transport etc. Use cases: High speed trains, urban vehicles, traffic jams etc.

This thesis addresses some of the major challenges in contemporary mobile communication systems and facilitates development of aforementioned vital characteristics of a 5G system, especially with respect to the features "Service in high mobility" and "Service in crowd", with following work:

- Exploiting Diurnal user mobility for predicting cell transitions and initiating suitable radio resource management schemes [KKSS13b] [KMS17a]
- Next cell and route prediction of Diurnal users for enabling Context-aware resource allocation in urban scenario [KZS16] [KS16a]
- Prediction of dynamic crowd formation in cellular networks to proactively trigger radio resource management [KKS15]
- Monitoring Vehicular user mobility to predict traffic status and manage radio resources [KMS17b]
- Framework to support mobility context awareness in cellular networks [KS16a]
- Location aware resource allocation for Device-to-Device (D2D) communications [SKL+15] and analysis of its robustness against positioning errors [KKL+15]
- Post-resource allocation power control in cellular networks to co-exist with D2D underlay [KS16b]

The above mentioned concepts rely highly on "Context Awareness", as a key enabler to invoke the 5G features in contemporary mobile systems. Context awareness was originally inspected in area of pervasive computing, where relevant information from user and his environment, such as location, user identity etc., were acquired and exploited by the computing system to adapt itself respectively. One of the earliest definitions for context awareness was provided by [ST94] as,*"The ability of a user's applications to discover and react to changes in the environment they are situated in"*. Since then, several researchers have adapted this concept to optimize system performance with context information ranging from user's location, identity, time, environment [RPM97], season, temperature [BBC97], to the user's focus of attention and emotional state [Dey98]. In recent times, these concepts are extended to mobile and wireless systems. Several research projects such as C-CAST [Pro10],

METIS-I [MP12], METIS-II [Pro15] have worked in the direction of implementing context awareness; so that mobile systems can adapt their configurations (actions) to the changing physical and application environment.

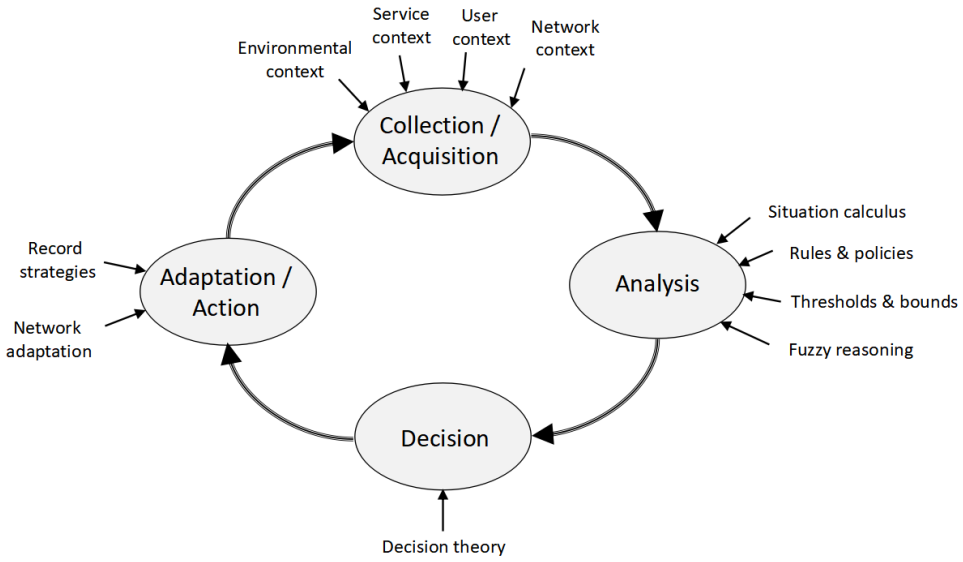

Figure 1.3: Working Principle of Context aware systems

Figure 1.3 (adapted and modified from [DDF+06]) demonstrates the working principle of a context-aware system. The acquisition of context-information is followed by their usage in an analysis, according to certain rules or algorithms. This is followed by a decision which leads to further reactive action or adaptation in the system. As a part of research project METIS-I [MP12] , the author has contributed in coming up with definitions of context information and context awareness [MP13b] as below:

Context information is the information enabling the perception of states and situations of network entities (e.g., network nodes, terminals, users, etc.) and their interactive relations.

A context-aware radio network is a network which utilizes the context information to assist its operation and optimization without explicit interaction with its users.

In detail, the following contributions are made by the author with respect to context-awareness and its incorporation in cellular networks:

- Cell transition prediction schemes are designed based on attributes such

as user position co-ordinates, movement inclination (degrees), distances to base stations and so on [KKSS13b]. Practical scenarios, where several mobile users co-travel in public transport forming data intensive moving user clusters or moving networks is considered. The aforementioned schemes are utilized to anticipate the arrival of these data intensive entities into a cell and trigger smart resource management even at high velocities. Further, an enhanced cell transition prediction scheme, based on user geometry (in dB) was also presented [KMS17a].

- Mobility prediction (next cell and route) is enhanced using additional context information about user's origin and destination, in a realistic urban scenario. The context information about future routes and possible coverage holes in them is used to devise efficient resource allocation, that sustains non real-time streaming/full buffer services, even in coverage holes [KZS16]. Further, the required signaling and interfaces to exchange context messages among various entities (in the vehicle as well as within the network) and overall architecture is outlined [KS16a].
- In day-to-day scenarios like open air festivals, shopping malls or stadiums, several mobile users gather to form a crowd, posing high load situation to respective base station. The concepts of next cell prediction, cluster detection and trajectory prediction are used in tandem to indicate the severity of crowd formation and to predict probable location of crowd formation. Subsequently, load balancing is triggered at the predicted site to free up resources. Alternatively, a small cell is activated at the predicted site of crowd formation, ensuring good service even in crowds [KKS15].
- Mobility behavior of vehicular users is analyzed and an algorithm is designed to predict traffic status of a cell. The proposed traffic prediction algorithm is a coalition strategy consisting of schemes to predict user cell transition, vehicular cluster/moving network detection, user velocity monitoring etc. The traffic status indication provided by the algorithm could be used to design efficient radio resource management (RRM) techniques e.g. pro-active initiation of load balancing at corresponding site. Further, appropriate small cells are activated/deactivated (energy efficiently) based on formation/dispersion of traffic jams respectively [KMS17b].
- Context aware procedures are investigated with a focus on user mobility. Commonalities are found among different procedures and a general

framework to support mobility context awareness is proposed. The new information and interfaces which are required from various entities (e.g., vehicular infrastructure) are discussed [KS16a].

- D2D communication is foreseen as a viable solution to provide service in crowded situations. A novel Resource Allocation (RA) concept for D2D users reusing the Physical Resource Blocks (PRBs) of the cellular UEs is proposed. The RA algorithm is based on a virtual sectoring concept that relies on network assisted positioning technologies [SKL+15]. In addition, the proposed location context aware D2D schemes are analyzed for their robustness against positioning errors [KKL+15]. Further, post-resource sharing power allocation at cellular users is proposed (three schemes namely, interference aware power control, blind power control and threshold based power control) to enable cellular links to co-exist with D2D underlay [KS16b].

In addition, the author has made contributions in developing concepts for interactive visualization and evaluation platform (IVEP)[CKS17], study of air interface harmonization as part of 5G pre-standardisation [PGM+17], [KER+17], [KMF+16] and supported other works in field of D2D [LKK+14], [LKKS14] and Mobility Robustness Optimization [KKSS13a]. Further, with his research work the author has contributed to European (FP7: METIS I, METIS II), national (BMBF: 5G-NetMobil), and other bilateral industry projects.

Outline: The remainder of this thesis is organized as follows: Chapter 2 deals with the background and related work in mobility context awareness and D2D scenarios. Chapter 3 presents various system modeling aspects and simulation framework considered in the thesis. Chapter 4 introduces various solutions incorporating context awareness in mobility scenarios such as data dense moving networks and vehicular mobility in urban environment to improve user Quality of Experience (QoE). Further, context aware schemes to improve user QoE in crowded situations (e.g. stadium) and vehicular traffic jams are presented. Finally, various signaling aspects, interfaces and architecture required to support aforementioned context aware procedures are highlighted. Chapter 5 illustrates use of context awareness to enhance D2D communication, in particular exploiting location awareness for D2D resource allocation and power control. Chapter 6 highlights the other contributions of the author which are supplementary to this thesis. Contributions made towards air interface harmonisation of 5G systems and development of an

interactive visualization and evaluation platform are outlined in this chapter. Finally, Chapter 7 is the conclusion of this thesis, which summarizes the major findings and provides an outlook on potential future applications and research directions.

Chapter 2

Background and Related Work

This chapter gives an insight into various works that exist in literature, related to the topics covered in this thesis. An overview of these works is presented outlining their influence on concepts developed in this thesis as well as the key differences. The chapter is divided into two sections namely, *Mobility Context Awareness* and *Context Aware D2D Communications*, former dealing with background and related work in context aware mobility scenarios and the latter discussing existing literature in D2D communications.

2.1 Mobility Context Awareness

2.1.1 Overview

The key differentiating feature between a cellular network and other wireless networks, is the capability of the former to provide continued service to its user, across its large service area. The large service area of the network is divided into smaller regions called 'cells' and the user mobility is supported across these cells by the means of 'Handover'. Figure 2.1 demonstrates the concept of handover (HO) in LTE systems. In LTE network, intra-frequency mobility is supported by HO process, which is triggered as soon as the condition for so-called A3 event [Net12] measurement reporting is satisfied. The following simplified HO condition needs to be met for a specific duration of successive time instances, referred to as time-to-trigger (TTT), until reporting is triggered:

$$RSRP_t + CIO_t > HYS + CIO_s + RSRP_s \tag{2.1}$$

$$RSRP_t - RSRP_s > HYS + (CIO_s - CIO_t)$$

$$RSRP_t - RSRP_s > HOM(s,t)$$

Where, $RSRP_t$ and $RSRP_s$ denote the Reference Signal Received Power (RSRP) of target eNB (TeNB) and serving eNB (SeNB) respectively. CIO_t and CIO_s represent Cell Individual Offset (CIO) of TeNB and SeNB, respectively. HYS denotes the *Hysteresis* value and HOM represents HO margin. When RSRP difference between TeNB and SeNB is greater than HOM for a period of TTT, then HO is initiated. After HO execution time has elapsed, the user is served by TeNB, marking the completion of HO process.

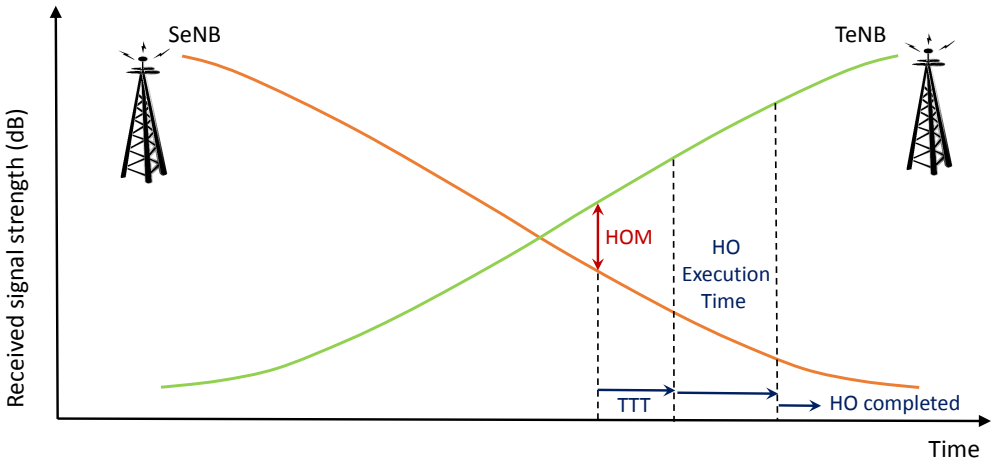

Figure 2.1: Concept of Handover

Over the last decade, number of users availing cellular services has increased drastically and the number of mobile subscriptions has reached 7.6 billion worldwide in year 2017 [CJC17]. This avalanche of users is a challenge for the network to support, as it demands more spectrum and other technological improvements to keep up with the ever increasing data traffic demand. As discussed in chapter 1, by the year 2020, the data traffic volume will be increased by 1000 fold and the number of connected devices will be 10-100 times more [OMM16], posing a major challenge to 5G system development.

Given the mobility of users in cellular networks, there is a fair probability of high user density at certain regions of the network, at certain times. Subsequently, the traffic load posed at some base stations is higher than the

others. When the bandwidth resources available at respective base station are not sufficient to sustain the needs of its users, then blocking and/or dropping of users occur. Due to the scarcity of radio resources, an access attempt made by a new user would be blocked. Further, when a user previously connected to the base station is disconnected due to the shortage of resources, then such an event is called dropping of a user. Dropping and blocking hamper the QoE of users, with former being more obnoxious than latter. A cell suffering from dropping and blocking of its users due to scarcity of resources is said to be in *Hotspot* situation.

2.1.2 Hotspot and high load situations

Understanding the effects of hotspot and studying the modeling aspects of it have been treated with high significance, since they assist in improved network design. Early works considered modeling hotspots by simply increasing the traffic in hotspot region [WCW98] [YW93]. Das et al. [DSJ98] define hotspot as a region consisting of several hot cells, where a hot cell is formed when tele-traffic demand exceeds a predefined threshold. In [KE89], hotspots are discussed in the form of asymmetric load. Here, a specific cell is modeled to have twice the load as neighboring cells. In [WCW98], hotpots are modeled in a CDMA network based on mobile user movement, such as overload caused by rush hour traffic. Further, hotspot situation is simulated by increasing traffic in hotspots by 1.1 to 2.7 times the full traffic. [YW93] defines hotspot as a cell with traffic load significantly larger than the design load. The hotspot situation is treated as permanent or temporary and is modeled by assuming load in hotspot cell as 66 Erlangs whereas the nominal peak load being 47 Erlangs. Thus, majority works consider high user arrival rate, low departure rate, or increased bandwidth demand of existing users leading to hotspot in a cell. However, [JFTK04] categorizes hotspots into three types based on their origin namely:

- Delay based: This type of hotspot occurs in a cell when average time spent by users in the specific cell is significantly higher than the time spent in other cells. For instance, due to an accident on a street, the traffic is held up and all the users in that cell are delayed.

- Capacity based: Such hotspot occurs when a base station is facing technical problems and can support a lower number of users. Such hotspot can be modeled by decreasing the capacity in a cell that is

deemed to be a hotspot. Here capacity is analogous to the number of users the cell can support.

- Preferential mobility based: This type of hotspot arises when several users move towards a given location thereby increasing the number of users in that region.

[JFTK04] also provides implementation details of the aforementioned hotspot types and demonstrates their effects on network performance. Further, an analytical model is developed to study hotspots by considering a cell as a M/M/B/B queue.

Having studied the effects of hotspot on network performance, the focus would now be diverted towards detection or prediction of such hotspots in a network. The earlier work to detect presence of hotspots in DS-CDMA systems was introduced in [TMC04]. The scheme consisted of *utilization trend detection* procedure and *aggregate utilization* procedure. The utilization trend detection procedure predicts traffic load of desired sector from its past utilization information. A set of past utilizations of specific sector is used to derive utilization trend and is represented as Exponential Moving Average (EMA). It shows whether the utilization was increasing or decreasing recently and smooths out the fluctuations. The utilization trend T_{EMA} is given by,

$$T_{EMA} = (1 - K_1)\ T_{EMAp} + K_1 t \tag{2.2}$$

Where T_{EMA} is the utilization trend, T_{EMAp} denotes previous value of smoothed trend, K_1 is a smoothing constant and t is the current trend value.

Similarly, the aggregate utilization is calculated using EMA, given as,

$$U_{EMA} = (1 - K_2)\ U_{EMAp} + K_2 U_c \tag{2.3}$$

Where U_{EMA} is aggregate utilization, U_{EMAp} is the previous aggregate utilization, K_2 is a smoothing constant and U_c is the current utilization.

Both utilization trend and average utilization are monitored simultaneously to determine the state of a cell. If the utilization trend shows that in a recent time period, there is an increase in average utilization for most of the time and the aggregate utilization surpasses the utilization threshold, then the desired sector is anticipated to be a hotspot in near future. If only one of the indicators is greater than the corresponding threshold, then no outcome is confirmed. However, if both indicators become greater than their corre-

sponding thresholds then the sector is said to be a hotspot. Further, there are several works in the literature for the detection of hotspots based on thresholds on traffic or cell load [YMT+02] [DBC03] [KV03] [KLH99] [XjHF+10] [AWVK10] or based on signaling costs and user experience factors [XCG10].

2.1.3 Solutions to high load situations

Once the hotspot is detected, there is a need to rectify such situation and improve the system performance and ease the congestion. Several solutions are investigated to satiate the hotspot cell and majority of these are proposed as an integral part of Self Organizing Networks (SON). A dynamic load balancing scheme based on ad-hoc relay stations is presented in [YMT+02] to efficiently balance traffic loads and to share channels between cells via primary and secondary relaying. These relay stations operate in the unlicensed ISM band and, hence, do not interfere with the cellular band. [DBC03] proposes a solution using smart antennas at the base station and dynamically controlling the cellular coverage size as per geographic traffic distribution. Another LB approach [KV03], makes use of bandwidth management algorithms for heterogeneous multimedia services in a cellular network which focus on minimizing the maximum available bandwidth among cells in an adaptive manner depending on current traffic conditions in the network. Such an approach is capable of resolving conflicting performance criteria while ensuring efficient network performance for QoS sensitive multimedia services.

Channel borrowing is one of the most common solution to a hotspot situation, however when a cold cell keeps lending out its free channels to fulfill demands of a hotspot, the lender cell eventually turns into a hotspot and starts borrowing channels from another cold cell and this process continues redundantly. A solution based on two thresholds is presented in [KLH99] to solve such a redundant channel borrowing problem by categorizing cell states as hot cell, warm cell and cold cell and strategically borrowing channels based on two thresholds. A cell-cluster based traffic load balancing procedure to solve congestion in co-operative cellular networks is presented in [XjHF+10] where a dynamic cell-construction method is studied based on traffic distribution among a hotspot cell and its neighboring cells. Following this, the traffic transferred from a hotspot cell to each of non-congested cells in a cell cluster is calculated and utilized to minimize blocking probability. Further, the spectral efficiency for both direct and co-operative relaying transmissions is studied and a mathematical model is formulated to jointly optimize routing and radio resource allocation in traffic load balancing, and a greedy based

Cell-cluster based Joint Routing and Radio resource allocation algorithm is proposed to find a suboptimal solution.

A self-organizing load balancing framework is proposed in [AWVK10] which provides self-optimizing LB policies to enhance adaptation and robustness of Fixed Relay Station (FRS) based cellular networks. A Self-organizing Co-operative Partner Cluster (SCPC) strategy is developed which dynamically selects optimal partners of each base station and relay station. A Comprehensive Load Balancing Policy Stack (CLBPS) to utilize merits of various load balancing policies is proposed as well. Both SCPC and CLBPS take into account LB performance, user experience and signaling costs. Another relay-based LB scheme is introduced in [YN10], in which relay stations are used to transfer over-loaded traffic from hotspot to neighboring cells which are cooler. An on-line algorithm is proposed which dynamically controls the association of relay stations with base stations and association of mobile users with relay stations and base stations. Additionally, Game theory is introduced as a methodology to analyze topics in mobile communications including subject of load balancing.

[AWVK10] studies load balancing problem using a game-theoretic approach, which, in the worst case, dictates each cell to decide independently on the amount of load that maximizes its payoff in an uncoordinated way. Further, it is investigated if the resulting Nash equilibrium would exhaust the gains achieved. Moreover, the behavior of the players is altered using linear pricing technique to obtain a more desirable equilibrium. In [KFP10], a LB technique based on delay tolerant message forwarding that relies on the store-carry and forward paradigm by utilizing the mobility of vehicular nodes in multi-cell wireless networks is proposed. A network flow problem is formulated with the objective of minimizing the weighted sum cost of three competitive parameters namely those of communication energy consumption, cell loading variance and message delivery delay. It was shown that by trading off message delivery delay considerable savings in energy are observed while at the same time cell load imbalance can be mitigated.

2.1.4 Mobility Load Balancing

3GPP considered Mobility Load Balancing (MLB) as an important use case of SON [3GP11], whose purpose is to cope up with the unequal traffic load in the system and to improve the system capacity. Since the implementation aspects of MLB are out of the scope of the 3GPP specification and only the generic operation is defined [SVV13], a considerable number of papers are

available that address various aspects of efficient MLB operation. [SVV13] proposes a general MLB framework that considers the transport load and congestion in addition to the radio load. The solution is not limited to any particular transport network topology and it also acts as a Traffic Steering (TS) mechanism as it resolves transport overload by redirecting users to neighbor cells with spare transport resources.

In general, during overload, serving eNB (SeNB) triggers LB to transfer some of its edge users to the neighboring cells or target eNBs (TeNBs), by adjusting a parameter called Cell Individual Offset (CIO). In [MMT13a], a LB scheme that adjusts the CIOs among SeNB and its neighboring cells by a fixed step θ is proposed. The best step θ is inferred to be depending on load conditions in both SeNB and neighboring cells, as well as user distribution in SeNB. Further, a Q-learning based algorithm is proposed that learns best θ values for various load conditions. In addition to this, various MLB schemes with adaptive performing periods are present such as Adaptive Periodic Mobility Load Balancing (APMLB) and Enhanced-APMLB [KS13], while inclusion of non-adjacent neighborhood cells in optimization area is considered in [ZMT13]. [ASN$^+$14] incorporates Multi-Objective Particle Swarm Optimization to perform MLB and dynamically adapt HO margin parameters in order to balance traffic of the network eNBs. An Enhanced MLB (ELB) scheme with the double threshold design including the common trigger threshold and the fairness-aware threshold is presented in [YSTL14]. This considers modeling the fairness metric during optimal target cell selection process, to yield fairness-aware threshold. There are few joint optimization methods present in the literature as well, such as [TK15] which considers joint load imbalance and inter-cell interference co-ordination, and [XCL$^+$15] which jointly considers MLB factor of the shifted user and that of hotspot cell along with proportional fairness scheduling factor. Specifically, the cell preferentially allocates radio resources to shifted users, which are suffering poor link quality or previously served by a hot-spot cell with large handover region.

Along with MLB, Mobility Robustness Optimization (MRO) is considered as another important use case in SON. In general, MRO sets the cell's HO Hysteresis and Time to Trigger, to select the optimum point at which a HO is initiated. However, MLB typically adjusts CIO and advances HOs from overloaded cells to less loaded ones. Thus, MLB affects HO metrics mainly because advancing HOs inadvertently increases radio link failures caused from too early HOs, Ping-Pong HOs etc. [MMT13b] introduces a Q-learning algorithm to learn the best MLB action to take in different load

states in order to attain a desired load transfer, but with minimal effect on HO performance. Similarly, [GCL+15] presents MLB to improve the efficiency of resource utilization, but in addition considers HO optimization to prevent ping-pong effect.

In this thesis, MLB implementation closely follows the investigations provided by SOCRATES project [SP10]. The LB process considers a parameter called "cell specific offset" indicated by HO_{offset} to force users to handover from SeNB to TeNB. The key focus of the MLB is to obtain the optimum value of HO_{offset} which makes provision for maximum number of users to change cell without being rejected by admission control at TeNB. The condition for HO process provided by equation 2.1, can be approximated for a user and condition for MLB handover is obtained, as given by equation 2.4.

$$RSRP_t + HO_{offset} > HYS + RSRP_s \tag{2.4}$$

The concept of MLB is demonstrated in figure 2.2

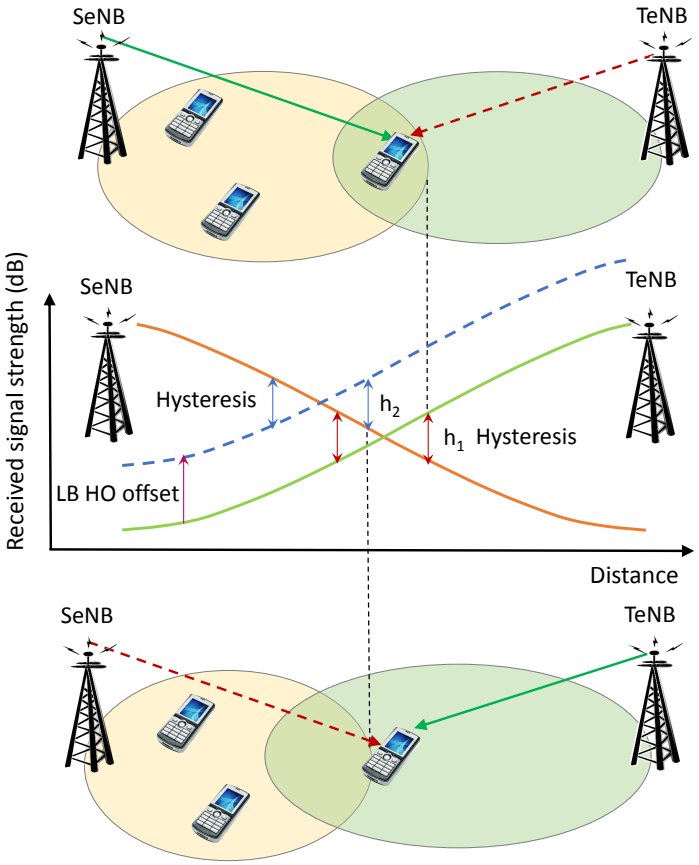

Figure 2.2: Concept of load balancing

The pseudo-code for the LB concept is presented in algorithm 2.1.1. In order to apply LB procedure, the SeNB has to create a list of potential targets for HO. For this, the SeNB needs to acquire measurement reports from the served UEs as well as collect reports about available resources from neighboring cells. For each adapted value of HO_{offset} denoted by T, the SeNB has to sort a list of potential TeNB considering the number of possible handovers by process of LB. Then, resulting load $\hat{\rho}_C$ is estimated for a given value of HO offset T and cell C from the list L. If the estimated load $\hat{\rho}_C$ is still lesser than the TeNB threshold $\rho_{Thld,C}$, then the HO offset to this cell is tuned to the value T and the virtual load at SeNB ρ_{SeNB} is subtracted by the value of a load resulted from handed over users with given T. This procedure continues as long as the load at SeNB ρ_{SeNB} is greater than threshold $\rho_{Thld,SeNB}$ and HO offset T does not surpass a maximum value T_{MAX} and neighboring cells are still able to cope up with the additional load due to LB handover.

2.1.5 Mobility Prediction

Majority of the work in literature deals with either detection of hotspot formation in a part of network and/or countermeasure for such situation. In this thesis, propagation of such hotspots from one part of a network to another is investigated, based on the realistic day-to-day scenarios giving rise to high load situations in a network (e.g., moving networks, moving user groups in public transport such as bus, train, tram etc.). The key focus is to efficiently predict the transition of such data traffic dense moving user groups to future cells and pro-actively trigger countermeasure (e.g., load balancing) at the anticipated cell site, well in advance.

Mobility prediction is one of the major investigation fields in cellular networks, which is utilized for performance optimization of various network functions such as resource management, scheduling, handover control, network selection to vehicular cloud computing in recent times. Mobility prediction can essentially be divided into two categories namely,

- History based: The user's next cell or path is predicted based on the statistics of user's mobility. The mobility history of the user is recorded and probability of user transition into next cell is derived.

- Measurement based: These schemes do not rely on the user mobility history rather they derive probability of user transition to next cell

Algorithm 2.1.1: Pseudo-code for the HO offset based LB algorithm [SP10]

Require: List L of potential Target eNB (TeNB) for LB HO obtained from OAM

Collect measurements from user; RSRP to potential TeNB,
Group users corresponding to the best TeNB for LB HO (criterion is the difference between SeNB and TeNB measured signal quality),
Get information from TeNB on available resources,
Estimate number of required PRBs after LB HO for each user within LB HO groups,
sort users U according to required HO offset regarding to TeNB,
$i \Leftarrow 1$; $T \Leftarrow 0$;
while $(\rho_{SeNB} > \rho_{Thld,SeNB}) \wedge (i \leq size(L)) \wedge (T < T_{MAX})$ *do*
    $T \Leftarrow T + step$
    $L \Leftarrow sort$(TeNB according to #users allowed to LB HO with given T, descending order)
    { find target cell C with the largest number of potential LB HO with given T,}
    while $(\rho_{SeNB} > \rho_{Thld,SeNB}) \wedge (i \leq size(L))$ *do*
    {Ability of load accommodation by cells from list L is verified}
        $C \Leftarrow L(i)$ { take next cell from sorted list}
        Estimate $\hat{\rho}_C$ and $\hat{\rho}_{SeNB}$ after HO for a given T
        {Based on SINR after the HO prediction}
        if $\hat{\rho}_C < \rho_{Thld,C}$ *then*
            $\rho_{SeNB} \Leftarrow \rho_{SeNB} - \hat{\rho}_{SeNB}$
            {update load in SeNB by subtracting handed over load}
            $T_{C,u} \Leftarrow T$
        end if
        $i \Leftarrow i + 1$
    end while
end while
Adjust to users U calculated HO offsets $T_{C,u}$.

based on radio network measurements such as signal strength, user velocity, user distance and angle etc.

Majority of the movement prediction schemes in literature are history based and a few rely on network measurements as well. [Laa05] considers a GSM system and applies concept of route clustering and prediction. The algorithm is assumed to be running on mobile user's phone. A cell cluster is assumed to be a group of nearby cells where most transitions happen within the cluster. Locations that are important to the user are treated as bases. Entire routes between two bases are considered and an attempt is made to learn all different physical routes as strings of cell identifiers. When the user completes a route r between bases a and b, it is determined if an existing route between a and b is similar to r. In such a case, the two routes are clustered together. Following such a clustering applied to the routes, a prediction is done using the previous base a and a history h of m most recently encountered cells. Predictions are based on matching the current cell history against known routes. The current time provides additional context to aid prediction.

In [LH05], neural networks are used to predict user location. The mobility model is designed considering inter cellular movement as a logical function of position, speed, acceleration and direction. Series of user's positions are used as they implicitly contain aforementioned attributes. Three nearest past positions $(x_1, x_2, x_3, y_1, y_2, y_3)$ are decomposed into two small groups (x_1, x_2, x_3) and (y_1, y_2, y_3). Thus, neural location predictor is designed in two group-complete-bipartite meshed feed-forward neural network. Three nearest past positions $(x_1, x_2, x_3, y_1, y_2, y_3)$ are fed into networks repetitively and location predictions $(x_4, x_5, x_6, y_4, y_5, y_6)$ and cell predictions (c_4, c_5, c_6) are obtained as the outputs.

A mobility prediction scheme based on two types of mobility profiles is used in [YV09], namely local and global mobility profiles. These profiles contain frequent paths, associated probabilities, information related to user's behavior, visited paths, time of the day and estimated time of each path. The local mobility profile is generated individually by each user's mobile phone and the global mobility profile is generated by the BS. BS has the responsibility of collecting mobility data of users passing through the cell. When a user is in a new BS, the mobile predicts next set of BSs based on current local profile of the user. If the mobile could not predict the next cell, then the current BS will use its global profile for prediction. The prediction information is used for adaptive bandwidth allocation and bandwidth borrowing.

Prediction of next network cell IDs for moving users with discriminative and generative models was proposed in [MPM12]. Keeping user privacy in mind, this work also considers generating models independently of the user. Feature generation is discussed using mobility traces and certain limitations are addressed as well. For prediction, the applied pattern recognition algorithm generating a discriminative model is based on Support Vector Machine (SVM), providing powerful solution to non-linear classification. The key principle is to separate the data points provided in training with a hyper plane and the prediction process can be simplified to detection on which side of the plane data could be found. During the training phase, SVM finds an optimal separating hyper plane maximizing the distance between plane and training samples. SVM considers training independently of a specific user as well as training with GPS positions, residence time and user IDs. Further, the incorporation of time with generative models is considered along with aforementioned concepts.

An adaptive method to determine registration and paging areas while reducing frequencies of location registrations and sizes of paging areas is designed in [IKY13]. The proposed method uses the interacting multiple model (IMM) algorithm to support multiple motion models of varied devices. IMM algorithm determines registration and paging areas using the estimated location of a mobile at the time of latest registration and a predicted location after a certain time frame, which is calculated from the most likely motion model in IMM filter. Since, IMM filter chooses the most likely one out of multiple motion models, it can track changes in the mobility patterns of a mobile user. Further, information on a BS on location registrations as well as information on a BS through which user communicates are utilized to enhance accuracy of prediction.

Most recently, mobility prediction has also been used for an efficient resource management in the vehicular cloud computing. In [MAA17], Artificial Neural Networks (ANN) is recognized as a key tool in prediction of vehicle's arrival. A hybrid model of Linear Regression (LR) aided by ANN is used in predicting the lifetime of resources (time through which resource is available before leaving the covered area), which intrinsically relies on mobility prediction. The training dataset fed to LR aided by ANN model contains features such as resource lifetime, position and speed of vehicle, number of vehicles ahead in same cloudlet area and in next cloudlet area, average speed of vehicles ahead in same cloudlet area as well as in next cloudlet area. Following this, ANN mobility prediction model enables vehicular cloud to apply pre-

planned procedures and to cope up with abrupt changes in vehicles locations while maintaining robust vehicular cloud performance.

Further, the Markov chain and several variations of it are widely investigated for its usage in mobility prediction. In a Markov chain, sequence of possible events are described, where probability of each event depends on the state attained in a previous event. This concept can be adapted for mobility prediction, where the Markov chain model regards user mobility as sequence of locations (or cell coverage area) and take the user's mobility patterns along with the probability of transition between locations (or cells) and the order of transitions into consideration. Typically, the Markov based prediction has two phases namely mobility history based learning and prediction. Figure 2.3 depicts mapping of a user's mobility in a two-tier cellular layout into a Markov chain. Here each state (node) represents the cell, arrows represent the state transition and $P_{i,j}$ represent the respective transition probability from $cell_i$ to $cell_j$.

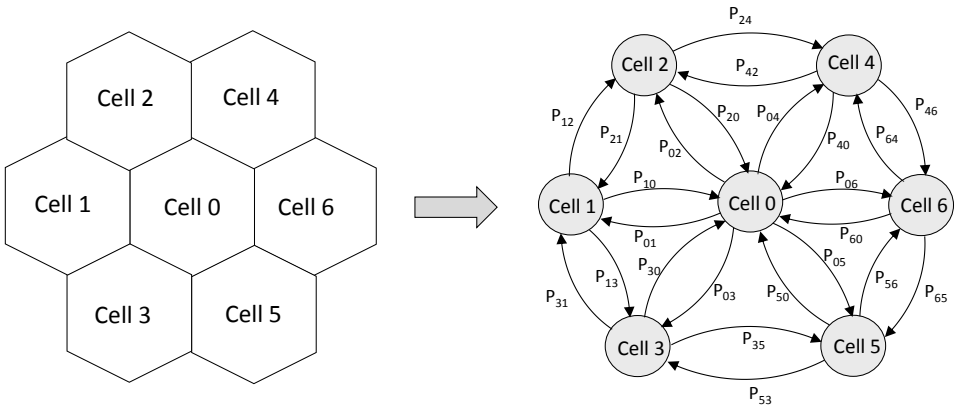

Figure 2.3: Markov Model for a 2-tier Cellular Layout

In [BK10], a Markov based mobility prediction is proposed with a Global Prediction Algorithm (GPA) that observes the movements of users using the mobility trace and a second order Markov chain. Further, a Local Prediction Algorithm (LPA) is considered that is implemented at the BS for tracking users more closely within the cell and predicting the next cell based on present cell sector. In case of GPA, history trace L consisting the identity of all cells crossed by the user during a time duration of T is stored. The movements of each user are modeled using a continuous-time Markov process whose

discrete states are the cells of the network. A second order Markov chain is considered, making probability of user transition into a next cell, dependent not only on its current cell but also on the previously visited cell. The estimated transition probability is given by,

$$P_e = P\left(X_{n+1} = Y / X_n = X\right) = \frac{M\left(X, C_{n+1}\right)}{N\left(C_{n-1}, C_n\right)} \tag{2.5}$$

Where, $X = C_{n-1}C_n$ is the sequence in L of previously visited cell and current cell of mobile user, and $Y = C_nC_{n+1}$. $M(C_kC_i, C_j)$ denotes the number of transitions of a user from cell C_i to cell C_j in the past, knowing that each time such a transition occurred the mobile user was previously in cell C_k.

Using the first order Markov model, the transition probability into a next cell is given by,

$$P_e = P\left(X_{n+1} = Y / X_n = X\right) = \frac{N\left(C_n, C_{n+1}\right)}{Z\left(C_n\right)} \tag{2.6}$$

Where $Z(C_n)$ is the number of times cell C_n appears in mobility history of the user. Furthermore, additional information called visit frequency is defined considering total number of adjacent cells (K) to cell C_n , given by,

$$H\left(C_n\right) = \sum_{j=1}^{K} Z\left(C_j\right) \tag{2.7}$$

Using this information, estimated transition probability is obtained as,

$$P_e = \frac{Z\left(C_n\right)}{H\left(C_n\right)} \tag{2.8}$$

[SWYS10] considers modeling mobility prediction using a Hidden Markov Model (HMM). HMM is a classical part of theory of Bayesian network which has two types of stochastic variables, a state variable which is hidden and an output variable which is observable. Figure 2.4 depicts the HMM concept which can be adapted for mobility prediction.

In the figure 2.4, $q_{1:T}$ denotes the factual situation of an observed object with underlining state-sequence forming a Markov chain. However, sometimes the measured value may not indicate practical situation because of impossibility, difficulty or distortion of observation. Thus, output variables $y_{1:T}$ are differentiated from state variables and their relationship is described by the one-to-one mapping between them. The two conditional independence

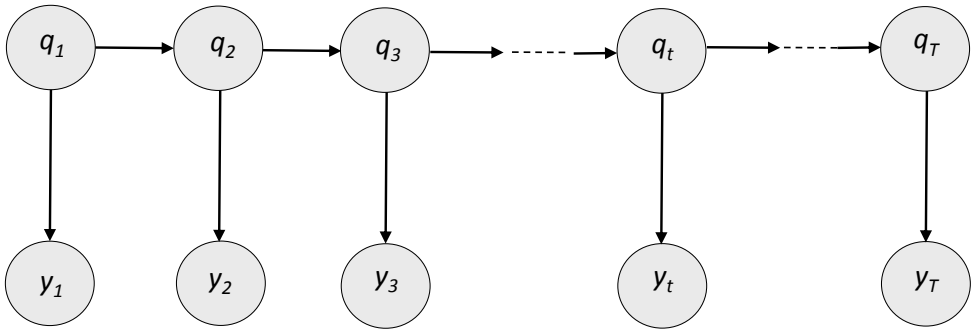

Figure 2.4: Hidden Markov Model Structure

statements that follow are as below,

$$q_t \perp \{q_{1:t-2}, y_{1:t-1}\} | q_{t-1} \tag{2.9}$$

$$y_t \perp \left\{q_{1:T\setminus t}, y_{1:T\setminus t}\right\} | q_t \tag{2.10}$$

Further, the joint probability distribution of all variables can be simplified as below,

$$p\left(q_{1:T}, y_{1:T}\right) = \prod_{t=1}^{T} p\left(q_t|q_{t-1}\right) p\left(y_t|q_t\right) \tag{2.11}$$

In addition, $\lambda \triangleq (\pi, A, B)$ is introduced where, π is a N X 1 vector and $\pi_{q_1} \triangleq p\left(q_1\right)$, A and B are N X N matrices with N being number of states, defined as,

$$A_{q_t, q_{t+1}} \triangleq p\left(q_{t+1}|q_t\right) \tag{2.12}$$

$$B_{q_t, y_t} \triangleq p\left(y_t|q_t\right) \tag{2.13}$$

The structure of a cellular network can be modeled as a graph with nodes representing cells and edges representing the neighboring relationship of cells. This graph is analogous to the state-transition graph of every stochastic variable, and each one-step transition must be from one node to another along the edge. Mobility prediction based on HMM consists mainly of two

parts namely, *Parameter learning* and *Prediction*. The objective of learning process is to determine optimal parameters fitting history data and it is given for considered HMM as,

$$\underset{\lambda}{max}\ p\left(y_{1:T}|\lambda\right) \tag{2.14}$$

Prediction is the inference process in HMM and three steps of inference are carried out here,

$$\underset{q_T}{max}\ p\left(q_T|y_{1:T}\right) = \underset{q_T}{max}\ \gamma\left(q_T\right) \tag{2.15}$$

$$\underset{q_{T+1}}{max}\ p\left(q_{T+1}|q_T\right) = \underset{q_{T+1}}{max}\ A_{q_T,q_{T+1}} \tag{2.16}$$

$$\underset{y_{T+1}}{max}\ p\left(y_{T+1}|q_{T+1}\right) = \underset{y_{T+1}}{max}\ B_{q_{T+1},y_{T+1}} \tag{2.17}$$

Where, $\gamma\left(q_t\right) \triangleq\ p\left(q_t|y_{1:T}\right)$ is deduced during decoding/recognition part of HMM. Eloborate details about these steps could be found in [SWYS10].

[FM12] also uses Hidden Markov chains and the related theory to design prediction of next cells in a cellular network and uses it to perform active/passive bandwidth reservation for mobile users. The system modeling considers sectorization of cell and subsequent mapping to a HMM is done. This is followed by procedures of learning, prediction and utilization. Each cell uses a specific Markov chain (with a limited number of states) to describe and predict local user movements. Training of Markov chains considers local trajectories and each predictor is tailored for the particular mobile environment.

In [XZZG12], a concept of collective behavioral patterns (CBP) is introduced based on the association patterns among locations of mobile users. CBP and association rules of CBP are investigated for their usage in inferring user's locations from others. The CBP-based scheme is integrated with a Markov-based predictor giving rise to a hybrid predictor that considers both individual's mobility patterns and collective behavioral patterns associated with crowds. A Markov chain of order 6 is trained for learning mobility of each user. A procedure called *prediction result fusion* is carried out in order to combine results from both the predictors.

A mobility prediction scheme incorporating the mobile positioning information as well as the knowledge of road topology is introduced in [SK04]. It is assumed that each BS keeps a database of the roads within its coverage area. The road between two neighboring junctions is considered a road

segment and identified using a junction pair (j_1, j_2). A junction is basically an intersection of roads, whose approximate co-ordinates are stored in database. The database also keeps information such as, identity of neighboring segments at each junction; statistical data of time taken to transit each segment and data about possible HOs at each segment; probability of user in a road segment transitioning to each neighboring segment, based on previously observed paths of users. The transition between these road segments is modeled as a second order Markov process. Thus, the transition of a user to a neighboring segment is assumed to be dependent only on the current segment and a segment traversed prior to it.

User mobility analysis in a LTE network based on a hidden Markov model is carried out in [LMQ+14] and its underlying principles are similar to those presented in [SWYS10]. HMM allows considering user mobility pattern with the dimension of serving eNB. Parameters of the predictor are defined and the effect of each parameter on prediction performance is evaluated using so called control variate method. Furthermore, real world mobility traces are collected and used for evaluation of HMM based prediction.

In recent years, a logical separation between control plane (CP) and data plane (DP) has been proposed, with an intention of providing high data rate services under the umbrella of a coverage layer in a dual connection mode and allowing significant savings in terms of signaling overhead. A predictive DP HO management that could minimize the out-of-band signaling related to the HO procedure is proposed in [MOH+15], which uses a mobility prediction scheme based on Markov Chains. The key principle followed by the learning process is to favor the most common routes followed by a user by giving them higher probabilities than other routes. The user's trajectory is predicted in terms of a HO sequence and is further used to minimize the interruption time and associated signaling when HO is triggered.

In the context of heterogeneous networks,[MLLZ12] proposes using mobility attributes of a user and network load to design a network selection scheme. For every user in an overlap area of two networks (e.g., cellular network and WLAN), parameters of user mobility namely, speed $v(m/s)$, distance of access point (AP) to the user $d(m)$ and angle $\theta(degrees)$ between user direction and a straight line from AP to user are extracted. Using the radius of overlap coverage (r) and aforementioned mobility attributes, the distance between a user and the boundary of overlap coverage is obtained

from relation,

$$cos\,(\theta) = \frac{s^2 + d^2 - r^2}{2sd} \tag{2.18}$$

Subsequently, the dwell time of a user in overlap area $t_d = s/v$ is obtained. The probability of a user moving out of overlap area is given as,

$$P_{out} = P\,(t_h > t_d) = \int_{t_d}^{\infty} g(t)dt = e^{\mu t_d} \tag{2.19}$$

Where, t_d is the call holding time which is exponentially distributed with mean $1/\mu$ seconds. The probability density function (PDF) of call holding time is given as,

$$g(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{2.20}$$

A performance parameter called Mobility threshold (M), corresponding to the optimal performance of heterogeneous network is defined. Considering blocking probability of the whole network as a network performance optimization target, an optimization problem is designed and M is obtained as its solution. The distribution of all the users' mobility and their arrival rates are taken into account for such an optimization problem. The network selection is done by comparing P_{out} with M; if $P_{out} < M$ the call will access AP of the overlap area (WLAN), else access is provided to the cellular network.

[PW12] presents a framework to design a generic connectivity map, which can be used to provide applications with predictions of the expected future connectivity. The data of all vehicles are transmitted and stored in a central database. Collected data mainly includes, position (GPS), current speed and angle, network and cell identities, RSSI (dBm), latency and bandwidth measurements. Based on the current and historical network data of vehicles, predictions are performed, taking into account influences of position, user mobility, available resource and time. Prediction relies on segmentation approach, where collected data is analyzed and assignment of the data to the digital map in data base is done. Map matching is carried out where the given route is analyzed and the position is matched to the probably traveled roads. The complexity of the procedure is reduced by subdividing the roads into segments. Predictions are made about the next cell, future signal prop-

erties (RSSI), available bandwidth and expected latency.

Knowing the correlation that exists between geographic location and received signal strength,[AzHV13] exploits user mobility patterns to predict their future data rates based on radio maps, where average values of historic signal strengths are stored on a map. The work emphasizes on the study that daily routes of users exhibit a high degree of temporal and spatial regularity, and user is inclined to traverse specific routes to and from frequently visited locations. Coupling mobility patterns with radio maps will allow prediction of average data rates of a user along its routes and this information is used to develop a long term predictive resource allocation scheme.

Quite similar to the concept in [AzHV13],[MSA+14] introduces coverage map prediction mechanism. [MSA+14] acknowledges proportional fair (PF) scheduling as the de-facto standard in cellular networks but highlights that it is not optimized for mobility. A Predictive Finite-horizon PF Scheduling is designed using data rate prediction, estimation of future channel allocations and slow-fading aware scheduling. This is enabled by coverage map prediction, where measurement traces are processed to construct channel quality maps. Localization is done based on user's GPS and at the sector level, user's location and velocity are determined, subsequently the data rate is predicted. To reduce computation due to querying of user's GPS, Channel History Localization Scheme is proposed which requires pre-knowledge of the user's trajectory.

Similar to [MLLZ12], [LH16] considers heterogeneous network set up and proposes RAN-cache HO, based on mobility prediction. The radio access network (RAN) cache is recently proposed to ease the pressure on the backhaul network and enhance quality of experience of users accessing content on the Internet. Unlike other works where users' mobility is assumed to be pre-known, here a parallel autoregressive (AR) based received signal strength (RSS) prediction approach is used to achieve mobility prediction. Measured RSS by the handover terminal is usually affected by the path loss, shadow fading and multipath fading. Filtering of measured RSS is carried out to avoid the effects of multipath fading and the shadow fading to some extent. Following this, the handover terminal measures the RSS of its HO candidate access nodes and predicts their future RSS simultaneously. This facilitates identification of the target HO access node.

2.1.6 Crowd and Traffic Jam Monitoring

In this thesis, not only the propagation of high load situation from one cell to another (e.g., moving networks, public transport etc.)is investigated but also dynamic crowd formation situations (e.g., public events, stadium, traffic jams etc.) in a network are examined. Hence, few prevailing techniques that exist in the literature for crowd surveillance are outlined here.

Crowd monitoring based on image processing techniques has been proposed as early as in year 1995 in [DY95]. A comprehensive study is done to understand the behavior of individuals in a crowd and formation of a crowd under various situations. The set up consists of a standard domestic camera to record scenes in a site of interest. The recordings are digitized and processed by a transputer-based image processor under the control of a standard workstation. Several well known image processing algorithms are used for edge detection, background removal and other temporal analysis of images. Relationship between number of people in an image and the number of picture elements is established after background removal. Similar to a human observer who can easily distinguish a dense crowd from the background (surrounding buildings, road surfaces etc.) and would be likely to use the ratio of 'crowd area' to 'background area' as a rough estimate for the crowd density, a computer-based density estimator is developed, where image pixels corresponding to the crowd are separated from those of the background. Further, estimation of crowd motion is also carried out.

[SSS09]proposes a method to model and learn the scene activity, observed by a static camera. A framework is presented to learn traffic patterns in the form of a multivariate nonparametric probability density function (pdf) of spatio-temporal variables that include the locations of objects detected in the scene and their transition times from one location to another. Learning is achieved by monitoring the trajectories of objects by a static camera over extended periods of time. It encodes the probabilistic nature of the behavior of moving objects in the scene and is utilized in activity analysis applications (e.g.,persistent tracking and anomalous motion detection). In addition, the model captures salient features of a scene, such as the usually adapted paths, frequently visited areas, occlusion areas, entry/exit points, etc. This framework is beneficial in general surveillance as well as crowd detection.

Further, few methods to study traffic jams which are caused by vehicular users are outlined. In [MLS08], an overview about the traffic congestion growth over last few decades is presented. The problems that arise due to a traffic congestion are discussed and the requirement of a real-time traffic

congestion prediction in the context of intelligent transport systems are highlighted. This position paper anticipates that network-enabled vehicles in the future will be able to both probe and predict near-term traffic congestion accurately and distribute their travel routes more evenly. It also asserts that current infrastructure-based traffic sensors do not scale well and do not capture vehicle trip-specific speed information. A programmable car is stated as vehicles of future, that can be networked for network-based active safety and real-time traffic congestion probing and prediction.

Similar to the video based analysis presented in [DY95],[SSS09], work presented in [PW14] assumes a fixed camera attached on high location, monitoring traffic flow in a specified road segment. The analysis is carried out on the frame of video stream to estimate the traffic density and traffic speed from the moving vehicles. Based on the combination of traffic density and traffic speed, traffic congestion level is classified into free flow, slow moving or congestion.

[HYW14]proposes a traffic jam prediction method based on two-dimension cellular automata model. The method is considered effective in describing various characteristics of the urban traffic networks, so as to predict the accurate positions of the traffic jams at the intersections. The considered model is initialized with the current traffic density of each road and the number of the vehicles traveling eastbound and northbound are not equal, emulating real situation of road network.

Further,[LSSK14] introduces mathematical models and algorithms to solve the problem of traffic jam prediction using statistics about the traffic parameters incoming from vehicle monitoring systems. The historical (statistical) time-stamped data on development of traffic situation and its dynamics are retrieved and the problem of road traffic forecasting is reduced to a search of set of records corresponding to similar conditions in a database. This problem is solved on the basis of fuzzy search methods and algorithms.

From these works on crowd monitoring and traffic congestion prediction, it could be noted that, they mainly depend on the visual surveillance, requiring a separate infrastructure (e.g., video camera, monitors, data base etc.) for the video/image analysis. These are costly and computationally intensive and are not necessarily developed from the perspective of a cellular network.

2.1.7 Contributions

Compared with existing work related to mobility context awareness, this thesis emphasizes more on developing concepts that require simpler data and

are easy to develop as well as integrate. The mobility prediction concepts are developed based on current measurements from the UE and network (e.g., user angle, distance, geometry (dB) etc.). Further, realistic day-to-day scenarios such as moving networks, large number of users commuting together in public transport (e.g., bus, train etc.) are considered for modeling high load situations in the network. In addition, practical situations of dynamic crowd formation (e.g., public events, stadium etc.) and traffic jams due to vehicular users are investigated. Addressing the limitations of existing methods for crowd monitoring and traffic prediction, aforementioned mobility prediction methods are further integrated with other concepts and system models to design methods for crowd and traffic status prediction, from the perspective of cellular networks. The concept of pro-active load balancing and activation/deactivation of small cells are used in tandem with these developed functions of context awareness to improve the system performance. In addition to this, history based mobility prediction (Markov model) is enhanced by using additional context information in an urban environment and is used concurrently with smart resource allocation scheme to improve service continuity of a user. Finally, a generic framework (architecture, signaling, interfaces etc.) to support all these context aware procedures is proposed.

2.2 Context Aware D2D Communications

2.2.1 Overview

In the contemporary times, telecom operators are facing difficulty to accommodate the existing demand of mobile users. In addition, various new data intensive applications are being unfolded to the mobile users for day-to-day usage (e.g., proximity-aware services). The 4G cellular technologies are found to be still lagging behind the prevalent high data demand posed by the mobile users. Device-to-Device (D2D) communications is one of the promising technical features in next generation cellular technologies that is envisioned to satiate these demands as well as lead to improvement in coverage, data rates, resource utilization, delay and so on [FDM+12] Device-to-Device (D2D) communications commonly refer to the technologies that enable devices to communicate directly without base stations or an infrastructure of access points, and with/without the involvement of wireless operators [LZLS12].

Conventionally, all the radio access communications in a cellular network go through the BS irrespective of whether communicating parties are in a

proximity of each other or not. This method is better suited for traditional low data rate mobile services (e.g., voice call, SMS), where the users are generally not in vicinity of one another. However, new services that have emerged in recent years (e.g., video sharing, proximity-aware social networking, gaming) demand high data rates and the users could potentially be located in a range for direct communication with each other. Thus, D2D can significantly improve spectral efficiency of the network, as well as user throughput, delay, etc. Some of the key use cases of D2D communication is depicted in figure 2.5.

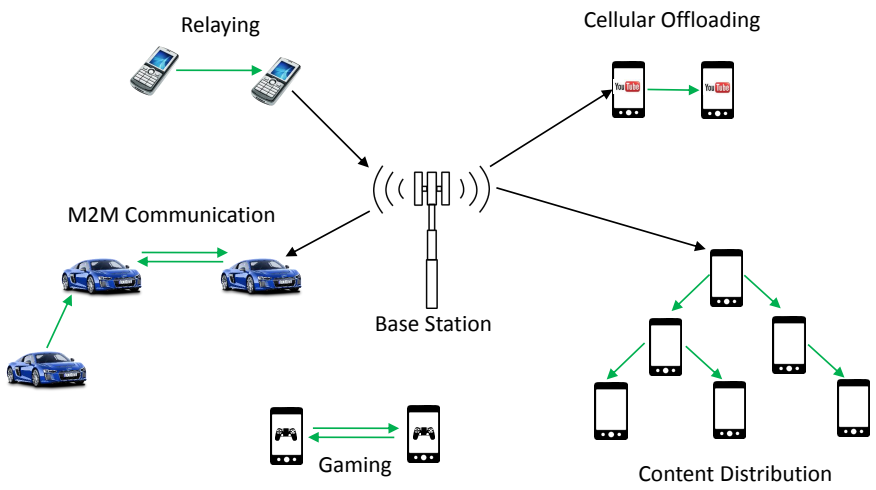

Figure 2.5: Key Use Cases of D2D Communication

D2D communications is classified into two categories namely Inband D2D and Outband D2D [AWM14], based on the spectrum used to carry out D2D communication, as ilustrated in figure 2.6. In case of Inband D2D, licensed cellular spectrum is used to carry out D2D communication. Whereas, Outband D2D allows cellular links in licensed band and D2D communication takes place in unlicensed spectrum (ISM band). Further, Inband D2D can be divided into Underlay and Overlay. The D2D overlay makes sure that the spectrum (licensed) used for D2D communication and conventional cellular communication are orthogonal to each other, whereas, D2D underlay allows reuse of the same spectrum for both D2D and cellular communication.

Although an Outband D2D seems like a simpler way to provision D2D communication, an Inband D2D is much preferred due to the fact that cellular operators will have high control over licensed cellular spectrum, whereas the interference in an unlicensed spectrum is uncontrollable, thereby pos-

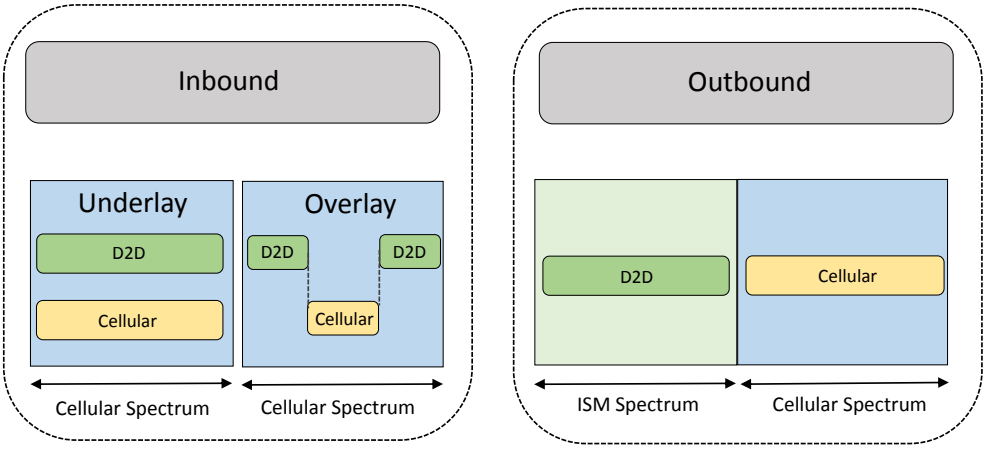

Figure 2.6: Classification of D2D Communication

ing restrictions for QoS provisioning [AWM14]. Specifically, Inband D2D underlay has the potential to improve spectral efficiency and resource utilization by allowing spectral resource reuse among D2D and cellular links. However, the major challenge here is the interference caused by D2D users to cellular communications and vice versa. This needs to be mitigated by methods of resource allocation, mode selection, power control and so on, to reap maximum benefits of D2D communications. Thus, the two major problems for D2D deployment in cellular networks, namely, *Power Allocation* and *Resource Allocation* are considered in this thesis. Particularly, in inband D2D, power transmitted requires to be properly regulated in order to manage interference from D2D to cellular links and vice-versa. Further, appropriate resource allocation can ensure that D2D links and cellular links that are in proximity of one another do not reuse resource blocks and pose high interference to each other.

2.2.2 Related Work

One of the first studies in D2D communications was done in [LH00], where D2D was used to enable multi-hop relays in cellular networks. Subsequently, the prospective of D2D communication in enhancing spectral efficiency of cellular networks was investigated in [DRJ+09][KA08][DRW+09]. Later, various possible use-cases of D2D were introduced such as peer-to-peer communication [LZLS12], machine-to-machine (M2M) communication [PP14],

multicasting[DZX+12] [ZHHC13], cellular offloading [BLRC10], video dissemination [DRW+09][GMD12] [GDM12] [LLG12] and so on. Some of these use cases are also depicted in figure 2.5. The various methods of D2D resource allocation, mode selection and power allocation are discussed in this subsection.

D2D Resource Allocation

Mutual interference between the cellular and D2D links reusing same resources is the most important issue in underlaying D2D communications. Thus, interference management schemes are important in order to deploy D2D underlay and achieve increased system capacity. An interference cancellation scheme is proposed in [XWC+10], where dedicated control channels are allocated to D2D users. Cellular users listen to this channel to measure SINR, and subsequently reports it to the eNB whenever SINR exceeds predefined threshold. In such cases, the eNB halts scheduling of cellular users on the resource blocks that are being used by D2D links. Further, the eNB broadcasts information about the location of users and respective allocated resource blocks. This assists D2D users to avoid using resource blocks which interfere with cellular users. Similar to [XWC+10], an interference minimization scheme is presented in [JKR+09]. Here, D2D users could also estimate signal power of cellular links and report these to a BS. Based on these reported values, the BS avoids allocating same resource blocks to D2D links and cellular links which can cause high interference to each other.

An interference management scheme is proposed in [MLPH11], where interference is not controlled by restricting the D2D transmission power as in majority of D2D interference management schemes. The proposed mechanism introduces an interference limited area where none of the cellular links and D2D links are allowed to share the same resources, there by avoiding the interference between D2D pair and cellular users. [CCZ+12] also considers a similar method, where interference limited areas are defined based on amount of tolerable interference and minimum SINR requirements for successful transmission. In such interference limited areas, cellular and D2D users cannot use the same resource. Further, the resource allocation is carried out in such a way that, D2D and cellular users located in same interference area are always allocated orthogonal resources. A method based on Han-Kobayashi rate splitting techniques [HK81] is proposed in [YT12], to improve throughput of D2D communications. In this work, a message is split into private and public parts. The private part can be decoded only

by the intended receiver, whereas public part is open to be decoded by any receiver. This facilitates the links prone to high D2D interference, to cancel the interference from public part of the message by means of a best-effort successive interference cancellation algorithm [RLJ00].

In [XSH+12b], a resource allocation scheme for D2D underlay based on sequential second price auction is presented. Here, a winner pays as much as the second highest bid. In case of D2D communications, every resource block is on auction and D2D pairs have to bid for the resource blocks they want to use. The bidding values are a function of achievable throughput of respective bidding D2D pair on the auctioned resource block. [DSS13]presents a method to select a cellular UE that can share radio resource with a D2D link, based of outage probability of the latter. The outage probability of D2D receiver is defined as the probability that the instantaneous SINR falls below a pre-determined SINR threshold. The scheme selects a cellular user that minimizes outage probability of the D2D link, to share common resource. [HYXY12] considers maximizing the number of D2D links while minimizing the average interference caused by D2D links, assuming uplink channel reuse in a single cell. The problem is formulated as a nonlinear programming and a heuristic algorithm based on Hungarian algorithm [Kuh56] is designed to solve it. A distance-constrained resource sharing criterion (DRC) is introduced in [WC12], where a BS makes sure that cellular to D2D interference is controlled by maintaining a minimum distance between a D2D transmitter and a cellular user reusing same resource. The BS must select a cellular UE with $L \geq L_{min}$ to allow resource reuse with a D2D pair, where L_{min} is a preselected distance constraint to keep interference under control (from cellular UE to D2D link). It is assumed that each UE reports its location (GPS) to the BS. Further, DRC does not intend to reduce transmission power of a cellular user, thereby avoids performance degradation of the cellular uplink.

In [DYRJ10], a mode selection problem is discussed, where a user can choose to transmit in a cellular mode or in D2D mode based on pre-defined constraints. The achievable transmission rate in each mode is estimated based on channel measurements performed by the users. This allows the user to choose the mode with a better transmission rate at each scheduling instance. A sum-rate optimization is considered in [XSH+12b], where iterative combinatorial auction game is adopted for resource allocation. The resources are treated to be the bidders who compete to gain business and D2D links are assumed as goods or services that need to be sold. A non-monotonic descending price auction algorithm is designed that can converge

in a finite number of iterations. The mode selection problem is considered along with the resource allocation in [SJWL13], where several pairs of D2D links co-exist with cellular users. The problem is formulated in order to maximize system throughput with minimum data rate requirements and particle swarm optimization [Ken10] is used for this purpose. Another proposal for mode selection and scheduling is proposed in [HKL12], the system time is assumed to be slotted and each channel is divided into sub-channels. The problem of maximizing the mean sum-rate of system with QoS satisfaction is formulated as stochastic optimization problem. The stochastic sub-gradient algorithm is used to obtain solution, based on which a sub-channel opportunistic scheduling algorithm is designed. This takes into account the CSI of D2D and cellular links along with QoS constraints on every D2D user. [HCLK10] attempts to obtain optimal communication mode for all users in terms of system equations, by capturing network information such as link gains, SINR, noise levels etc. The scheme allows two users to choose D2D mode only if their pathloss is lesser than the pathlosses between each user and the BS.

Power Control in D2D Communications

Power control is an important function that is able to limit the system interference. Typically, D2D transmissions take place in a local area resulting in a better SINR than conventional cellular communication. In such a scenario, appropriate power control for D2D communication is considered as a flexible way to maintain or improve overall system performance. The low power D2D transmissions result in low interference to conventional cellular links and low power consumption. [XH10] investigates various power control schemes which combine D2D mode and normal cellular communication. Mainly, power control in the D2D mode is emphasized. Three different methods of power control are studied; First method with D2D UEs using same fixed transmission power, second method where D2D UEs have the same fixed SNR target and use LTE uplink open loop fractional power control (OFPC) [3GP09], and third method where D2D UEs use close loop power control scheme based on OFPC with a single feedback item added. These methods are used to restrict the D2D transmission power to a tolerable level. In [KA08], cellular uplink resources are reused by D2D links. Thus, D2D causes interference to the cellular uplink transmissions at the BS. An algorithm is designed where D2D users monitor the received power of downlink control signals and estimate pathloss between D2D transmitter and BS. This

assists D2D users to restrict their transmission power below a threshold, such that high interference to the cellular users is avoided. In case, the required transmission power for a D2D link exceeds the minimal interference threshold, then respective D2D transmission is not permissible.

In [YTDR09] [YDRT11], a single cell scenario is considered with one cellular user and two D2D users sharing same radio resources. The BS is assumed to know instantaneous CSI of all the links and be able to control transmit power and radio resources of D2D links. The sum-rate is required to be optimized with energy/power constraint in three methods of link sharing namely, non-orthogonal sharing, orthogonal sharing and cellular mode. The optimal solution is given either in closed form or is chosen from a set. Another method for power allocation and mode selection in underlay D2D scenario is proposed in [JHC12]. Here, power efficiency is formulated as a function of transmission rate and power consumption, and power efficiency of the users is measured in different modes they can assume (cellular and D2D modes). Subsequently, each user chooses the mode with higher estimated power efficiency. [PLW+09] studies the uplink interference between D2D and cellular users to propose interference avoidance methods. In order to minimize the interference from cellular users to D2D links, D2D users make use of the resource block information from control channel. This assists D2D users to choose resource blocks that are not used by cellular users in their proximity. Further, as a step to counter interference from D2D users to cellular links, the expected interference from D2D link on a cellular resource block is broadcast to all D2D users. Based on this information, D2D users can adapt their transmission power and the choice of resource block, such that D2D interference to a uplink transmission does not exceed a permitted threshold.

[XTL11] proposes a heuristic algorithm for power allocation in OFDMA-based cellular networks. A heuristic that performs power allocation and mode selection is proposed using existing subcarrier and bit allocation algorithms [Zha04],[YHPH06]. The resource allocation is first done for cellular users and then resource allocation and mode selection for D2D are performed. If required transmission power for D2D link exceeds a pre-defined threshold then such a D2D pair has to communicate through the BS. In order to improve the user spectrum utility, [CDY+12] proposes a mode selection and power allocation, where spectrum utility is defined as combination of users' data rates, power expenditure and bandwidth. The optimal transmission power is derived for each mode (D2D or conventional BS mode), and subsequently an evolutionary game [Wei97] is used for mode selection. The mode selection is

carried out individually and independently at each user. Further, the BS accumulates mode selection decisions of all involved users and broadcasts this to all users, in order to assist their future mode selections. In [MBM+14], a power control mechanism with variable target known as Soft dropping is applied to cellular and D2D users. The soft dropping scheme adapts the transmit power to reach a variable target SINR and pose tolerable levels of interference. The target SINR in such a scheme decreases gradually with an increase in required transmit power. This improves the probability of setting up a reasonable power allocation, where target SINR values of all co-channel links could be met. This is because, links with poor quality that would need higher power, aim at lower SINR targets whereas, links with better quality requiring low power, aim at higher SINR targets.

[CFF+16] studies severe interference posed by densely deployed D2D pairs to cellular links and proposes a discrete location aware power control. Here, the total cell area is divided into several regions with total power budget for each region being conducted with weighted allocation. The weighted factor is set by the system to weigh fairness among D2D pairs and network throughput. This is done in order to meet the constraint on the outage probability of cellular user. In this setup, power control scheme only needs positions of D2D pairs which are active instead of channel information. A new interference management strategy for D2D underlay cellular networks is proposed in [BY12] without the exact location information of cellular UEs. A case is assumed where one D2D pair reuses resource from cellular users in both uplink and downlink. The BS has knowledge of distances at which D2D and cellular UEs are present (away from BS). A power control mechanism is presented to restrict the maximum transmit power of D2D and limit harmful interference to cellular links. Then, an interference limited ring (ILR) is introduced, in which the D2D users cannot reuse resources from cellular UEs. The radius of ILR is derived as a closed form and it ensures that the outage probability of D2D (caused by cellular interference) is lesser than a pre-defined threshold. [ZJLZ14] addresses the joint optimization problem of D2D mode selection, modulation and coding schemes (MCSs) assignment, radio resources and power allocation. The problem is formulated in order to minimize the overall power consumption while ensuring minimum rate guarantee. The problem is decoupled into a PRB allocation sub-problem, designed to obtain minimum overall power consumption under given set up and a Mode selection and MCS assignment sub-problem. These problems are solved by Lagrangian relaxation and tabu search methods, respectively.

2.2.3 Contributions

Many of the existing schemes for D2D interference management via resource allocation, mode selection or power control require channel and link information of entire system to find an optimal resource, mode or power assignment. Though theoretically feasible, it is costly and complex for deployment in practical situations. Further, several proposals use advanced mathematical techniques, such as game theory [XSH+12a], nonlinear programming [ZCYJ13] etc., for D2D interference management. However, such schemes tend to introduce a higher computational overhead that should also be taken into account while evaluating system performance. Moreover, vast majority of the existing schemes try to restrict either the number of allowed D2D connections in a system, the amount of resources reused/assigned for D2D users or transmission power of D2D links. Regulating D2D transmissions to mitigate the aforementioned problem would mean sub-optimal exploitation of D2D communications, hence maximum benefit cannot be reaped from D2D communications at all times. Considering these issues with the existing work in D2D communications, this thesis emphasizes on exploiting location awareness to design a D2D resource allocation scheme without requiring link conditions in entire system. Acknowledging the high interference caused by densely distributed D2D users in an D2D underlay cell, a location based D2D resource allocation scheme is proposed to maximize the number of connections in a cell. Further, the robustness of location based schemes is evaluated with respect to positioning errors. In addition, post-resource allocation power control at cellular users is proposed as a solution for cellular users to co-exist with densely distributed D2D links in a D2D underlay cell, without imposing restrictions on D2D transmissions.

Chapter 3

System Modeling

This chapter deals with various system modeling assumptions and simulation parameters considered in this thesis. The chapter gives generic overview about simulations in the field of mobile communications and discusses in detail different simulation models used in this thesis.

3.1 Overview

With the evolution of mobile communication technology over several decades, there has been significant improvement in hardware components, chip-sets as well as algorithms such as link adaptation and scheduling. Further, cooperative multi-antenna operation modes have led to increased data throughput and system capacity. With the introduction of 4G technologies (e.g., LTE-A, WiMAX) and extensive research leading to 5G technologies, it becomes imminent to carry out performance comparisons across various Radio Access Technologies (RATs). Thus, it becomes crucial to have similar base line assumptions to evaluate performance, since use of different radio propagation models, mobility assumptions, interference conditions etc., will lead to varying results of system-level simulations.

In general, system-level simulations are used to simulate the global behavior of entire system. Some key building blocks of a system-level simulator are as below:

- Cell layout and deployment scenarios
  - Dictates number of tiers in cellular layout, site-to-site (s2s) distances, antenna type (omnidirectional or three sector) and so on.

  - Radio propagation characteristics including large scale effects such as path loss, shadowing and small scale effects such as fast fading are taken into account.

- RAT characteristics
  - Typical RAT parameters such as carrier frequency, bandwidth, radio resources, transmit power etc are defined.

- Use mobility and distribution
  - Number of users in the system and their distribution are considered.
  - Mobility models (e.g random walk, predefined trajectories for user groups, direction-oriented road-like movements etc) are taken into account.

- Network events and procedures
  - Network events such as handover events, resource scheduling etc. are specified.
  - Radio resource management schemes or other schemes under assessment are defined.

- Key performance indicators
  - KPIs are logged periodically or aggregated with respect to simulation time.
  - Typical examples of KPIs include throughput, spectral efficiency, E2E latency, number of links and so on, which are used for performance comparisons.

Figure 3.1 depicts the general framework of simulation methodology used. The simulation methodology comprises of realistic environment models (3D) for shadowing, fading and path losses, realistic user mobility models, link-to-system interfacing, additional context information, network events, and procedures. Subsequent sections discuss various simulation models used in this thesis, based on differences in cell layout and deployment scenarios for different use cases.

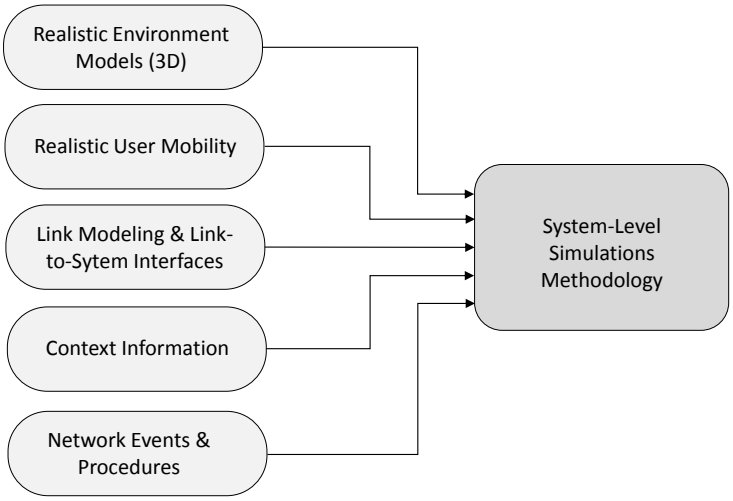

Figure 3.1: Simulation methodology

3.1.1 LTE Overview for Simulation Modeling

Long term evolution (LTE) [HT09] is the key RAT considered for simulation in this thesis. LTE (3GPP release 8) transmissions are based on Orthogonal Frequency Division Multiplexing (OFDM). The radio resources are organized as a 2D grid with respect to time and frequency. The resource assignment is carried out in terms of Physical Resource Blocks (PRBs), each with 180 KHz bandwidth (12 subcarriers) [HT09]. Figure 3.2 shows the LTE frame structure and organization of radio resources as a grid. Each LTE frame has a size of 10 ms and consists of 10 sub-frames (SF), 1 ms each. Each sub-frame has 2 time slots, each time slot being 0.5 ms. A resource block (RB) consists of 12 subcarriers and 7 OFDM symbols, accommodated in one time slot. A resource element corresponds to a modulation symbol, hence a resource block has 84 resource elements (12×7). It is important to note that minimum bandwidth available for allocation is 180 KHz and allocation resolution in time domain is 1 ms, though the resource block specification refers to 0.5 ms time slot. The scheduling is done on TTI basis (1 TTI = 1 ms), facilitating link adaptation techniques such as AMC and HARQ retransmissions [HT09].

The DL peak bit rate is given as [Kle08],

$$R = \Delta f \times N_{SC} \times \frac{log_2 M}{1 + G_{CP}} \tag{3.1}$$

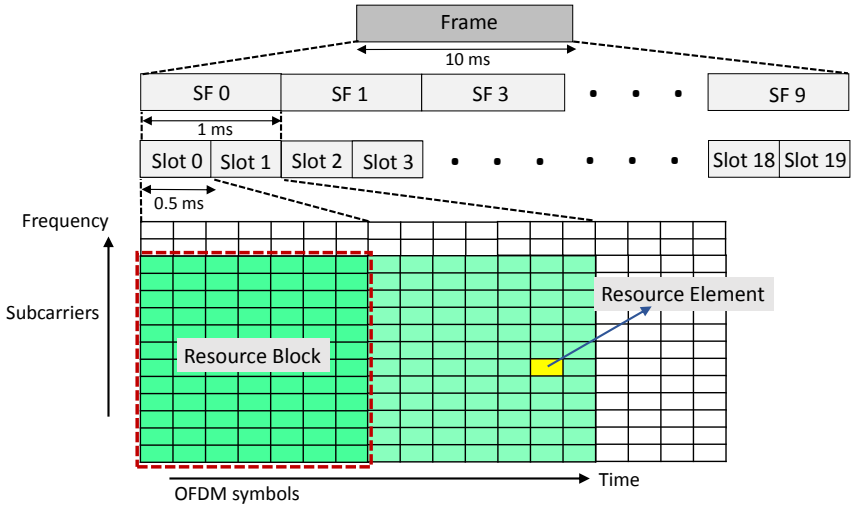

Figure 3.2: LTE frame structure and resource grid [3GP12a]

$$= 15KHz \times 1200 \times \frac{6bit}{1 + 1/14}$$

$$\approx 100Mbps$$

Where the subcarrier frequency Δf is 15 KHz, the total number of subcarriers N_{SC} is 1200, assuming 20 MHz bandwidth. Further, with 64 QAM modulation scheme, size of modulation symbol M is 6 and the ratio of cyclic prefix time to effective symbol time G_{CP} is 1/14. The aforementioned figures are considered to showcase the best possible peak bit rates [3GP12b]. However, in real world scenarios data rates of a user often depend on available number of resources for scheduling, modulation and coding schemes used, which in turn depend on SINR.

Further, the DL load of a LTE system is defined as the ratio of number of PRBs allocated in a system to total number of PRBs available in that system, given by,

$$\rho_{DL} = \frac{NPRB_{use}}{NPRB_{tot}} \tag{3.2}$$

Where, $NPRB_{use}$ and $NPRB_{tot}$ are number of used PRBs and total number of available PRBs respectively. The signaling overhead due to control signaling, pilot based channel estimation, synchronization etc impacts $NPRB_{tot}$. For instance, 6 PRBs are used per TTI to account for signaling overhead, leading to $NPRB_{tot}$ of 44 PRBs, assuming 10 MHz bandwidth. This accounts for 12% of total resources or 1.08 MHz of total bandwidth.

For system level simulations of the LTE system, a generic link-to-system mapping is required to determine achievable data rates at a system level given current SINR conditions. Initiating a dedicated link level simulation for each active link in the system, during a system level evaluation leads to drastic increase in overall simulation time and computational efforts required. Thus, use of a link-to-system interface will greatly reduce computational and processing overhead. In order to deduce receive conditions of a user for system level evaluation, the so-called *geometry* or *G-factor* is used. It is defined as the ratio of average power of own cell of the user (P_O) to interference power from other cells plus noise (P_{IN}), which in a wide system bandwidth corresponds to the average wide-band Signal-to-Interference-plus-Noise-Ratio [MNK+07],

$$G = \frac{P_O}{P_{IN}} = \frac{RSRP_k}{\sum_{i=1, i \neq k}^{N} RSRP_i + P_N} \tag{3.3}$$

Where $RSRP_k$ and $RSRP_i$ denote Reference Signal Received Power (RSRP) of connected and neighboring base stations, respectively. $P_N = kT_0B$ represents thermal noise power determined by the Boltzmann constant $k = 1.3806504X10^{-23}$, ambient temperature $T_0 = 290K$, and system bandwidth B.

3.2 Synthetic Simulation Model

This section deals with one of the key simulation methods used in this thesis. The simulation set-up borrows methodologies that emanate from organizations such as NGMN [NGM07], 3GPP[3GP06], IEEE [IEE08],[IEE05] and 3GPP2 [3GP04]. The simulation set-up considers cellular layout consisting of hexagonal cells with base station at center of each cell and base station is equipped with LTE technology. The details about considered parameters, deployment scenarios, assumptions and overall simulation methodology are provided in this section.

3.2.1 Related Parameters

For evaluating the performance of developed concepts in Chapter 4, sections 4.2, 4.4 and 4.5, the Downlink (DL) characteristics of LTE system is considered and modeled. The typical baseline parameters for performance evaluation are tabulated in table 3.1.

Table 3.1: LTE specific DL parameters [3GP10], [3GP06], [IEE08], [HT09]

Parameter	Assumption
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
Cell layout	Hexagonal grid, 19 cell sites
Site-to-site distance	$500m$
BS transmission power	46 dBm
Antenna pattern (Horizontal)	Omni-directional, $A_H(\varphi) = 0$
Antenna pattern (Vertical)	$A_V(\theta) = -min\left[12\left(\frac{\theta-\theta_{etilt}}{\theta_{3dB}}\right)^2, SLA_V\right]$ $\theta_{3dB} = 10, SLA_V = 20dB$, $\theta_{etilt} = 15°$ $\theta = tan^{-1}\left(\frac{h_{BS}-h_{MS}}{D}\right)$ h_{BS}, h_{MS} are heights of BS & MS respectively D is the distance between BS & MS
3D antenna pattern combining	$A(\varphi,\theta) = -min\{-[A_H(\varphi) + A_V(\theta)], A_m\}$ $A_m = 25dB$
Antenna gain	17 dBi
BS noise figure	5 dB
BS cable loss	2 dB
UE noise figure	7 dB
Thermal noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B)$

3.2.2 Deployment Modeling

The organizations such as 3GPP, IEEE, NGMN and 3GPP2 have consensually proposed the use of a common site layout for evaluations. The same methodologies are followed in this thesis with slight modifications as necessary. The site layout considered for evaluations is a hexagonal grid made of multiple tiers of cells. A 3-tier cellular layout with 19 cells being the most common set up as shown in figure 3.3. Typically, when a user leaves evaluation area, it again enters considered area from exactly the opposite side. This is called as wrap-around approach [ZKAQ01]. However, for simplicity, the simulation used here considers bounce-back approach, where a user reaching the limits of evaluation area will bounce back into evaluation area with a randomly chosen direction. Boundary of evaluation area is defined by a cir-

cle, encapsulating desired service area for inspection. Both these approaches make sure that number of users considered for the simulation remain intact within considered evaluation area.

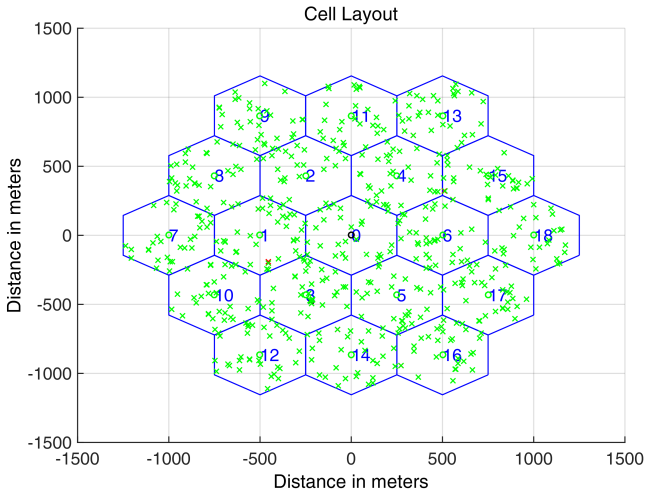

Figure 3.3: Synthetic simulation setup

After hexagonal grid is laid out with multiple base stations in a predefined evaluation area, the next vital step is to define radio propagation characteristics in this considered site. The radio propagation law considers path loss which depends on separation between transmitter and receiver as well as frequency of the carrier. For scenarios in urban and sub-urban areas outside the high rise core, where buildings are of nearly same height [IR97], path loss is given by,

$$L = 40\ \left(1 - 4 \cdot 10^{-3}\ \Delta h_b\right)\ log_{10}R - 18\ log_{10}\Delta h_b + 21\ log_{10}f + 80, \quad (3.4)$$

Where R denotes distance between the BS and a UE in km, f denotes the carrier frequency in MHz, Δh_b represents height of BS in m, measured from average rooftop level. For instance, BS antenna height (Δh_b) is $15m$ indicating BS is fixed at 15 m above the average rooftop. The carrier frequency considered is 2 GHz. Applying the aforementioned values for f and Δh_b in equation 3.4 leads to simplified path loss given by [3GP10], [3GP06],

$$L = 128.1 + 37.6\ log_{10}R \quad (3.5)$$

However, L shall in no circumstances be less than free space loss (between isotropic antennas) [IR97],

$$
\begin{aligned}
L &= -10\ log_{10}\left(\frac{\lambda}{4\pi R}\right)^2 \qquad (3.6)\\
&= 20\ log_{10}\left(\frac{4\pi R}{\lambda}\right)\\
&= 20\ log_{10}\left(\frac{4\pi R f}{c_0}\right)\\
&= 20\left[log_{10}\left(\frac{4\pi}{c_0/10^9}\right) + log_{10}\left(R/1000\right) + log_{10}\left(f/10^6\right)\right]\\
&= 20log_{10}\left(\frac{4\pi}{c_0/10^9}\right) + 20log_{10}\left(R'\right) + 20log_{10}\left(f'\right)\\
&\approx 32.45 + 20log_{10}\left(R'\right) + 20log_{10}\left(f'\right)
\end{aligned}
$$

Where R' is the distance between transmitter and receiver in km, f' denotes the carrier frequency in MHz and c_0 is the speed of light.

Following the modeling of path loss, the other two important characteristics of propagation law namely *Shadowing and Fast fading* need to be modeled. The wireless channels operate through electromagnetic radiation from a transmitter to corresponding receiver. In case of free space, signal propagation almost follows a straight line (line-of-sight). However, in the real world, a signal radiated from the transmitter (e.g., Base station) in a given cell will often encounter a large number of obstacles, such as buildings, trees, mountains etc., before reaching the receiver. This will lead to physical effects inherent to radio wave propagation namely, reflection, refraction, diffraction and scattering [TV05]. The presence of tall buildings in urban areas leads to severe diffraction losses, and usually mean no line-of-sight transmission between transmitter and receiver. Further, there would be multiple weaker copies of same signal from transmitter, due to reflection or scattering from obstacles, which travel different paths of various lengths and superimpose constructively or destructively at receiver's location. In addition to this, a wireless receiver is often mobile in nature, inducing *Doppler effect*, where speed of mobile receiver affects how rapidly the signal power level fades as it moves and the direction of relative movement between receiver and transmitter impacts received signal frequency [TV05]. Thus, due to variety of aforementioned phenomena in wireless channel, the channel strength varies

randomly over time and frequency. These variations can be roughly divided into two types,

- *Large-scale fading or Slow fading* : This is due to attenuation of a signal as function of the distance between transmitter and receiver. This is analogous to the path loss and if the signal attenuation is very high then such a signal will vanish. Further, the large-scale fading also includes shadowing effects caused by large objects like buildings and hills. Such large-scale fading occurs when a mobile receiver travels distance of the order of a cell size, and is typically frequency independent [TV05].

- *Small-scale fading or Fast fading* : This is due to multiple signal paths that exist between a transmitter and a corresponding receiver, leading to constructive and destructive interference at the receiver's location. The small-scale fading occurs at a spatial scale of the order of a carrier wave length, and is frequency dependent [TV05].

The Gudmundson model [Gud91] is the standard shadow fading model recommended by organizational units such as 3GPP [3GP10], [3GP06]. In this model, shadowing is modeled as log-normally distributed random variables with zero mean and a standard deviation of 8 dB. The model is proven to be good fit for large to moderate cell sizes [Gud91]. In macro cell scenarios, correlation distance is around 50 m and shadowing correlation between cells is assumed to be 0.5 [3GP06]. The shadowing realizations vary with distance k and shadowing values of spatially adjacent locations (within correlation distance) are correlated. This implies that if a user travels a distance greater than correlation distance away from his original position, then shadowing values at his original location and present location are uncorrelated. To model the correlation properties a simple decreasing correlation function is used, given by [Gud91],

$$R_A(k) = \sigma^2 \; a^{|k\|} \tag{3.7}$$

$$a = \varepsilon_D^{\upsilon \; T/D} \tag{3.8}$$

Where, σ^2 is the variance, usually in range of 3 dB to 10 dB [Gud91], ε_D is the correlation between two random variable realizations separated by distance D, υ is the user velocity and T is the sampling interval. A two-dimensional shadowing grid is precomputed with a resolution of correlation

distance (50 m) and shadowing factors for any intermediate location is obtained via interpolation. This procedure simplifies simulation process and accelerates it significantly. The shadowing factor ζ at a specific position $k_{x,y}$ is obtained from linear interpolation as [IEE08],

$$\zeta\left(k_{x,y}\right) = \sqrt{1-\frac{x_{pos}}{d_{corr}}}\left[S_{0,l}\sqrt{\frac{y_{pos}}{d_{corr}}} + S_{3,l}\sqrt{1-\frac{y_{pos}}{d_{corr}}}\right] + \sqrt{\frac{x_{pos}}{d_{corr}}}\left[S_{1,l}\sqrt{\frac{y_{pos}}{d_{corr}}} + S_{2,l}\sqrt{1-\frac{y_{pos}}{d_{corr}}}\right] \tag{3.9}$$

Where, $\{S_{0,l}, S_{1,l}, S_{2,l}, S_{3,l}\}$ are shadowing grid values of the corresponding square where the considered BS or UE is located, x_{pos} and y_{pos} denote the offset values with respect to $x-$ and $y-$axis, within the considered shadowing grid square and d_{corr} is the pre-defined correlation distance. An illustration of this procedure is presented in figure 3.4.

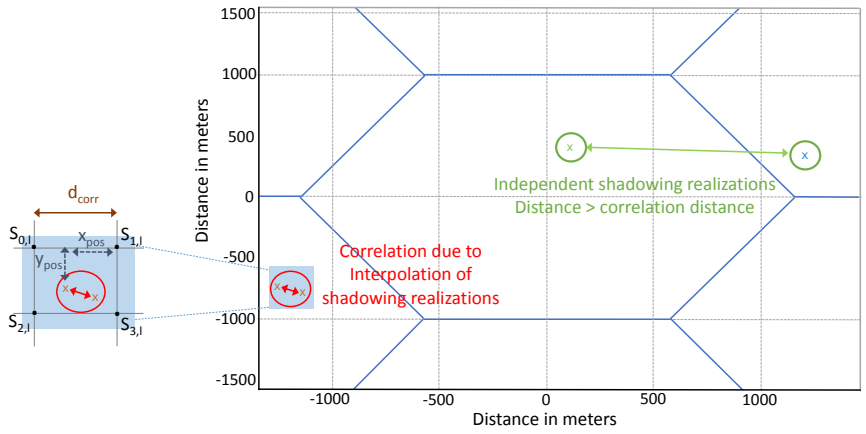

Figure 3.4: Illustration of shadowing realizations via interpolation

The effective shadowing factor ζ_{eff} is deduced as [IEE08],

$$\zeta_{eff} = a\ \zeta_{UE} + b\ \zeta_{BS} \tag{3.10}$$

Where $a = b = \sqrt{1/2}$. Further, to incorporate the variations in received signal strength due to frequency and velocity dependent fading, a precomputed fast fading trace [YB00] is used during simulation and link-level evaluation of instantaneous SINR values is carried out. Figure 3.5 depicts an

exemplary geometry intensity map that is derived, considering path loss, shadow fading and other losses incurred along transmission path.

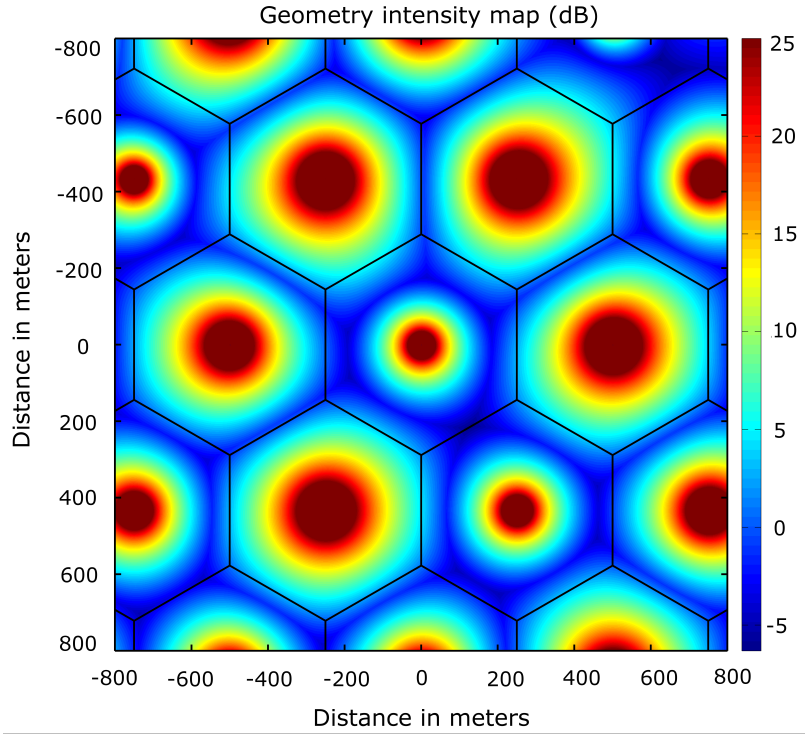

Figure 3.5: Geometry intensity map

3.2.3 Scheduling and L2S Modeling

Once the deployment of cell sites with base station and UEs is done and radio propagation characteristics are defined, the radio resources need to be allocated to enable transmissions between BS and UEs. The system resources such as radio channels, time slots, PRBs, etc., need to be assigned by a scheduler that resides in MAC layer and is part of Radio Resource Control (RRC). The resources are assigned for PHY layer transmissions for different service types based on predefined scheduling metrics such as, Channel Quality Indicator (CQI). Further, prioritization of certain service types could be done based on their QoS requirements (e.g., minimum delay, GBR etc.,) and a corresponding policy could be implemented. Typically, it is assumed that a scheduler takes into account all the control channel and protocol overhead to calculate the total number of resources available to be scheduled [IEE08]. In the considered simulator the following widely known scheduling methods are used:

- Round Robin (RR): In this scheme, resources are shared among users in the time domain in a circular fashion, without priorities. The scheme is hence starvation free, however the resources allocated, may be fixed for the whole service duration or impacted by additional service requirements or MNO policies.

- Maximum Carrier-to-Interference (max C/I): In this scheme, the users that experience best receive conditions (in terms of C/I ratio) are prioritized over other users in the system and served.

- Proportional Fair (PF): In such scheduling strategy, a metric is computed for all the active users that demand resources in a specific scheduling interval. The users are then sorted according to this scheduling metric and users that have higher value of this metric are allocated with available resources. Before the next scheduling interval, metrics for all users are updated and the process is repeated [IEE08].

For PF-scheduling, the scheduling metric $M_i(t)$ for a user i at scheduling instant t is given by [IEE08],

$$M_i\left(t\right) = \frac{T_i(t)}{\left(\overline{T}_i(t)\right)^{\alpha}} \tag{3.11}$$

Where, $T_i(t)$ represents the instantaneous data rate that can be provided for a user i at the scheduling instant t. $T_i(t)$ depends on the CQI feedback, which in turn determines the modulation and coding scheme used to meet PER requirement. $\overline{T}_i(t)$ is the average throughput of user i smoothed by a low-pass filter at scheduling instant t. α is the fairness exponent factor which has default value of 1. $\overline{T}_i(t)$ of a scheduled user is given by [IEE08],

$$\overline{T}_i(t) = \frac{1}{N_{PF}}\ T_i(t) + \left(1 - \frac{1}{N_{PF}}\right)\ \overline{T}_i(t-1) \tag{3.12}$$

Where, N_{PF} denotes the latency scale of PF scheduler calculated as [IEE08],

$$N_{PF} = T_{PF}\ \frac{N_{partition}}{T_{frame}} \tag{3.13}$$

Where, T_{PF} is the latency time scale in seconds, T_{frame} is frame duration of the system and $N_{partition}$ denotes the number of resource partitions, with each partition assignment considered as a separate packet transmission. For an unscheduled user $\overline{T}_i(t)$ is given by,

$$\overline{T}_i(t) = \left(1 - \frac{1}{N_{PF}}\right) \overline{T}_i(t-1) \tag{3.14}$$

As discussed before, a link-to-system interface is required to deduce achievable data rates at a system level given the current SINR conditions. The link-to-system approach used here is based on SISO system operation, i.e., only a single transmit and receive antenna is used for user data transmission and reception, respectively. The G-factor described in equation 3.3 is used to describe the receive conditions of a user at system level. The L2S interface incorporates technology specific features like frequency domain packet scheduling (FDPS), and accounts efficiency losses such as, control signaling overhead, protocol overhead etc., to yield a mapping of G-factor values onto spectral efficiency [MNK+07]. Specifically, efficiency losses in terms of bandwidth (e.g., in 10 MHz bandwidth, 50 PRBs of 180 KHz are present that leads to 50×180 KHz = 9 MHz of usable bandwidth), cyclic prefix (normal cyclic prefix length yields $G_{CP} = 1/14$), and overhead caused by pilot symbol assisted channel estimation and dedicated as well as common control channels are considered. The effective user data rate R is calculated as [MNK+07],

$$R = S \times B \tag{3.15}$$

Where, B is the system bandwidth, S is the spectral efficiency in $bits/s/Hz$ given by,

$$S = B_{eff}\ log_2\left(1 + \frac{\xi}{\xi_{eff}}\right) \tag{3.16}$$

Here, B_{eff} adjusts for the the system bandwidth (BW) efficiency of LTE set to 0.6, and ξ denotes SINR, ξ_{eff} adjusts for the SINR implementation efficiency of LTE and is set to 1 [MNK+07].

3.2.4 Overall Methodology

Figure 3.6 depicts the framework of system level simulations used. The simulator is implemented in C-programming language. In order to track the BS and UE related events, specific data structures are created. The parameters specific to a scenario are configured, algorithms that need to be evaluated are selected and other initializations are done at simulation setup. As the simulation starts running, user locations, respective fading values , geometry values etc., are updated accordingly. Based on various measurements and

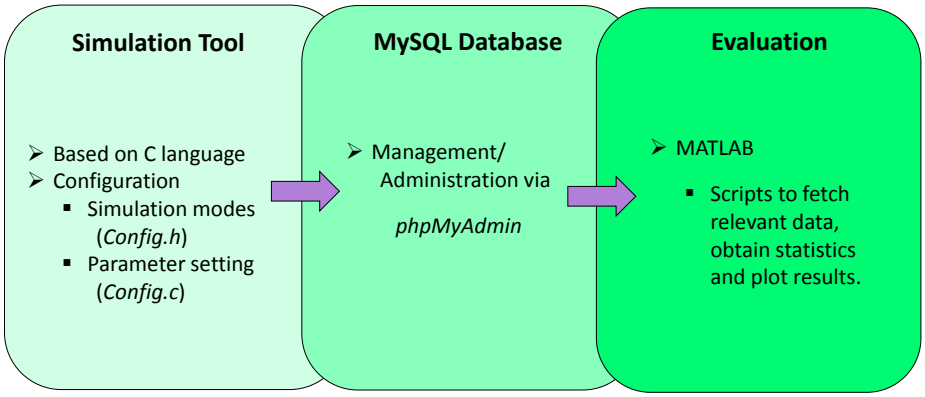

Figure 3.6: Simulation framework

statistics, suitable RRM or other procedures are initiated. The parameters specific to the UEs, BSs or other events are logged periodically. In order to reduce simulation run time, user mobility traces and fast fading traces can be precomputed and accessed during simulation. The number of monitored parameters and events in the simulation can increase easily, due to the complexity and dynamics of considered scenarios. Thus, performance of regular desktop computers would be poor while handling such large amounts of data. Therefore, a framework which allows running of multiple instances of simulation, while separating simulation process from data management and evaluation is designed and used. The simulations are handled by remote servers or dedicated powerful computers, whereas simulation data is hosted by MySQL based databases. The evaluation of simulation results can be performed offline using MATLAB scripts.

3.3 Realistic Simulation Model

3.3.1 Overview

This section presents modeling of a simulation set up that is closely aligned with real-world deployment. For this purpose, the so-called *Dense Urban Information Society* scenario proposed by the METIS project [MP12] is used. The simulation set up features a 3D environmental model with tall buildings, streets, cross-roads, parks and so on, mimicking a typical European urban city (e.g., Madrid), hence the setup is also called *Madrid grid* [MP13c]. The

radio propagation, user mobility behavior, system modeling and performance evaluation are tailored to suit the attributes of chosen environmental model. Figure 3.7 depicts an exemplary signal propagation in considered dense urban scenario.

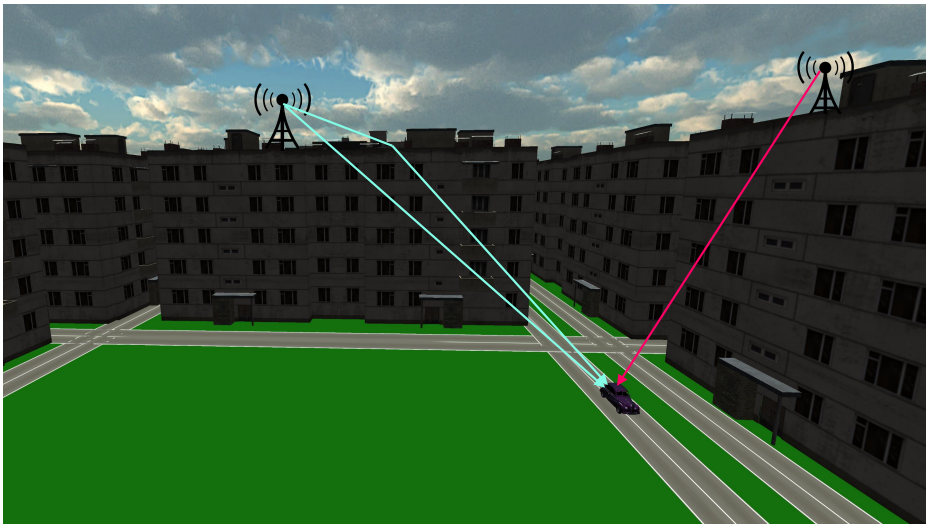

Figure 3.7: Exemplary signal propagation in dense urban scenario [MP13c]

This simulation setup is used extensively for the concepts proposed in Chapter 4, section 4.3.

3.3.2 Environmental Model

The urban environmental model consists of buildings, roads, parks, sidewalks, cross-roads etc., capturing various aspects within the Madrid grid. This model has been developed within the METIS consortium [MP13c] and is based on observations regarding the city structure of Madrid. It is an example of typical European city environment capturing way more aspects than Manhattan grid layout [Mol11],[MP13c]. More realistic and non-homogenous building layout is proposed in order to capture, for e.g., real life mobility behavior of users, heterogeneity of cellular network deployment or diversity of SINR distribution.

A top view of the Madrid grid layout is shown in figure 3.8 and a 3D view is depicted in figure 3.9. The key elements of the layout which are important in the scope of this thesis are outlined as below [MP13c],

- Base plane with 387 m from east to west and 552 m from south to north.

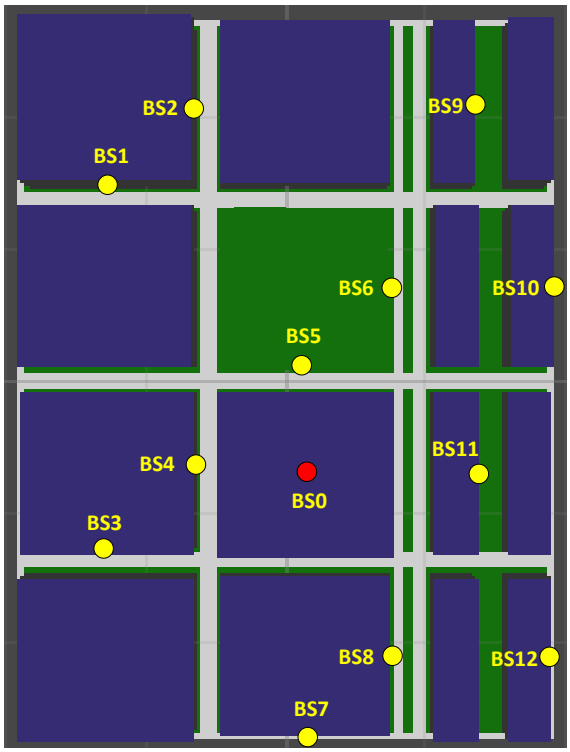

Figure 3.8: A 2D visualization of the Madrid grid [MP13c]

- Buildings (7 square shaped and 8 rectangular shaped): square buildings have length and width of 120 m with varying heights, whereas rectangular ones have length of 120 m and width of 30 m. The building height is uniformly distributed between 8 and 15 floors with 3.5 m per floor.
- Park with length and width of 120 m.
- Roads for vehicular mobility with width of 3m and 25 cross-roads in complete layout.

3.3.3 Propagation Model and Parameters

The Madrid grid layout comprises of 12 LTE micro base stations (BS1-BS12) and a LTE macro base station (BS0) as shown in figure 3.8. The macro BS has three sector antennas, with one antenna steering into northern direction and other antennas oriented with 120 degrees offset with respect to north. Micro BS antennas are steered toward streets perpendicular to building walls. Radio propagation models follow [MP13c] and takes into account

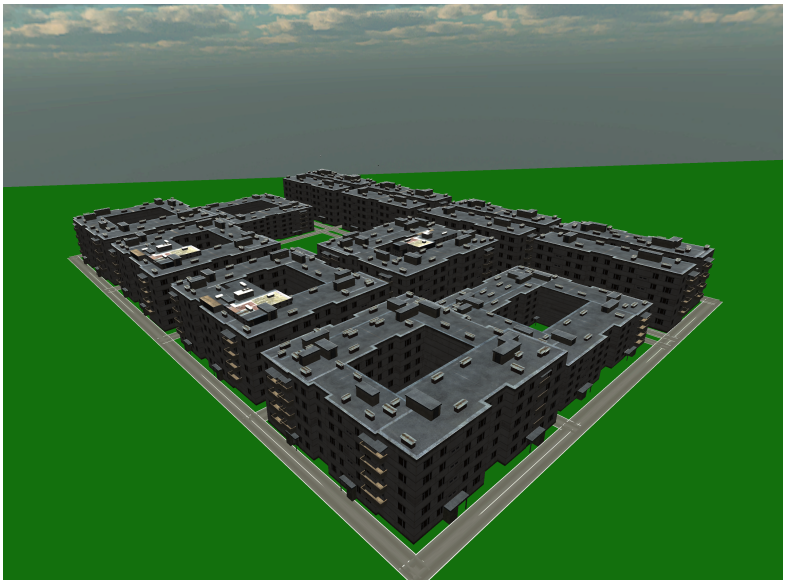

Figure 3.9: A 3D visualization of the Madrid grid [MP13c]

urban macro and micro cell outdoor-to-outdoor propagation characteristics. The propagation model takes into account effects of diffraction, refraction and reflection, appropriate to the situation. This thesis mainly considers connectivity between vehicular users and micro BSs, thus the propagation model only relevant to outdoor micro BS (Urban Micro O2O) [MP13c] is described here. The key consideration here is that the transmitter (micro BS) is situated much below the mean height of building. This implies that there is lack of dominant visibility of users and main propagation is carried out via reflection between buildings. The model is based on ITU-R UMi path loss model for Manhattan grid layout [MP13c]. The model can distinguish among the main street where a transmitter is located, parallel streets and perpendicular streets as depicted in figure 3.10.

When user is in the same street as BS, i.e Receiver (Rx) and transmitter (Tx) are in same street (main street), LOS path loss is obtained as [MP13c],

$$PL_{LOS}(d_1) = 40\ log_{10}(d_1) + 7.8 - 18\ log_{10}(h'_{Tx}) \qquad (3.17)$$
$$- 18\ log_{10}(h'_{Rx}) + 2\ log_{10}(f_c)$$

Where, f_c is carrier frequency in GHz, d_1 is the distance between a transmitter and corresponding receiver in meters, h'_{Tx} and h'_{Rx} are effective antenna heights in meters, of transmitter and receiver, respectively. The effec-

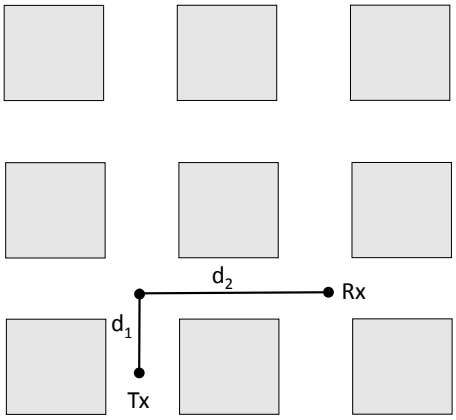

Figure 3.10: Geometry model of propagation scenario [MP13c]

tive antenna heights are deduced by subtracting effective environment height in urban environments ($\approx 1m$) from actual antenna heights [MP13c]. If a user is in perpendicular street then path loss is given by [MP13c],

$$PL_{NLOS} = min\left(PL\left(d_1, d_2\right),\ PL\left(d_2, d_1\right)\right) \tag{3.18}$$

Where,

$$\begin{aligned} PL\left(d_k, d_l\right) = PL_{LOS}\left(d_k\right) + 17.9 - 12.5 n_j \\ + 10\ n_j\ log_{10}\left(d_l\right) + 3\ log_{10}\left(f_c\right) \end{aligned} \tag{3.19}$$

$$n_j = max\left(2.8 - 0.0024\ d_k,\ 1.84\right) \tag{3.20}$$

In the perpendicular street, if the distance between a transmitter and its receiver is less than 10 m then LOS conditions apply and the path loss is obtained using equation 3.17. In case, a user is in parallel street then the path loss is assumed to be infinite.

Table 3.2 summarizes the simulation parameters and assumptions considered. In L2S mapping, user specific transmission bandwidth B_u is deduced based on service requirements and defined as multiples of PRBs, with each PRB being 180 KHz wide. The overall system efficiency factor η_{eff} is assumed to be 0.57 [MNK+07].

Table 3.2: Simulation parameters (Madrid grid) [3GP10], [3GP06], [IEE08], [MP13c]

Parameter	Assumption
Carrier frequency	2.6 GHz
System bandwidth	10 MHz (44 PRBs)
BS transmission power	46 dBm
Minimum coupling loss	53 dB
BS noise figure	5 dB
BS cable loss	2 dB
Thermal noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B)$
UE noise figure	7 dB
Antenna gain	17 dBi
L2S mapping [MNK+07]	SISO, $C = \eta_{eff}\ B_u\ log_2(1 + SINR_{eff})$

3.3.4 Overall Methodology

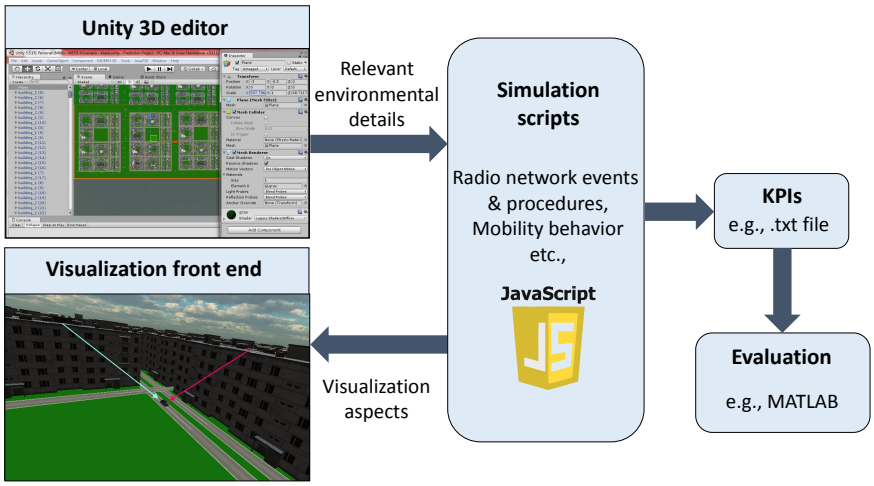

Figure 3.11: Simulation framework

The simulator makes use of a powerful game engineering tool set called *Unity 3D* [Tec15]. The intrinsic features and assets of the tool greatly assists the specific modeling requirements of simulation scenario. The tool has a visualization front end (editor), which allows the modeling of environmental aspects of simulation scenario such as buildings, roads, foliage, cross-roads etc. The radio network related aspects such as propagation modeling, user mobility modeling, radio resource management and other network procedures are implemented using JavaScript. The simulation progresses with respect to

time scale, taking into account the mobility of users and changing network dynamics accordingly. The key performance indicators are written into a text file at periodic logging intervals. The data in text file could be used at the end of entire simulation time to evaluate system-level performance and plot figures, using for e.g., MATLAB. The overall methodology is illustrated in figure 3.11.

3.4 Single Cell Monte Carlo Simulations

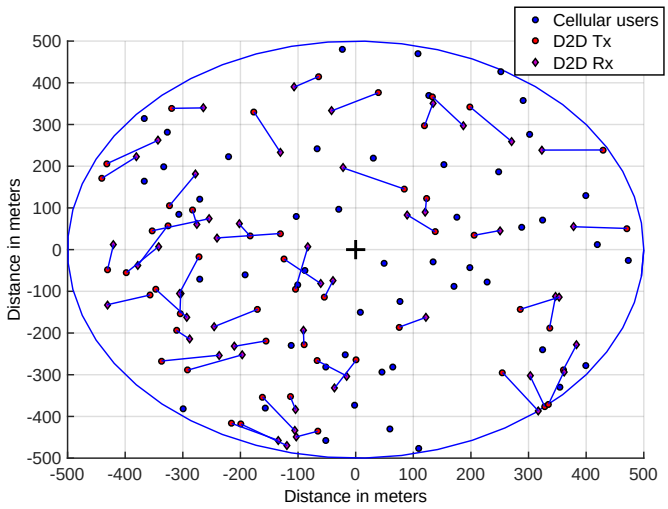

Figure 3.12: An exemplary simulation snapshot

This section deals with simulation methodology used exclusively for D2D solutions presented in Chapter 5. The simulation considers a single cell scenario and uses *Monte-Carlo* method of evaluation [Sid83]. Monte-Carlo simulation is a stochastic method, where a large number of similar random experiments are carried out and obtained results are averaged to deduce the behavior or performance of an algorithm/system considered [Sid83].

The simulation set up consists of a single cell with BS at its center. The cell is equipped with LTE technology. The cellular users and D2D users are dropped as per uniform distribution in the area of cell. The LTE uplink (UL) transmission is considered and parameters are configured accordingly. L2S interfacing follows [JF12] and does a mapping between SINR ranges and channel quality indices. Subsequently, corresponding code rate and bits per resource element is obtained, assisting in deduction of user throughput. The KPIs are calculated per each simulation run (snapshot) and once the

considered number of Monte-Carlo runs are complete, KPIs are averaged or other statistical tools are used to evaluate system-level performance. An exemplary simulation snapshot of this methodology is shown in figure 3.12, where D2D users and cellular users are dropped in a single cell following uniform distribution. The simulation parameters and assumptions considered are tabulated in table 3.3.

Table 3.3: Simulation parameters (D2D)

Parameter	**Assumption**
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
UE transmission power	24 dBm (max), -40 dBm (min)
Noise figure	5 dB (C-mode), 7 dB (D-mode)
Thermal noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B)$
UL power control	Open loop [3GP09]
Path loss (UE-to-UE)	$PL(d) =$ $\begin{cases} PL_{LOS}(d) = 32.45 + 20\ log_{10}(f) + 20\ log_{10}(d/1000) \\ PL_{NLOS}(d) = 16.3 + 45\ log_{10}(f) + 40\ log_{10}(d/1000) \end{cases}$ d in meters, f in MHz
Path loss (UE-to-BS)	$PL_{LOS}(d) = 30.8 + 24.2\ log_{10}(d)$ $PL_{NLOS}(d) = 2.7 + 42.8\ log_{10}(d)$ $prob(d) = min\,(18/d, 1) \times ((1 - e^{-d/63})) + e^{-d/63}$ for 2 GHz and d in meters
Shadowing standard deviation	UE-to-UE : $\begin{cases} LOS : \sigma = 7\ dB,\ log - Rayleigh\ distribution \\ NLOS : \sigma = 7\ dB,\ log - normal\ distribution \end{cases}$ UE-to-BS : $\begin{cases} LOS : \sigma = 4\ dB,\ log - normal\ distribution \\ NLOS : \sigma = 6\ dB,\ log - normal\ distribution \end{cases}$

Chapter 4

Mobility Context Awareness

4.1 Overview

In the contemporary world, cellular broadband services are availed by users while traveling around from one place to another. By the year 2020, the information society is anticipated to include various scenarios such as moving networks, dynamic crowds, traffic jams etc., which involve high mobility of users as well as dynamically changing network load. It becomes an implicit task for the network to provide a superior QoE to users irrespective of high mobility or load. This chapter deals with building context awareness in such situations where the user mobility plays a major role. Mobility context awareness is built in several practical scenarios foreseen in 5G mobile communications namely,

- Data dense moving networks and mobile user groups
  - Cell transition prediction methods are designed to anticipate incoming high load situations.
  - Proactive load balancing procedure is triggered.
  - Elaborated in section 4.2 *Cell Transition Prediction and RRM techniques*
- Vehicular user in urban environment
  - Enhanced route prediction methods are designed to anticipate vehicular user running into coverage holes.
  - Context aware resource allocation is carried out to provide uniform service quality.

  - Elaborated in section 4.3 *Mobility Context Aware Resource Allocation*

- Dynamic mobile user crowds
  - Crowd formation prediction method is presented based on user mobility behavior.
  - Load balancing and activation of small cells are proactively carried out.
  - Elaborated in section 4.4 *Dynamic Crowd Formation Prediction and Resource Management*

- Vehicular traffic jams
  - Technique for traffic status prediction is proposed.
  - Proactive load balancing and activation/deactivation of small cells are triggered correspondingly.
  - Elaborated in section 4.5 *Radio Resource Management in Vehicular Traffic Jams*

Thus, context awareness is built and exploited in these scenarios to design a smart and pro-active radio resource management intending to improve system performance and provide mobility support. The chapter also proposes architecture to support mobility context awareness, outlining required functional decomposition, signaling and interfaces detailed in section 4.6.

4.2 Cell Transition Prediction and RRM techniques

4.2.1 Introduction

In day-to-day scenarios such as public transportation (e.g. buses, trains), groups of mobile users travel together and the data traffic demand is massive due to growing popularity of mobile multimedia services [Net11]. A moving network is a vehicle with advanced communication and networking capabilities, where a mobile router is present within the vehicle managing all user connections in the vehicle, there by reducing required UE transmit powers [MP13a]. In contrast, a moving user cluster is formed by a group of users traveling in conventional vehicle, where connections are individually

managed by the serving base station. When such moving networks or moving user groups travel within service areas of mobile network providers, they lead to dynamically changing and potentially high traffic demands. One of the most significant problems caused by data intensive moving user clusters or moving networks is congestion due to "hotspot" situation in a cell. Increase in dropping of users already connected to the base station and blocking of the new access attempts made by users are the immediate effects in such scenarios. This leads to a poor QoE in the system.

The work presented here emphasizes on the hotspot problem resulting from moving networks/user groups. Techniques based on mobility estimation and user geometry (dB) are presented to predict arrival of moving networks/user groups into a cell. User cell transitions are predicted well in advance and this context is beneficially applied for pro-actively triggering load balancing mechanisms as potential countermeasures for combating congestion.

Figure 4.1 depicts the concept of next cell prediction and initiation of load

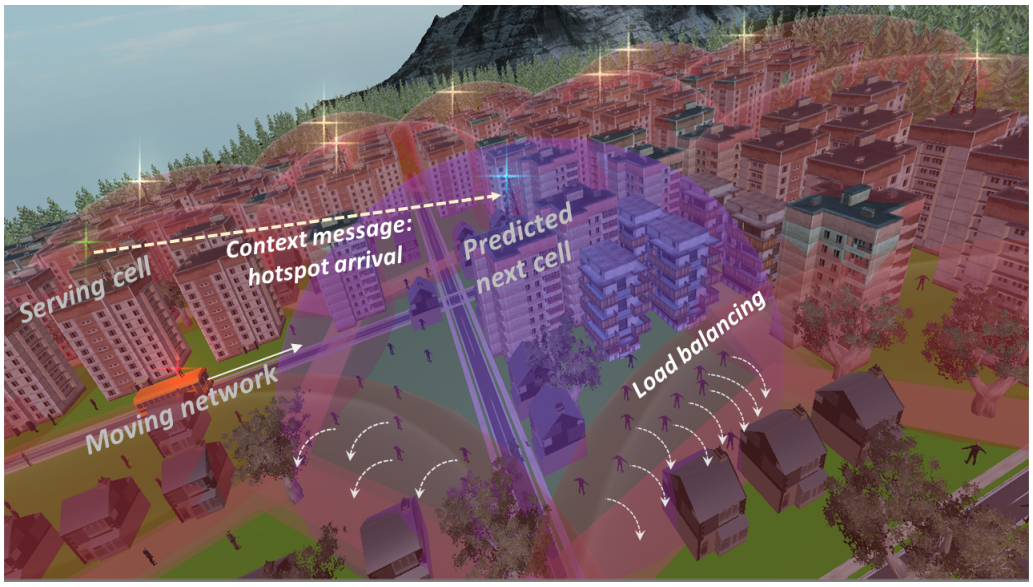

Figure 4.1: Cell transition prediction triggering load balancing

balancing. The next cell for transition of data dense moving network is predicted and context information about arrival of hotspot is conveyed to the predicted cell. Load balancing is subsequently triggered in predicted cell to free-up resources and accommodate moving network eventually, when it transits into the cell.

4.2.2 Diurnal Mobility Model

The mobility of commuters is not purely random but rather direction oriented characterized by the origin and destination of a user [SMBS10]. Such a mobility is termed as Diurnal mobility. Almost all types of public transportation, such as trains, buses, and trams, follow diurnal mobility [SMBS10]. If a random walk mobility model is considered, then a user can travel in all

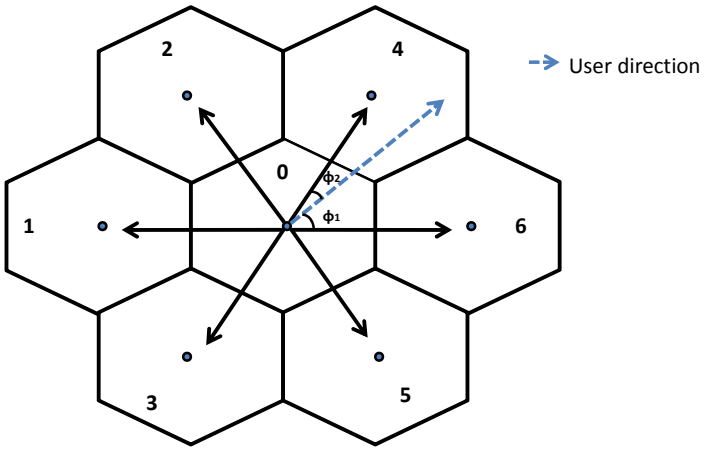

Figure 4.2: Diurnal mobility model

six directions with an equal probability from its current cell. On the contrary, in a diurnal mobility model, the user transition is possible in only two directions (e.g. streets, train tracks) and in other directions probability of transition is zero [SMBS10], as depicted in figure 4.2. Thus, user transition is possible into one of the two adjacent cells from its current cell. If ϕ_1 and ϕ_2 are the angles of user's direction with respect to the closest directions leading towards center of neighboring cells, then probability of user transition into those neighbors are [SMBS10]:

$$p_1 = \left(1 - \frac{\phi_1}{60}\right) \tag{4.1}$$

$$p_2 = \left(1 - \frac{\phi_2}{60}\right). \tag{4.2}$$

4.2.3 User Direction Estimation

This section presents various approaches for sampling user positions and estimating user directions. The model employs a "virtual" circle inscribed in each cell. This "virtual" circle corresponds to either a certain signal strength threshold derived from the radio propagation data or fixed radius in meters

from the cell center. The position estimation of a user is carried out at fix, but velocity-dependent intervals. The user positions within the area of virtual circle are recorded and used to estimate direction. The user angle is calculated at each position as,

$$\phi = \tan^{-}1\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \tag{4.3}$$

where (x_2,y_2) and (x_1,y_1) are present and previous positions, respectively.

Different modes of estimating a user direction are depicted in figure 4.3. In method (1) average of all user angles recorded in the circle is used to estimate user direction, whereas in method (2) only average of user angles recorded within the circular strip is considered. In method (3) exponential moving average (EMA) [Ach00] is performed on the user angles recorded in the circle and in method (4) only the instantaneous angle of user before leaving the "virtual" circle is used. The users may not enter the virtual

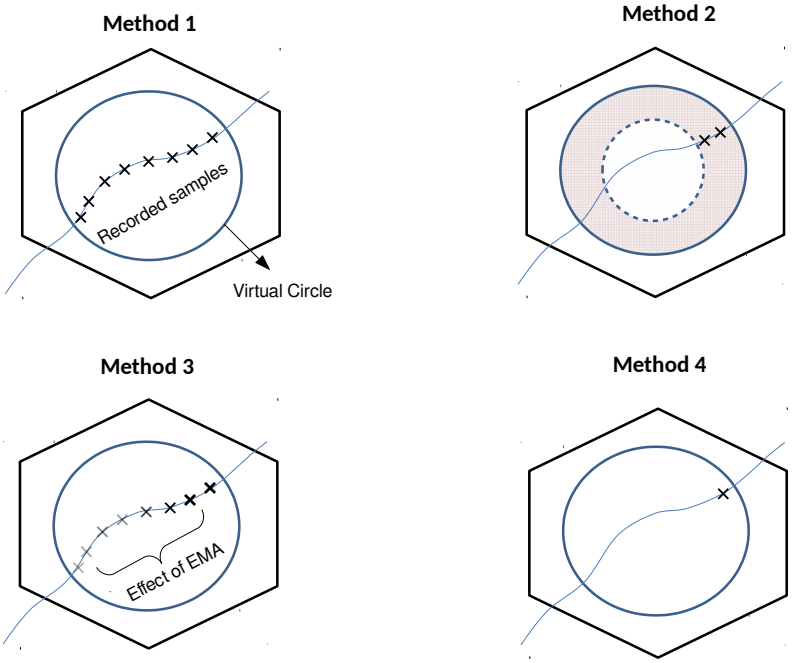

Figure 4.3: Direction estimation methods

circle in some special cases, as shown in figure 4.4. In order to estimate user direction in such cases, the following algorithm is employed:

1. The distance of user from cell center is monitored at fixed logging intervals.

2. IF the distance is greater than radius of circle: GOTO step 3, ELSE: user is in the circle, STOP.

3. IF present distance is greater than previous distance: user will not enter the circle, ELSE: repeat step 2.

4. Instantaneous user angle is used to estimate direction.

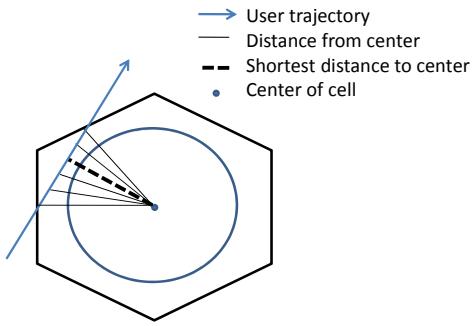

Figure 4.4: User not entering recording circle

4.2.4 Prediction of Cell Transition Based on Angular Deviation

Table 4.1 lists potential next cells of a user in cell 0, as shown in figure 4.2, based on user angle. Similarly, next cells can be deduced for each cell in the cellular layout. The probability of transition to next cells is given by equation 4.1.

Table 4.1: Next cell based on user angle

Average angle ϕ	**Set of Next Cells**	ϕ_1	ϕ_2
0-60	6,4	ϕ	$60 - \phi$
60-120	4,2	$\phi - 60$	$120 - \phi$
120-180	2,1	$\phi - 120$	$180 - \phi$
180-240	1,3	$\phi - 180$	$240 - \phi$
240-300	3,5	$\phi - 240$	$300 - \phi$
300-360	5,6	$\phi - 300$	$360 - \phi$

4.2.5 Comparison of Direction Estimation Methods

All modes mentioned in figure 4.3 yield same prediction results, when users are at a high velocity and follow a straight line motion. When the deviation

of user trajectory is within angular range for same next cells (e.g. $0 - 60°$), all approaches predict same set of next cells, although estimated directions are different. Trajectory A in figure 4.5 illustrates such case and the results at two different velocities (20 and 100 km/h) are listed in Table 4.2. The sampling interval is 3 s and actual next cell is 6.

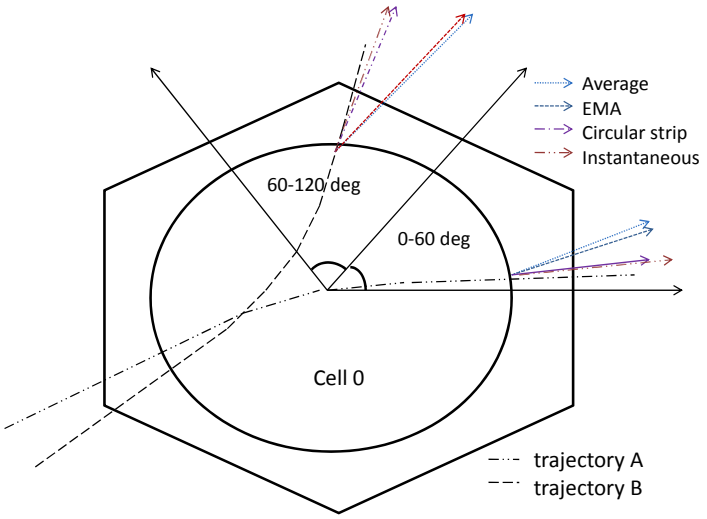

Figure 4.5: Illustration of predicted directions for various user trajectories

In case the user trajectory deviates outside angular range for same next cells, instantaneous and circular strip-based approaches lead to prediction of actual next cells. However, EMA and average lead to prediction of different set of next cells, due to consideration of history values of angles. Trajectory B in figure 4.5 illustrates such scenario and the results at two different velocities (20 and 100 km/h) are listed in Table 4.3 with cell 4 being the actual next cell. Sampling rate is to be adapted in order to yield robust estimation at high velocities, especially for history-based approaches.

4.2.6 Prediction of Cell Transition Based on Distance

The user trajectories indicated in green and blue lines in figure 4.6 have the same estimated user direction (angle). However, the future cell for user transition depends on position of user trajectory at the circumference of the circle. If d_1 and d_2 denote the distances of user from centers of cell 1 and cell 2, at the point of prediction, then transition probabilities to these cells based on distance are,

$$p_1 = 1 - \frac{d_1}{d_1 + d_2}, \tag{4.4}$$

Table 4.2: Comparison of Estimation Methods (Within Range)

Estimation Method	**Velocity (km/h)**	**Angular Deviation**	**Set of Next Cells**	P_1	P_2
Average	20	24.12	6,4	0.598	0.402
EMA		23.39	6,4	0.611	0.389
Circular Strip		18	6,4	0.7	0.3
Instantaneous		18	6,4	0.7	0.3
Average	100	24.6	6,4	0.59	0.41
EMA		21.4	6,4	0.643	0.357
Circular Strip		18	6,4	0.7	0.3
Instantaneous		18	6,4	0.7	0.3

Table 4.3: Comparison of Estimation Methods (Outside Range)

Estimation Method	**Velocity (km/h)**	**Angular Deviation**	**Set of Next Cells**	P_1	P_2
Average	20	55	6,4	0.084	0.916
EMA		56	6,4	0.067	0.933
Circular Strip		63	4,2	0.95	0.05
Instantaneous		65	4,2	0.917	0.083
Average	100	55	6,4	0.084	0.916
EMA		59	6,4	0.017	0.983
Circular Strip		65	4,2	0.917	0.083
Instantaneous		65	4,2	0.917	0.083

$$p_2 = 1 - \frac{d_2}{d_1 + d_2}, \tag{4.5}$$

4.2.7 Combined Approach

For both angle-based and distance-based methods, obtained transition probabilities are different. These probabilities can be combined by taking average of the two. However, distance based approach is observed to have more impact on transition probability than user angle. Thus, in the probability equation distance component is weighed by $\alpha > 1$. The probability equations of the combined approach are,

$$p_1 = 1 - \frac{\phi_1}{60(1+\alpha)} - \frac{\alpha}{1+\alpha} \cdot \frac{d_1}{d_1 + d_2}, \tag{4.6}$$

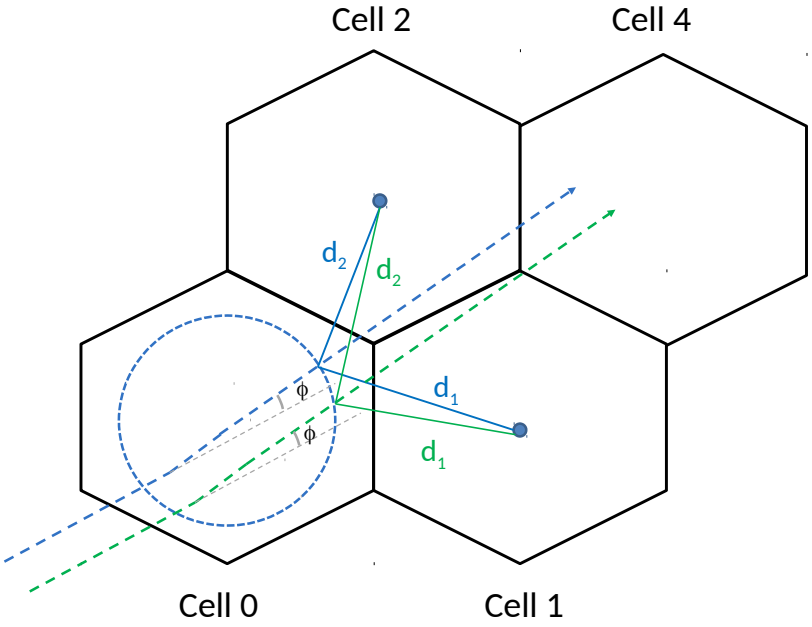

Figure 4.6: Cell transition prediction based on distance

$$p_2 = 1 - \frac{\phi_2}{60(1+\alpha)} - \frac{\alpha}{1+\alpha} \cdot \frac{d_2}{d_1 + d_2}, \tag{4.7}$$

4.2.8 Special Case

In figure 4.7, a user is traveling from cell 3 following diurnal mobility. Typically, a user transition is always into one of the two next cells based on its direction [SMBS10]. In special case, a brief transition (indicated in red) would happen to a third cell before moving to probable next cell. For instance, in cell 6: predicted next cells based on the user angle are cell 15 and cell 18. However, user briefly transits to cell 4 which is untraceable. Nevertheless, by considering three potential next cells instead of two for each user direction range, the brief transition of user to a third cell could be predicted. The probabilities of transition in these cells based on angle [SMBS10] are:

$$p_1 = \left(1 - \frac{\phi_1}{60}\right) \tag{4.8}$$

$$p_2 = \left(1 - \frac{\phi_2}{60}\right) \tag{4.9}$$

$$p_3 = 0 \tag{4.10}$$

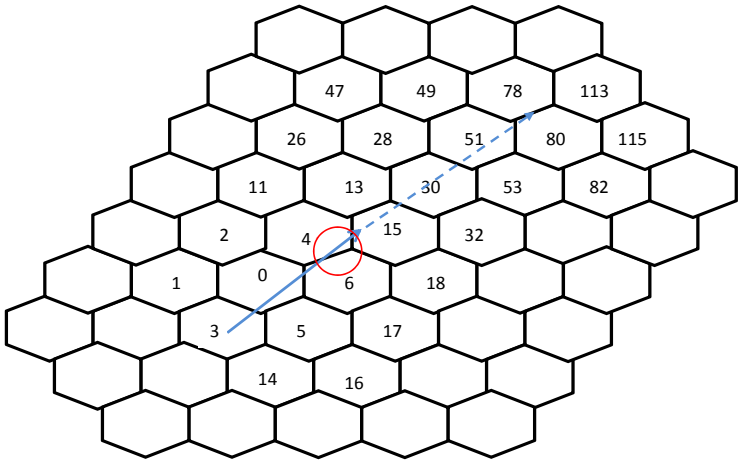

Figure 4.7: User trajectory in special case

If d_1, d_2, and d_3 are distances of user from centers of next cells, at point of prediction, then probabilities based on distance are:

$$p_1 = \frac{2}{3} - \frac{d_1}{d_1 + d_2 + d_3}, \tag{4.11}$$

$$p_2 = \frac{2}{3} - \frac{d_2}{d_1 + d_2 + d_3}, \tag{4.12}$$

$$p_3 = \frac{2}{3} - \frac{d_3}{d_1 + d_2 + d_3}, \tag{4.13}$$

Combined probabilities similar to section 4.2.7 are given by:

$$p_1 = \frac{3 + 2\alpha}{3(1 + \alpha)} - \frac{\phi_1}{60(1 + \alpha)} - \frac{\alpha}{1 + \alpha} \cdot \frac{d_1}{d_1 + d_2 + d_3}, \tag{4.14}$$

$$p_2 = \frac{3 + 2\alpha}{3(1 + \alpha)} - \frac{\phi_2}{60(1 + \alpha)} - \frac{\alpha}{1 + \alpha} \cdot \frac{d_2}{d_1 + d_2 + d_3}, \tag{4.15}$$

$$p_3 = \frac{2\alpha}{3(1 + \alpha)} - \frac{\alpha}{1 + \alpha} \cdot \frac{d_3}{d_1 + d_2 + d_3}, \tag{4.16}$$

4.2.9 Geometry Based Cell Transition Prediction

In this section, a scheme to predict user-cell transition based on the user geometry (dB) is proposed. Typically, in day to day scenarios, vehicular users follow a direction based mobility [SMBS10] as opposed to a random waypoint mobility (RWP) as depicted in figure 4.8.

The proposed scheme considers the geometry of a user with respect to its neighboring cells. Geometry is defined as the average carrier to interference

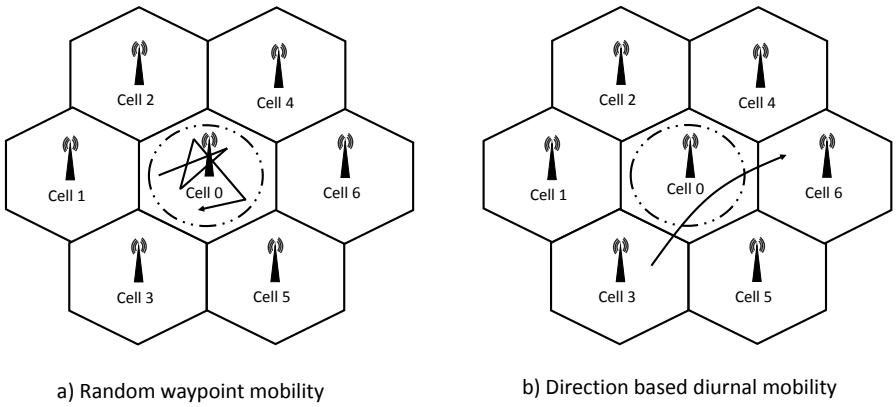

Figure 4.8: RWP mobility and Direction based diurnal mobility

ratio and is given by,

$$Geometry\,(dB) = 10log_{10}\left(\frac{P_k}{\sum_{i=1,i\neq k}^{n} P_i}\right) \tag{4.17}$$

where P_k is the power received from considered base station and P_i are the interference from other base stations.

Figure 4.8.a) depicts the trajectory of a user following random waypoint

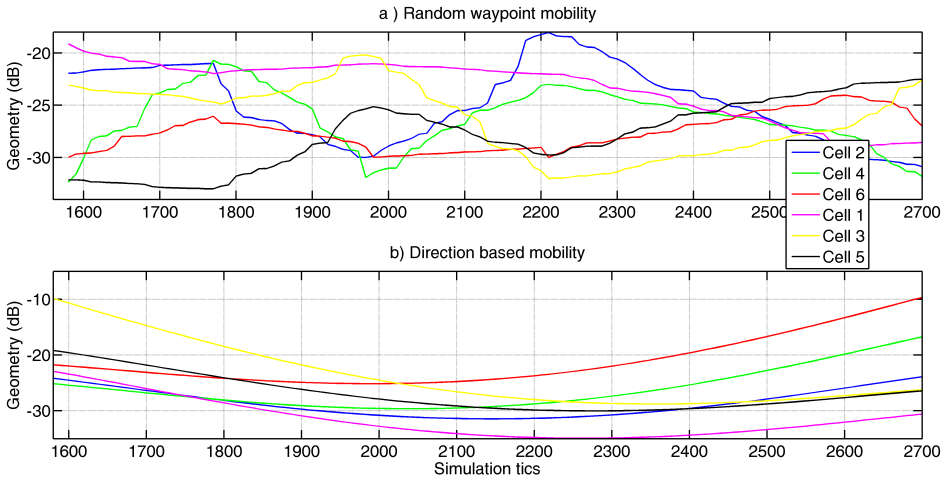

Figure 4.9: Exemplary geometry patterns

mobility. Figure 4.9a). illustrates the recorded geometry pattern of such a user with respect to its neighboring cells. In case of random waypoint mobility, such geometry pattern is obscure and is non-trivial to use it for cell

transition prediction. However, if a user with direction based diurnal mobility (e.g. commuter in public transport) is considered as shown in figure 4.8.b) and its geometry with respect to neighboring cells are recorded, then geometry pattern is as shown in figure 4.9b). A positive gradient is exhibited by the geometry pattern for cells which are being approached (cell 6, cell 4, cell 2), however a negative gradient is exhibited by geometry pattern for cells which the user is moving away from(cell 5, cell 3, cell 1). This behavior can be utilized for user's next cell prediction.

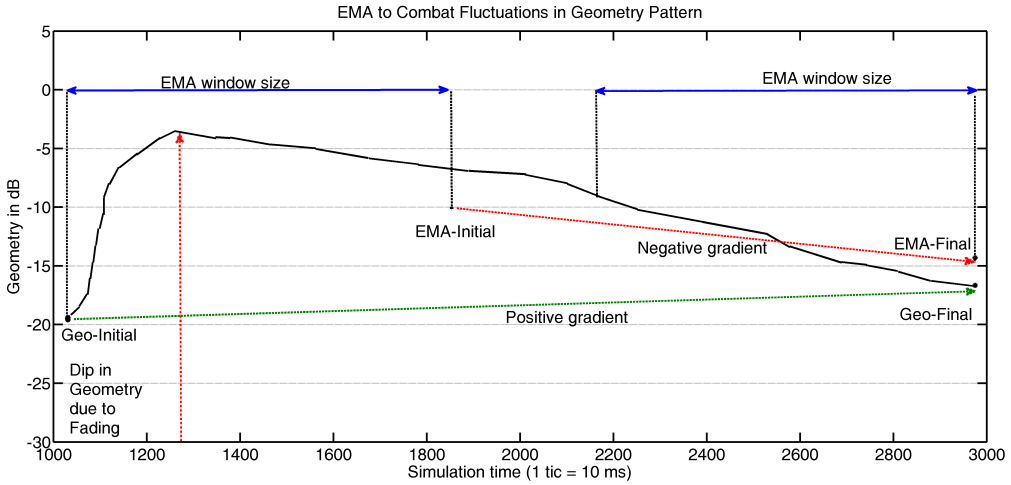

Figure 4.10: EMA to combat fluctuations in geometry pattern

The characteristics of wireless channel such as shadowing, fast-fading etc., highly impact the geometry values. As a result, geometry pattern might experience certain fluctuations. The next cell prediction can be potentially influenced by this. An extreme case is considered in figure 4.10, which depicts the user geometry pattern with respect to a cell from which user is moving away. The initial recordings of geometry pattern suffer due to fading which is reflected as large dip in geometry values. While determining the gradient, if instantaneous initial value (Geo-Initial) and instantaneous final value (Geo-Final) are considered, then a positive gradient is obtained in extreme cases. This leads to wrong inference that the cell is being approached.

In such situations, instead of using instantaneous values exponential moving average filtering (EMA) could be used. EMA applies weighting factors which decrease exponentially, so that the most recent values get higher weightage than the older values [Ach00]. Figure 4.11 depicts the setup used to obtain EMA. Corresponding to two different signal strength thresholds (e.g., 8

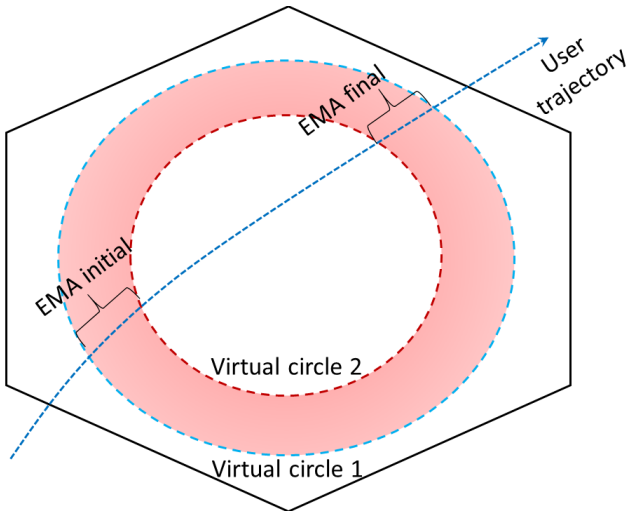

Figure 4.11: Recording of samples to calculate EMA

dB and 5 dB respectively), two virtual circles inscribed in a cell,. User geometry samples recorded in the region between these circles are used to calculate EMA. If EMA is used (EMA-Initial, EMA-Final) instead of instantaneous values as in figure 4.10, a negative gradient is obtained as desired. This infers that user is moving away from the cell. Even though aforementioned scenario is rare, using EMA as a precaution is recommended. In principle, the number of samples (N_{EMA}) used to obtain EMA is given by,

$$N_{EMA} = \frac{2 \times d_{corr}}{T_s \times V} \tag{4.18}$$

where d_{corr} is the correlation distance of underlying shadowing map, T_s is the rate at which samples are recorded and V is the user velocity.

The set of potential next cells are determined dynamically based on the geometry patterns which have positive gradient. The probability of transition into these next cells based on geometry values are given by:

$$p_1 = \frac{EMA_{geo1}}{EMA_{geo1} + EMA_{geo2} + EMA_{geo3}} \tag{4.19}$$

$$p_2 = \frac{EMA_{geo2}}{EMA_{geo1} + EMA_{geo2} + EMA_{geo3}} \tag{4.20}$$

$$p_3 = \frac{EMA_{geo3}}{EMA_{geo1} + EMA_{geo2} + EMA_{geo3}} \tag{4.21}$$

where EMA_{geo1}, EMA_{geo2} and EMA_{geo3} are the EMA of geometry values of potential next cells. These are obtained by performing EMA [Ach00] on geometry values recorded in the region between two virtual circles, before user leaves present cell (same as EMA-Final).

4.2.10 Evaluation

A multi-cell scenario is simulated using a LTE system-level simulator as illustrated in figure 4.7 with a base station at center of each cell. The cluster of 60 users moving together at high velocity (120 km/h) constitutes a mobile hotspot. All the users in the system are assumed to have a full buffer traffic [3GP10]. The evaluation methodology follows [3GP10]. Figure 4.7 shows the trajectory followed by a moving user group. Table 4.4 summarizes simulation parameters.

Table 4.4: Simulation Parameters (Mobile hotspot)

Parameter	**Assumption**
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
Total transmit power	40 W/250 mW ($s2s = 500\text{m}/200\text{m}$)
Control channel overhead	12%
Shadowing	log-normal Standard deviation: 8 dB Correlation distance: 50 m
Fast fading	2-tap Rayleigh fading channel
Noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B) + 7$
Background users per cell	30
Hotspot users	60 at 120 km/h

Table 4.5 shows results of next cell prediction for user trajectory in figure 4.7, from cell 3 to cell 15.

The incapability of the angle based scheme to trace brief transitions into a third cell is demonstrated in table 4.5. In the third cell transition probability is always zero. Further, even with high values of α, prediction accuracy of combined approach suffers. This is due to the limitation of angle based approach, that is reflected in combined approach as well. Figure 4.12 describes the affect of α on prediction of next cell during brief transition (special case). With increase in value of α, prediction in actual next cell by combined approach improves but does not exceed the probability of prediction in other

Table 4.5: Next Cell Prediction Results

Present Cell	**Next Cells**	P_{angle}	$P_{distance}$	$P_{combined}$	$P_{geometry}$
Cell 3	Cell 5	0.25	0.346	0.327	0.084
	Cell 0	0.75	0.410	0.478	0.847
	Cell 1	0	0.243	0.194	0.067
Cell 0	Cell 6	0.25	0.447	0.408	0.798
	Cell 4	0.75	0.353	0.433	0.166
	Cell 2	0	0.198	0.158	0.036
Cell 6	Cell 18	0.25	0.202	0.211	0.001
	Cell 15	0.75	0.332	0.416	0.004
	Cell 4	0	0.465	0.372	0.995
Cell 4	Cell 15	0.25	0.468	0.424	0.978
	Cell 13	0.75	0.311	0.399	0.002
	Cell 11	0	0.219	0.175	0.020

cell (e.g. next cell 2 in figure 4.12).

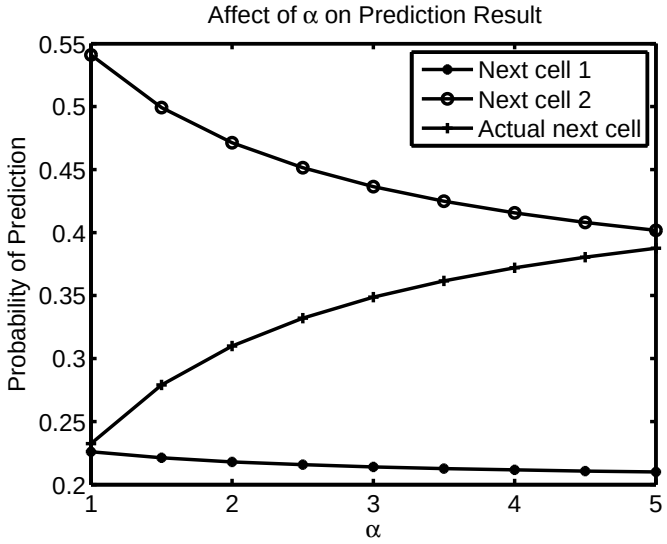

Figure 4.12: Affect of α on Prediction Probability

The simulations are carried out for site-to-site distances (s2s) of 500m and 200m (dense deployment), respectively. The same distance is traversed by the moving user group in both cases. Load balancing is triggered proactively in predicted next cell and reductions in blocked access attempts, dropped connections and blocked access attempts are observed, thereby improving QoE of users.

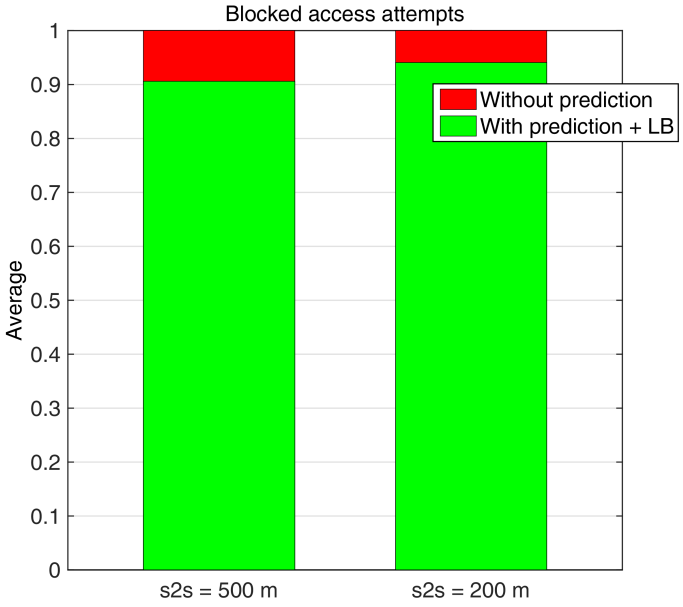

Figure 4.13: Blocked access attempts(distance based)

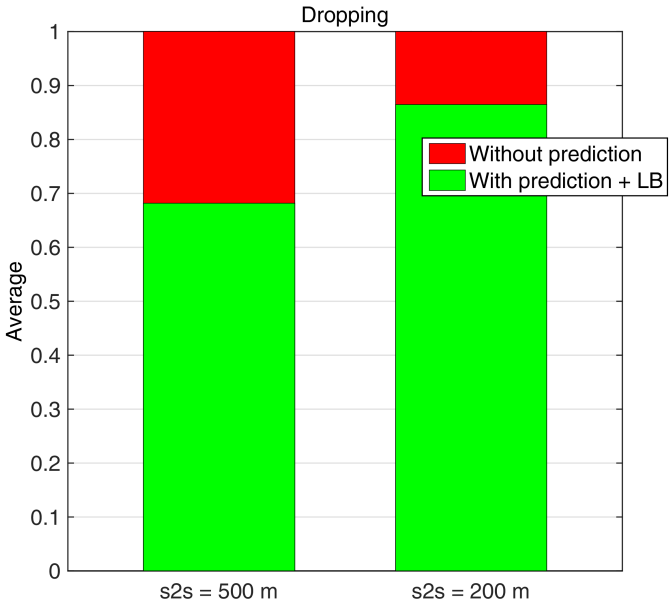

Figure 4.14: Dropped connections(distance based)

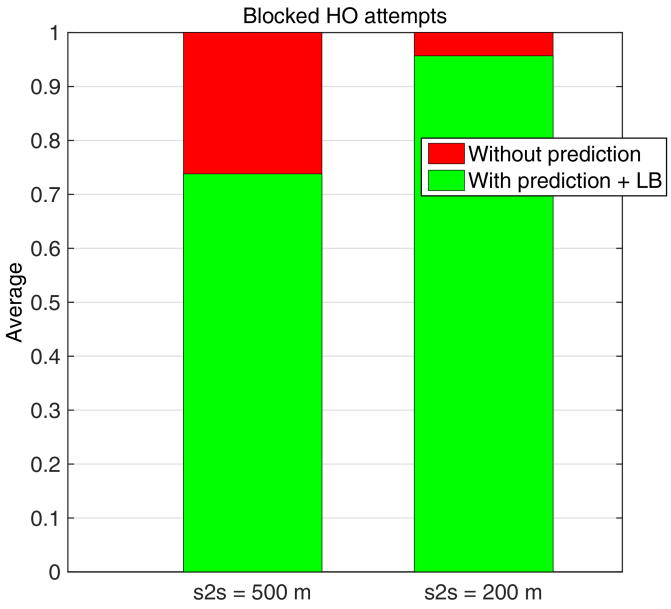

Figure 4.15: Blocked HO attempts(distance based)

From figures 4.13, 4.14, and 4.15, it could be seen that blocked access attempts, dropped connections and blocked HO are reduced by 10%, 32% and 26%, respectively for s2s = 500 m. For s2s = 200 m, blocked access attempts, dropped connections and blocked HO are reduced by 6%, 13% and 4%, respectively. These results are obtained when LB is triggered based on distance (GPS) based prediction.

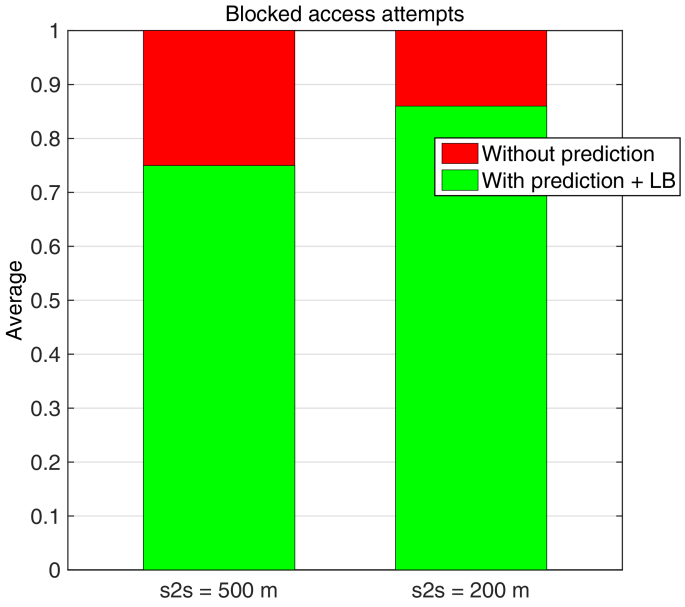

Figure 4.16: Blocked access attempts

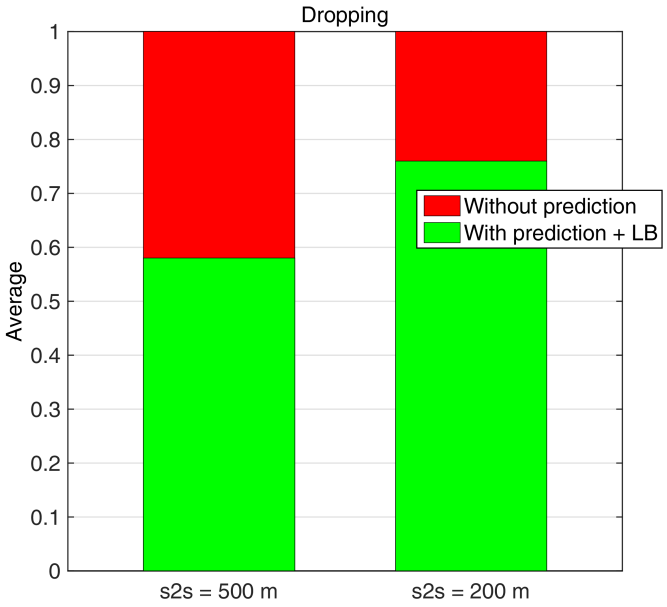

Figure 4.17: Dropped connections

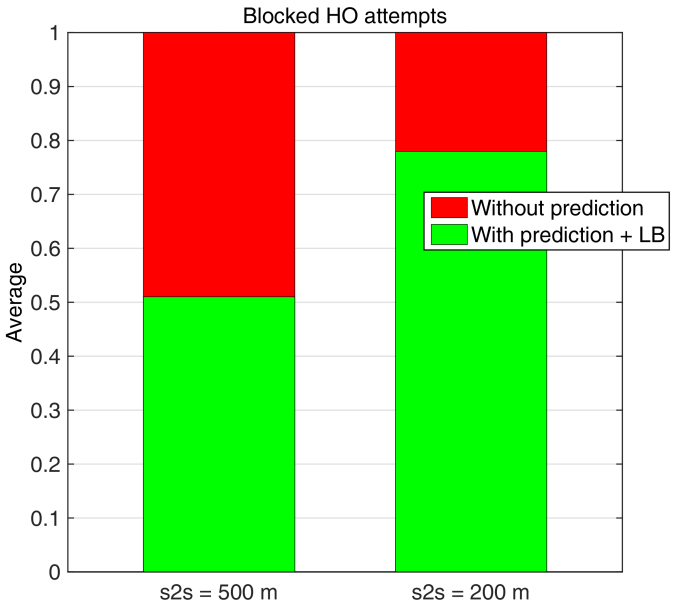

Figure 4.18: Blocked HO attempts

Figures 4.16, 4.17, and 4.18, show that blocked access attempts, dropped connections and blocked HO are reduced by 25%, 42% and 49%, respectively for s2s = 500 m. For s2s = 200 m, blocked access attempts, dropped connections and blocked HO are reduced by 15%, 25% and 22%, respectively. These results are obtained when LB is triggered based on geometry (dB) based prediction.

The results of this work have been published in **[KKSS13b]** ***and*** **[KMS17a]**.

4.3 Mobility Context Aware Resource Allocation

In the previous section, context awareness was built based on user attributes (e.g., angle, distances, geometry) to predict the next cell for user transition. The QoE of users was improved by triggering pro-active RRM and combating congestion issues. This section deals with utilization of context information (e.g., user origin, destination) to enhance mobility prediction. Subsequently, a context aware resource allocation is performed to sustain non-real time streaming services in an urban environment with coverage holes.

4.3.1 Introduction

In real world scenarios, a large proportion of users exhibit similar mobility patterns on daily basis (e.g., office goers, public transport). They tend to regularly traverse a limited set of trajectories, comprising of specific landmarks. Such a mobility can be referred to as Diurnal mobility, which constitutes a major portion of mobile users. Further, in day to day life, there are several instances of users running into a coverage hole (e.g., tunnel), where the user throughput will be nil. By anticipating the encounter of such a coverage hole in the near future, it is possible to allocate more resources suitably and buffer the data before running into a coverage hole. In the coverage hole, this buffered data can be used to support streaming/full buffer service. This will enable a uniform service experience for the user, even in deep shadow regions or coverage holes.

In this work, the information arising from aforementioned mobility (e.g., origin, landmarks, destination) is used to enhance the accuracy of mobility prediction and the context information gained from route prediction assists in anticipating a user running into a coverage hole. Context aware resource allocation is triggered accordingly. Thus, substantial improvements can be observed in user throughput and the service interruption time is reduced, sustaining streaming/full buffer service even in case of coverage holes.

4.3.2 System Model

Majority of the user movements are direction oriented, depending on the origin and destination of a user. As an example, consider figure 4.19, where a user traverses through following 5 landmarks on a daily basis. The user starts from his home in the morning, travels to his office, then to a restaurant at afternoon, then to the library and to the gym in evening, before returning home. More than one path might exist between two landmarks. Further, in each path the sequence of landmarks encountered by the user might be different too. Nevertheless, the user's trajectories are confined to a specific set of landmarks. By monitoring the users for a specific period of time (e.g., several business days), they can be classified as diurnal mobile users.

Figure 4.20 depicts the system model used here, referred to as the Madrid grid [MP13c]. Elaborate details of the system model could be found in section 3.3. Contrary to conventional hexagonal cell layout, the considered realistic model has 12 micro base stations positioned at various sites. There are 25 crossroads present in the considered layout with 7 landmarks at dif-

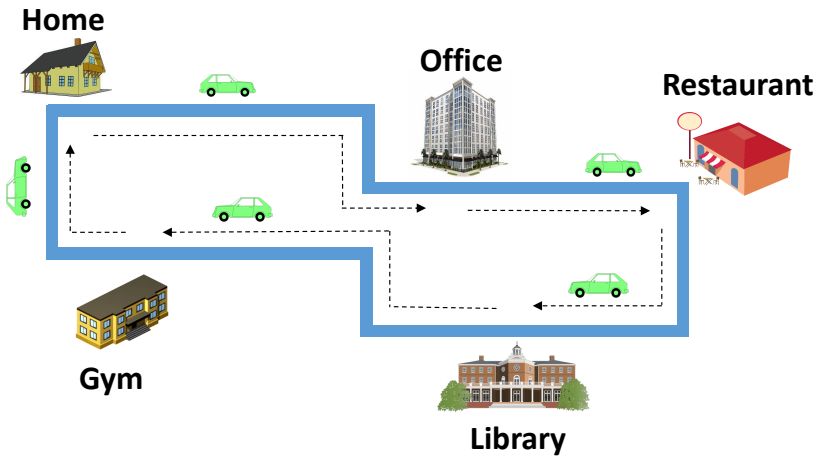

Figure 4.19: Diurnal mobility

ferent crossroads namely: home, university, gym, office, and corner points (North East, South East and South West) as shown in figure 4.20.

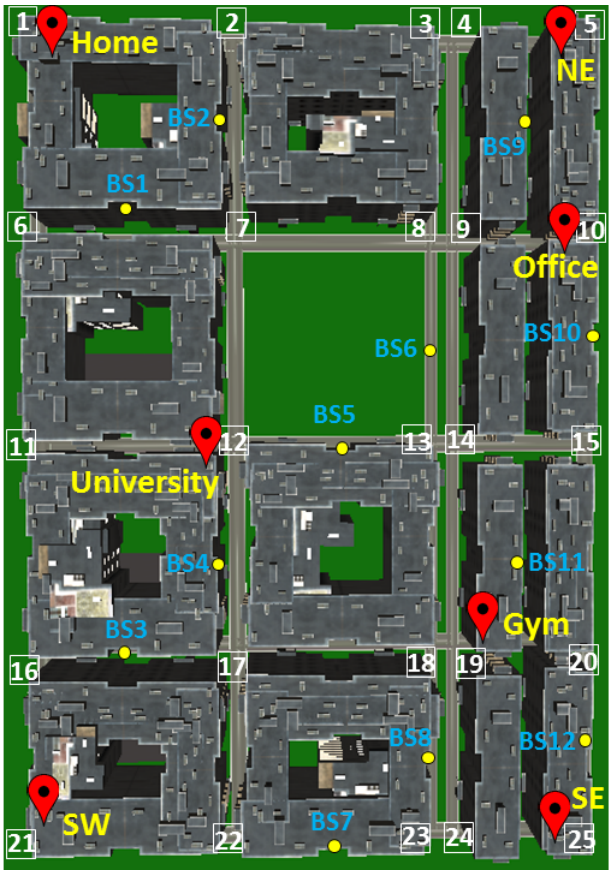

Figure 4.20: System model

The broader view of the considered urban scenario is as in figure 4.21.

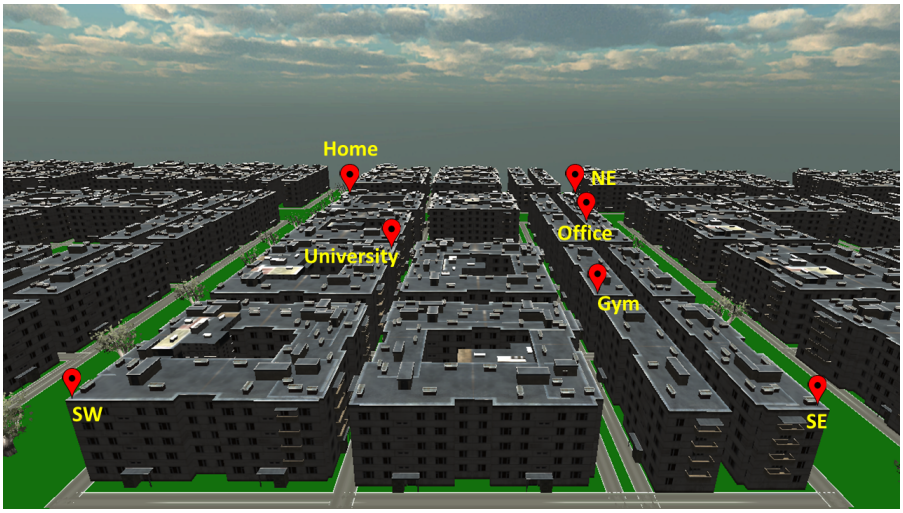

Figure 4.21: Front view

Table 4.6 depicts the 10 different paths considered diurnal mobile user can take with respective probabilities. Each path is defined by sequence of crossroads, including the landmarks.

4.3.3 Next Cell Prediction

With each path, there exists a unique sequence of base stations to which the user was connected. Such sequences of connected base stations analogous to paths in table 4.6 are tabulated in the table 4.7.

A user can be monitored for a specific duration, to obtain statistics on connected base stations and their sequence. Subsequently, a simple first order Markov model [BK10] can be used to derive the probability of transition into a next cell, given by,

$$P\left(BS_n \rightarrow BS_{n+1}\right) = \frac{N\left(BS_n \rightarrow BS_{n+1}\right)}{N\left(BS_n\right)} \tag{4.22}$$

Where, $BS_n \rightarrow \mathrm{BS}_{n+1}$ indicates the user transition from $cell_n$ to $cell_{n+1}$, $N\left(.\right)$ indicates the number of times the user was found in a specific cell $(cell_n)$.

Assuming a diurnal mobility, specific landmarks are traversed by the user on a daily basis. These landmarks are associated with specific cells as shown in table 4.8.

The Markov based prediction can be extended by using the information about origin of a user (user originating from a specific landmark) as,

Table 4.6: Paths taken by the mobile user

Nr.	Sequence of crossroads	%age
1	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 10$ $\rightarrow 15 \rightarrow 20 \rightarrow 25 \rightarrow 24 \rightarrow 19 \rightarrow 18 \rightarrow 23$ $\rightarrow 22 \rightarrow 21 \rightarrow 16 \rightarrow 17 \rightarrow 12 \rightarrow 11 \rightarrow 6 \rightarrow 1$	10%
2	$1 \rightarrow 6 \rightarrow 11 \rightarrow 12 \rightarrow 7 \rightarrow 8 \rightarrow 9$ $\rightarrow 4 \rightarrow 5 \rightarrow 10 \rightarrow 15 \rightarrow 14 \rightarrow 19 \rightarrow 20 \rightarrow 25$ $\rightarrow 24 \rightarrow 23 \rightarrow 22 \rightarrow 21 \rightarrow 16 \rightarrow 11 \rightarrow 6 \rightarrow 1$	8%
3	$1 \rightarrow 2 \rightarrow 7 \rightarrow 12 \rightarrow 11 \rightarrow 16$ $\rightarrow 21 \rightarrow 22 \rightarrow 23 \rightarrow 18 \rightarrow 19 \rightarrow 24 \rightarrow 25$ $\rightarrow 20 \rightarrow 15 \rightarrow 10 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$	7%
4	$1 \rightarrow 6 \rightarrow 7 \rightarrow 12 \rightarrow 13 \rightarrow 14 \rightarrow 19 \rightarrow 18$ $\rightarrow 17 \rightarrow 16 \rightarrow 21 \rightarrow 22 \rightarrow 23 \rightarrow 18 \rightarrow 19 \rightarrow 24$ $\rightarrow 25 \rightarrow 20 \rightarrow 15 \rightarrow 10 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$	5%
5	$1 \rightarrow 6 \rightarrow 11 \rightarrow 16 \rightarrow 21 \rightarrow 22 \rightarrow 23 \rightarrow 18$ $\rightarrow 19 \rightarrow 24 \rightarrow 25 \rightarrow 20 \rightarrow 15 \rightarrow 10 \rightarrow 5 \rightarrow 4$ $\rightarrow 3 \rightarrow 8 \rightarrow 13 \rightarrow 12 \rightarrow 11 \rightarrow 6 \rightarrow 1$	5%
6	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 10 \rightarrow 15$ $\rightarrow 20 \rightarrow 25 \rightarrow 24 \rightarrow 19 \rightarrow 18 \rightarrow 17 \rightarrow 16$ $\rightarrow 21 \rightarrow 22 \rightarrow 17 \rightarrow 12 \rightarrow 7 \rightarrow 2 \rightarrow 1$	30%
7	$1 \rightarrow 6 \rightarrow 7 \rightarrow 12 \rightarrow 17 \rightarrow 16$ $\rightarrow 21 \rightarrow 22 \rightarrow 17 \rightarrow 18 \rightarrow 19 \rightarrow 24 \rightarrow 25$ $\rightarrow 20 \rightarrow 15 \rightarrow 10 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$	15%
8	$1 \rightarrow 2 \rightarrow 3 \rightarrow 8 \rightarrow 7 \rightarrow 12 \rightarrow 17 \rightarrow 18$ $\rightarrow 19 \rightarrow 24 \rightarrow 23 \rightarrow 22 \rightarrow 21 \rightarrow 16 \rightarrow 17 \rightarrow 23 \rightarrow 24$ $\rightarrow 25 \rightarrow 20 \rightarrow 15 \rightarrow 10 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$	10%
9	$1 \rightarrow 6 \rightarrow 11 \rightarrow 16 \rightarrow 21 \rightarrow 22 \rightarrow 17$ $\rightarrow 18 \rightarrow 19 \rightarrow 24 \rightarrow 25 \rightarrow 20 \rightarrow 15 \rightarrow 10$ $\rightarrow 9 \rightarrow 14 \rightarrow 13 \rightarrow 12 \rightarrow 7 \rightarrow 6 \rightarrow 1$	5%
10	$1 \rightarrow 6 \rightarrow 11 \rightarrow 16 \rightarrow 17 \rightarrow 12 \rightarrow 7 \rightarrow 8 \rightarrow 9$ $\rightarrow 10 \rightarrow 5 \rightarrow 4 \rightarrow 9 \rightarrow 14 \rightarrow 19 \rightarrow 24 \rightarrow 25 \rightarrow 20$ $\rightarrow 24 \rightarrow 23 \rightarrow 18 \rightarrow 17 \rightarrow 22 \rightarrow 21 \rightarrow 16 \rightarrow 11 \rightarrow 6 \rightarrow 1$	5%

Table 4.7: Sequence of base stations analogous to path

Path Nr.	Sequence of base stations	%age
1	$\text{BS}_2 \rightarrow BS_9 \rightarrow BS_{10} \rightarrow$ $\text{BS}_{12} \rightarrow BS_8 \rightarrow BS_7$ $\rightarrow BS_3 \rightarrow BS_4 \rightarrow BS_1$	10%
2	$\text{BS}_1 \rightarrow BS_4 \rightarrow BS_1 \rightarrow BS_5 \rightarrow BS_9$ $\rightarrow BS_{10} \rightarrow BS_5 \rightarrow BS_{11} \rightarrow BS_{12}$ $\rightarrow BS_8 \rightarrow BS_7 \rightarrow BS_3 \rightarrow BS_1$	8%
3	$\text{BS}_2 \rightarrow BS_6 \rightarrow BS_4 \rightarrow BS_3$ $\rightarrow BS_7 \rightarrow BS_8 \rightarrow BS_{12}$ $\rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_2$	7%
4	$\text{BS}_1 \rightarrow BS_6 \rightarrow BS_5 \rightarrow BS_{11}$ $\rightarrow BS_8 \rightarrow BS_4 \rightarrow BS_3 \rightarrow BS_7$ $\rightarrow BS_8 \rightarrow BS_{12} \rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_2$	5%
5	$\text{BS}_1 \rightarrow BS_3 \rightarrow BS_7 \rightarrow BS_8$ $\rightarrow BS_{12} \rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_5$ $\rightarrow BS_6 \rightarrow BS_4 \rightarrow BS_1$	5%
6	$\text{BS}_2 \rightarrow BS_9 \rightarrow BS_{10} \rightarrow BS_{12}$ $\rightarrow BS_8 \rightarrow BS_4 \rightarrow BS_3 \rightarrow BS_7$ $\rightarrow BS_3 \rightarrow BS_4 \rightarrow BS_1 \rightarrow BS_2$	30%
7	$\text{BS}_1 \rightarrow BS_6 \rightarrow BS_4 \rightarrow BS_3$ $\rightarrow BS_7 \rightarrow BS_3 \rightarrow BS_8 \rightarrow BS_{12}$ $\rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_2$	15%
8	$\text{BS}_2 \rightarrow BS_9 \rightarrow BS_5 \rightarrow BS_2 \rightarrow BS_6$ $\rightarrow BS_4 \rightarrow BS_8 \rightarrow BS_7 \rightarrow BS_3 \rightarrow BS_8$ $\rightarrow BS_{12} \rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_2$	10%
9	$\text{BS}_1 \rightarrow BS_3 \rightarrow BS_7 \rightarrow BS_3$ $\rightarrow BS_8 \rightarrow BS_{12} \rightarrow BS_{10} \rightarrow BS_9$ $\rightarrow BS_5 \rightarrow BS_6 \rightarrow BS_1$	5%
10	$\text{BS}_1 \rightarrow BS_3 \rightarrow BS_4 \rightarrow BS_1 \rightarrow BS_5$ $\rightarrow BS_{10} \rightarrow BS_9 \rightarrow BS_5 \rightarrow BS_{11} \rightarrow BS_8$ $\rightarrow BS_{12} \rightarrow BS_8 \rightarrow BS_4 \rightarrow BS_7 \rightarrow BS_3 \rightarrow BS_1$	5%

Table 4.8: Mapping landmarks to base stations

Landmark	Home	Office	University	Gym	NE	SW	SE
Crossroad	1	10	12	19	5	21	25
Base station	BS1	BS10	BS4	BS8	BS9	BS3	BS12

$$P\left(BS_n \rightarrow BS_{n+1}/O\right) = \frac{N\left(BS_n \rightarrow BS_{n+1}/O\right)}{N\left(BS_n/O\right)} \tag{4.23}$$

Where O indicates origin. This case considers only statistics of the user when originated from a specific landmark (in $cell_n$ (C_n)), as opposed to considering the complete transition history.

As an example, consider figure 4.22 a), where user has arrived at cell 2 from office (cell 3). When cell transition is predicted at cell 2 using first order Markov model, all the user statistics are considered. However, this case considers only the statistical history of user in cell 2 when he had arrived from cell 3.

Similarly, the Markov based prediction can be further extended by using information about both the origin and destination of a user (user originating from a certain landmark and traveling to a specific landmark) as,

$$P\left(BS_n \rightarrow BS_{n+1}/O\&D\right) = \frac{N\left(BS_n \rightarrow BS_{n+1}/O\&D\right)}{N\left(BS_n/O\&D\right)} \tag{4.24}$$

Where O indicates origin, D indicates destination. Only the statistics of the user when originated from a specific landmark and traveling to a certain landmark is considered in this case, instead of considering the entire transition history. For instance, figure 4.22 b), shows a user who has arrived at cell 2 from office (cell 3) and will be further traveling to gym (cell 6). When cell transition probabilities are deduced at cell 2 using simple Markov case, complete user statistics are considered. Where as, this case considers only the statistical history of user in cell 2 when he had arrived from cell 3 and traveled further to cell 6. With advent of autonomous cars and technological advancements in vehicular positioning, acquiring information about user destination will be more feasible. With appropriate interface, such information can be imported by cellular networks and can be utilized to improve mobility prediction.

4.3.4 Route Prediction

Here, more localized mobility prediction is carried out and immediate future route of user is predicted. Considered routes and crossroads are shown in figure 4.23. Assume a user traveling in road R_n, who can traverse further roads R_{n+1} or R_{n+2} or R_{n+3}, at crossroad C_n. In this work, at each crossroad C_n, the next road R_n to which user would go into is predicted.

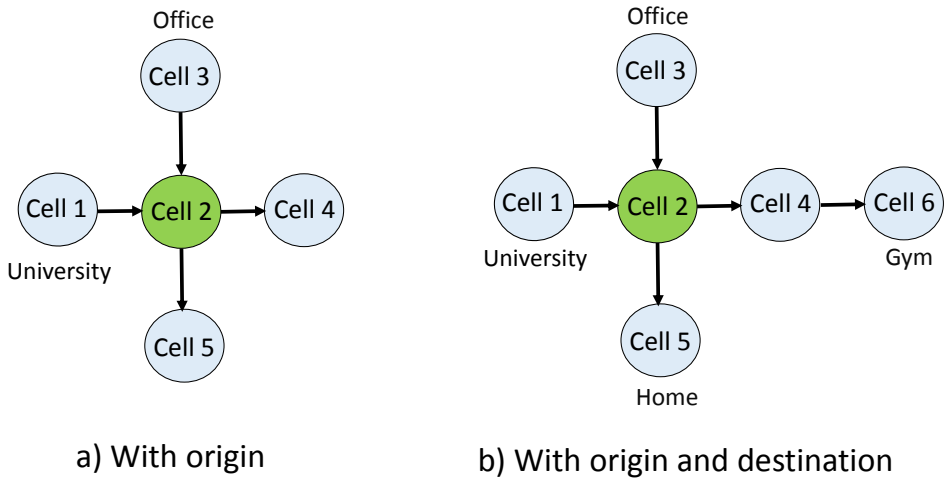

Figure 4.22: Markov chain with additional information

Based on statistics of different routes (sequence of crossroads) a user traverses and using first order Markov model [BK10], the route transition probability is obtained as,

$$P\left(R_n \rightarrow R_{n+1}\right) = \frac{N\left(R_n \rightarrow R_{n+1}\right)}{N\left(R_n\right)} \tag{4.25}$$

Where, $R_n \rightarrow R_{n+1}$ indicates transition from $road_n$ to $road_{n+1}$, $N\left(.\right)$ indicates the number of times user was found in a specific crossroad ($road_n$).
The Markov based route prediction can be extended using information about the origin of a user (user originating from a specific landmark), given as,

$$P\left(R_n \rightarrow R_{n+1}/O\right) = \frac{N\left(R_n \rightarrow R_{n+1}/O\right)}{N\left(R_n/O\right)} \tag{4.26}$$

Where O denotes a specific origin. Here, prediction considers the statistics of user only when it originated from a specific landmark.
Further, Markov based route prediction can be further extended using information about the origin as well as the destination of a user, as,

$$P\left(R_n \rightarrow R_{n+1}/O\&D\right) = \frac{N\left(R_n \rightarrow R_{n+1}/O\&D\right)}{N\left(R_n/O\&D\right)} \tag{4.27}$$

Where O denotes a specific origin and D denotes a specific destination.

In this case, the prediction considers only statistics of the user when

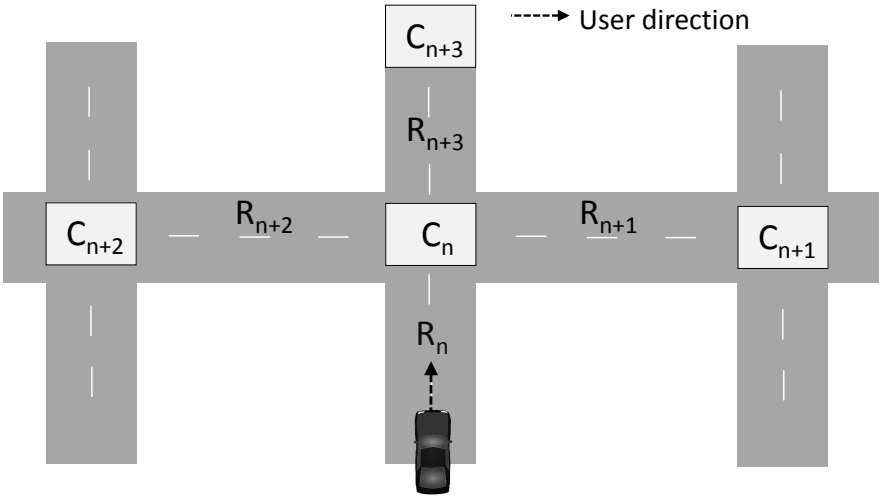

Figure 4.23: Route map

it originated from a particular landmark and further traveled to a specific destination. Here, the assumption is that user has a-priori knowledge of origin and destination (one specific landmark out of all considered landmarks).

4.3.5 Context Aware Resource Allocation

There are several instances in the day-to-day life, when the signal reception becomes very poor due to coverage holes (e.g., tunnels). In such scenarios, SINR of the user is very low and the throughput of user drops to zero. This becomes problematic for certain services such as streaming and full buffer services. Figure 4.24 depicts a scenario where a coverage hole (tunnel) is located in the road. As the user travels through coverage hole, his SINR and throughput are diminished, as demonstrated by graphs (a) and (b) respectively. The probabilities (P_A,P_B and P_C) with which a user might travel into next route is provided by route prediction. Context aware resource allocation is triggered proactively, when the route prediction anticipates that the user is most likely to take the road with a coverage hole in it. Subsequently, more PRBs are allocated to such a user and the buffering of excess data is done as depicted in graph (c) of figure 4.24. As soon as the user is in coverage hole, this buffered data is used (depicted in graph (d)). Thus a uniform service quality for streaming services could be maintained by this procedure. The functioning of a context aware resource allocation scheme is illustrated in figure 4.25. In order to reduce unnecessary buffering of data due to false prediction (false alarm), the actual user traversal in a road with coverage is monitored and confirmed. If the user does not transit to such a road, then

the buffering action is stopped and a reallocation of resources is done.

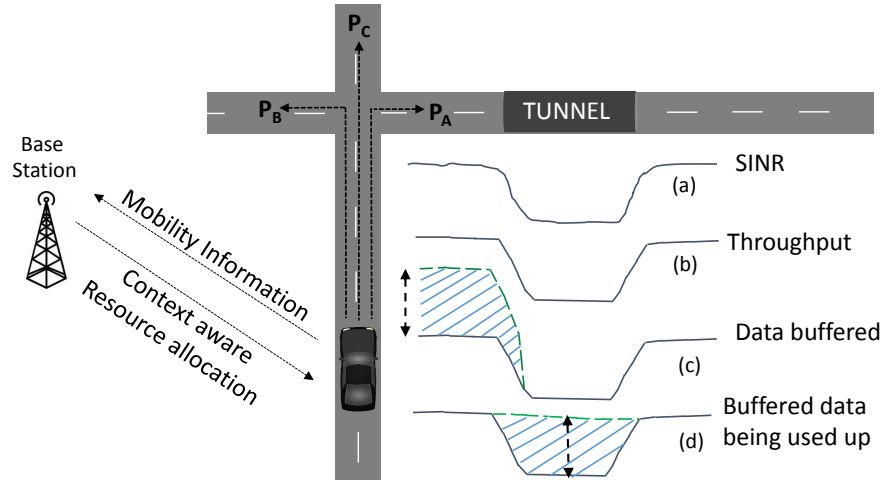

Figure 4.24: Context aware resource allocation

The downlink data rate depends mainly on the number of PRBs allocated to a user and corresponding SINR. Further, the usage of MIMO influences the data rate as well. Consider a streaming service requiring 2 Mbps date rate. Based on SINR to throughput mapping [JF12], throughput of a user is given as,

$$Throughput = \sum_{i} \frac{12 \times 7}{0.5} \times BRE_i \times \#PRB \; Kbps \tag{4.28}$$

Where, i indicates the number of MIMO layers, BRE_i indicates the number of bits per resource element and $\#PRB$ represents number of physical resource blocks allocated to the user.

Assuming a SISO ($i = 1$) system and $\#PRB = 1$, maximum throughput is possible if $SINR \geq 19.5$ dB for the corresponding PRB [JF12] given by,

$$Throughput_{max} = 12 \times 14 \times 5.5546 \; Kbps \; = 933 \; Kbps$$

With $i = 2$, $\#PRB = 1$ and $SINR \geq 19.5$ dB, maximum possible throughput is,

$$Throughput_{max} = 2 \times 12 \times 14 \times 5.5546 \; Kbps \; = 1.866 \; Kbps \; \approx 2 \; Mbps$$

Hence, available number of MIMO layers has clear impact on achieved data rate with fixed number of allocated PRBs at given SINR. Further, with differing SINR due to channel variations, the number of PRBs required to

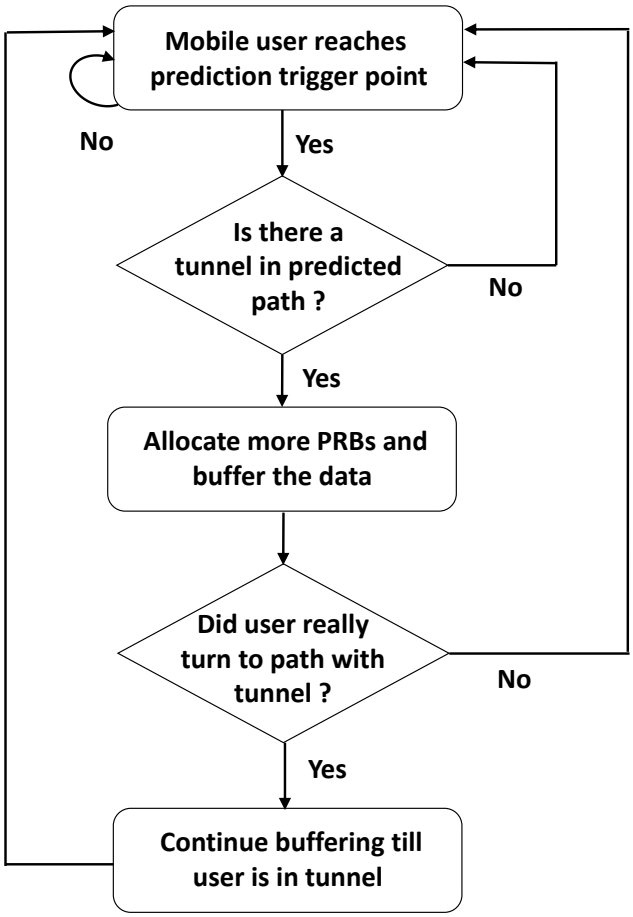

Figure 4.25: Context aware resource allocation algorithm

achieve 2 Mbps of data rate vary too. For example, at $SINR = 6$ dB and $i = 1$, $\#PRB$ required is 8 to obtain 2 Mbps data rate. Whereas, at $SINR = 12$ dB and $i = 1$, $\#PRB$ required is 4 to achieve the same data rate. For the same two cases of $SINR = 6$ dB and 12 dB, it would require $\#PRB$ of 4 and 2 respectively to achieve 2 Mbps data rate with $i = 2$. Thus, the resource allocation scheme needs to allocate $\#PRB$ suitably based on aforementioned constraints to ensure a guaranteed data rate for the service under consideration.

In case if a user is predicted to run into a coverage hole, then the context aware resource allocation needs to allocate $\#PRB$ such that 2 Mbps data rate is ensured in that moment as well as an additional data is buffered to guarantee 2 Mbps data rate even when the user is in a coverage hole. Consider figure 4.26 which depicts a vehicular user traveling at 30 km/h. The route prediction predicts that a user will turn into a road with coverage hole.

In the considered example, coverage hole is effective for ≈ 60 m. With the

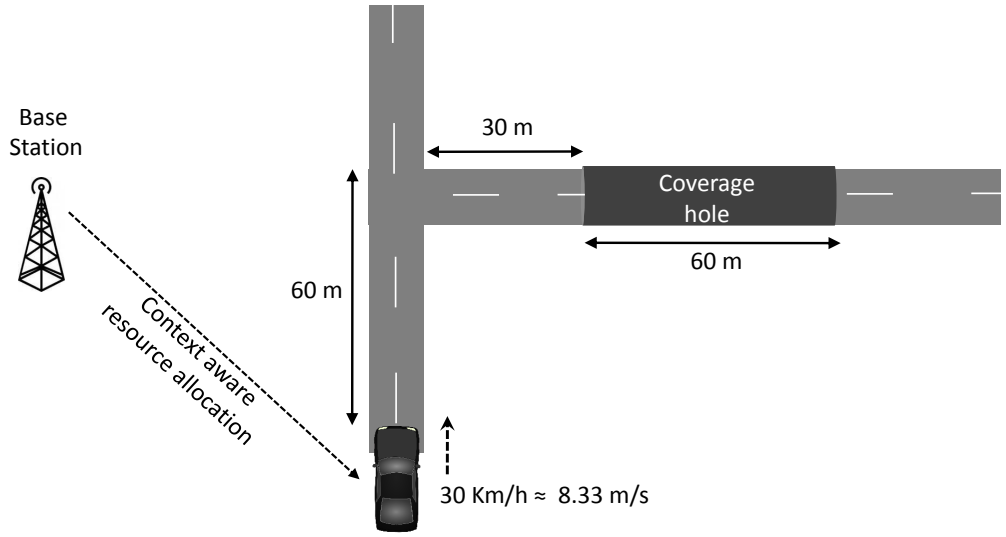

Figure 4.26: Context aware resource allocation example

average speed of the user being 8.33 m/s, the user is estimated to be in coverage hole for around 7.2 seconds. Thus, the data required to be buffered in order to serve the user at 2 Mbps in coverage hole turns out to be 14.4 MB. The user will have to travel 90 m before entering coverage hole which gives user 10.8 seconds to buffer excess of 14.4 MB data. The context aware resource allocation needs to allocate $\#PRB$ to user such that ≈ 2Mbps data rate is ensured for next 10.8 seconds and excess 14.4 MB data is buffered in the same duration. The resource allocator can either buffer all data at once at 14.4 Mbps before entering coverage hole or buffer at rate of 1.44 Mbps for next 10 seconds, based on availability of resources and SINR conditions.

Thus, the key idea of context aware resource allocation is to allocate resources such that whenever a user is anticipated to run into a coverage hole buffering could be carried out appropriately to provide guaranteed data rate in coverage holes as well.

4.3.6 Evaluation

The simulation set up follows system model described in section 4.3.2. Figure 4.20 shows considered 3D model following Madrid Grid layout which is implemented in UNITY 3D. The base stations are enabled with LTE technology, having bandwidth of 10 MHz (50 PRBs) and carrier frequency of 2 GHz. The simulation parameters (e.g.,antenna parameters, pathloss models

etc.) follow ITU-R, METIS guidelines [MP13c]. Elaborate details about considered simulation assumptions are provided in section 3.3. The set up has a car traveling with average velocity of 30 km/h and it can assume trajectories as described in table 4.6. The simulation is run for 100 times to acquire statistics of user mobility and his connections to various base stations. As shown in table 4.6 the user can traverse 10 set of trajectories with different probabilities. After acquiring the aforementioned statistics, the prediction schemes are executed for 100 simulation runs and average accuracy of prediction is deduced. The user traverses the same 10 routes in table 4.6 with similar probabilities, in learning phase (obtaining statistics) as well as prediction phase. In figure 4.27 a) accuracies of implemented next cell prediction schemes are compared. Use of additional information about origin and destination, enhances the accuracy of next cell prediction, and it is around 85% surpassing accuracy of other schemes under comparison. The accuracies of route prediction methods are compared in figure 4.27 b). Accuracy of prediction is shown to be improved when additional information (origin, origin & destination) is used. The accuracy of prediction with information about origin & destination is around 90%.

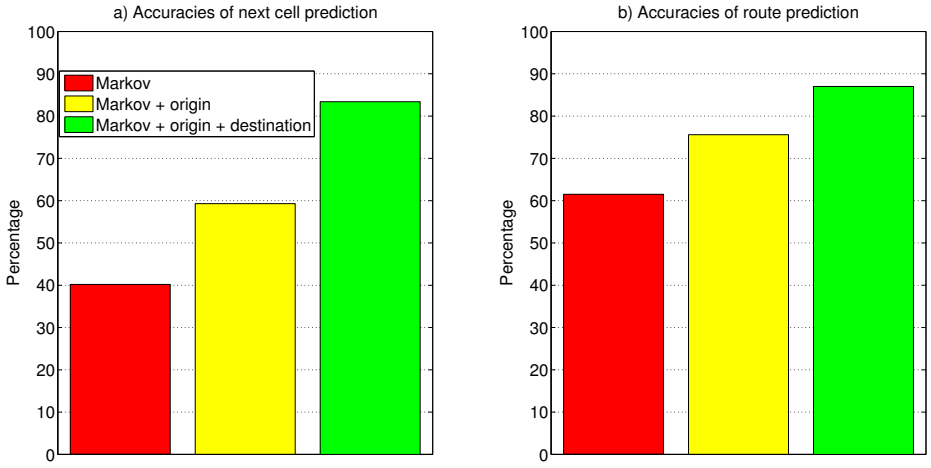

Figure 4.27: Results of mobility prediction

Figure 4.28 shows 6 coverage holes positioned at various sites in simulation set up. The SINR to bit resource per element (BRE) mapping and throughput derivations follow [JF12]. The simulations are carried out for 100 runs and the average user throughput is obtained.

A non-guaranteed bit rate service is assumed (e.g., file transfer) with the user initially allocated with 1 PRB and additional PRBs are allocated based on the route prediction. Figure 4.29 a) shows an average user throughput

Figure 4.28: Simulation set up with coverage holes

in roads with coverage hole and PRB = 2 due to triggering of the context aware resource allocation. Figure 4.29 b) depicts the same with PRB = 3.

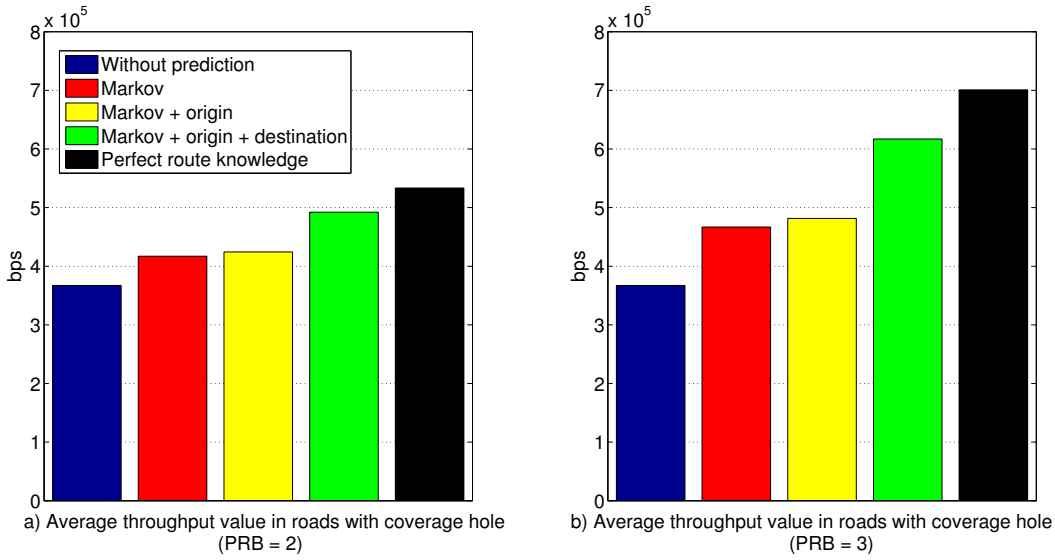

Figure 4.29: Average throughput in paths with coverage hole

Figure 4.30 a) compares percentages of user throughput improvement in roads with tunnel (PRB = 2) and figure 4.30b) analyzes the same with PRB = 3. The average user throughput in roads with tunnel is improved when the route prediction is deployed. The results get better with usage of

additional information (origin, destination) for route prediction. In case of PRB = 2, around 35% improvement is seen and with PRB = 3, improvement of 65% is witnessed. The maximum possible improvement with perfect route knowledge is also depicted in these figures.

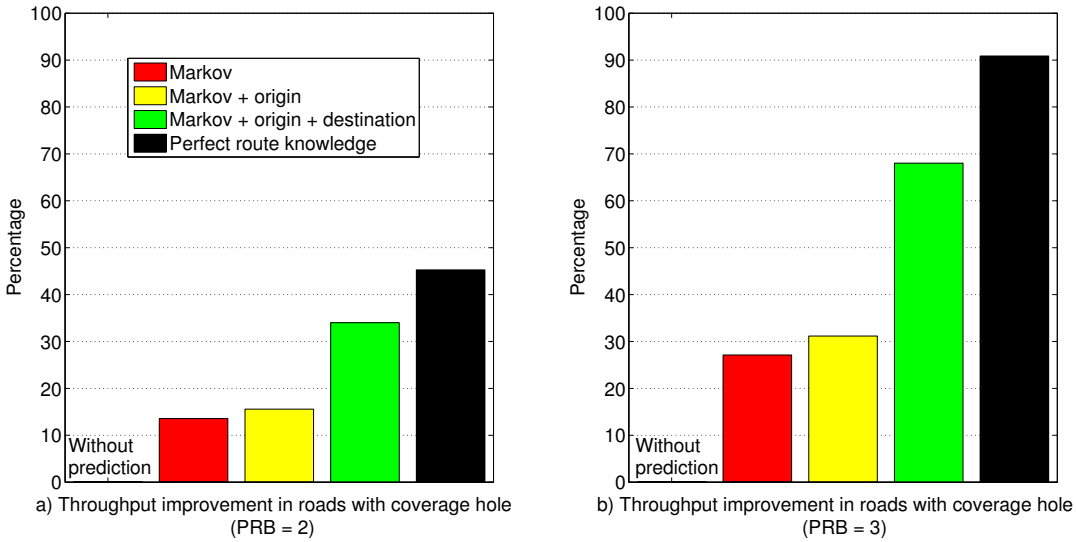

Figure 4.30: Throughput improvement in paths with coverage hole

Now a non-real time streaming service requiring 2 Mbps of data rate is considered. Figure 4.31 depicts the average throughput achieved in roads with coverage hole and figure 4.32 shows throughput improvement in roads with coverage hole. The Markov prediction improves throughput by circa 23%, use of the origin information improves throughput by around 50% and use of the origin and destination information leads to throughput improvement of around 80% in roads with coverage hole.

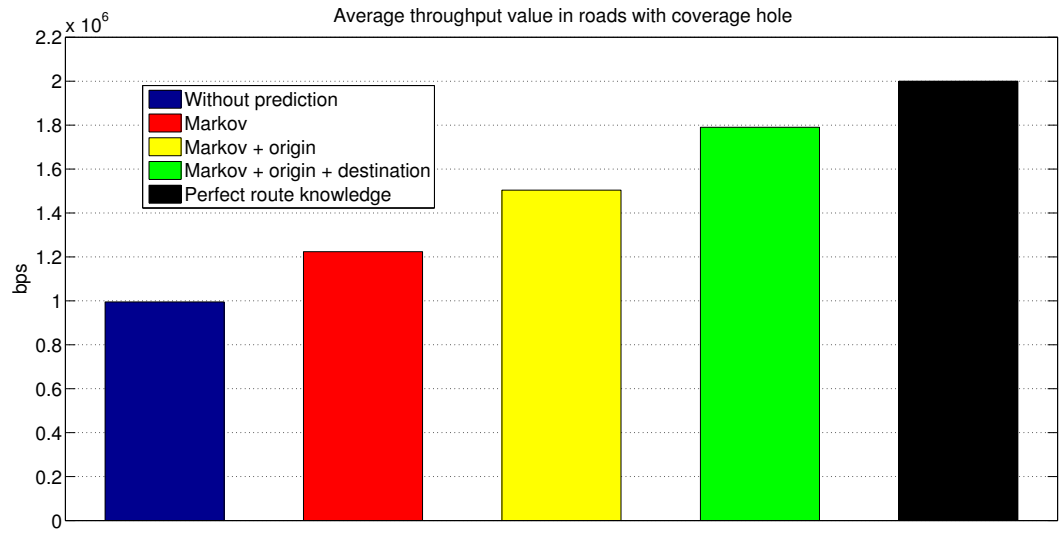

Figure 4.31: Average throughput in paths with coverage hole (streaming service)

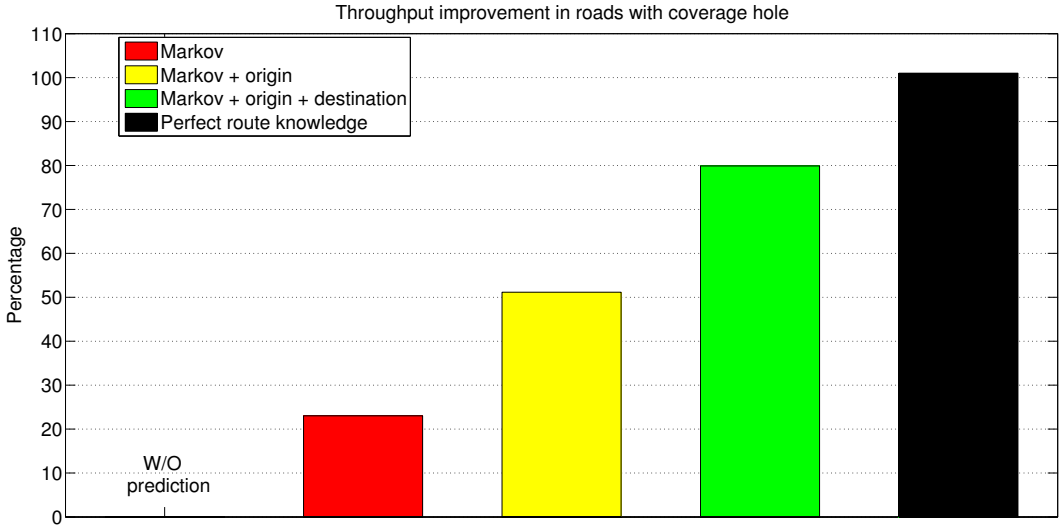

Figure 4.32: Throughput improvement in paths with coverage hole (streaming service)

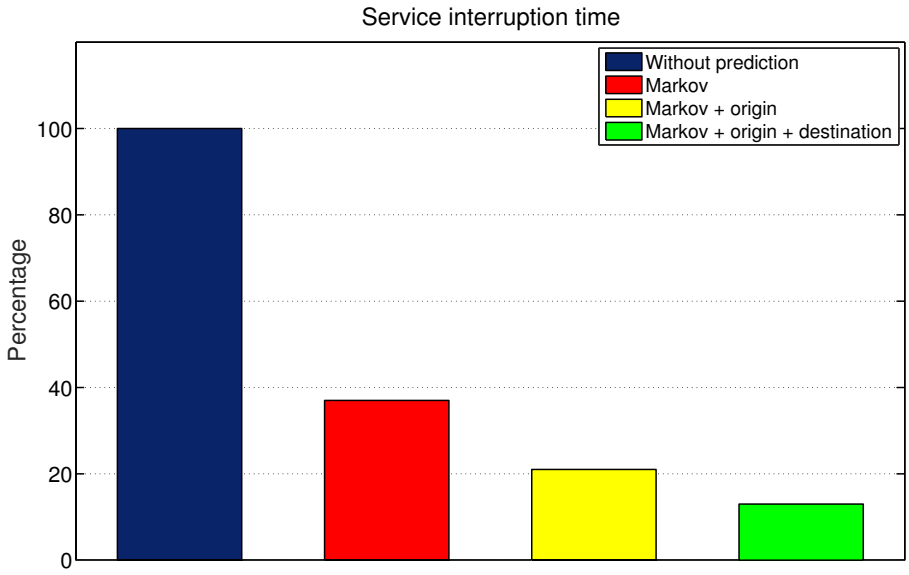

Figure 4.33: Service interruption time

With the context aware resource allocation enabling a sufficient buffering, figure 4.33 depicts an average reduction in service interruption time. It could be seen that a simple Markov prediction reduces the service interruption by around 60%, Markov prediction with origin reduces the service interruption by circa 70% and Markov prediction with origin and destination reduces the same by around 90%. Thus, mobility context aware resource allocation improves service continuity and helps in providing a uniform service quality.

The results of this work have been published in **[KZS16]** ***and*** **[KS16a]**.

4.4 Dynamic Crowd Formation Prediction and Resource Management

In section 4.2 and section 4.3, the mobility of users was predicted to initiate load balancing or carry out context aware resource allocation. Whereas, this section presents the prediction of a dynamic crowd formation based on user mobility statistics in a cellular system to initiate suitable radio resource management.

4.4.1 Introduction

In practical scenarios, such as open air festivals, stadiums or public events, a large number of mobile users gather to form a crowd. This event increases load at the serving base station. This situation when coupled with scarcity of radio resources at SeNB, results in congestion leading to an increase of dropped users and blocked access attempts. As a consequence, users in a crowd suffer from poor QoE. Such situation is common in legacy networks, but mobile users in a 5G system will expect an improved QoE even in crowded situations irrespective of high traffic density [MP13a]. A-priori knowledge of crowd formation at a particular location proves to be helpful in designing smart resource management schemes that aims to improve QoE of users in a crowd. A framework to predict crowd formation in a cell is presented here, which is based on user's cell transition prediction, detection of moving user cluster and trajectory prediction. This context information about crowd formation is used to proactively trigger load balancing or activate small cell at anticipated site of frequent crowd formation (e.g., stadium, convention halls etc.). This helps in mitigation of congestion due to crowd and in turn QoE of users is improved.

Figure 4.34 depicts a typical example of crowd formation, where an event location (e.g. stadium, city hall) is situated in cell 0. To this location, users would arrive from several sides as per their origin. For instance, from bus stops (located in cell 1, cell 6), from train stations (located in cell 2), from highway (located in cell 5) or pedestrian users distributed in cells neighboring to cell 0 and cell 0 itself. Sooner or later, these aforementioned users enter cell 0, proceed to the event location, giving rise to a crowd. Identifying and monitoring such users will assist in acquiring substantial information to predict the formation of a crowd. Subsequently, context aware load balancing could be proactively triggered to free some resources or small cell can be activated at the site of crowd, to alleviate congestion.

4.4.2 System Model

The system model as shown in figure 4.35 is considered in order to monitor users moving towards the site of interest and to predict the formation of a crowd. The site of interest is cell 0 where formation of crowd is assumed to be frequent (due to presence of stadium, convention hall, etc.). The mobility of users in the cells neighboring to cell 0 are monitored. A transition region is defined at the boundaries of neighboring cells as shown in figure 4.35. If

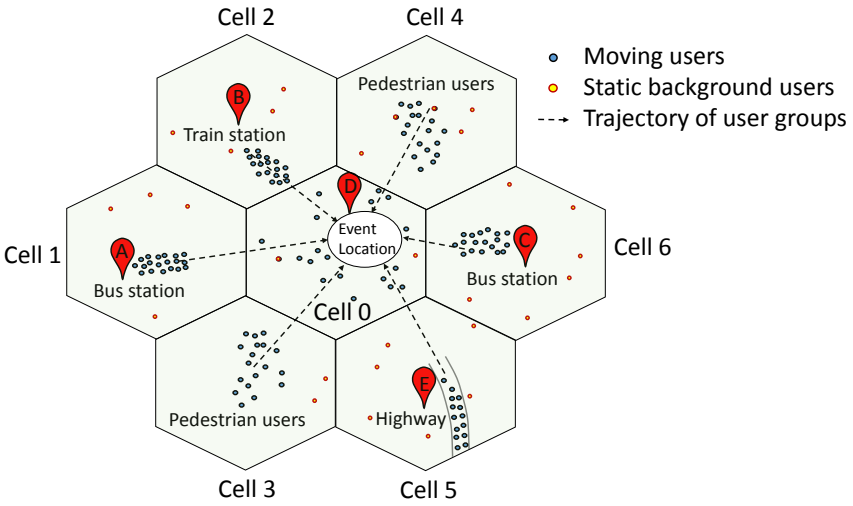

Figure 4.34: Crowd formation scenario

equation 4.29 is satisfied then such a user is said to be in transition region.

$$\gamma_s < \theta_t, \tag{4.29}$$

where γ_s is the geometry (in dB) of the user with serving cell, θ_t is a threshold value derived from the radio propagation data to define a cell transition region. The geometry is defined as a ratio of the received power from serving base station (P_i) to the interference power received from other base stations (P_j), given by

$$\gamma_s = \frac{P_i}{\sum_{j \neq i} P_j} \tag{4.30}$$

In the cells neighboring to cell 0, users that are traveling towards cell 0 are monitored. In order to obtain the number of users that are approaching the site of interest, equation 4.31 is used:

$$\frac{\sqrt{(x_0 - x_m(n))^2 + (y_0 - y_m(n))^2}}{\sqrt{(x_0 - x_m(n-1))^2 + (y_0 - y_m(n-1))^2}} < \quad 1, \tag{4.31}$$

where, (x_0,y_0) denotes the position of base station (cell 0), ($x_m(n)$, $y_m(n)$) is the present location of a moving user m, and ($x_m(n-1)$, $y_m(n-1)$) is the past position of a mobile user m.

From the set of users satisfying equation 4.31, following statistics are derived and monitored to predict severity of crowd formation at the site of interest (cell 0):

1. The number of users in transition region of neighboring cells, having

cell 0 as predicted next cell (N_t).

2. The number of user clusters in transition region of neighboring cells, having cell 0 as predicted next cell (N_{Ct}).

3. The number of users already transited to cell 0 from neighboring cells (N_0).

4. The number of user clusters already transited to cell 0 from neighboring cells (N_{C0}).

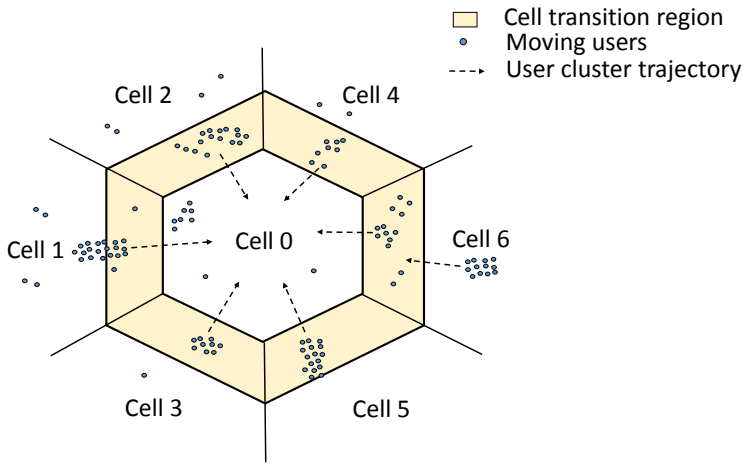

Figure 4.35: Crowd formation model

4.4.3 Cluster Detection

In general, the users traveling to an event location tend to move in groups. For instance, as shown in figure 4.34, users would begin simultaneously from same location (e.g., bus station, train station) and travel in groups towards the event location from same side. Therefore, existence of several user groups/clusters traveling to same destination acts as a valuable indicator of crowd formation. The following algorithm is used to detect the presence of user clusters approaching site of interest,

1. In each neighboring cell, consider users approaching site of interest, i.e., users satisfying equation 4.31.

2. In each cell, find distances among all the moving users approaching site of interest as below

For $i = 1$ to N
  For $j = 1$ to N
    If $i \neq j$, obtain d_{ij}

where, N is the number of users approaching site of interest, $i, j \in (1, 2, ..., N)$, d_{ij} is the distance between users i and j.

3. Define cluster radius (R) and minimum number of users required to form a cluster (θ_{NR}).

4. For all $k \in (1, 2, ..., N)$, where k represents the index of potential cluster center, check which k maximizes the number of times equation 4.32 is satisfied.

$$d_{kj} <= R \tag{4.32}$$

This step yields the maximum number of users N_R present in a radius R around user k and traveling together.

5. A moving user cluster approaching site of interest exists if equation 4.33 is met:

$$N_R >= \theta_{NR}, \tag{4.33}$$

4.4.4 Next Cell Prediction

Once a user is found to be in transition region (satisfying equation 4.29), prediction of user's next cell for transition is carried out. Thereby information about the number of users/user clusters that are likely to transit into cell 0 imminently is acquired. This acts as a valuable information in predicting crowd formation in cell 0. There are various future cell prediction methods, which are based on neural networks [LH05], machine learning [MPM12], route clustering [Laa05] etc. However, such schemes have limitations in terms of computational complexity and cost. Thus, this work uses unsophisticated schemes for next cell prediction based on geometry values of a user, which is discussed in detail in section 4.2. The probabilities of transition based on geometry are given by,

$$p_1 = \frac{\gamma_1}{\gamma_1 + \gamma_2}, \tag{4.34}$$

$$p_2 = \frac{\gamma_2}{\gamma_1 + \gamma_2}, \tag{4.35}$$

where γ_1 and γ_2 are geometries of user with respect to two potential next cell candidates.

4.4.5 Crowd Formation Prediction and RRM

The statistics which are discussed in section 4.4.2 are used to design algorithm to predict the severity of crowd formation. The users have a full buffer data traffic (the buffers of the users' data flows always have unlimited amount of data to transmit [3GP10]), thereby constituting a worst case scenario. The so called crowd formation indicator (CFI) is introduced to classify the system operation into three states namely:

1. Green: indicating that crowd formation is not likely in near future.
2. Yellow: indicating that crowd formation is likely in near future.
3. Red: indicating that crowd formation is imminent.

The conditions for the system to be in each state and inferences are elaborated as follows:

1. CFI is green if equation 4.47 is satisfied. This state infers that the formation of a crowd is not likely at a site of interest (cell 0). It is made sure from equation 4.47 that there are not enough users/user clusters in the transition region or users/user clusters already moved to cell 0, that would induce a crowd formation:

$$\begin{aligned}(N_t < \theta_N) \wedge (N_{Ct} < \theta_{Ct}) \wedge \\ (N_{C0} < \theta_{C0}) \wedge (N_0 < \theta_0),\end{aligned} \tag{4.36}$$

   where $\wedge$ indicates the logical AND operation.

2. CFI is yellow when equation 4.48 is satisfied. This state indicates the likeliness of crowd formation at the site of interest (cell 0). This infers that there exists ample users/user clusters in transition region which are set to move into cell 0 soon and incite a crowd formation. Equation 4.48 also analyzes if the aggregate of users in transition region

and users that already moved to cell 0 are adequate enough to trigger formation of a crowd:

$$(N_t > \theta_N) \vee (N_{Ct} > \theta_{Ct}) \vee (N_t + N_0 > \theta_{N0}), \tag{4.37}$$

where $\vee$ indicates the logical OR operation.

When CFI turns yellow, load balancing is triggered proactively in cell 0 in order to free-up resources. These resources will be used to accommodate users in transition region which enter cell 0 in near future, subsequently forming a crowd. Load balancing has to be carried out only on static background users in cell 0 and not on moving users. Otherwise, there would be high number of ping-pong handovers and LB failures.

3. CFI is red if equation 4.49 is satisfied. This signals the impending formation of a crowd at site of interest (cell 0). This condition arises when the number of users/user clusters that have already moved into cell 0 are sufficient to prompt a crowd formation, i.e.,

$$(N_0 > \theta_0) \vee (N_{C0} > \theta_{C0}). \tag{4.38}$$

The threshold values θ_N, θ_{Ct}, θ_0, θ_{C0} and θ_{N0} have to be fixed by the network operator on the basis of available bandwidth and maximum connections that could be served at the site of interest. These threshold values are dependent on bandwidth/requested service rate of the users. As every user in this scenario is requesting full buffer data traffic, thresholds are set in terms of number of users.

When the CFI becomes red (equation 4.49 is satisfied), potential location of a crowd formation can be deduced by using equation 4.39 and equation 4.40, solving for x and y. Subsequently, a small cell closest to the predicted location of crowd formation is activated, in order to prepare the site under consideration to accommodate users in a crowd. Figure 4.36 illustrates such scenario.

$$ax + by = c, \tag{4.39}$$

$$px + qy = r, \tag{4.40}$$

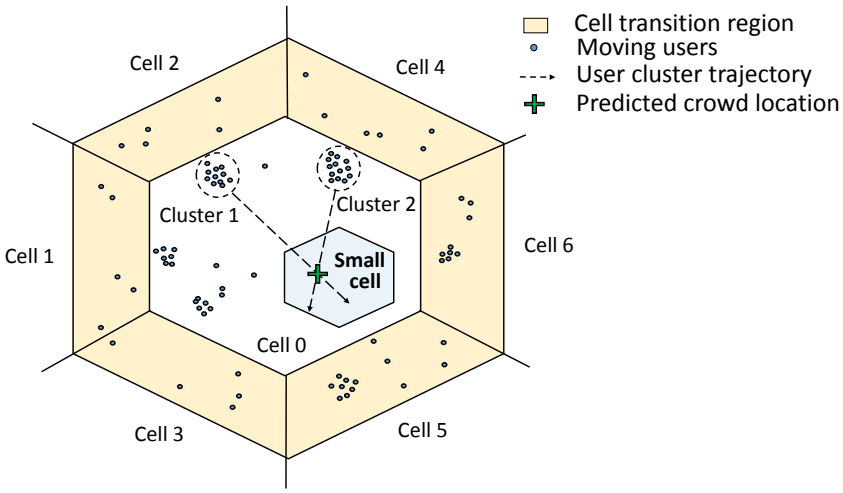

Figure 4.36: Prediction of crowd formation location

where a, b, c, p, q and r are known values derived using present and previous locations of any two user clusters, given by,

$$a = y_1^p - y_1, \tag{4.41}$$

$$b = x_1 - x_1^p, \tag{4.42}$$

$$p = y_2^p - y_2, \tag{4.43}$$

$$q = x_2 - x_2^p, \tag{4.44}$$

$$c = ax_1 + by_1, \tag{4.45}$$

$$r = px_2 + qy_2, \tag{4.46}$$

where (x_1, y_1) and (x_2, y_2) denote the centers of two moving user clusters that have moved to cell 0, (x_1^p, y_1^p) and (x_2^p, y_2^p) are previous positions of the aforementioned moving user clusters.

When the CFI is red and crowd formation location is anticipated, then a small cell closest to it is activated. The activated small cell starts serving the users in crowd and the macro cell (cell 0) is off-loaded.

4.4.6 Evaluation

Figure 4.37 illustrates a multi-cell scenario that is set up using a LTE system level simulator [3GP10]. In figure 4.37, cell 0 is bound to be the site of frequent crowd formation. The moving users emerge from cells neighboring to cell 0 and subsequently proceed to a certain point in cell 0 to incite a crowd. Table 4.9 summarizes simulation parameters. The thresholds are set

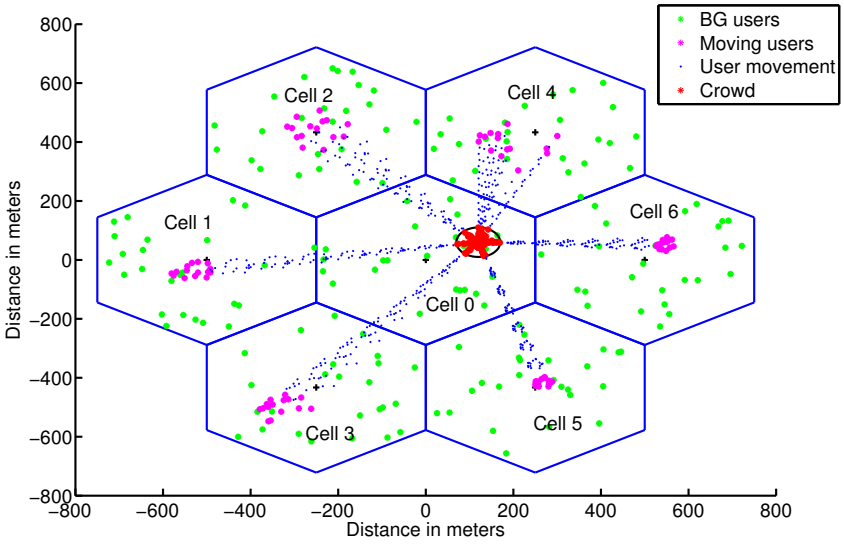

Figure 4.37: Crowd formation simulation

as $\theta_N = 30$, $\theta_{Ct} = 3$, $\theta_0 = 45$, $\theta_{C0} = 3$, $\theta_{N0} = 45$, $\theta_{NR} = 5$ and $R = 30m$.

Table 4.9: Simulation Parameters (Crowd)

Parameter	**Assumption**
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
Total transmit power	40 W (s2s=500 m) 10 W (s2s=250 m)(small cell)
Shadowing	log-normal Standard deviation: 8 dB Decorrelation distance: 50 m
Fast fading	2-tap Rayleigh fading channel
Noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B) + 7$
Background users per cell	30
Moving users	105 at velocity ranging from $3 - 7$ km/h

Various states assumed by cell 0 with the progress of simulation is demonstrated in figure 4.38. At the beginning of simulation time, system state is green, as there is not adequate number of users/user clusters in transition region that would move into cell 0 or already transited to cell 0, in order to induce crowd formation. Thus, formation of a crowd is unlikely. With the progress in the simulation time, mobile users in neighboring cells start traveling at different velocities (from $3-7$ km/h) towards cell 0. Clusters are also formed by many mobile users that move together based on their relative

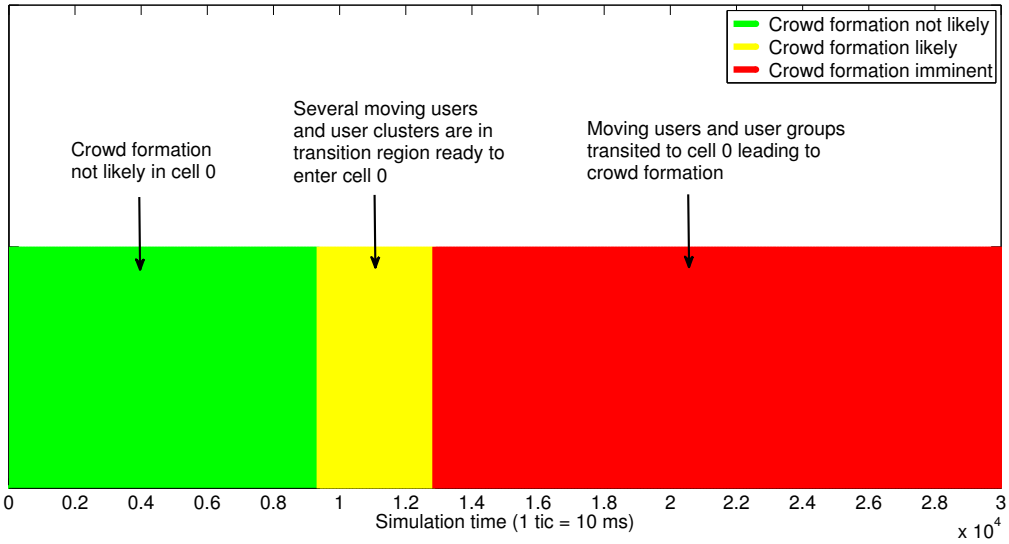

Figure 4.38: Crowd formation indicator

positions and velocities. The system state changes to yellow, when enough number of users/user clusters reach transition region of their respective cells and become ready to move into cell 0. This indicates that many users/user clusters would transit into cell 0 imminently and proceed subsequently to form a crowd. The system state changes to red as soon as adequate number of users/user clusters have already moved to cell 0 and are proceeding towards same destination. This indicates the high likeliness of crowd formation in cell 0. By solving equation 4.39 and equation 4.40, the probable location of crowd formation could be found.

Load balancing is triggered as soon as the system state is yellow at the site of cell 0. Load balancing is performed on the background users located near cell edge of cell 0. Figure 4.38 a) demonstrates the improvements in KPIs. Considerable reductions are observed in dropping of users (23%), blocking of users (10%) and blocked handover attempts (15%). Further, activation of small cell at site of frequent crowd formation (e.g., stadium, convention hall, etc.) leads to substantial reductions in dropping of users (60%), blocking of users (56%) and blocked handover attempts (59%), as depicted in figure 4.38 b). Thus, QoE of users can be significantly improved even in presence of a crowd.

The results of this work have been published in **[KKS15]**.

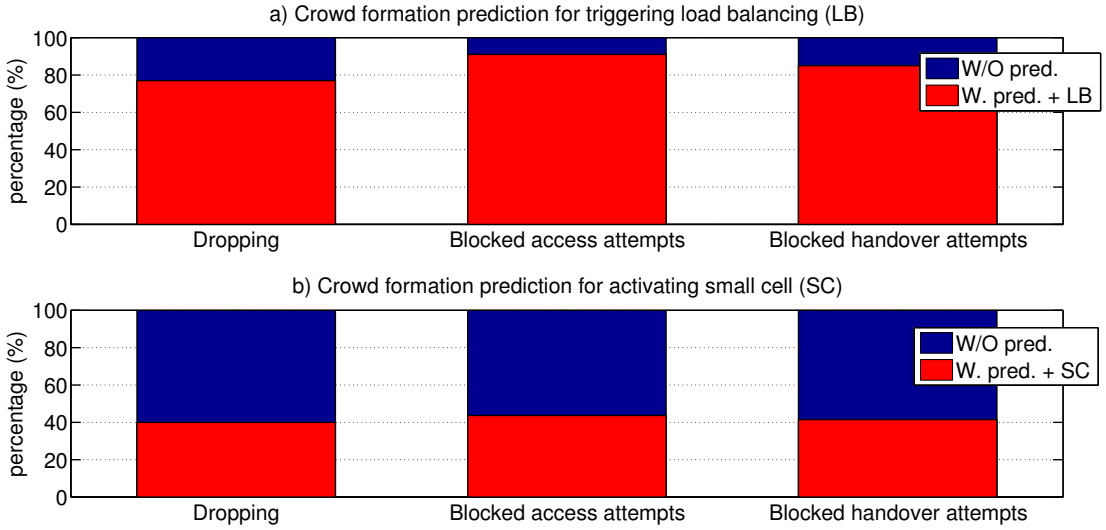

Figure 4.39: Improvements in KPIs

4.5 Radio Resource Management in Vehicular Traffic Jams

This section is similar to section 4.4 but deals with vehicular traffic jams. The vehicular users with comparatively higher velocities than pedestrian users (as in section 4.4) are considered in this section. Vehicular user mobility is monitored to design algorithm to indicate traffic status and suitable RRM is initiated in tandem.

4.5.1 Introduction

In everyday scenarios, cellular broadband services are availed by mobile users while moving in vehicles (e.g. infotainment in car, individual commuter usage in public transport etc.). Based on the road topology, arrival/departure rate etc., there is a dynamic variation in the volume of such vehicular users in a cell. Further, the frequency of a traffic jam occurrence is high at certain sites (e.g. signal post, crossroads). A sizeable number of vehicular users stop momentarily at these junctions, and cause a high load situation at the serving base station. The congestion prompted due to this event, causes a rise in dropping of users connected to the serving base station and blocking of new access attempts made by the users. This hinders the QoE of users in the cell. Deducing the traffic status in a cell is beneficial for efficient radio resource management and to improve QoE of users even in highly congested traffic jams.

A framework is presented in this section to predict vehicular traffic status from the perspective of a cellular network and to study its influence on cellular network. Mobility behavior and related statistics of vehicular users are observed. Subsequently, methods for next cell prediction, identification of vehicular cluster/moving networks and vehicular velocity estimation are collaborated in order to develop a traffic status prediction algorithm. The prior knowledge about emergence of traffic jam in near future, assists in proactive triggering of load balancing at serving base station or activation of small cell at the predicted site of a traffic jam. Consequently, congestion due to traffic jams are alleviated and users have a better QoE despite presence of traffic jams. An illustration of traffic jam formation is shown in figure 4.40, from

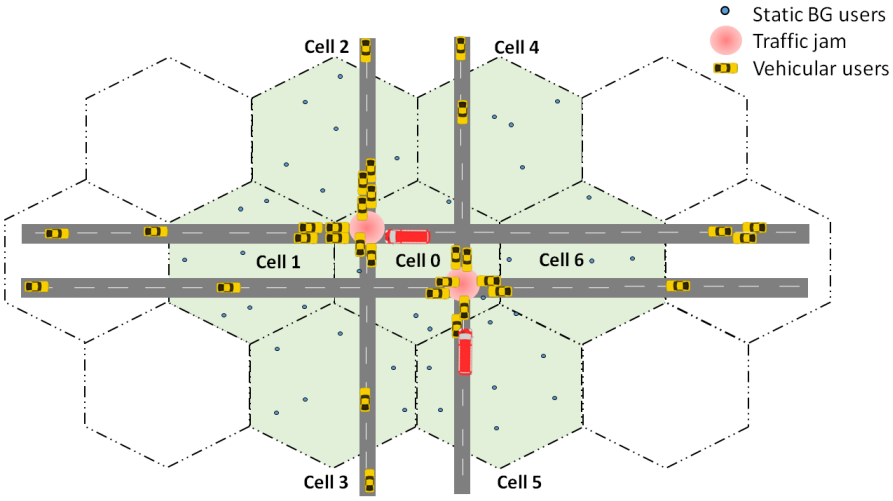

Figure 4.40: Traffic jam formation scenario

the standpoint of a cellular network. Cell 0 harbors signal posts where traffic jams are recurrent. The vehicular users transit from cells neighboring to cell 0 and subsequently halt at these junctions causing traffic jams. Gradually, vehicular users resume their travel again and traffic jam is dispersed. Corresponding to these events, there is a variation in traffic status of cell 0. Thus, information significant for traffic status prediction can be obtained by tracking the mobility of vehicular users in cells adjacent to cell 0 (site of frequent traffic jams). The traffic status can be used to trigger load balancing or duly actuate small cells in the vicinity of traffic jam site as depicted in figure 4.41.

4.5.2 System Model

The system model considered for monitoring vehicular users and developing traffic status prediction algorithm is as shown in figure 4.42. The traffic jams

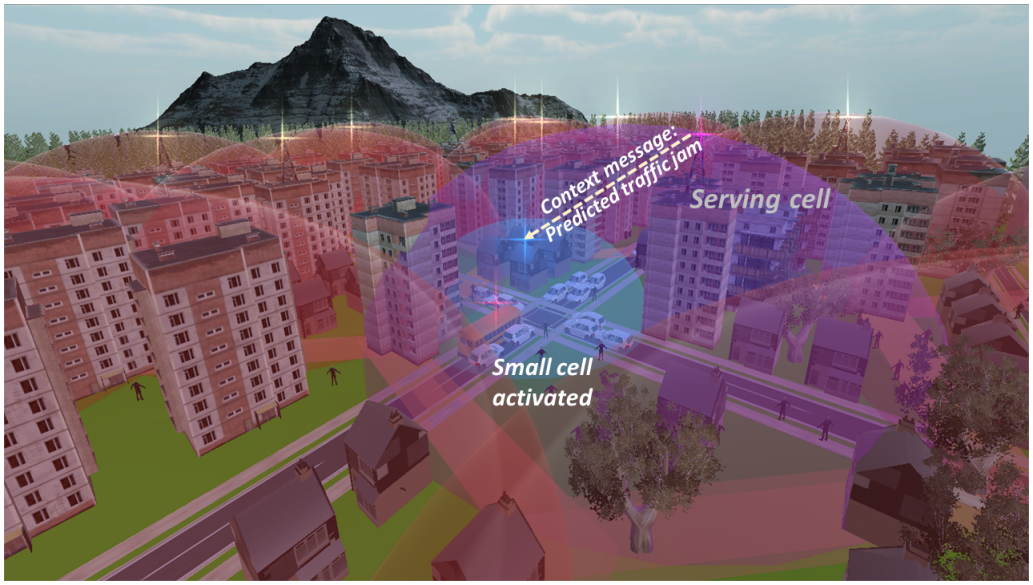

Figure 4.41: Context aware RRM in traffic jams

are recurrent in cell 0 (due to presence of signal posts etc.) and thereby it is the site of interest. Mobility behavior of vehicular users in cells adjacent to cell 0 are tracked and users approaching cell 0 are found using equation 4.31. Further, in the considered system model, a transition region is defined at boundaries of neighboring cells similar to section 4.4. Figure 4.43 depicts vehicular user cluster and moving network that are possibly present in the considered system.

The group of vehicular users satisfying equation 4.31 are monitored and below statistics are obtained to design algorithm to predict traffic severity at the site of interest (cell 0):

1. $N_t \rightarrow$ number of vehicular users in transition region of neighboring cells, having cell 0 as next cell for transition.

2. $N_{Ct} \rightarrow$ number of vehicular user clusters in transition region of neighboring cells, having cell 0 as predicted next cell.

3. $N_0 \rightarrow$ number of vehicular users already transited to cell 0 from neighboring cells.

4. $T_t \rightarrow$ total data traffic demand of vehicular users in transition region of neighboring cells.

5. $N_{\Delta vj} \rightarrow$ number of vehicular users with negative velocity gradient, nearby frequent jam location.

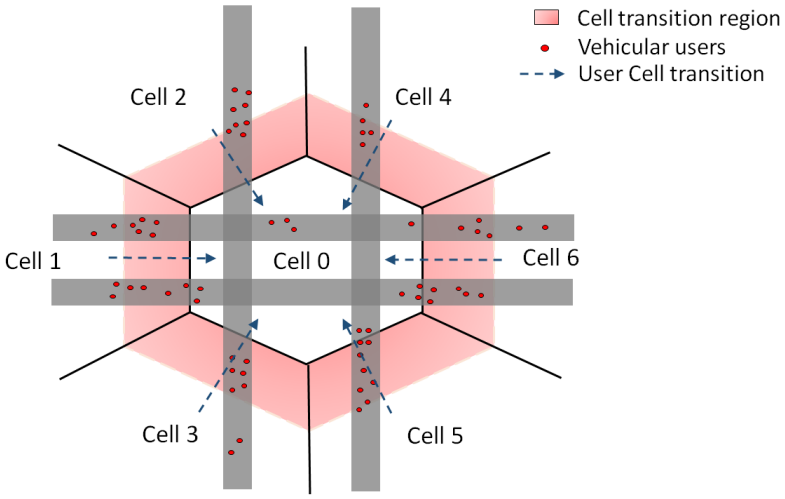

Figure 4.42: Traffic jam formation model

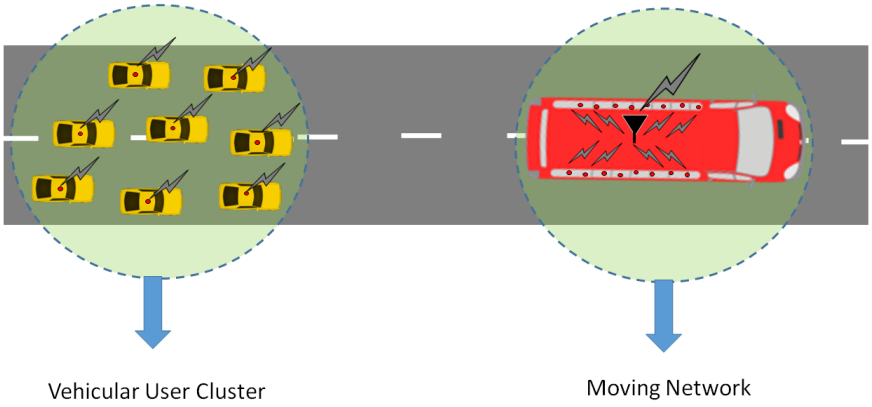

Figure 4.43: Vehicular user cluster

6. $N_{C\Delta vj} \rightarrow$ number of vehicular user clusters with negative velocity gradient, nearby frequent jam location.

7. $T_{\Delta vj} \rightarrow$ cumulative data traffic demand of vehicular users with negative velocity gradient, nearby frequent jam location.

4.5.3 Traffic Status Prediction and RRM

The algorithms for next cell prediction and vehicular cluster detection are similar to those described in section 4.4.4 and section 4.4.3 respectively. The statistics acquired regarding mobility of vehicular users in and around the site of interest are used to design traffic status prediction algorithm. The

vehicular users have full buffer data traffic (the buffers of the users' data flows always have unlimited amount of data to transmit [3GP10]), and pose a worst case scenario. The traffic status indicator (TSI) can have the following states:

a) Green: The density of vehicular users/clusters in cell 0 or in transition region of adjacent cells is not adequate to cause traffic jam imminently. The collective data traffic request and amount of access attempts to cell 0 by vehicular users are minimal. Hence, there are no symptoms of impending congestion. In certain scenarios, although a large number of users are not present in the traffic jam, the cumulative data demand of respective users is large enough to cause high load situation. Therefore, statistics regarding data traffic have to be considered as well. The condition for TSI to be green is given by equation 4.47,

$$\begin{aligned}(N_t < \theta_N) \wedge (N_{Ct} < \theta_{Ct}) \wedge (N_0 < \theta_0) \\ \wedge (T_t + T_{\Delta vj} < \theta_{T\Delta vj}),\end{aligned} \tag{4.47}$$

where $\wedge$ indicates the logical AND operation.

b) Yellow: TSI in this state signals the likeliness of an upcoming high load situation in the cell. An ample number of vehicular users/moving networks predicted to enter cell 0 are present in transition region or the cumulative data volume of vehicular users in transition zone is large enough to cause a hotspot in cell 0. Equation 4.48 gives the condition for TSI to be yellow. Equation 4.48 also considers if the aggregate of vehicular users in transition zone and users that have transited earlier to cell 0 are sufficient to cause congestion.

$$\begin{aligned}(N_t > \theta_N) \vee (N_{Ct} > \theta_{Ct}) \\ \vee (N_t + N_0 > \theta_{N0}) \vee (T_t > \theta_{Tt}),\end{aligned} \tag{4.48}$$

where $\vee$ indicates the logical OR operation.

When TSI is yellow, load balancing is triggered proactively to enable cell 0 to accommodate newly entering vehicular users. The load balancing procedure is illustrated in figure 4.44, where static background users at the cell edge (of cell 0) are forced to be served by appropriate neighboring base stations. LB should not be performed on vehicular users, as their motion would cause ping-pong handovers and LB failures.

c) Red: When conditions in equation 4.49 are met, then it is an indication that traffic jam is imminent at frequent jam site. Typically, when the vehicles

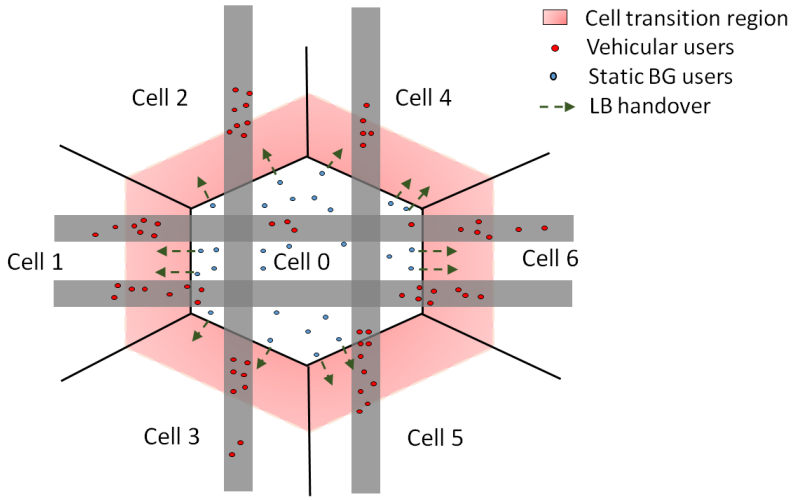

Figure 4.44: Proactive load balancing

have to stop at signal posts, brakes are applied to slow the vehicle down before gradually reaching a halt. Such behavior is taken into account in equation 4.49 and statistics of vehicular users/moving networks in vicinity of jam site, with negative velocity gradient are obtained. The corresponding cumulative data traffic demand is taken into account as well. In order to get these statistics, a predefined circle is considered around probable jam site and vehicular users present in it are examined. Doppler processing [Fre12] is assumed to be available in the considered set up, to estimate velocity of vehicles.

$$\begin{aligned}(N_{\Delta vj} > \theta_{\Delta vj}) \vee (N_{C\Delta vj} > \theta_{C\Delta vj}) \\ \vee (T_{\Delta vj} > \theta_{T\Delta vj}).\end{aligned} \tag{4.49}$$

The network operators can observe and fine tune the threshold values θ_N, θ_{Ct}, θ_0, $\theta_{T\Delta vj}$, θ_{N0}, θ_{Tt}, $\theta_{\Delta vj}$ and $\theta_{C\Delta vj}$ based on available resources and the maximum number of connections that could be served.

As soon as TSI turns red, a small cell (SC) in the vicinity of traffic jam is switched on. The activated small cell will start to serve vehicular users in traffic jam. Figure 4.45 illustrates the switching on of SC at frequent traffic jam site.

Eventually, when vehicular users resume their travel and dissolve the traffic jam, TSI changes correspondingly to yellow and later to green. Small cell is switched off as soon as the traffic jam disperses, to minimize energy consumption of small cells. Figure 4.46 depicts overall framework of the traffic

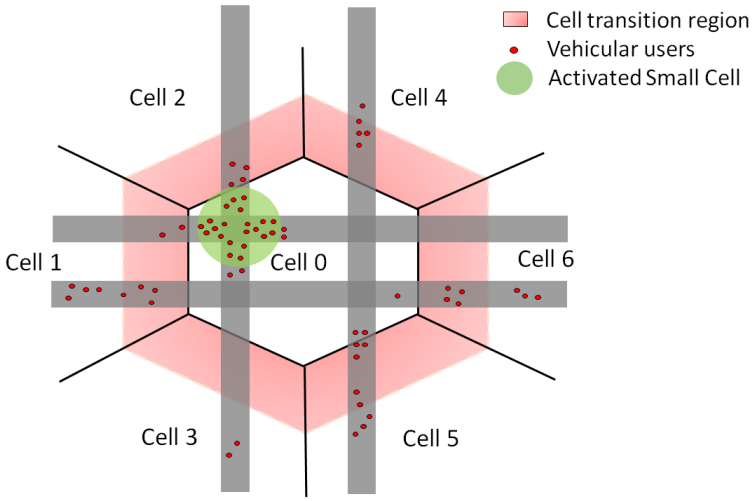

Figure 4.45: Small cell activation

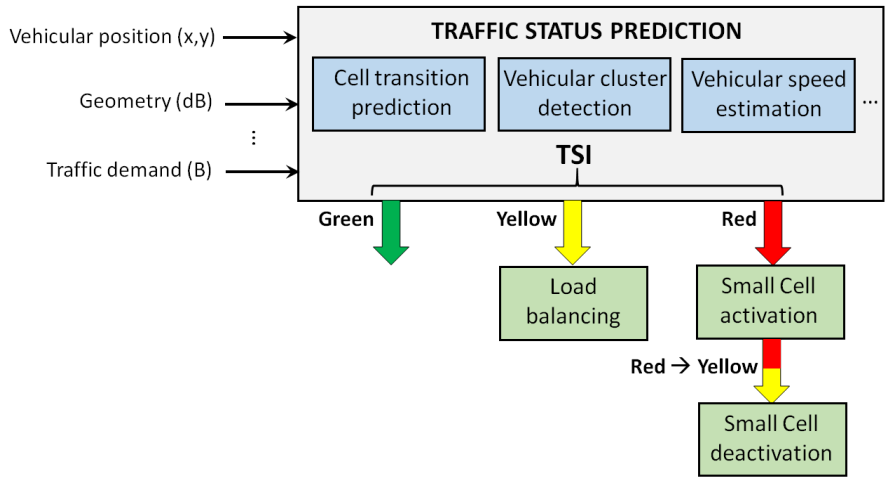

Figure 4.46: Overall framework

status prediction and proactive RRM.

4.5.4 Evaluation

A multi-cell scenario as shown in figure 4.47 is set up using LTE system level simulator that follows [3GP10]. Traffic jams are recurrent in cell 0. The vehicular users emanate from cells adjacent to cell 0, proceed into cell 0 following the road topology and induce traffic jams. Table 4.10 summarizes simulation parameters. The thresholds are set as $\theta_N = 30$, $\theta_{Ct} = 3$, $\theta_0 = 45$, $\theta_{N0} = 45$, $\theta_{C\Delta vj} = 3$, $\theta_{\Delta vj} = 30$, $\theta_{T\Delta vj} = 250$ MB, $\theta_{Tt} = 250$ MB $\theta_{NR} = 5$and $R = 30m$.

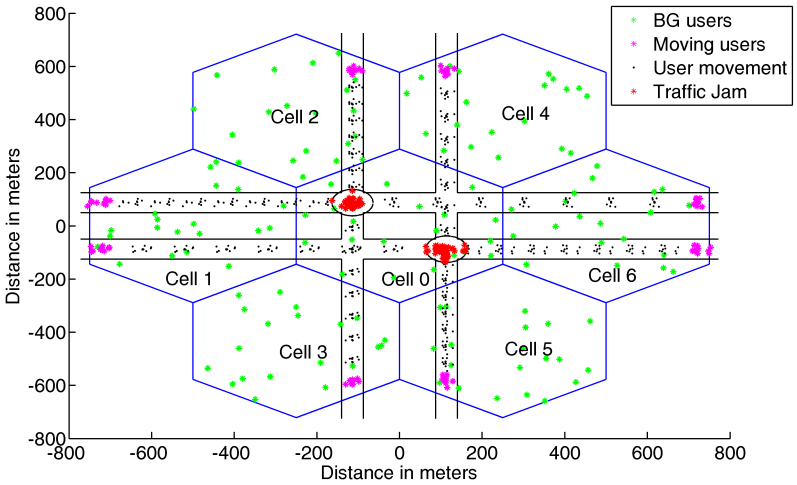

Figure 4.47: Simulation of traffic jam formation

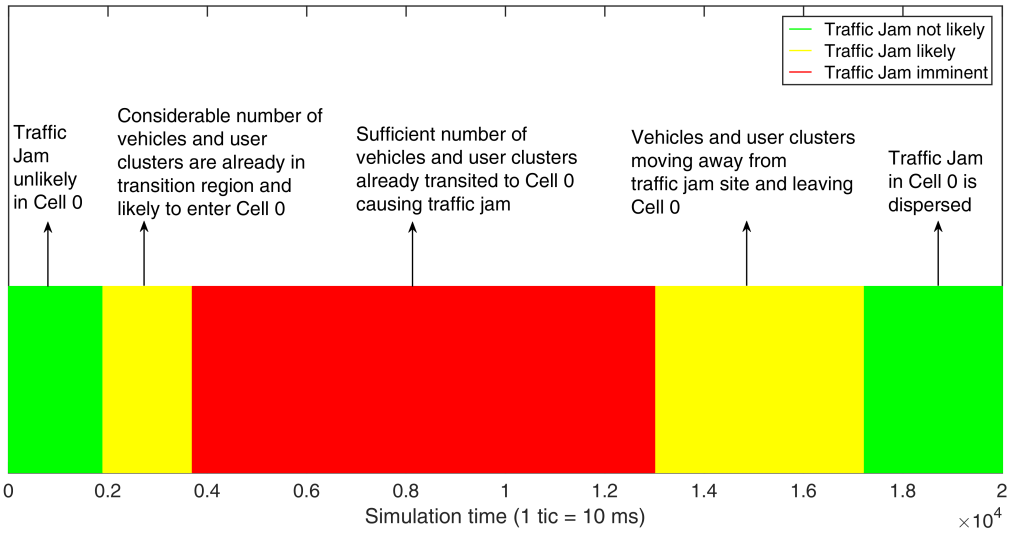

Figure 4.48: Traffic status indicator

Figure 4.48 illustrates evolution of TSI with simulation time. There is neither adequate number of vehicular users in cell 0 nor in transit zone of adjacent cells at the beginning of simulation time. Thus, TSI is initially green. However, TSI becomes yellow, when an ample number of vehicular users/clusters prepared to move into cell0, are found in the transition region. Further, TSI turns red, when adequate amount of vehicular users/moving networks transit to cell 0 and exhibit signs of stopping near traffic jam site (negative velocity gradient). The TSI stays red throughout the course of traffic jam. Once the vehicular users start dispersing, then TSI switches

Table 4.10: Simulation Parameters (Traffic jam)

Parameter	Assumption
Carrier frequency	2 GHz
System bandwidth	10 MHz (50 PRBs)
Total transmit power	40 W (s2s=500 m) 10 W (s2s=250 m)(small cell)
Shadowing	log-normal Standard deviation: 8 dB Decorrelation distance: 50 m
Fast fading	2-tap Rayleigh fading channel
Noise power	-174 dBm/Hz $+ 10 \cdot \log_{10}(B) + 7$
Background users per cell	30
Vehicular users	135 at velocity ranging from $30 - 80$ km/h
Monitoring interval	1 second

from red to yellow again. Finally, TSI resumes green state, when most of the vehicular users travel out of cell 0.

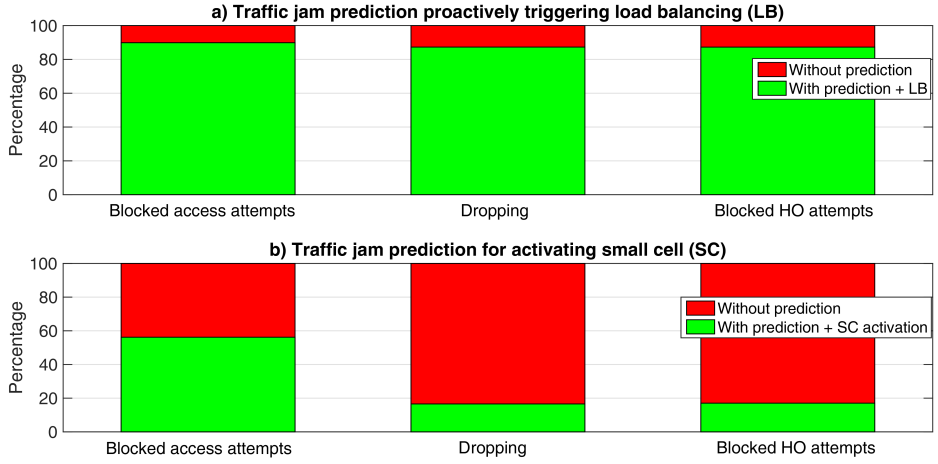

Figure 4.49: Improvements in KPIs

Load balancing is proactively triggered at cell 0, when TSI is yellow (Note: LB has to be triggered only once, when TSI switches from green to yellow). This procedure reduces dropping of users by ≈18%, blocking of new access attempts by ≈10% and blocked handover attempts by ≈18% (shown in figure 4.49 a)). Further, appropriate small cells are switched on at respective traffic jam sites, when TSI turns red. In the presented evaluation, traffic jams occur nearly at the same time at two sites as depicted in figure 4.47. Following the activation of small cells, dropping of users is reduced by ≈82%, blocked

access attempts is reduced by ≈42% and blocked handovers are reduced by ≈81% (shown in figure 4.49 b)). The reduction of these KPIs, indicate that users have an improved QoE even during traffic jams.

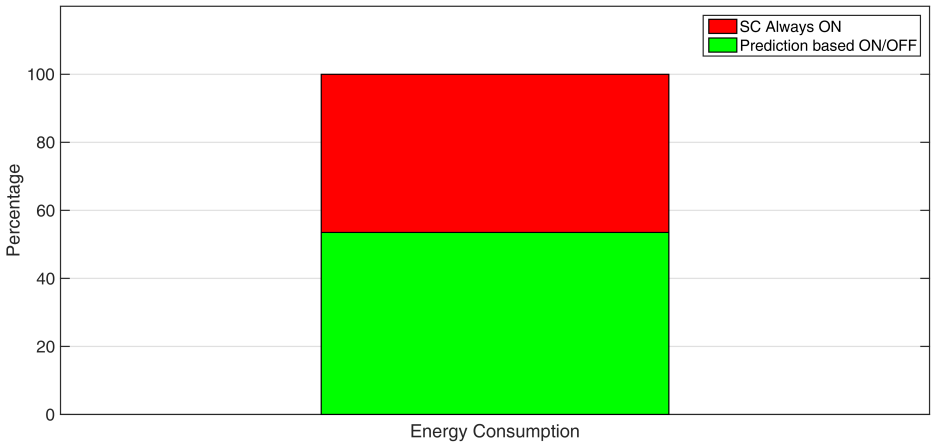

Figure 4.50: Comparison of energy consumption

Further, as traffic jam disperses (TSI change from red to yellow), the small cell which is switched on is now deactivated. The energy consumptions of small cell being always ON and prediction based ON/OFF method, are compared in figure 4.50. With prediction based scheme, energy consumption is minimized by ≈ 45% for the considered simulation set up. Thus, prediction based small cell activation/deactivation not only improves user QoE but also is energy efficient compared to always ON strategy.

The results of this work have been published in **[KMS17b]**.

4.6 Mobility Context Awareness Framework

Following the previous sections, which emphasized on importance of mobility context awareness to enable smart resource management in cellular systems, this section deals with various signaling aspects, corresponding interfaces and overall architecture of such a system.

4.6.1 Introduction

The presented framework deals with context awareness in cellular networks and mobility of users being the focal point. Mobility prediction is one of the

crucial elements in enabling context awareness. Development of several radio resource management schemes, handover optimization procedures and so on are assisted by anticipating/learning mobility behavior of the user. There exists various schemes in literature exploiting user mobility behavior to build context awareness. This section presents a framework to support mobility context awareness in cellular networks by analyzing the commonalities among several context aware schemes (as in figure 4.51). Further, the set of required inputs, signaling and interfaces are also studied.

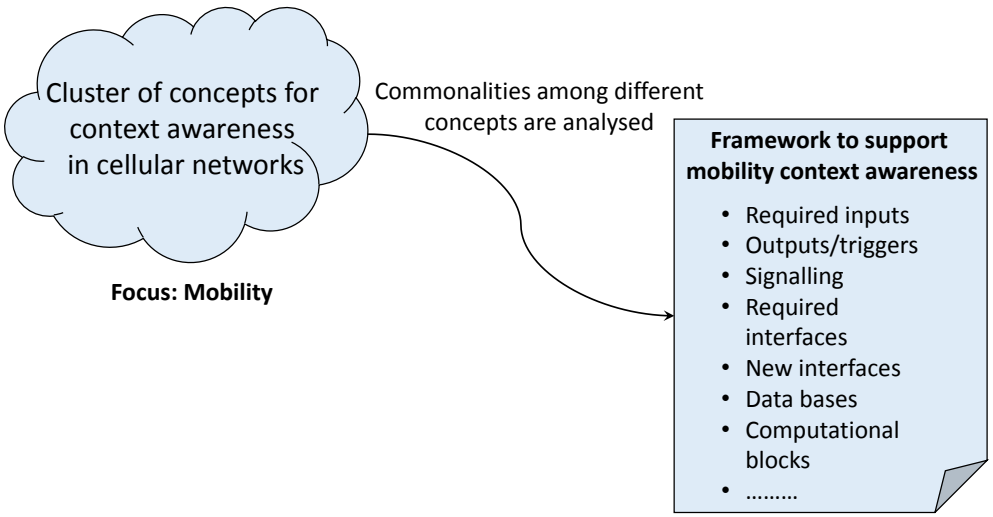

Figure 4.51: Motivation

4.6.2 Required Information Set

As a part of research project METIS [MP12], the author has contributed in analyzing various context aware schemes and broadly categorizing the context information based on the following criteria[MP13b]:

- Relation to network entities:
  - UE context information: information collected from the UE, e.g., position, speed, UE battery status, UE settings,device capability, user profiles, buffer status etc.
  - Context information of network layer and lower: e.g., radio propagation map, current or historical interference status, network load etc.

  - Context information of higher layer: e.g., QoS requirement of the applications and services, service profile etc.

- Levels of abstraction:
  - Primary context information: information that can be measured or collected from the network directly, e.g. GPS position.
  - Secondary context information: knowledge that needs to be inferred from the primary context information, e.g., energy efficiency, UE behavior pattern, traffic pattern etc.

- Time scale:
  - Static information: e.g. UE capability.
  - Quasi-static information: e.g., UE settings, handover history.
  - Dynamic information: e.g., UE location, traffic expectations etc.

In particular to the mobility context aware schemes presented in this chapter, common information set required are listed below:

- User position: The user position is typically represented as (x,y,z) coordinates. Global positioning system (GPS) or network assisted positioning can be used to obtain this information.

- User velocity: Doppler measurements are conventionally used to estimate velocity of the user. Such information can also be acquired from speedometer of vehicle via appropriate interface between vehicular infrastructure and cellular network.

- User geometry (dB): The measurements of geometry are carried out by the user terminal and subsequently communicated to base station.

- Neighboring cells list (NCL): This list is maintained by operations support system (OSS) [AL09].

- Route maps: These are similar to NCL but contain information about roads, cross roads, possible coverage holes in them and other vital radio/non-radio information of the routes under consideration.

In addition to the aforementioned information set, the following information from vehicular infrastructure is beneficial to enhance the mobility prediction and context awareness, discussed in section 4.3.

- Origin: The initial position (location/landmark) from where the user started his journey.
- Destination: The final position (location/landmark) to where the user intends to travel.

4.6.3 Framework

Mobility context aware procedures can be functionally decomposed into three blocks as in figure 4.52 namely:

1. Context extraction: The functionalities of this block include acquiring locations of the user, his signal strength/geometry measurements, information of user origin and destination (from vehicular infrastructure) and so on. Velocity estimate is necessary to fix the sampling interval of aforementioned information.
2. Communication: This block is concerned with signaling/communication of vital context information among users and base stations, as necessary. The key parts are: a) Signaling of user positions, geometry etc., to the base station. b) Signaling context message (e.g., trigger for load balancing/resource management, cell activation/deactivation message etc.) from serving base station to target (predicted) base station.
3. Prediction/Decision: This block makes use of the extracted/exchanged context information and facilitates prediction of future cells, routes and other event predictions. In tandem to these predictions, decisions of resource management, cell activation/deactivation etc., are also made here.

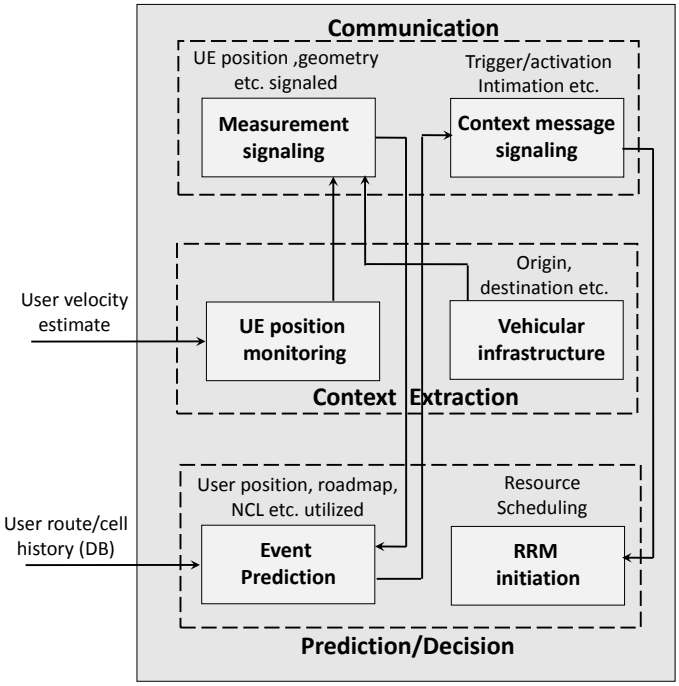

Figure 4.52: Functional Decomposition

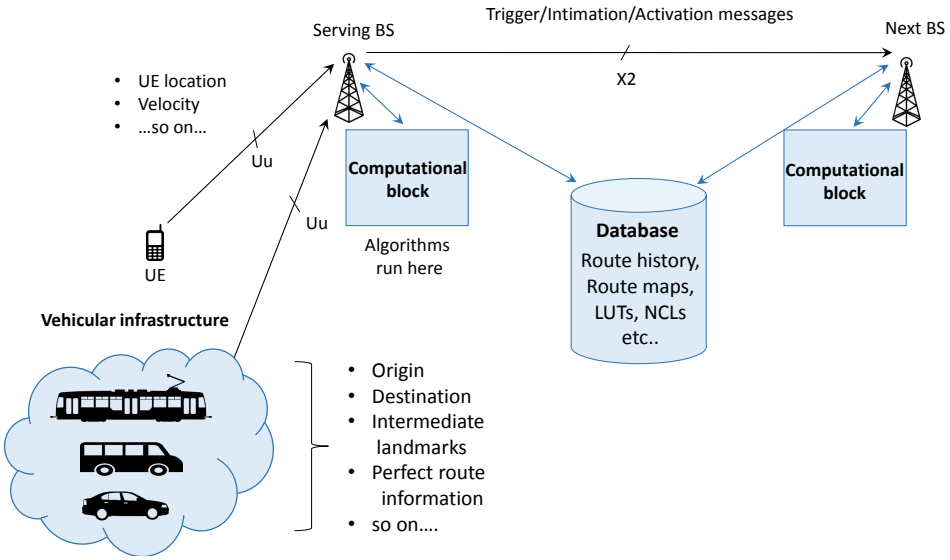

Figure 4.53: Overall Architecture

Figure 4.53 depicts the overall architecture based on the functional decomposition analysis, to facilitate mobility context awareness. The Uu interface is used to signal/communicate information about user positions, geom-

etry(dB), velocity etc., to serving base station [AL09]. Further, Uu interface is also usable to convey information from vehicular infrastructure to serving base station. The typical information includes user origin, destination, intermediate landmarks, perfect route information (which is envisioned to become feasible due to advancements in autonomous driving). In order to maintain history of user mobility, neighboring cell list, route maps of respective areas and other suitable look up tables (LUT), a database has to be set up. Further, an appropriate interface is essential to send user mobility related information to database. Here, it is assumed that mobility statistics of a user are stored in a database via serving base station. The computational blocks at serving base station are in charge of prediction (e.g. next cell) after obtaining context information from a user and interfacing with database. The successive context message (e.g., triggering LB, cell activation) needs to be communicated to the target cell via X2 interface [AL09].

It could be observed from section 4.3, that the context information about user origin and destination are vital for enhancing mobility context awareness. For the efficient extraction of such information from vehicular infrastructure and subsequent exploitation in prediction schemes, new interfaces are essential. Figure 4.54 illustrates the new interface foreseen between navigation module in the vehicle and the user equipment. This interface assists in acquiring information regarding user origin, destination, route information etc., and further relays them to the serving base station via Uu interface.

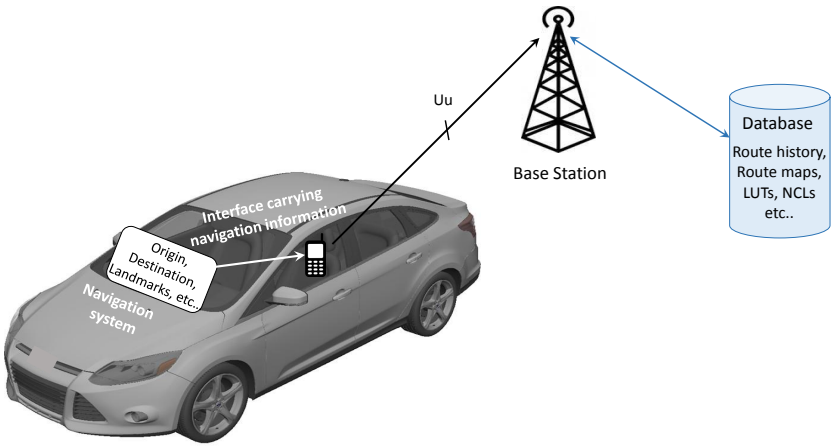

Figure 4.54: Interface in Vehicular Infrastructure

When many users are traveling together (e.g. public transport like bus, train etc.), it is possible to reduce the amount of interfaces and signaling

by having an anchor node manage user connections in the vehicle. This constitutes a moving network [KKSS13b]. In this case a new interface is needed between anchor node and navigation system of vehicle as shown in figure 4.55.

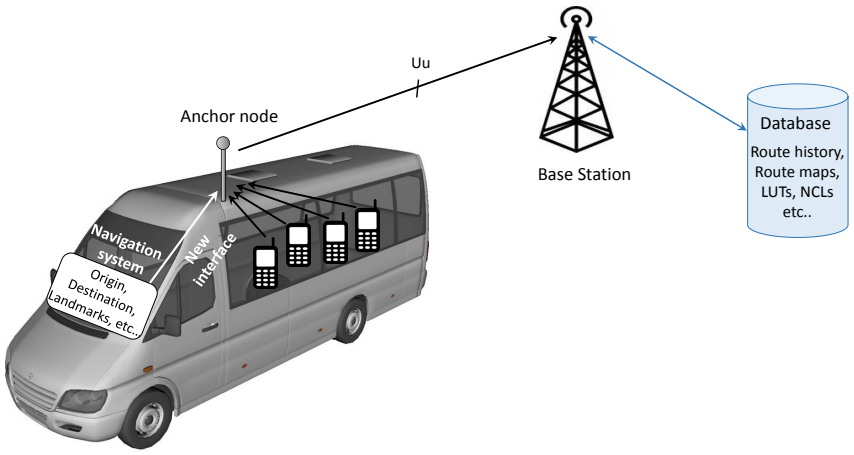

Figure 4.55: Interface in Moving Networks

With the proposed signaling, interfaces and functional blocks shown in in figures 4.53, 4.54 and 4.55, it is viable to render maximum assistance for mobility context awareness in cellular network.

***The results of this work have been published in* [KS16a].**

4.7 Summary

The key challenges in 5G mobile communications are to accommodate more traffic volume (1000 times) and higher number of connected devices ($10-100$ times). Providing good quality of experience to the users becomes a major issue with emergence of practical day-to-day scenarios such as traffic dense moving networks, mobile hotspots, dynamic crowds, traffic jams and so on. This chapter proposed mobility context awareness as a key enabler to design smart resource management schemes to provide good QoE to the users amidst aforementioned high mobility and high load situations. The following are the highlights of the chapter:

- Developed concepts for next cell prediction based on user attributes

such as geometry, user angle and distances.

- The developed mobility context awareness was used to proactively manage radio resources via load balancing for practical scenarios of data dense moving networks and mobile hotspots. Substantial reductions in dropping of users, blocking of users and blocked HOs were observed.
- Enhanced mobility prediction scheme was developed in an urban scenario, using additional context information of a user origin and destination.
- Aforementioned scheme was used to design a context aware resource allocation algorithm, thereby improving service continuity in coverage holes.
- Schemes were developed to anticipate dynamic crowd formation and to predict traffic status from a cellular network perspective.
- Load balancing and small cell activation/deactivation were triggered pro-actively based on the developed schemes to manage radio resources in crowded situations and traffic jams.
- QoE in the system was improved due to reduced number of blocked users, dropped users and blocked HOs.
- Finally, an overall framework was presented with signaling and interfaces facilitating incorporation of mobility context awareness into cellular systems.

Chapter 5

Context Aware D2D Communications

5.1 Overview

Mobile systems of today are designed to provide an acceptable broadband experience to users when being sparsely distributed in public areas e.g., parks, bus-stops etc. However, when the user density increases, the quality of user's experience deteriorates. On the other hand, a base station experiences overload due to increasing number of connected devices and traffic volume demand. Handling a large number of mobile connections and corresponding traffic volume is a major challenge for network operators. Conventionally, the UE traffic is routed via a base station, which handles connection establishment between two UEs (depicted in figure 5.1 a)). However, there are scenarios where UEs wanting to communicate are in close proximity. In such a case, it is possible to assign a direct link between them for communication, hence relieving base station from traffic routing (depicted in figure 5.1 b)). Device-to-Device (D2D) communications commonly refer to the technologies that enable devices to communicate directly without base stations or infrastructure of access points, and with/without the involvement of wireless operators [LZLS12]. The physical proximity of devices is exploited for D2D communications which leads to improvement in coverage, data rates, resource utilization, QoS [FDM+12] etc.

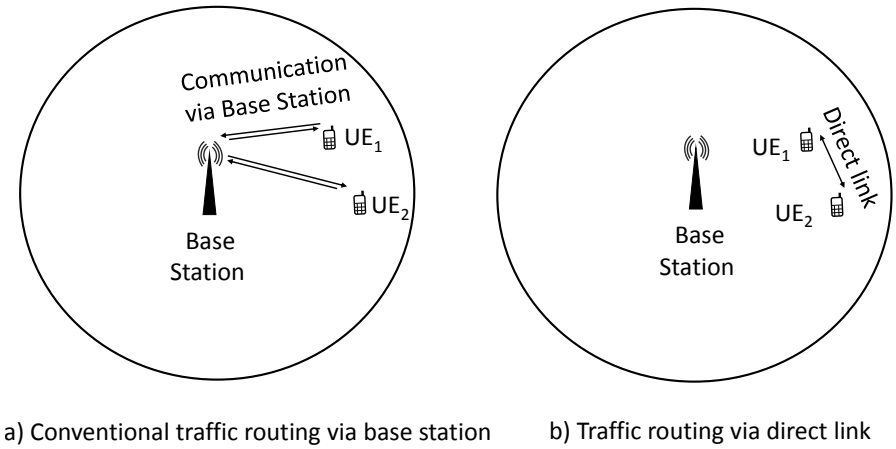

Figure 5.1: Conventional traffic routing vs. D2D communication

This chapter proposes D2D communications as a viable solution to improve system capacity and makes the following contributions,

- Resource allocation (RA) for D2D communication
  - PRB reuse scheme among D2D and cellular links is presented.
  - Location awareness (Virtual cell sectoring) is exploited for smart resource reuse.
  - Elaborated in section 5.2 *Location aware D2D Resource Allocation*
  - Robustness of location based D2D RA schemes are tested against positioning errors.
  - Elaborated in section 5.3 *Robustness of D2D RA against Positioning Errors*
- Power allocation in cellular network to assist D2D communication
  - Post-RA power control schemes are presented for cellular networks with D2D underlay.
  - Elaborated in section 5.4 *Post-RA Power Allocation in Cellular Networks*

5.2 Location aware D2D Resource Allocation

5.2.1 Introduction

In order to obtain maximum benefits from network-controlled D2D communication, enablers/solutions are essential for procedures namely; mode selection, resource allocation, power and interference control, etc. Resource allocation is one of the key procedures required for the set up of D2D communication, so that peers can communicate with each other. In the context of D2D resource allocation, it is possible to let D2D users reuse/share resources with cellular UEs (C-UE) to increase spectrum utilization [DSS13]. A single D2D link could also be allocated multiple PRBs to obtain high throughput [LKKS14]. Since the same resources are set to be reused by C-UE and D2D pair, they interfere with each other. Assuming uplink resources being reused, D2D transmission interferes with cellular transmission (at base station) and cellular transmission interferes with D2D transmission (at D2D receiver). In order to control the interference from D2D transmitter, either maximum transmission power should be restricted to a predefined value or D2D transmissions must not be permitted in a specific area near base station, where interference caused would be high [HCLK10]. Resource allocation should make sure that physical distance between C-UE and D2D transmitter is large enough to restrict interference from cellular UEs to D2D receivers. Therefore, user position serves as a vital information for such resource allocation methods. Regarding positioning technologies, network assisted techniques such as Angle of Arrival (AoA) and Time Difference of Arrival (TDOA) [ZGFW12], are more favorable than UE-based method, such as Assisted GPS (A-GPS) or Global Navigation Satellite System (GLONASS). This is because former requires less signaling efforts and less battery power consumption than latter. This section presents a novel Resource Allocation (RA) procedure for D-UEs reusing/sharing the Physical Resource Blocks (PRBs) of the C-UEs. The RA algorithm relies on *sectoring* considered cell into *virtual* sectors based on UE geo-location and density. D2D RA is then carried out as per pre-defined vectors illustrated later in this section.

5.2.2 System Model

A single cell scenario is considered with eNB at its center. N_{Cell} C-UEs and N_{D2D} D-UEs (links) are present in the cell (Figure 5.3). Figure 5.2 depicts three types of interferences (depicted by red and orange connectors) that can

occur when D-UEs and C-UEs reuse the same resources: D2C interference from the D-UE Tx to the C-UE Rx (eNB), C2D interference from C-UE (Tx) to the D-UE Rx, D2D interference from D-UE Tx to another D-UE Rx using the same resources.

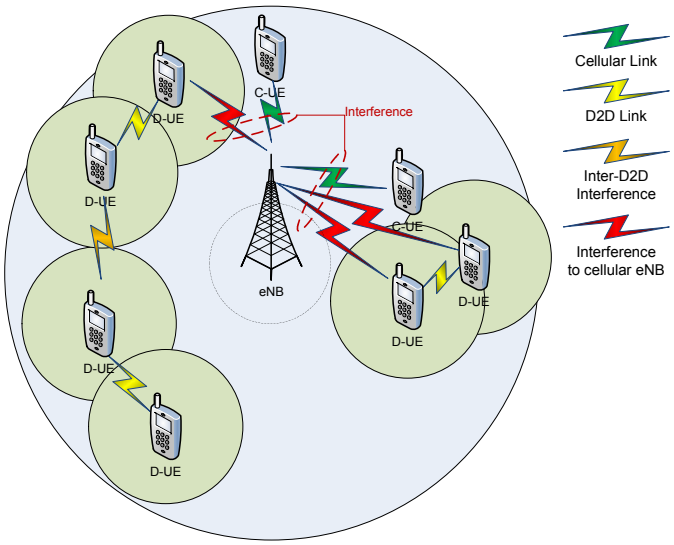

Figure 5.2: System Model

5.2.3 D2D RA based on positions

In general, an interference aware RA scheme considers sufficient physical separation between the UEs before they can reuse same resources. Sectoring based RA procedure tries to find pairs of sectors that can share the same set of PRBs, making sure that they potentially have maximum spatial distances between them. In general, the sectoring may be performed with respect to D-UEs, C-UEs, or both UE categories. A pair of sectors is composed of a reference sector and corresponding peer sector. Further, UE specific geo-location information, such as angular information is used to establish a sector pair.

The RA approach can be divided into two steps

1. Sectoring the cell into k number of sectors
2. Sector-wise RA vector for PRB allocation

There are various methods to partition a specific cell into sectors. The straightforward ones being Fixed angle and UE density based. Other complex clustering methods like k-means [FS08] and Neural Gas [Fri95] are also applicable for sectoring a cell.

Fixed Angle

A cell is partitioned into fixed number of sectors with uniform sizes. In this simplest case, the required number of sectors determines the angle of each sector. A sector pair consists of diagonally opposite sectors which will be assigned similar resource vectors. A cell divided into 8 equal sized sectors (denoted by N_S) is shown in figure 5.3. The D2D RA (for D2D Tx) begins in S_1. The D2D transmitters residing in S_1 are allocated with PRBs of C-UEs affiliated to its peer sector, i.e., S_5. UEs in S_2, S_3 and S_4 are allocated the PRBs from S_6, S_7 and S_8 respectively. This procedure is executed in a circular manner covering 360 degrees.

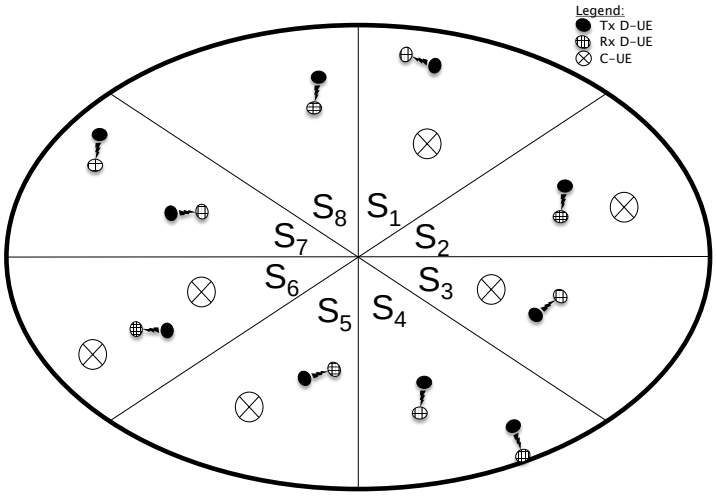

Figure 5.3: Fixed Angle Sectoring with 8 Sectors

User-Density based Sectoring

Fixed Angle sectoring could give rise to empty or too crowded peer sectors making it difficult for RA procedure. Considering this issue, a cell can also be sectored based on UE density (C-UE and D-UE) resulting in equal number of UEs in any sector. Figure 5.4 illustrates determination of reference angle of

each sector, e.g., mean of all user angles in specific sector. In essence, a sector reference angle describes the user distribution and potential concentration in a cell sector.

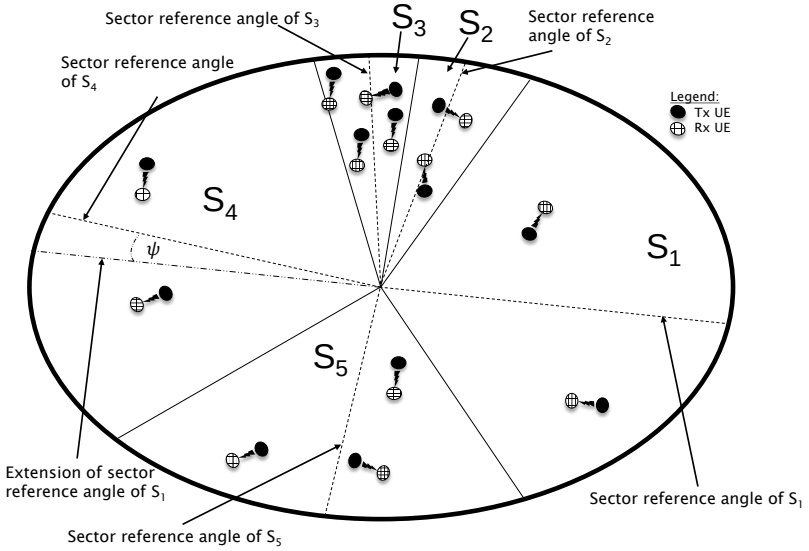

Figure 5.4: D-UE Density based Sectoring

An exemplary sectoring illustration is depicted in figure 5.4. A peer sector is determined for every sector by considering angular difference between the reference angle of given sector (plus 180 degrees) and reference angles of remaining sectors.

$$\psi_k = |mod(\psi_{S_i} + 180°, 360° - \psi_{S_k})| \; i \neq k,$$
$$\forall i, k \in \{1, ..., N_{Sectors}\}$$

ψ_k ($\psi_k \in [0°, 360°]$) is sorted in ascending order. The sector exhibiting least angular difference (ψ_k) is chosen as a peer sector. Figure 5.4, exemplifies sectors and their peer sectors , e.g., sector S_4 is recognized as the peer sector of sector S_1. However, there can be instances where two sectors would select the same sector as their peer sector, e.g., both S_2 and S_3 pick S_5 as their peer sector. In such scenarios, it can be assumed without loss of generality that a D2D pair is chosen in counter-clockwise direction from a specific sector and the D2D pair for resource reuse is selected from corresponding peer sector in a random fashion.

Resource Allocation Algorithm

The type of RA method based on sectoring is influenced by several factors such as C-UE D-UE ratio, UE priorities, QoS requirements etc., which could be addressed through adaptive RA algorithms. Nevertheless, the generic sectoring based RA algorithm is depicted as follows.

Algorithm 5.2.1: Sectoring based RA Algorithm

Data: AoA's of all UEs
Result: RA Vector for each Sector
Initialization ;
while *Un-Allocated D-UEs exist* **do**
  1. Allocate N_{Cell} PRBs to N_{Cell} C-UEs ;
  2. Sector total D2D users with $N_{RB_{tot}} - N_{Cell}$ sectors ;
  3. Assign orthogonal PRBs (if available) to each of these sectored D2D pairs ;
  4. For the remaining D2D pairs $(N_{D2D} - (N_{RB_{tot}} - N_{Cell}))$
  if $(N_{D2D} - (N_{RB_{tot}} - N_{Cell})) \leq (N_{RB_{tot}} - N_{Cell})$ **then**
    assign un-allocated D2D pairs the PRBs in the respective peer sectors determined in step 2. Stop (RA process is done) ;
  else
    Select an un-allocated D2D pair from each sector and allocate the PRB of the D-UE located in the peer sector determined in step 2 ;
  5. As there are still $(N_{D2D} - 2(N_{RB_{tot}} - N_{cell}))$ D-UEs left that are not yet assigned any resources, perform sectoring with respect to C-UE density/locations ;
  6. Sector the un-allocated $(N_{D2D} - 2(N_{RB_{tot}} - N_{cell}))$ D-UEs with the corresponding number of sectors ;
  7. Assign each un-allocated D2D pair (sectored in step 6)) the corresponding cellular radio resource that is used in the respective peer sector as determined in step 5). Stop (RA process is done).

Figure 5.5 illustrates the sectoring example with $N_{cell} = 2$, $N_{D2D} = 6$, and $N_{RB_{tot}} = 4$. The C-UEs C_1 and C_2 are assigned 2 out of the 4 available PRBs. Subsequently, the two D-UEs displaying smallest angular difference to mean D2D angles - P_2 and P_5, are assigned with the remaining 2 PRBs. Then, reuse of these 4 PRBs is carried out for remaining 4 D-UEs based on peer sectors.

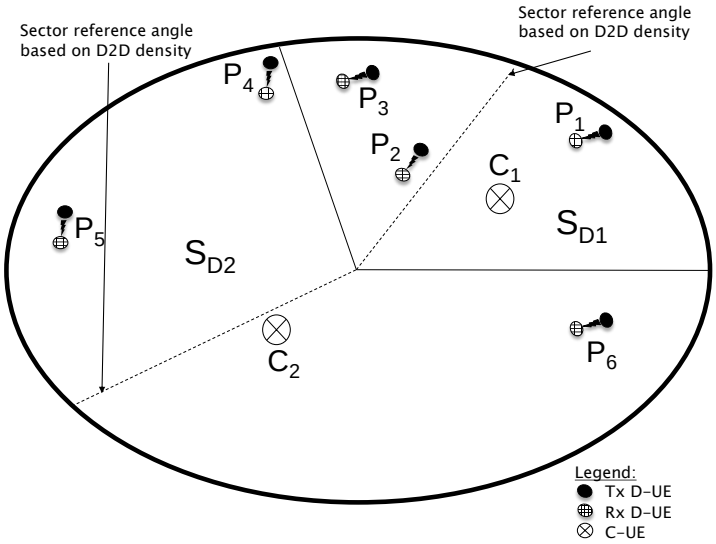

Figure 5.5: RA after user-density based Sectoring

5.2.4 Performance Evaluation

A cell with radius 500m and a LTE eNB at its center is considered. The D2D communication is assumed in the LTE uplink reusing uplink resources. A fixed number of C-UEs and D-UEs are dropped randomly (uniform distribution) in the cell and maximum permitted distance for D2D communication is set as 120m. The total number of PRBs considered is 50, total bandwidth being 10 MHz.

Table 5.1: Simulation Assumptions (D2D RA)

Parameter	C-Mode	D-Mode
Number of UEs	25	50
Transmit Power (eNB)	46 dBm	46 dBm
Transmit Power (UEs)	24 dBm	24 dBm
Noise Figure	5 dB	7 dB
Interference margin (for TPC)	1.5 dB	1 dB
Channel Model	3GPP[3GP10]	3GPP[3GP13]

The following four D2D RA schemes are compared with respect to the number of dropped connections and system capacity.

Distance based RA

In this case, a D2D link is allocated with the PRB of the C-UE such that the distance between them $L > L_{min}$ as depicted in figure 5.6. Here, L_{min}

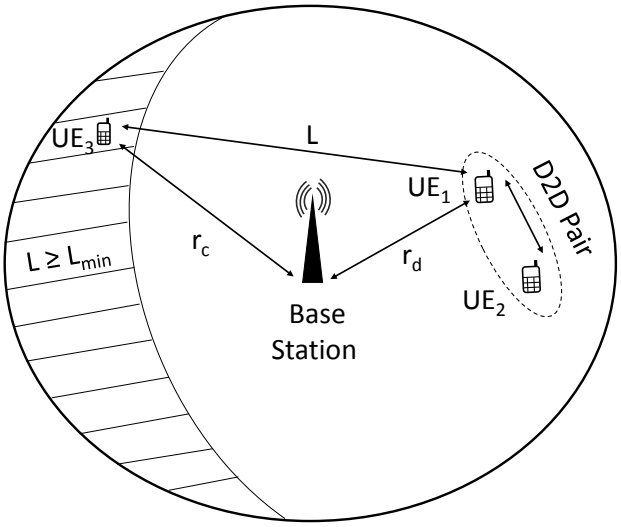

Figure 5.6: Illustration of Distance based RA

is the pre-defined distance constraint to limit the interference from chosen C-UE to the D-UE, resusing PRBs [WC12]. The D2D transmission takes place in cellular uplink (UL[HCLK10]). D2C interference is not entirely eliminated since the spatial seperation between BS and D2D Tx is not taken into account for the RA. Nevertheless, with proper selection of the distance constraint L_{min}, C2D interference can be restricted. In this evaluation, a fixed angle sectoring with 4 sectors is performed. Following this, PRB reuse is enabled between UE in a sector and a UE which is at the largest distance in corresponding peer sector (from considered UE). In case PRB of the UE at largest distance is already assigned, then UEs in the adjacent sectors are considered and so on. Thus, maximal physical distance among UEs reusing same PRBs is ensured in this case.

User-Density based Sectoring Scheme

The cell is divided into sectors based on UE density and first, the C-UEs are allocated orthogonal PRBs. The D-UEs are then allocated remaining orthogonal PRBs. If there are no PRBs left, then PRB reuse in the peer sectors is carried out as described in section 2.2.2.

Fixed Angle Sectoring

The RA procedure is similar to Density based sectoring but the number of sectors and angles of each sector are pre-defined. As shown in figure 5.3, 4

and 6 equal sized sector configurations are used. Since the number of UEs is each sector are non uniform, some of the D2D UEs might not be assigned any PRBs.

Random RA Scheme

Here the PRBs are reused in a random manner without any regards to distance / geo-location constraints. The PRBs are allocated to C-UEs in orthogonal manner and remaining PRBs are then assigned to D-UEs in a First in First Out (FIFO) manner. Any remaining D-UEs are allocated the same PRBs of any allocated UE randomly.

5.2.5 Evaluation Results

The cell with a central base station has N_{cell} number of C-UEs and N_{D2D} D-UEs. The total number of PRBs is $N_{RB_{tot}}$. A case with more number of D-UEs than C-UEs is considered such that,

$$N_{cell} < N_{RB_{tot}} < N_{D2D} + N_{cell} \leq 2 \times N_{RB_{tot}} \tag{5.1}$$

The number of sectors is set as 4 for both fixed angle and distance based RA schemes. Monte Carlo system level simulations are caried out in two scenarios. The first scenario has low SINR requirements with target SINR of C-UEs and D-UEs is set as -7 dB and 7 dB, respectively. These values are derived from practical perspective that, the minimal SINR value required for a cellular link to be established in LTE is -7 dB [JF12] . Further, in D2D mode, considering receiver sensitivity level at UE as -107.5 dBm, noise figure of 7 dB and a thermal noise of -121 dBm, the minimum SINR value at UE is found to be 6.5 dB [3GP10]. Figure 5.7 and figure 5.8 show respectively, the number of established links ($SINR \geq SINR_{th}$) and system capacity normalized to 100% with respect to Random RA scheme. The sectoring scheme with low signaling and positioning overhead is shown to perform on par with the distance based RA. Further, sectoring scheme gains $\approx 15\%$ more established links, $\approx 25\%$ more capacity with respect to random scheme.

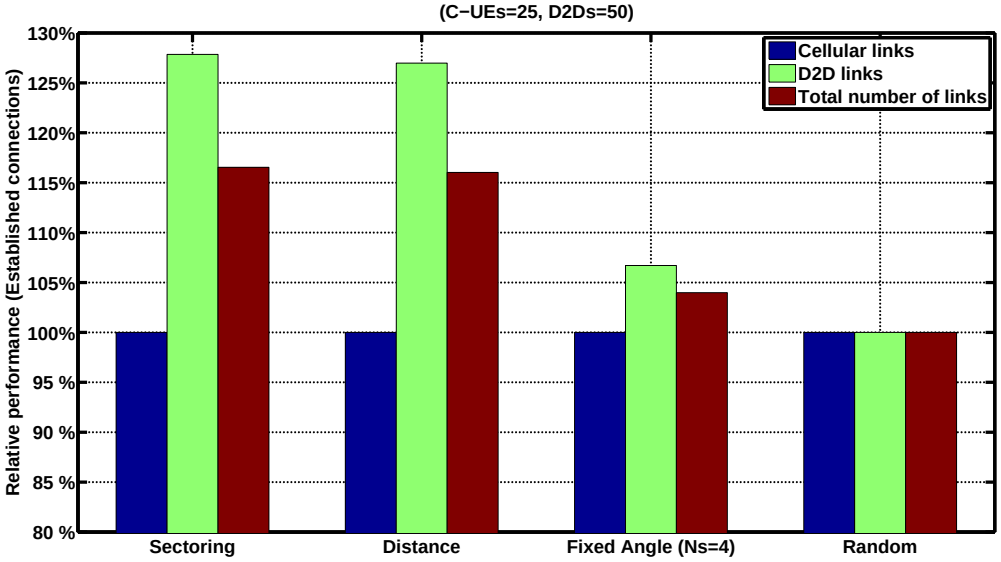

Figure 5.7: Percentage of Established Connections (low SINR)

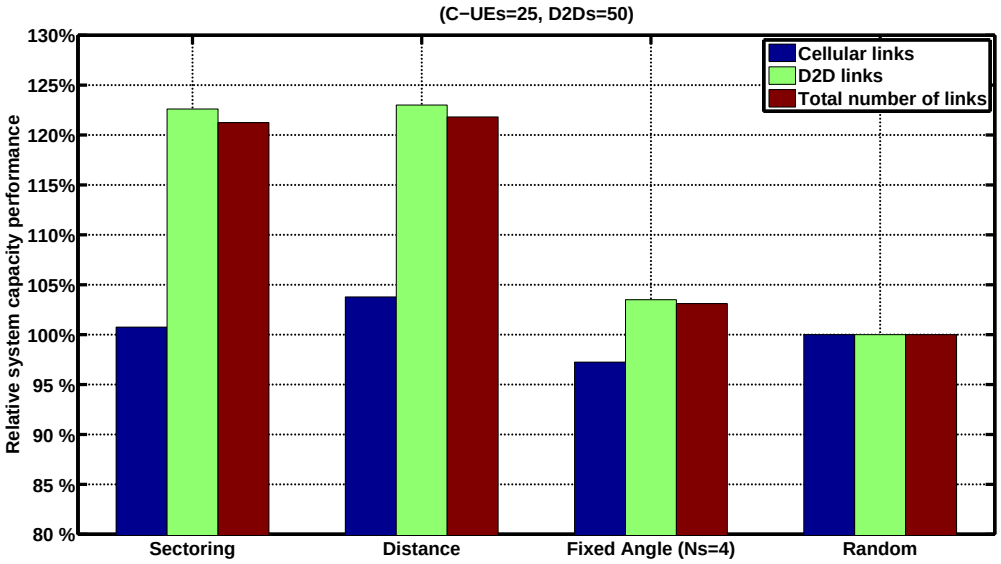

Figure 5.8: System Capacity (low SINR)

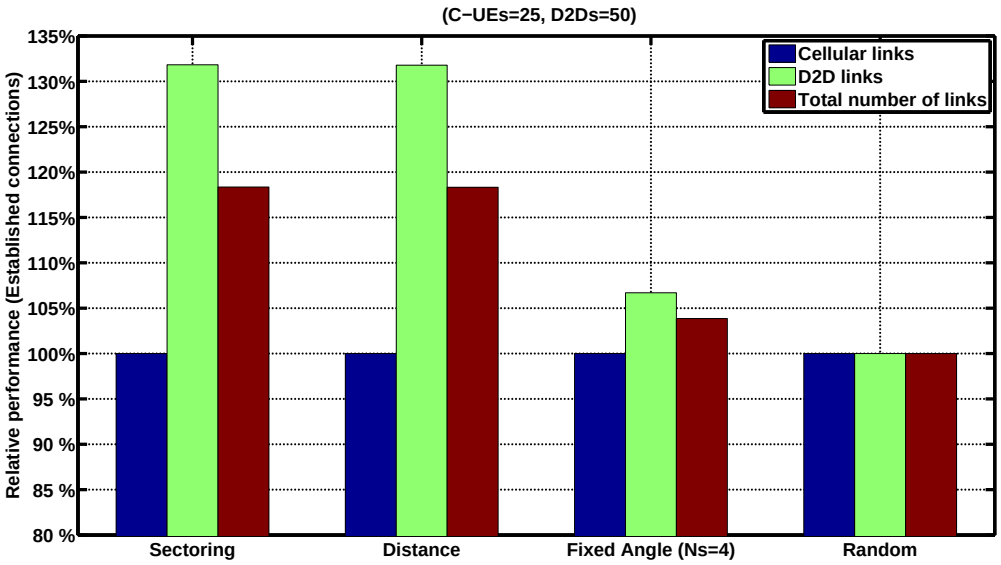

Figure 5.9: Percentage of Established Connections (high SINR)

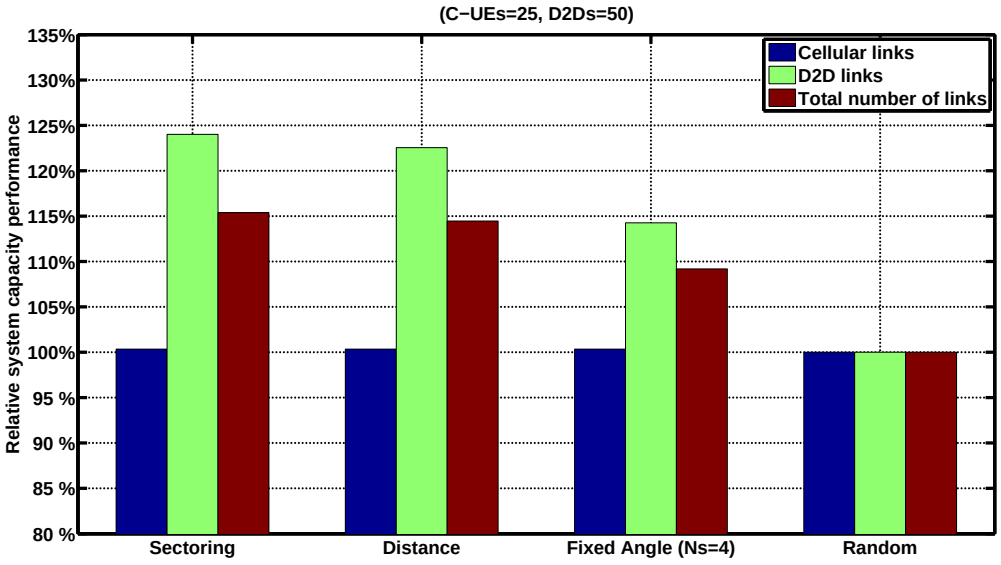

Figure 5.10: System Capacity (high SINR)

The second scenario with high SNR has target SINR for C-UEs and D-UEs as high as 8 and 10 dB respectively. From figure 5.9 and figure 5.10 respectively, it could be observed that sectoring scheme outperforms Random RA with $\approx 20\%$ more established connections and $\approx 15\%$ more system capacity. In comparison, the performance enhancement of the UE density based sectoring scheme is analogous (+-2%) to that of static sectored distance based approach. There can be two reasons for such a pattern; 1) Some of the UEs in considered sector may not be allocated any PRBs as a result

of fixed number of sectors(4) in the distance based scheme, 2) Diagonally opposite sectors does not necessarily mean farthest sectors. This may cause higher interference and outage if the UEs in vicinity are allocated the same set of PRBs. The virtual sectoring procedure ensuring uniform number of UEs per sector proves to be more robust and tolerant towards high cellular interference for the considered UE constellation.

***The results of this work have been published in* [SKL+15].**

5.3 Robustness of D2D RA against Positioning Errors

5.3.1 Introduction

In section 5.2, D2D RA schemes were presented which relied on location awareness in one form or another (either user distance or user angle). This section studies the robustness of such location aware D2D RA schemes against positioning errors. From this study the limitations of location based D2D RA algorithms are inferred and necessity of power control in tandem with D2D underlay is hinted. Elaborate study of power allocation in cellular networks with D2D underlay is presented in section 5.2.
The following location aware D2D RA schemes are considered to investigate their robustness with respect to positioning errors :

Distance Based RA

This scheme is similar to the one presented in section 5.2.4. The D2D links are allocated with PRBs from C-UEs that are farthest and separated by a distance of at least L_{min}. Where, L_{min} is the pre-selected distance constraint to control interference from selected C-UE to D-UE using same PRB [WC12]. In this section, the distance based resource allocation is extended such that, if PRBs from C-UEs at distances larger than L_{min} are already shared by D-UEs, then unallocated D-UEs can be assigned PRBs from remaining C-UEs. This makes sure that maximum number of D2D pairs are established in a cell and resource reuse factor of 1 is attained.

Virtual sectoring based resource allocation

The D2D RA scheme based on virtual sectoring has already been discussed in section 5.2. In section 5.2, the PRB reuse was allowed among C-UEs and D2D links as well as among D2D links themselves. However, the PRB reuse in this section is limited only between C-UEs and D2D links. This is done in order to allow (maximum) PRB reuse of 1 between C-UEs and D2D links. Further, this facilitates study of maximum impact of D2D interference to C-UE transmission in the uplink.
The following steps are carried out in this resource allocation scheme:

1. The angle of each user (C-UE and D-UE) in the cell are estimated based on Angle of Arrival (AOA) measurements.

2. Based on the specified number of virtual sectors in the cell, the total number of users is clustered according to observed user angles.

3. A D2D pair will reuse the radio resource from a C-UE belonging to the vertically opposite sector.

4. This procedure is carried out in each sector starting in counter-clockwise direction for the whole cell.

5. If there are any D2D pairs, for which no C-UE can be found in opposite sector, then search for C-UE is performed in sectors adjacent to opposite sector. If C-UE for resource reuse is found: STOP, else continue search.

6. The search is carried out exhaustively in all sectors except in the sector to which the corresponding D2D pair belongs. STOP.

The virtual sectoring based on user density is depicted in figure 5.11. Here, each sector exhibits an equal density of users. The inner circle represents the exclusion zone where no D2D transmissions are allowed.

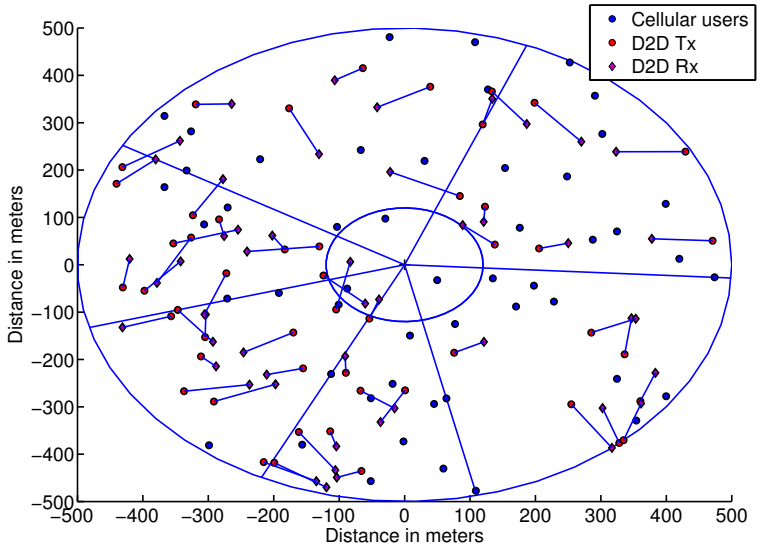

Figure 5.11: Virtual sectoring based on user density

Alternatively, only D2D transmitters can be used to find angular limits of each sector, instead of considering all the users. Therefore, uniform number of D2D transmitters will be present in each sector. Figure 5.12 illustrates virtual sectoring based on D2D density.

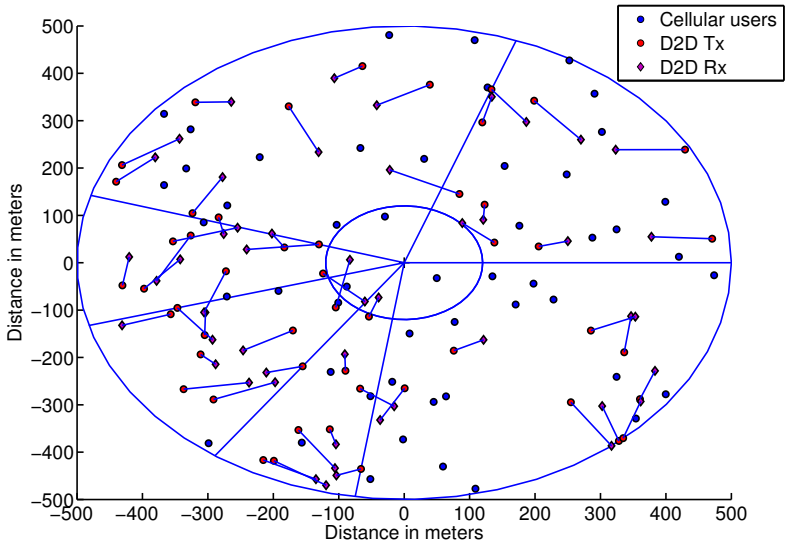

Figure 5.12: Virtual sectoring based on D2D density

The simplest method of virtual sectoring is to partition a cell into sectors using fixed angles as shown in figure 5.13. The RA procedure in both D2D density based sectoring and fixed angle sectoring is similar to virtual sectoring based on user density.

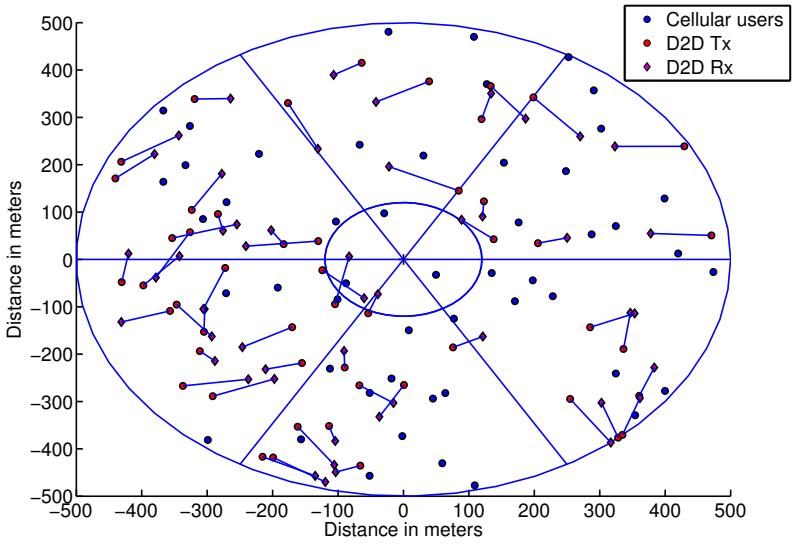

Figure 5.13: Virtual sectoring based on fixed angles

5.3.2 Evaluation

The location information of users (D-UE and C-UE) are important for the aforementioned schemes. These methods assume perfect position information. However, real world deployments face positioning inaccuracies which are influenced by position estimation method used, environmental factors, signal strength and so on. A single cell with radius of 500m is considered with central LTE base station. The C-UEs and D-UEs are uniformly distributed in this cell. Table 5.2 shows simulation parameters used. Bandwidth of 10 MHz bandwidth and 50 physical resource blocks (PRBs) are assumed [3GP10]. The resource reuse factor is 1, indicating that a D2D pair can share a PRB from only one unique C-UE. In order to evaluate the effect of positioning inaccuracies, position estimation errors are generated following uniform distribution. This is subsequently added to (x,y) co-ordinates of the users. The generation of positioning errors follow below steps:

- Choose at random err_x such that,
  $err_x \in uniform - distribution(a, b)$
- Choose at random err_y such that,
  $err_y \in uniform - distribution(a, b)$
- $x_{err} = x + err_x$
- $y_{err} = y + err_y$

- The values a and b are chosen such that $\sqrt{(x - x_{err})^2 + (y - y_{err})^2} \leq d$
- d is the intended inaccuracy in terms of distance.

The position inaccuracies are varied from 0 m to 120 m in steps of 10 m. The comparison is made between performance of the location based RA schemes for cases with perfect position information and instances with erroneous location estimates. Monte Carlo simulations is carried out with 1000 runs at each error step.

Table 5.2: Simulation Assumptions (D2D Robustness)

Parameter	Value (C-Mode)	Value (D-Mode)
Number of UEs	50	100 (50 pairs)
Transmit Power (eNB)	46 dBm	46 dBm
Max Transmit Power (UEs)	24 dBm	24 dBm
Power Control	Open Loop [3GP09]	Open Loop [3GP09]
Noise Figure	5 dB	7 dB
SINR Target (Rx)	-4 dB (eNB)	-2.5 dB
Channel Model	3GPP [3GP10]	3GPP [3GP13]

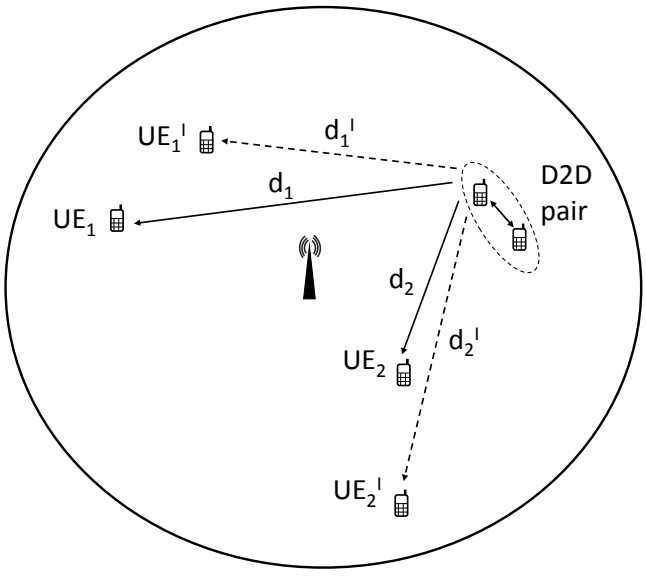

Figure 5.14: Effect of position estimation error in distance based RA

The impact of positioning error on distance based RA is shown in figure 5.14. d_1 and d_2 represent the distances of cellular UE_1 and UE_2 from D2D pair respectively, such that $d_1 > d_2$. However, in presence of positioning error, the erroneous distances are perceived to be $d_1^{'}$ and $d_2^{'}$, where $d_2^{'} > d_1^{'}$.

Therefore, distance based RA scheme permits PRB reuse between D2D pair and UE_2 instead of UE_1. Thus, interference to D2D receiver might be higher resulting in sub-optimal performance.
The influence of location estimation errors on virtual sectoring based RA is illustrated in figure 5.15. d_1 and d_2 are the distances of UE_1 and UE_2 from D2D pair respectively, with $d_1 > d_2$. Further, the D2D pair resides in a sector which is vertically opposite to sector where UE_1 is present. However, because of positioning error, D2D pair is identified in sector vertically opposite to the sector of UE_2. Thus, as per RA based on virtual sectoring, D2D pair is allowed to reuse PRB from UE_2 in the place of UE_1. This may negatively impact interference to D2D receiver resulting in poor performance.

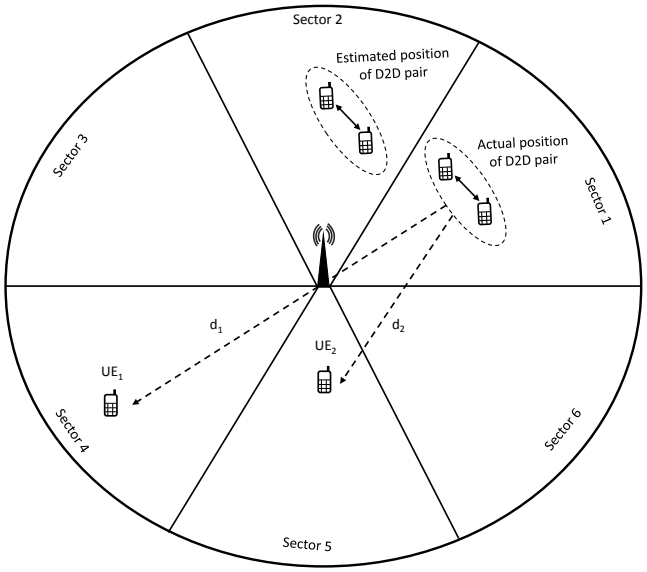

Figure 5.15: Effect of position estimation error in virtual sectoring based RA

Throughput of D2D users are compared in figure 5.16 for RA schemes with and without positioning errors. The key inference is that D2D links are found to be more robust to the interference from cellular users, than the other way around. Due to this, although a wrong judegment is made in PRB allocation to a D2D pair arising from position error, divergence in D2D throughput is minimal. The distance based scheme exhibits higher robustness but displays marginal deviation above position error of 90 meters. At error step of 100 meters, distance based scheme exhibits maximum degradation which is however around 1 Kbps.

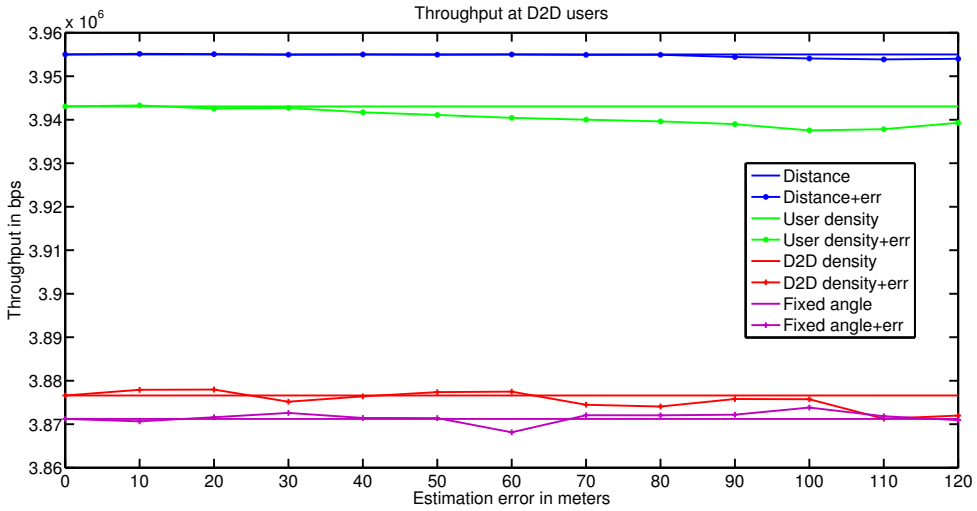

Figure 5.16: Throughput at D2D users

The User density based sectoring shows deviation above position error of 40 meters. When error step is 100 meters, the divergence is utmost and is around 5 Kbps. Since the RA schemes based on virtual sectoring are intuitive in nature rather than being based on complex calculations, positioning errors could also result in better assignments than originally intended. This pattern can be observed in instances of fixed angle and D2D density based sectoring schemes. Here, at certain error samples, D2D throughput is improved slightly. The maximum boost in D2D throughput due to D2D density based sectoring is 1 Kbps at error of 20 meters and the maximum degradation is 5 Kbps at error of 110 meters. The D2D throughput is enhanced utmost by 3 Kbps at error of 100 meters when sectoring is based on fixed angle. In the same context, maximum degradation is 3Kbps at error of 60 meters. The D2D links are found to be very stable and are least affected by the interference from C-UEs. Further, the estimation error exists at both D2D and C-UEs. Thus, when a certain RA scheme is used, throughput plot for D2D users display marginal deviations from original throughput curve. The behavior of specific RA scheme is followed even when estimation error is present.

The deviations in throughput at base station due to positioning errors are depicted in figure 5.17. There exists mutual interference between D2D users and C-UEs sharing the same PRB. Typically, whenever the throughput at D2D users degrade as a result of wrong assignments due to positioning errors, throughput at base station (for C-UEs) exhibits improvement. The maximum degradation in throughput at BS exhibited by distance based scheme

is around 5 Kbps at error of 100 meters. For user density based sectoring, maximum reduction is around 4 Kbps at error of 120 meters, whereas maximum improvement is around 6 Kbps at error of 90 meters. In case of D2D density based sectoring, maximum addition to throughput at BS is 3 Kbps at error of 60 meters and the maximum degradation is found to be 5 Kbps at error of 120 meters. The maximum reduction in throughput at BS displayed by fixed angle based sectoring is 9 Kbps at error of 90 meters.

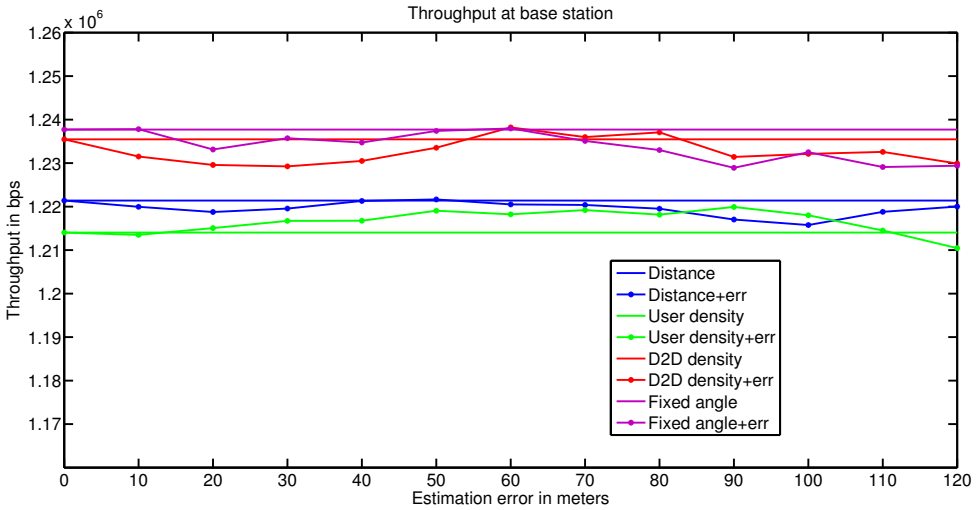

Figure 5.17: Throughput at base station

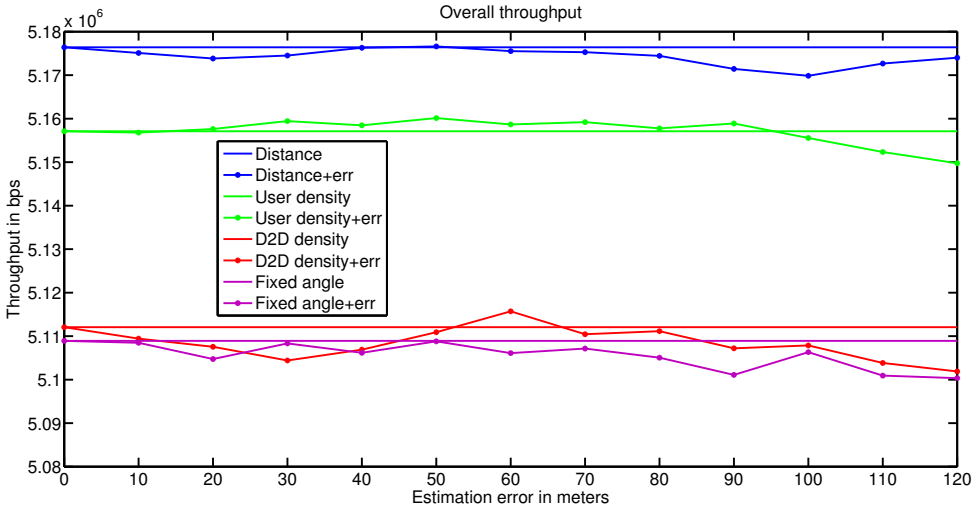

Figure 5.18: Overall throughput

The deviations noted in overall throughput due to positioning errors is shown in figure 5.18. The utmost deterioration in overall throughput shown

by distance based scheme is around 6 Kbps at error of 100 meters. For user density based sectoring it is around 7 Kbps at error of 120 meters and enhancement is around 3 Kbps at error of 50 meters. The maximal addition to overall throughput exhibited by D2D density based sectoring is 4 Kbps at error of 60 meters and the utmost degradation is 5 Kbps at error of 110 meters. The highest degradation in overall throughput displayed by fixed angle based sectoring is 9 Kbps at error of 120 meters. When distance error is incorporated, the error values are added to (x, y) co-ordinates of users to get $(x^{'}, y^{'})$ such that maximum difference in terms of distance between (x, y) and $(x^{'}, y^{'})$ is a predefined value 'd'. (d is varied from 0 to 120 meters in steps of 10 meters). When angular error is considered, the error component is added to user angle ($theta$) giving rise to $theta^{'}$, such that maximum difference in terms of angle between $theta$ and $theta^{'}$ is a predefined value 'a' (a is varied from 0 to 40 degrees in step of 5 degrees). In both cases, the position of user is changed and it influences the PRB allocations.

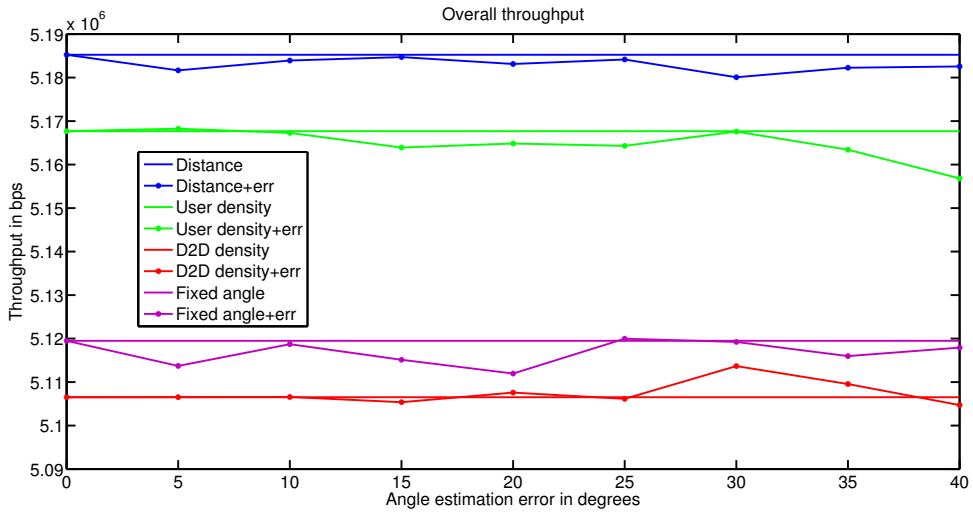

Figure 5.19: Overall throughput (Angular estimation error)

The variations observed in total throughput due to angular estimation errors are shown in figure 5.19. The user angles are appended intentionally with angular errors and performance comparison of D2D RA schemes is carried out with and without angular errors. The angular error is varied from 0 degrees to 40 degrees in steps of 5 degrees. Monte Carlo simulation is done with 1000 runs at each error step. The overall throughput for distance based RA scheme shows maximum degradation of 5 Kbps at error of 30 degrees. The user density based sectoring exhibits maximum deterioration of 11 Kbps at error of 40 degrees. The D2D density based sectoring shows utmost en-

hancement of 7 Kbps at error of 30 degrees and highest degradation is around 2 Kbps at angular errors of 15 degrees and 40 degrees. The fixed angle based sectoring method displays maximum degradation of 7 Kbps at error of 20 degrees.

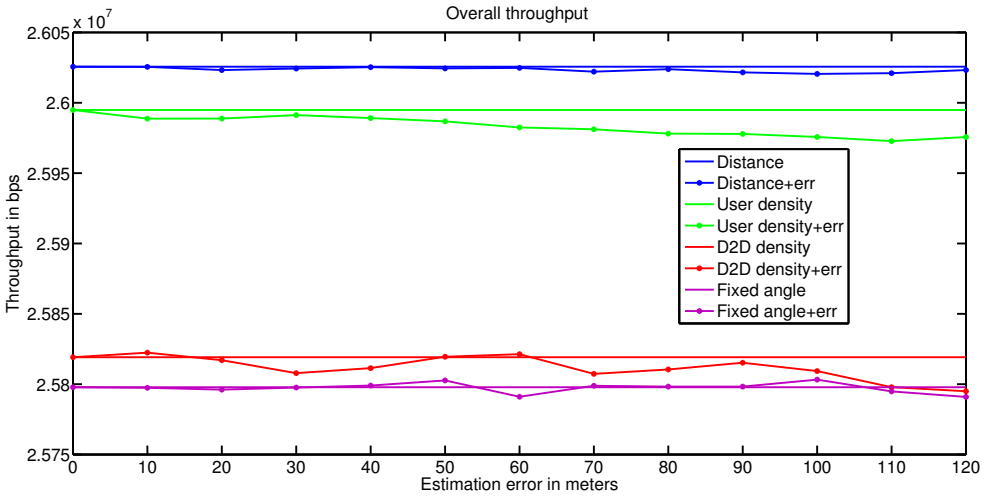

Figure 5.20: Overall throughput (High SINR targets)

The variations noted in total throughput due to positioning errors is depicted in figure 5.20, when high SINR targets are set for C-UEs and D-UEs . The SINR target for C-UEs is 8 dB and SINR target for D-UE is 10 dB. The distance based RA scheme is observed to be most robust with largest deterioration in total throughput being 10 Kbps at error of 100 meters. For user density based sectoring, utmost reduction is observed at error of 110 meters which is around 20 Kbps. D2D density based sectoring exhibits largest degradation of 30 Kbps at error of 120 meters. The highest reduction in overall throughput observed in fixed angle based sectoring is around 10 Kbps at error of 60 and 120 meters.

From figures 5.16, 5.17, 5.18, 5.19 and 5.20 it could be noted that location based D2D RA methods exhibit non linear variations due to positioning errors. However, the maximum degradation recorded is in order of 11 (low SINR targets) and 30 Kbps (high SINR targets), which are marginal in comparison to overall throughput which is in order of Mbps. The sectoring scheme is intuitive and designed to be simple. The sophisticated inputs such as pathloss, power levels or radio conditions of user are not regarded for making PRB assignments. Influenced by positioning error, PRB assignments would vary and at some instances they may perform better than original PRB allocations without errors. Nevertheless, the noted enhance-

ments are marginal.

The evaluation yields average performance results that do no reflect the situation of a particular user in a very bad receive situation. At every error step, evaluation considers 1000 simulation runs. This results in combination of several possible scenarios of good and bad receive conditions. Thus, evaluation reflects an average performance of location based D2D RA methods under influence of positioning errors. Evaluation is done with low and high SINR targets for C-UEs and D-UEs. The recorded divergence from actual throughput plots are in the order of Kbps and is minimal in comparison to overall throughput which is in the order of Mbps. Thus, it is inferred that as long as the users are contained in the same cell (e.g. connected to same base station,or bear same cell-id), positioning errors do not have any major impact on performance of D2D RA algorithms. Further, during this study it also led to the observation that D2D interference to uplink transmission of C-UE is not treated with higher priority in any of the location based D2D RA scheme. Thus, power allocation for C-UEs to co-exist with D2D underlay is of high importance and discussed in detail in section 5.4.

***The results of this work have been published in* [KKL+15].**

5.4 Post-RA Power Allocation in Cellular Networks

5.4.1 Introduction

D2D communication can be carried out in two major ways, namely: 1) Overlay: D2D communication takes place over dedicated spectrum and is orthogonal to frequencies used by conventional cellular users in a cell [LZLS12]. 2) Underlay: D2D links and conventional cellular links share the available spectrum in a cell for their transmissions [LZLS12]. Although D2D overlay is simple, it doesn't allow spectral re-usability. Whereas, D2D underlay allows for better spectrum utilization. However, the key issue in D2D underlay is mutual interference among D2D pairs and cellular users while reusing the same frequency resources. This interference would hamper the SINR (signal to interference plus noise ratio) at receivers, there by reducing the performance.

There are several resource allocation schemes in literature that discuss resource reuse between D2D pair and a cellular user that would cause minimum

interference to each other [WC12][SKL+15][KKL+15][ZJLZ14]. Further, several proposals are made to regulate transmission power of D2D users to mitigate interference [LLAH15][MBM+14][ZJLZ14]. Some proposals emphasize on selection of either D2D or conventional modes of transmission based on constraints of SINR, throughput etc,[ZJLZ14], and some allow the usage of D2D only in a specified region [BY12] fairly away from the base station.

However, majority of these schemes impose restrictions on usage of D2D communication, confining them to certain power levels or allowing them only in certain cases. Thus, maximum benefit cannot be reaped from D2D communications. On the contrary, D2D links would pose high interference to cellular links with which they reuse physical resource blocks (PRBs) and cause significant dropping of cellular connections due to interference.

In this work, a post-resource allocation (RA) power control for cellular users is proposed as a solution against interference caused by D2D transmissions on shared PRBs. Three post-RA power control (PC) schemes namely: Interference aware PC, blind PC and threshold based PC are introduced and feasibility of their implementation are discussed in detail. Simulation results demonstrate that the proposed schemes reduce dropping of cellular connections due to interference, improve cellular throughput while having no negative impact on D2D performance. The proposed schemes makes use of context information about PRB reuse between D2D transmitters and cellular users to trigger power control mechanism. Further, interference aware PC exploits the location information of D2D transmitter.

5.4.2 System Model

Figure 5.21 describes the interferences caused due to D2D underlay in a cell. In the considered scenario D2D users share uplink resources with cellular users (CUE). Figure 5.21a) depicts the interference at base station caused by a D2D transmitter (UE_1), to uplink transmissions from a cellular user (UE_3). Both D2D pair (UE_1 and UE_2) and cellular user (UE_3) are sharing the same physical resource block (PRB_i). Simultaneously, uplink transmission from cellular user (UE_3) poses interference at D2D receiver (UE_2), as shown in figure 5.21b). However, interference from D2D transmitter to base station is observed to be stronger than the interference from a CUE to a D2D receiver. This was inferred based on simulations incorporating 3GPP channel models for UE to UE (D2D) [3GP13] and UE to base station [3GP10]. This is also attributed to the fact that D2D users are in close proximity to one another, generally at line of sight with better reception and more robust

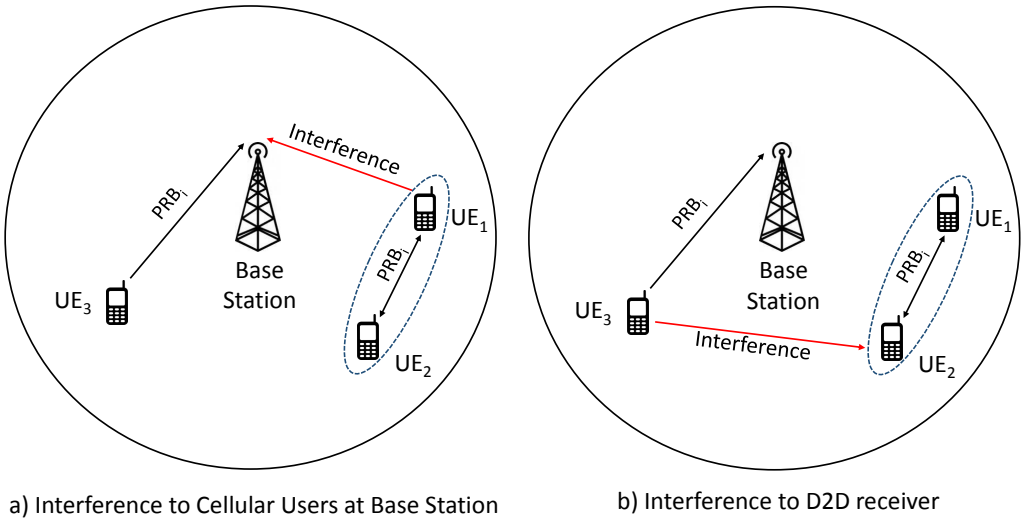

Figure 5.21: Interferences in D2D scenario

to interference from CUE. Hence, CUEs suffer more than D2D users due to interference. When received SINR levels at base station, of transmissions from CUEs are very low due to interference, the corresponding CUEs are considered dropped.

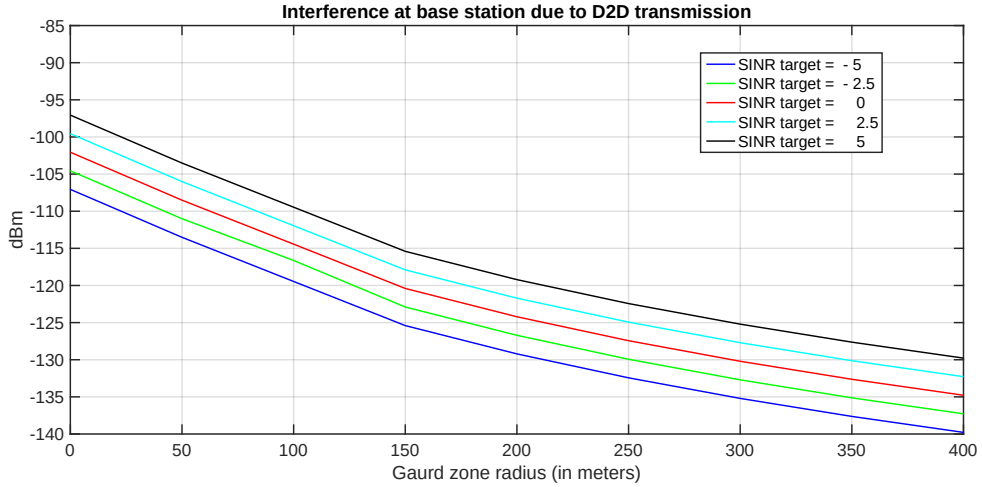

Figure 5.22: Effect of guard zone

One of the simplest ways to reduce interference caused by D2D transmissions is to make sure that a D2D transmitter is sufficiently away from the base station [BY12]. By defining such guard zones around base station where

no D2D transmissions are allowed, the interference at base station could be regulated [BY12]. This is demonstrated in figure 5.22, where the guard zone radius is varied from 0 to 400 m, in a cell of radius 500 m. The interference caused at base station reduces as origin of D2D transmission move away from the base station. This example is for a single D2D pair separated by 50 m. However, the effects will be non linear with many D2D pairs and would vary with D2D proximity range, power levels, SINR targets, channel conditions etc. Further, restricting D2D transmissions outside the guard zone will not allow maximal exploitation of D2D transmission opportunities. Thus, as a solution to overcome D2D interference, we propose power control mechanisms at CUEs, after the resource allocation for D2D users is carried out. The resource allocation (RA) mechanism specifies which CUEs and D2D pairs can reuse PRBs among them. In this work, distance based resource allocation for D2D transmission is used [WC12]. This scheme allows D2D pair to reuse PRB from a CUE which is at the farthest distance from it [WC12][KKL$^+$15].

5.4.3 Interference Aware Power Control

The interference aware power control follows the same principle as LTE uplink power control [3GP09]. This scheme is carried out on the CUEs sharing PRBs with D2D users. The CUEs which have their SINR lesser than the target SINR due to interference, will be boosted with power levels sufficient to overcome the interference. Equations 5.2 and 5.3 are similar to LTE open loop power control [3GP09], but has been designed to overcome interference from D2D transmission along with pathloss compensation. The logic of the scheme is described in algorithm 5.4.1.

Denotations of terms used in algorithm 5.4.1 are as below,

- $SINR_{cue} \Rightarrow$ SINR of relevant CUE
- $SINR_t \Rightarrow$ intended target SINR
- $\alpha \Rightarrow$ path loss compensation factor
- $PN_0 \Rightarrow$ average noise power per PRB
- $NF \Rightarrow$ noise figure at base station
- $M_0 \Rightarrow$ number of PRBs
- $PL \Rightarrow$ pathloss of the CUE w.r.t base station

Algorithm 5.4.1: Post-RA Interference Aware Power Control

Data: Obtain CUEs sharing PRBs with D2D pairs

if $SINR_{cue} < SINR_t$ **then**

$$P_0 = \alpha * (SINR_t + IN) + (1-\alpha) * (P_{max} - 10log_{10}(M_0)) \tag{5.2}$$

where, $IN = (PN_0 + NF)_{dB} + \left(\widetilde{I}_{D2D}\right)_{dBm}$

$$P_{cue} = P_0 + \alpha * PL + 10log_{10}(M_0) \tag{5.3}$$

if $P_{cue} > P_{max}$ **then**

$P_{cue} = P_{max}$

else if $P_{cue} < P_{min}$ **then**

$P_{cue} = P_{min}$

else

Continue

else

Continue without changing P_{cue}

- $P_{max} \Rightarrow$ maximum allowed power level at CUE
- $P_{min} \Rightarrow$ minimum allowed power level at CUE
- $P_{cue} \Rightarrow$ power allocated to CUE
- $\widetilde{I}_{D2D} \Rightarrow$ estimated interference from D2D pair

Interference from a D2D pair on a PRB shared with a CUE is given as,

$$\widetilde{I}_{D2D} = P_D - \widetilde{PL}_{Dbs} \tag{5.4}$$

where,
$P_D \Rightarrow$ Power level of D2D transmitter
$\widetilde{PL}_{Dbs} \Rightarrow$ Estimated pathloss between D2D transmitter and base station

The key component for this algorithm is interference estimate ($\widetilde{I}_{D2D}$) from the D2D transmission on a PRB. The assumption here is that the power level of the D2D transmitter (P_D) is known by base station, since control signaling is still carried out via base station. Further, based on the location information of D2D transmitter, pathloss estimation ($\widetilde{PL}_{Dbs}$) could be made

at base station. This can be helpful in obtaining an interference estimate, that needs to be overcome by a CUE to reach the SINR target.

Further, this algorithm is more feasible to be used, when resource reuse agreement between D2D pairs and CUEs is valid for a multiple number of TTIs (e.g., low mobility/static scenario such as crowded stadium, open air festivals etc.). This enables, easier estimation of $\widetilde{I}_{D2D}$ and its incorporation in subsequent TTIs. The effectiveness of this algorithm is enhanced with better interference estimation techniques. Efficient D2D interference estimation is an open problem that needs to be further investigated.

5.4.4 Blind Power Control

As a simpler alternative to interference aware PC, the blind power control is proposed. This algorithm doesn't require an estimation of interference caused by D2D transmissions, thus highly reducing complexity. The algorithm follows similar steps as LTE open loop power control [3GP09]. However, when SINR of a CUE sharing its PRB with a D2D pair is found to be lesser than target SINR at base station, then a predefined power step (in dB) is blindly added to transmission power of CUE. This is depicted in equation 5.6. The idea of the scheme is described in algorithm 5.4.2.

Although this algorithm is simple to implement, care should be taken while choosing power step ($step_{blind}$) to obtain CUE transmission power (P_{cue}). Choosing larger power step would lead to wastage of power at CUEs, which might reach SINR target with rather smaller power step than chosen one. Choosing smaller power step would make CUEs unable to reach SINR target and chances of CUEs still being dropped due to interference might arise.

5.4.5 Threshold Based Power Control

A third approach towards post-RA power control at CUEs called threshold based PC is described in this section. This approach neither requires interference estimation nor chooses power step blindly. The algorithm design is similar to blind power control but the power step is not blind and is derived based on certain logic. The algorithm is carried out on CUEs sharing PRBs with D2D pairs, when SINR of respective CUE drops below target SINR. The principle is shown in algorithm 5.4.3.

Unlike blind power control, power step chosen in this algorithm is correlated with SINR of corresponding CUE. If the SINR of a CUE is drastically

Algorithm 5.4.2: Post-RA Blind Power Control

Data: Obtain CUEs sharing PRBs with D2D pairs

if $SINR_{cue} < SINR_t$ **then**

$$P_0 = \alpha * (SINR_t + PN_0 + NF) + (1 - \alpha) * (P_{max} - 10log_{10}(M_0)) \quad (5.5)$$

$$P_{cue} = P_0 + \alpha * PL + 10log_{10}(M_0) + step_{blind} \quad (5.6)$$

  if $P_{cue} > P_{max}$ **then**
    $P_{cue} = P_{max}$
  else if $P_{cue} < P_{min}$ **then**
    $P_{cue} = P_{min}$
  else
    Continue

else
  Continue without changing P_{cue}

below the SINR target, then a larger power step is chosen to overcome high interference situation. Else if the SINR of a CUE is slightly below the target SINR, then a smaller power step is used. This way, instead of using a blind power step, a power step based on SINR thresholds can be used. This minimizes the wastage of power resources while still boosting power levels based on interference situation. An example logic to derive a power step, based on SINR thresholds is described in algorithm 5.4.4. The resolution of SINR thresholds and power steps corresponding to different interference situations can be customized by network operators as required, by monitoring D2D scenario and resulting SINR degradations over a period of time.

5.4.6 Evaluation

Uniform user distribution in a single cell

This section discusses the simulation set up used and evaluates the performance of aforementioned PC algorithms. A single cell scenario is considered with a base station at center, enabled with LTE radio access technology. A simple depiction of the scenario is shown in figure 5.23. The cell radius

Algorithm 5.4.3: Post-RA Threshold Based Power Control

Data: Obtain CUEs sharing PRBs with D2D pairs

if $SINR_{cue} < SINR_t$ **then**

$$P_0 = \alpha * (SINR_t + PN_0 + NF) + (1 - \alpha) * (P_{max} - 10log_{10}(M_0)) \quad (5.7)$$

$$P_{cue} = P_0 + \alpha * PL + 10log_{10}(M_0) + step_{th} \quad (5.8)$$

where $step_{th}$ is deduced from algorithm 5.4.4

if $P_{cue} > P_{max}$ **then**

$P_{cue} = P_{max}$

else if $P_{cue} < P_{min}$ **then**

$P_{cue} = P_{min}$

else

Continue

else

Continue without changing P_{cue}

Algorithm 5.4.4: Example SINR-threshold based power step deduction

if $(SINR_t - 3) \leq SINR_{cue} < SINR_t$ **then**

$step_{th} = 1\ dB$

else if $(SINR_t - 6) \leq SINR_{cue} < (SINR_t - 3)$ **then**

$step_{th} = 4\ dB$

else if $(SINR_t - 9) \leq SINR_{cue} < (SINR_t - 6)$ **then**

$step_{th} = 7\ dB$

else if $SINR_{cue} < (SINR_t - 9)$ **then**

$step_{th} = 10\ dB$

else

$step_{th} = 0\ dB$

is 500m and has 10 MHz bandwidth (50 PRBs) [3GP10]. 50 cellular users (CUEs) and 50 D2D users (25 D2D pairs) are uniformly distributed in the cell. The resource reuse factor is 1, which means a D2D pair can reuse a PRB from only one unique CUE. A simple distance based resource allocation [WC12] is implemented, allowing PRB reuse between CUEs and D2D pairs. The simulation parameters are listed in table 5.3. The key parameters for power control are chosen as $\alpha = 1$, $M_0 = 1$ and $PN_0 = -121dBm$ [3GP09]. The power control for cellular users (C-Mode) in table 5.3 refers to the open loop power control of CUEs while not sharing resources with D2D pairs (also prior to D2D resource allocation).

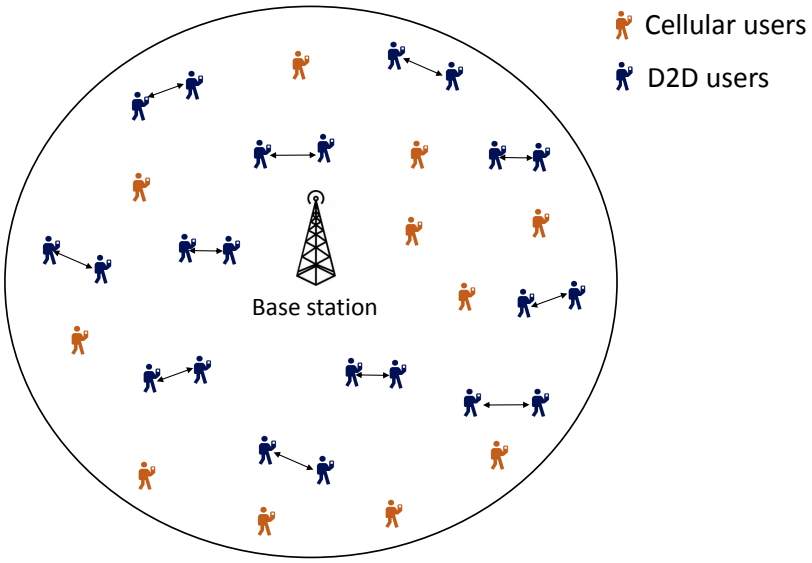

Figure 5.23: Simulation scenario

Table 5.3: Simulation Assumptions (D2D PC)

Parameter	Value (C-Mode)	Value (D-Mode)
Number of UEs	50	50 (25 pairs)
Transmit Power (eNB)	46 dBm	46 dBm
Max Transmit Power (UEs)	24 dBm	24 dBm
Min Transmit Power (UEs)	-40 dBm	-40 dBm
Power Control	Open Loop [3GP09]	Open Loop [3GP09]
Noise Figure	5 dB	7 dB
SINR Target (Rx)	-4 dB (eNB)	2.5 dB
Channel Model	3GPP [3GP10]	3GPP [3GP13]

Performance of interference aware PC at CUE is evaluated against D2D SINR targets ranging between -5 dB to +5 dB. The simulations are executed

for 100 runs and average performance metrics are obtained. Figure 5.24 shows an enhancement in throughput at the base station resulting from interference aware PC at CUEs. It can been seen that with increasing SINR target at D2D, interference posed for cellular transmission is higher. This is reflected as declining throughput at base station as D2D SINR targets increase. However with post-RA interference aware PC, the throughput performance proves to be better than without post-RA PC.

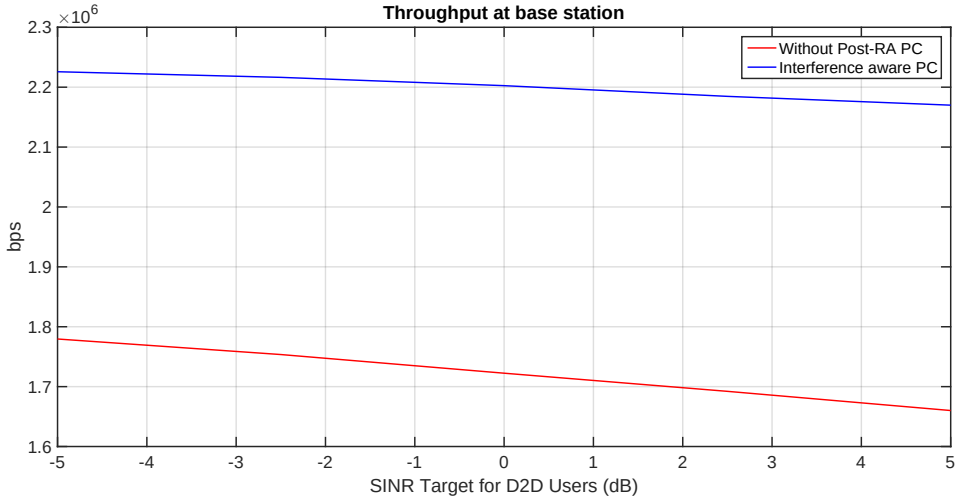

Figure 5.24: Improvements in throughput at base station after post-RA interference aware PC

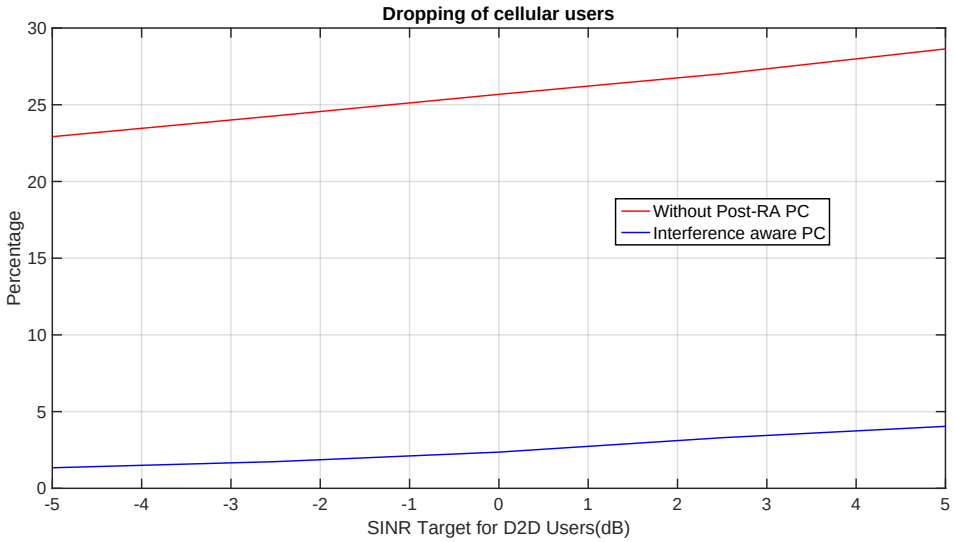

Figure 5.25: Reductions in cellular dropping after post-RA interference aware PC

Figure 5.25, shows reductions in dropping of cellular users due to high interference from D2D transmissions. It can be seen that post-RA interference aware PC keeps the cellular dropping below 5% for an entire range of considered D2D SINR targets. Further, figure 5.26 compares the overall throughput and it could be seen that interference aware PC impacts the overall throughput positively. It is also observed that the rise in interference at D2D receivers due to post-RA interference aware PC at CUE, is so low that there is no negative impact on the performance of D2D transmissions.

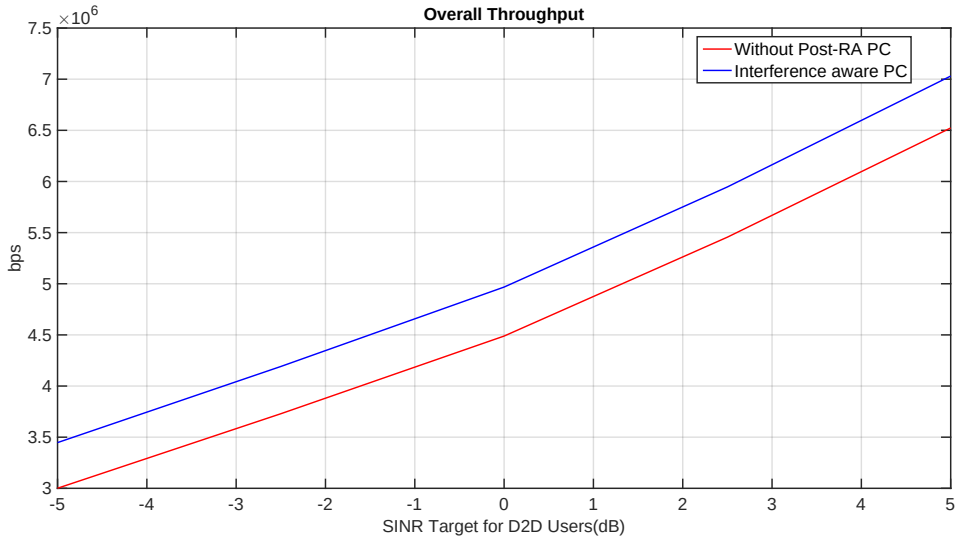

Figure 5.26: Improvements in overall throughput after post-RA interference aware PC

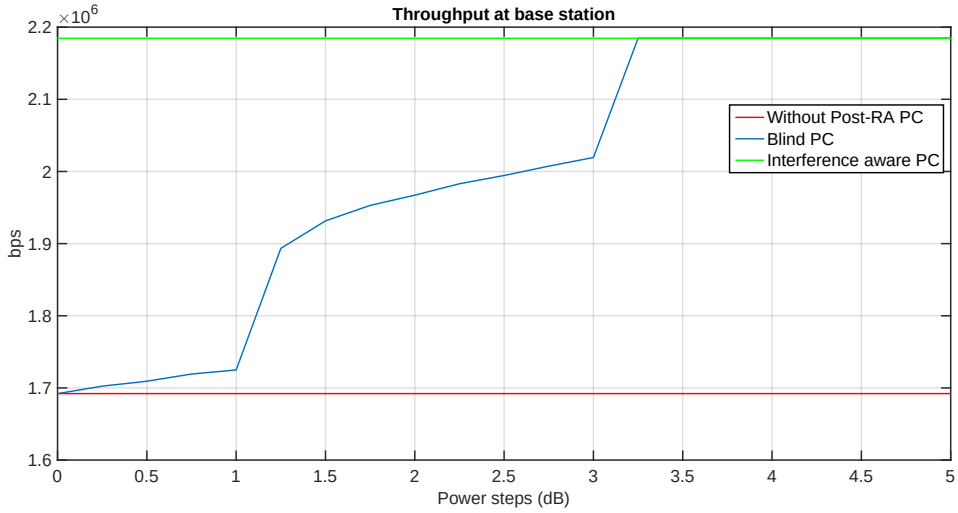

Figure 5.27: Improvements in throughput at base station after post-RA blind PC

In order to evaluate the performance of blind power control, power steps ranging from 0 to 5 dB are chosen. The simulations are run for 100 times and average performance metrics are obtained at each power step. Figure 5.27 depicts the throughput at the base station. It could be seen that with an increase in the value of a power step, throughput at the base station increases. However, for the considered simulation parameters, target data rate would have been reached (indicated in green) at a power step value of 3.25 dB. The

power steps above this are unnecessary and are used here to reflect wastage of power resources. Therefore, choice of power step is a key concern in blind power control.

Figure 5.28 shows dropping of cellular users due to D2D interference. Blind PC reduces dropping by up to 5% for all the considered range of power steps. This is because blind PC would add power in surplus than required for some users, where as in some cases added power would not be sufficient to overcome high interference and CUE would still be dropped.

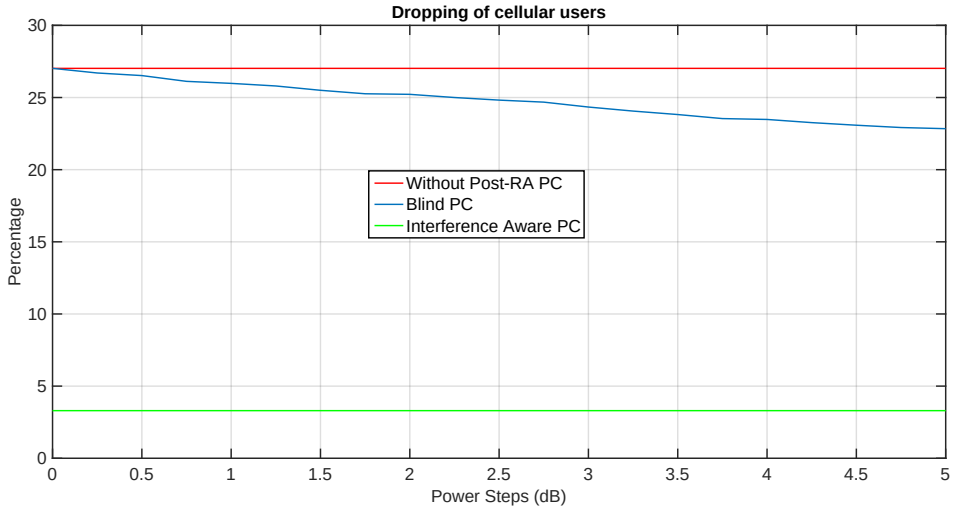

Figure 5.28: Average cellular dropping after post-RA blind PC

Finally, the threshold based PC is incorporated and is compared with the interference aware PC and blind PC. Based on the results in figure 5.27, a power step of 3.25 dB is chosen for blind PC. 100 simulation runs are carried out and average performance metrics are derived. It could be seen from the figure 5.29 that dropping of CUEs are reduced by around 6% due to threshold based PC and it performs better than blind PC. Figure 5.30 compares the throughput at base station. Interference aware PC performs better than the rest. Although blind PC has slightly better throughput than threshold based PC, there is no guarantee that all CUEs are connected as reflected in figure 5.29 and surplus power than required at some CUEs have contributed to better average throughput at base station.

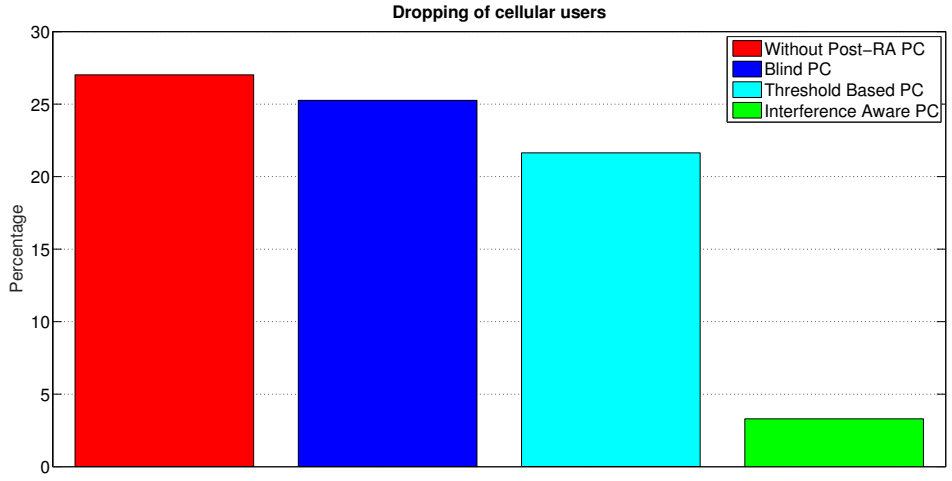

Figure 5.29: Comparison of average cellular dropping

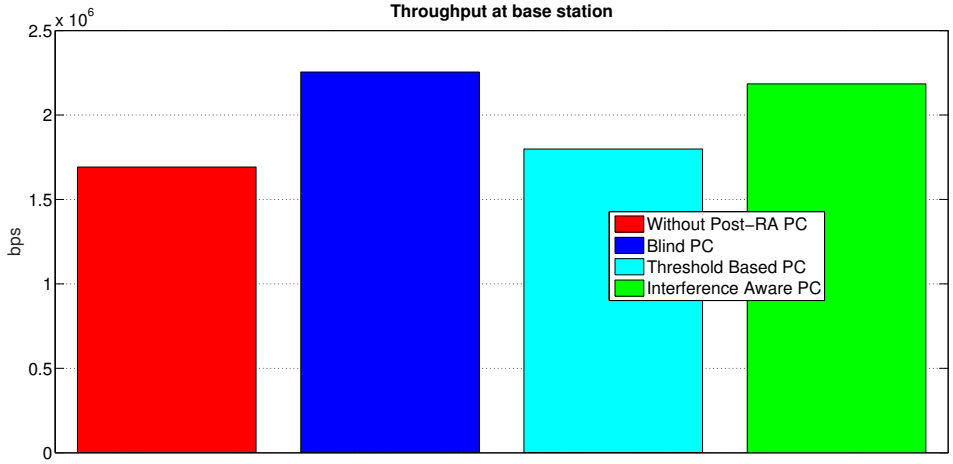

Figure 5.30: Comparison of throughput at base station

Mobile Crowd in a Macro Cell

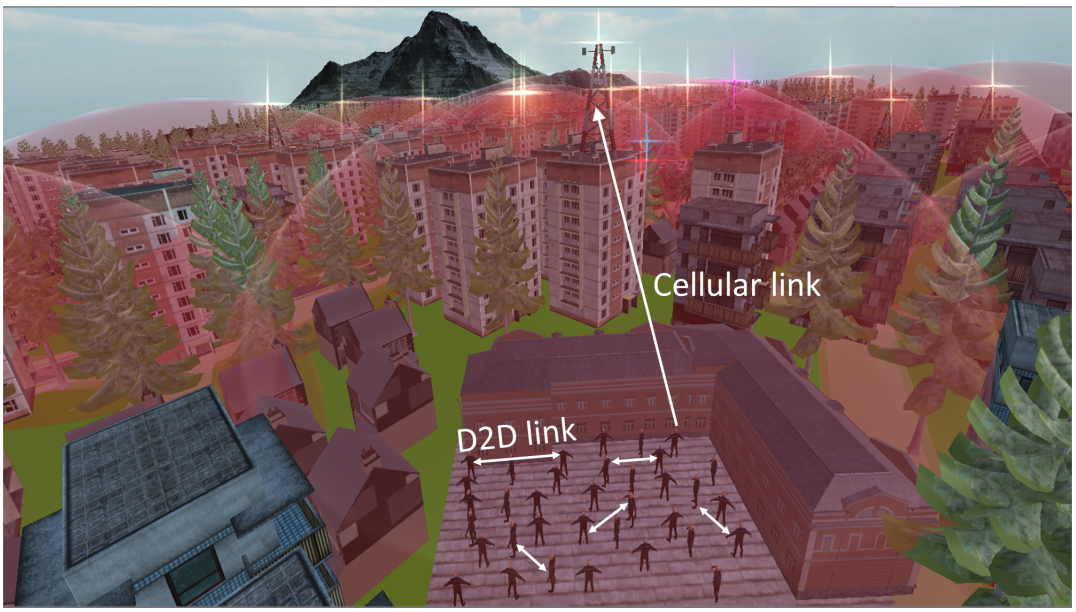

Figure 5.31: Illustration of crowd scenario

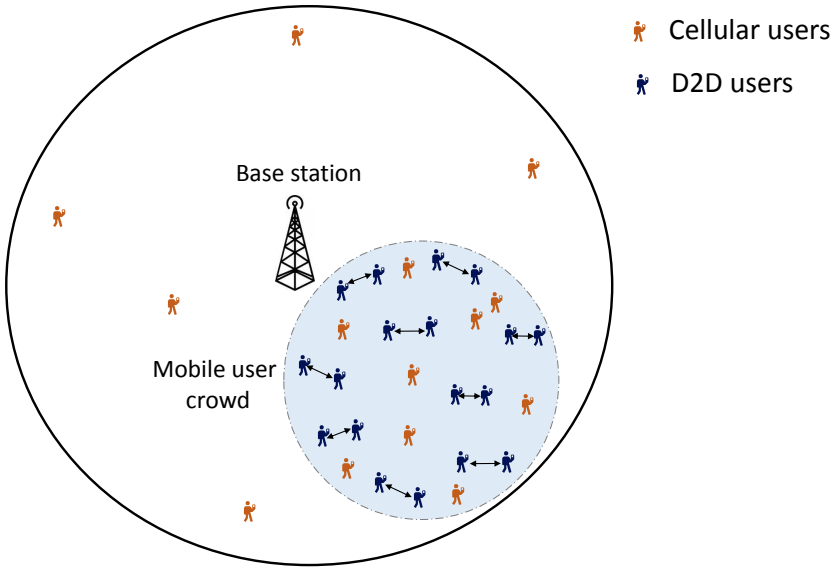

Figure 5.32: Crowd scenario in macro cell

The post-RA PC schemes are also implemented and studied in a crowd scenario in macro cell. A simple illustration of such a scenario is depicted in figure 5.31, where a crowd of mobile users are gathered in a specific region of a cell (e.g., city hall, shopping mall, public event etc.). D2D underlay is

enabled in such scenario, leading to establishment of D2D links and reuse of PRBs from cellular users. The simulation parameters follow table 5.3. A LTE macro cell of radius 1000 meters with 10 MHz bandwidth (50 PRBs) [3GP10] is considered. A crowd of 50 cellular users (CUEs) and 50 D2D users (25 D2D pairs) is uniformly distributed in a circular area (radius 500 m) in a section of the cell. The exemplary scenario is as shown in figure 5.32. In each presented case of PC, simulations are executed for 100 runs and average performance metrics are obtained.

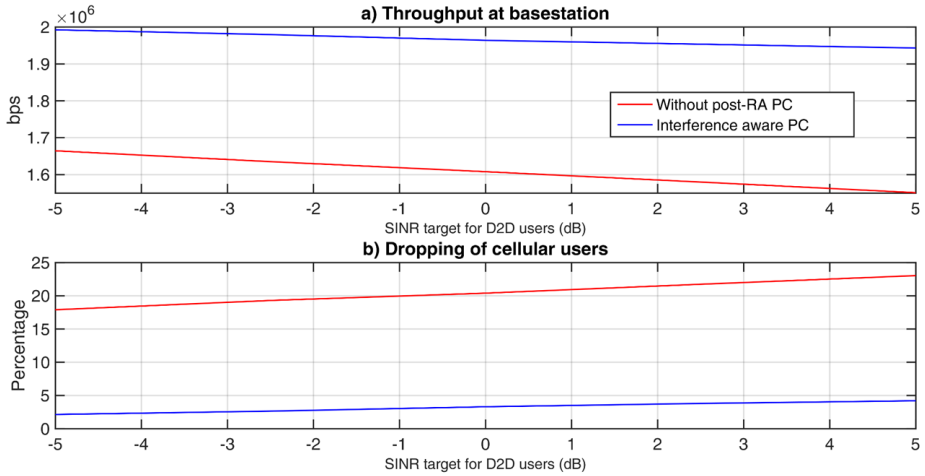

Figure 5.33: Improvements after interference aware PC

D2D SINR targets are varied between -5 dB to +5 dB. Figure 5.33 a) shows improvements in throughput at base station due to interference aware PC at CUEs. With interference aware PC, throughput performance is stabilized. It can be seen from figure 5.33 b) that for whole range of considered D2D SINR targets, interference aware PC reduces the cellular dropping by around 20%.

Power steps ranging from 0 to 5 dB are considered in order to evaluate the performance of blind PC. Figure 5.34 a) depicts that with increase in value of power step, throughput at base station increases. However, for the considered simulation parameters, at power step value of 3 dB, target data rate would have been reached (indicated in green). The power steps above this are unnecessary and reflects wasted power resources. Figure 5.34 b) shows that, for all the considered range of power steps, blind PC reduces dropping by only 2-3%.

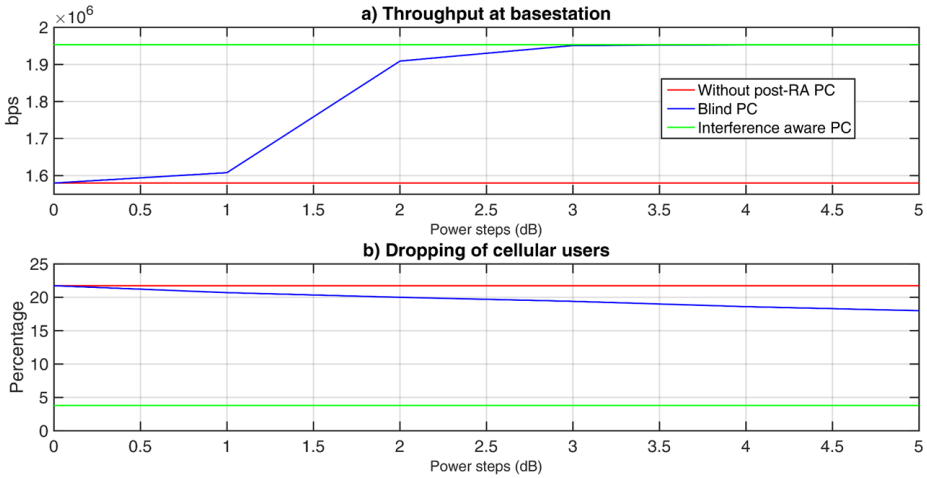

Figure 5.34: Improvements after blind PC

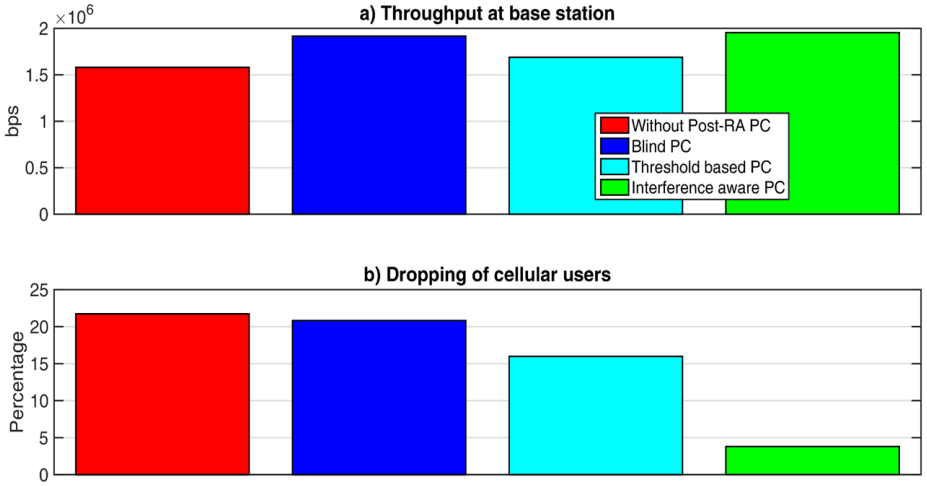

Figure 5.35: Improvements after threshold based PC

Finally, a threshold based PC is incorporated and is compared with the interference aware PC and blind PC. Based on the results in figure 5.34, a power step of 3 dB is chosen for blind PC. Figure 5.35 b) shows that the dropping of CUEs are reduced by around 8% due to threshold based PC. Figure 5.35 a) compares the throughput at base station. It could be seen that post-RA power control improves cellular performance without hampering D2D transmissions. Further, interference aware PC outperforms rest of the PC schemes. However, it is suggestible to use interference aware PC when there is room for an efficient interference estimation (cost and complexity wise) and switch to a threshold based dynamic power control if not.

Whenever, neither interference estimation nor efficient SINR estimation is possible, system can fall back on blind PC.

***The results of this work have been published in* [KS16b].**

5.5 Summary

Device-to-Device communication is considered as a key technological feature in 4G (LTE-A Release 12 onwards) and upcoming 5G mobile communication systems. Some of the major challenges to be addressed by 5G systems are satiating high traffic demand and accommodating large number of connected devices. D2D communications is a key technological enabler in this direction, which offloads base station and provisions direct communication between devices in proximity. This chapter discussed location aware resource allocation and power control schemes for such D2D communication. The following are the highlights of the chapter:

- Developed concepts for D2D resource allocation based on virtual sectoring of a cell, requiring only partial location information and lesser signaling.
- Algorithm for PRB reuse for D2D communication was designed taking into account high mutual interference between D2D link and cellular uplink.
- Substantial improvements were observed in system throughput and number of connected links.
- Robustness of the location based D2D resource allocation schemes against positioning errors was studied in a crowded scenario.
- Taking into account the high interference caused by D2D links to cellular users, power allocation schemes were proposed.
- As a result of power control, significant improvements were exhibited in terms of dropped cellular links due to high D2D interference, without hampering D2D performance.

Chapter 6

Other Contributions

This is a supplementary chapter highlighting the other research contributions of the author in addition to the key topics discussed in chapters 4 and 5. The author has participated in EU-research projects such as METIS [MP12], METIS II [Pro15], BMBF project 5G-NetMobil [Pro17] and other bilateral industry projects. In this tenure, contributions were also made towards topics related to Mobility Robustness Optimization, Mobility Context Awareness in Vehicular use cases and D2D communications that are along the same line as subject matter discussed in chapters 4 and 5. In addition, the author contributed to the topic of *Air Interface Harmonization in 5G Systems* as well. Further, the author also worked towards concept and development of an *Interactive Visualization and Evaluation Platform*, capable of demonstrating 5G concepts to a wider range of audience.

6.1 Mobility Robustness Optimization, Vehicular use cases and D2D solutions

6.1.1 Mobility Robustness Optimization

The growing popularity of smartphones and mobile PCs is expected to increase wireless data traffic in the order of 1000 times by 2020 [OMM16]. However, the current situation of Mobile Network Operator (MNO)s is characterized by increasing margin pressure due to declining revenues and an increasing cost base, especially Operational Expenditure (OPEX). Self-optimization functionalities, e.g. for Mobility Robustness Optimization (MRO), are essential means for reducing OPEX. In particular, mobile user groups or moving networks at high speeds impose challenges and may severely degrade network

performance as well as user experience.

A major solution would be to enable access nodes (e.g. base station) to learn the effectiveness of applied actions, perform optimal handover parameter adaptations according to locally observed conditions, thus establishing context awareness. A Fuzzy Q-Learning-based approach is presented that aims at providing a generic basis for enabling self-optimizing and self-healing network operations. The designed concept consists of the following key components: Fuzzy Inference System (FIS), heuristic Exploration/Exploitation Policy (EEP), and Q-Learning (QL). Its performance in a reference scenario is compared with a trend-based handover (HO) optimization scheme presented in [JBT+10] and a scheme that assigns time-to-trigger (TTT) values based on velocity estimates.

The presented FQL-based MRO scheme enables self-tuning of HO parameters using a limited rule set that reflects the situations the system may be exposed to. It learns optimum parameter adaptations using an EEP. Further, this scheme is able to simultaneously account for several KPI performance targets and by increasing its knowledge base self-optimize its operation. Simulation results indicate significant improvements with respect to call dropping ratio (CDR) up to 65%, handover failure ratio (HFR) up to 100%, ping-pong handover ratio (PHR) up to 90%, satisfied Users up to 8%, overall performance indicator (OPI) up to 170% and Too Early HOs up to 54%. More details about this contribution could be found in [KKSS13a].

6.1.2 Mobility Context Awareness in Vehicular use cases

Mobile communication has become one of the key technologies in contemporary world which is on the verge of its fifth generation (5G). In parallel, Intelligent Transportation System (ITS) is an emerging sector whose focus is safe, smart and efficient transportation. In recent years, various vehicular use cases are studied within the scope of cellular networks, to provide reliable and low latency services as required. With contemporary advancements such as Mobile Edge Clouds (MEC), ability of cellular networks to support vehicular use cases with low latency requirements (e.g., platooning of vehicles)has also increased. In this work, mobility context awareness is built in a cellular network (similar to section 4.2) and is subsequently exploited to assist migration of service or application that is being used by vehicular use case (e.g., vehicular platoon). Further, solutions based on Cooperative Multi-Point transmission (CoMP)and sidelink transmission are proposed to be used in tandem with service migration, in order to reduce data/service interrup-

tion that would arise due to high mobility of vehicles (e.g., handovers (HO)). Message sequence charts are also presented for aforementioned solutions and functional framework is presented for inspected vehicular scenario. Simulation and analysis show the potential of proposed mobility context awareness schemes in assisting service migration among MECs in timely fashion and reducing the amount of data transfer and data interruption times, for the considered vehicular use case. Elaborate details of this work could be found in [KMS18].

6.1.3 Mobility Context Awareness for Multi-path TCP

Cellular networks of today comprises several Radio Access Technologies (RAT) (e.g, LTE, HSPA, (soon) 5G) with various cell sizes, transmission frequencies etc., co-existing with each other to form heterogeneous networks (HetNets). Multipath-TCP (MPTCP) enables simultaneous use of multiple network interfaces (through multiple RATs) available at the UE to maximize resource usage. The application however is oblivious to the presence of multiple TCP connections (or subflows) and perceives it as a regular TCP interface. However, MPTCP faces several challenges among which handling of user handover (HO) is a key issue. This work focuses on the effects of handovers and coverage loss on the performance of MPTCP and proposes Mobility Context Awareness (MCA) as a potential solution. The prediction of HOs in secondary RAT will be beneficial to take appropriate actions (e.g., Coordinated Multipoint (CoMP) transmission) in order to sustain scheduled subflows on such RATs. The simulation results show key improvements in goodput, assimilation time of ordered packets during handovers, which become crucial for delay sensitive applications. Detailed information of this work is found in [KS19].

6.1.4 Optimization of D2D Communications

As discussed earlier in chapter 5, D2D communication offers several benefits, e.g. low end-to-end latency, traffic offload etc. In particular, network controlled D2D operation underlaying the primary cellular system, e.g. LTE, and reusing its radio resources, is seen as an appealing option for increasing system capacity. However, a D2D link that reuses radio resources of another link introduces mutual interference in between these two links. A network controlled radio resource management (RRM) algorithm that maximizes overall user satisfaction with the assistance of context information

is proposed. The task of the algorithm is to decide on appropriate operation mode and allocate radio resources to both cellular and D2D links with respect to their channel state information (CSI) and service requirements. Further, the RRM algorithm is extended to take into account the link priority information. Simulation results demonstrate substantial performance improvement in terms of links satisfaction ratio. Detailed information about this work could be found in [LKK+14].

Further, a network controlled algorithm with low computation complexity is proposed to efficiently maximize the reuse of cellular network spectrum with the help of using channel information at base station. This algorithm can be used to select the transmission mode of potential D2D pairs. Besides, an auction algorithm for maximizing system capacity is also proposed. The low complexity of these two algorithms reduces the setup delay of D2D links. Simulation results demonstrate improved system performance in terms of overall system capacity and number of established D2D links. Elaborations of these algorithms could be found in [LKKS14]. Additional studies were made with respect to Post-RA-power allocation in D2D underlay cellular networks and a regulated-Interference aware power allocation (R-IAPA) strategy was proposed. This scheme takes into account the power consumption at the cellular users and regulates it accordingly to minimize UL power consumption. More details could be found in [KHS19].

6.2 Air Interface Harmonization in 5G Systems

The standardization of 5G, referred to as new radio (NR), has started with 5G air interface (AI) envisioned to comprise a number of AIs including the evolution of Long Term Evolution (LTE) and new NR AIs operating at frequencies from below 1GHz and up to 100GHz. One of the key challenges is the design of an overall 5G AI that efficiently combines these AIs into one 5G AI framework. The author contributed to the inspection of how the AIs can be harmonized into a single 5G AI framework, while minimizing the complexity of standardization and implementation as part of METIS II project [Pro15].

The harmonization among the 5G AIs could take place at any layer: at the physical layer (PHY), at the Medium Access Control layer (MAC) or at the Packet Data Convergence Protocol (PDCP) layer. Among various detailed harmonization options that are being investigated in this context, for the

integration of the new AIs at the physical layer, cyclic prefix-orthogonal frequency division multiplexing (CP-OFDM) waveform with various numerologies is considered as a basis and the MAC layer harmonization can be built on top of the PHY layer, where the CP-OFDM-related AIs of different numerologies are catered for different future services, bands, and cell types. The harmonized MAC layer would allow a single MAC scheduling for the new AIs, therefore, it is particularly suitable for co-located deployments. In cases where the PHY and the MAC layer-based harmonization and aggregation options are not desirable, e.g., in the case of tight interworking between evolved LTE and NR, the PDCP /RLC layer based aggregation can be proven a satisfactory alternative. Since PDCP does not have an access-specific function, it seems to be a reasonable design for aggregation to have at least one common PDCP layer for the all AIs that are envisioned to be aggregated. The results of the first conducted 5G AI analysis constituted discussion topics for other 5G-PPP projects as well as 3GPP. Additional details of this work can be found in [KMF+16].

Further, the author also contributed to the investigations on a high performing solution for intra-RAT mobility in NR and inter-RAT mobility between evolved LTE and NR. Inter-RAT hard handover is a state of the art technique to integrate multiple RATs in order to support mobility and reliability across different RATs. However, the hard handover incurs a transmission interruption which stands as an obstacle along the way of accomplishing 5G design. Besides that, in order to minimize the standardization and implementation complexity; the 5G UP instances related to evolved LTE and NR needs to be aggregated. Therefore, LTE-NR interworking needs to be supported to provide reliable communication as soon as 5G is commercialized. According to simulation results, the user plane aggregation on a common PDCP layer, similar to the LTE dual connectivity feature can provide large user throughput gains. For the case when both NR and LTE are using similar carrier frequencies and same bandwidth, the user throughput is also almost doubled at low load compared to hard handover which is the state of the art solution. Elaborate information of this work is found in [KER+17].

Furthermore, contributions were also made towards service-tailored UP design framework and architecture considerations in 5G Radio Access Networks, identifying the major challenges in the legacy user-plane approaches and highlighting the up-to-date 5G standardization activities in this area. Further analysis were made regarding new functional requirements related to service-oriented design and the introduction of new mechanisms to ad-

dress them. Subsequently, discussions were presented on how various user plane design decisions related to the control/user plane split options, network slicing, and radio access network (RAN)- core network (CN) interfacing can potentially impact the overall 5G architecture. More information about this work could be found in [PGM+17].

6.3 Interactive Visualization and Evaluation Platform

6.3.1 Overview

Various 5G concepts are being developed in order to satiate the challenges such as 1000 times more traffic volume, 10-100 times more connected devices and different service requirements, that are anticipated in 5G. One of the key tasks for 5G consortium however, is to present the developed technology and concepts to a wider audience (academia, industry) and even convince decision makers from non-information and communications technology (ICT) industries. Thus, it is desired to have easy to understand illustrations of envisioned 5G use cases and proposed technical solutions, targeting the non-experts. A visualization tool allowing viewer to interact with 5G enabled scenarios is hence desired for better understanding of the concepts from both industrial marketing and academic perspectives.

The conventional approach for a cellular technology evaluation typically considers a hexagonal cell layout [IR08]. Several proprietary but calibrated link and system level simulators exist, which are used for evaluation of a certain technology or technical solution. The source code for selected parts might be shared or made public (e.g., channel models in WINNER project [II07]). However, the hexagonal cell layout might not be sufficient in some cases, when it is required to reflect the intricate details of the scenarios (e.g. buildings, roads, parks, mobility pattern etc., in a scene) that allow for a realistic evaluation of the given technology. The contemporary simulation framework allows only for evaluation of the considered technical solution with specified simulation parameters and outputs key performance indicators (KPI) at the end of simulation, which are then used to compare the performance of considered technical solutions. These results are subsequently used for marketing of the technical solutions and for academic purposes.

The existing simulators do not have provision or have limited provision for visualization of a considered scenario and dynamic run-time interaction

with a simulator. These limitations of existing simulation framework pose a challenge for the 5G experts to provide a non-expert audience with a better understanding of 5G concepts and convince the decision makers. Another key issue is that some of the 5G concepts discussed in 5G-PPP projects are abstract and non trivial to be demonstrated in the form of experimentation, without showing interactions of different network elements in a vast area (e.g., virtualization, cloudification, etc.,) [Pro16]. Thus, it is desired to use a new, more realistic approach for evaluation and visualization, while reusing data from existing technology evaluation methods whenever it is applicable.

In this work, a simulator agnostic interactive visualization/evaluation platform (IVEP) is proposed, which allows a dynamic 3D visualization of the simulation scenario. The proposed framework allows the viewer to interact with network elements and non network objects in considered scenario, there by immersing the viewer into the simulated use case. Further, the framework provides a real time visualization of KPIs, both system specific (global) and user specific (local), also allows viewer to vary some of the key simulation parameters from visualization front end, and experience interaction with considered use case. These features allow the proposed visualization/evaluation platform to demonstrate the 5G use cases better, paving way for easier understanding and evaluation of the concepts under consideration, even by non-experts.

6.3.2 Architecture of Visualization/Evaluation Framework

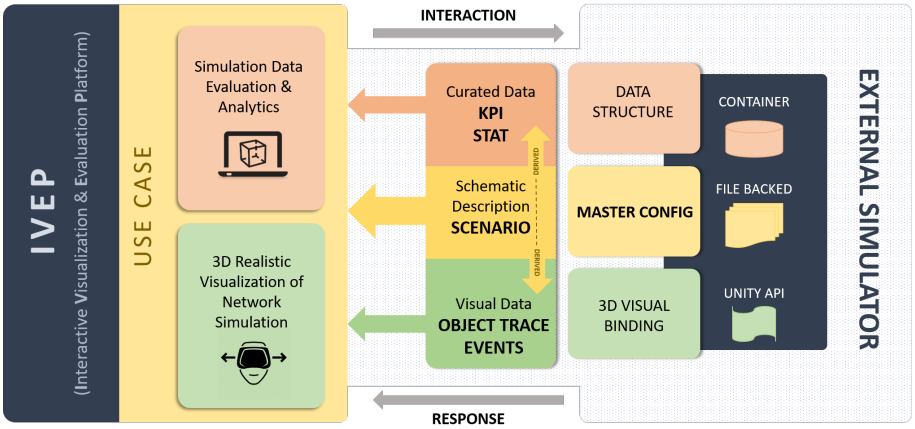

Figure 6.1: IVEP architecture

Architecture of proposed platform is as depicted in figure 6.1. The building blocks of the architecture can be functionally decomposed into three major groups namely:

1. External simulator: The block where use case and concepts are simulated.

2. Logic and data back end: The functional block enabling interfacing of simulator with visualization front end.

3. Visualization front end: The front end of platform to visualize use case being simulated. A game engine called UNITY 3D [Tec15] is chosen for this.

The key features of each block and dynamics among them to enable interactive visualization is discussed in detail in subsequent subsections.

Simulator

The external simulator in the framework can be any standardized open source or proprietary simulator. The framework is agnostic to the external simulator being plugged in. However, the simulator should require simple modification to allow data generated (e.g. KPIs, mobility information etc) to be bundled as simple data structure and be curated (e.g. Java script object notion-JSON).

Logic and Data Back End

Logic and data back end constitutes the most important part of IVEP. This facilitates bundling data and logic from external simulator in a way understandable at visualization front end. Figure 6.2 describes the intricate details of logic and data back end. A file backed master configuration will allow the platform to choose a desired simulation scenario from the list of possible scenarios available at simulator. Further, each scenario which needs to be plugged into visualization front end requires to provide the following information:

1. Dashboard data: This comprises of data from simulator which needs to be displayed on the dashboard of visualization front end. This consists of KPIs from simulator (global and local). Global denotes KPIs of entire scenario and local is specific to a user in the simulated scene. These KPIs can further be grouped based on their logical similarities.

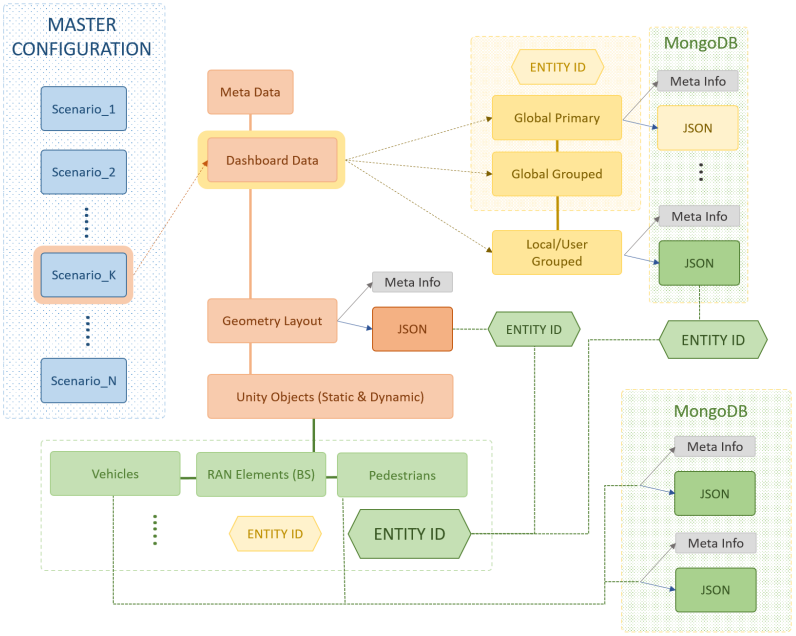

Figure 6.2: Logic and data back end

2. Geometry layout: The features of visualization layout such as buildings, parks, roads etc., their positions and physical dimensions are contained here.

3. Unity objects: The features of static and dynamic objects in scene namely: vehicles, base station and pedestrians are encompassed here.

Meta data and meta info are used to control non-data features of visualization (e.g. color, texture etc). All these three logic and data entities are curated as JSON structure and communicated along with meta data to visualization front end. This communication can be carried out via files or data base. In our architecture, we recommend usage of NoSQL data base (MongoDB), since it provides easier parsing of simulation data (JSON) and is easier/agile to handle than files.

Visualization Front End

A UNITY 3D game engine is devised as visualization front end. The curated content from logic and data back end is fetched from file/database to this block. This content is subsequently processed, and utilized to create and control visual elements accordingly. Figure 6.3 describes an example of visualization front end, guided by curated logic and data back end information. It could be seen that dashboard data dictates the numerical display (KPI)

Figure 6.3: Logic and data back end interfaced to visualization front end

and meta data determines type of graph (pie chart, bar graph etc), color of dashboard and other features. Curated information of geometry layout and unity objects together dictates visualization scene, thus enabling dynamic visualization of simulated use case.

It is further possible to immerse into the visualization scene by zooming in and out of the scene or rotating the scene in any desired direction. The front end also allows clicking on a specific unity object (e.g. pedestrian user) to enter *ego mode* and experience visualization from that user's perspective. Local (user specific) KPIs can also be visualized in this mode.

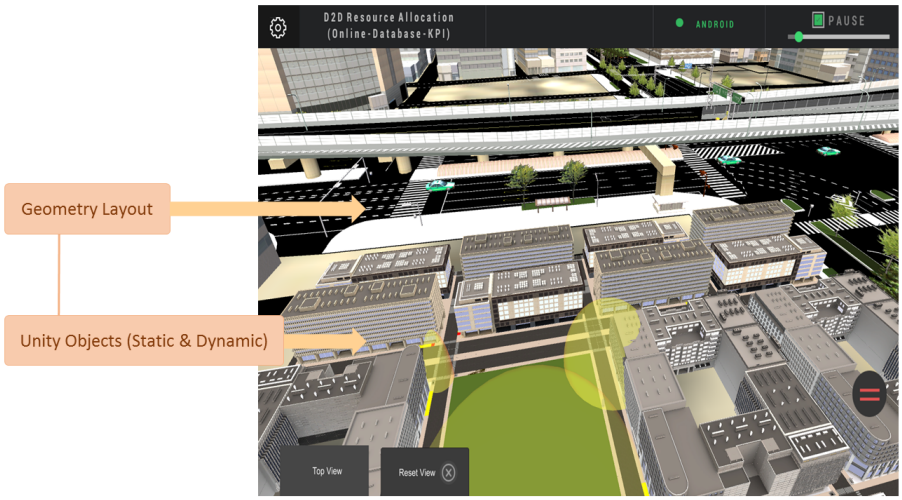

Figure 6.4: Plug and play feature of visualization front end

Figure 6.4 shows another example of interfacing logic and data back end to visualization front end. Based on contained geometry layout and unity objects' information, it could be seen that a completely new scenario is resulted compared to figure 6.3. This demonstrates the plug and play capability of IVEP.

Another key feature made possible in IVEP is to dynamically interact with the simulator from visualization front end. Provision could be made to change few simulation parameters from front end (e.g. via pre-defined sliders). Changes made to these parameters are communicated to the simulator via file/database and subsequently one can observe the changes in KPIs of further simulation.

6.3.3 Concept and Example Scenarios

In order to visualize the scenario being simulated, following three important traces are required at visualization front end:

1. Geometry trace: This describes the visualization scene layout with position and/or features of network elements (e.g. base station, mobile users) and non network elements (e.g.buildings, roads, park etc).

2. Mobility trace: This describes the mobility of users in the scene with respect to time.

3. KPI trace: This trace comprises of KPIs generated from the simulator with respect to time.

Once the data traces from simulator are available at the visualization front end, the scene being simulated can be visualized along with network and non network elements in it.

The simulated scene can be zoomed into and rotated in any angle (0-360), thereby giving an experience of immersion into the simulated scene. The top view of the scenario is shown in figure 6.6.

Further, a user can be clicked to enable *ego-mode*, which allows viewer to visualize the scene from perspective of clicked user. This is shown in figure 6.7.

Figure 6.8 shows the global KPIs associated with simulation scene as graphs, as well as numerical display. When in ego-mode, it is possible to visualize user specific local KPI graphs as shown in figure 6.9.

Further, viewer can interact with few simulation parameters from visualization front end. However, this feature is not allowed with pre-generated

Figure 6.5: View of simulated scenario at IVEP front end

Figure 6.6: Top view of simulated scenario at IVEP front end

simulation traces. Any changes in these parameters are fed back to the simulator via a file. Based on these interactions, the resulting change in system KPIs can be visualized in subsequent time steps. For instance, figure 6.8 shows a spike in total D2D throughput (highlighted in yellow) following an interactive increase in D2D SINR target at visualization front end.

6.3.4 Summary

As part of METIS-II project [Pro15], the author contributed to the concept and development of an interactive visualization and evaluation platform (IVEP) that is capable of show casing the developed 5G concepts and use

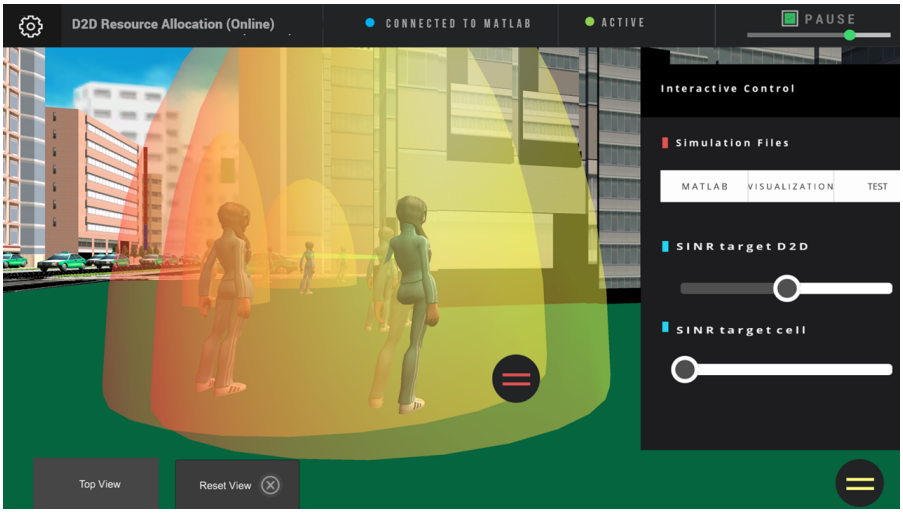

Figure 6.7: Ego user view

Figure 6.8: Dynamic Global KPI graphs

cases to wider audience, including non-ICT industries. The architecture of this simulator agnostic framework was discussed, highlighting key building blocks of the platform. The key procedure for such integration was discussed in detail. The platform was shown to provide an experience of immersion into the simulated scene, thereby making it better suited for providing easy to understand demonstrations of 5G concepts and use cases. Extended version of this work is available in [CKS17]. Further, the author also contributed to the development methodology of visualization platform for overall project, showcasing various use-cases, service-types and developed concepts in the project.

Figure 6.9: Dynamic Local KPI graphs

Chapter 7

Conclusion

7.1 Mobility Context Awareness

Mobile communication of today's world is facing key challenges due to increasing data traffic demand and number of connected devices. Providing a better QoE to the users is an important aspect of the present and future mobile communication systems (e.g., 5G), especially in densely crowded and high mobility situations. There are various such day-to-day scenarios (e.g., crowded public transportation, moving networks, dynamic crowds, traffic jams etc.), where providing better QoE is challenging. In the scope of thesis, mobility context awareness is explored as a feasible solution to enhance QoE of users despite high mobility and crowded situations.

Cell Transition Prediction and RRM

Exploiting mobility data and context information of diurnal user movements (e.g., public transportation, vehicular users, etc.) allows for predicting cell transitions and lays the basis for designing efficient resource management schemes. In day-to-day scenarios, several mobile users co-travel in public transport forming data intensive moving user clusters or moving networks. The primary step in relieving congestion problems prompted by moving user groups/moving networks is to predict their advent into a cell. Emphasizing on mobility behavior of moving user groups/networks, techniques for hotspot prediction is proposed (in section. 4.2)based on user attributes such as user angle, geometry (dB) and distance metrics. Corresponding context information is exploited to trigger load balancing as a potential countermeasure against forthcoming congestion. Through simulation results, it was demonstrated that proposed methods are capable of predicting arrival of hotspot

in advance, even at high velocities. In specific, geometry based prediction resulted in reductions of blocked access attempts by 5% and 15% for site-to-site (s2s) distances of 500 m and 200 m, respectively. In terms of dropped connections, reductions were around 42% (s2s = 500 m) and 25% (s2s = 200 m). Furthermore, number of blocked HO attempts were also reduced by 49% (s2s = 500 m) and 22% (s2s = 200 m). The considerable reductions observed in dropping and blocking of the users, indicate better QoE in the overall system.

Mobility Context Aware Resource Allocation

One of the main challenges for mobile communication systems is to satisfy high data demand while providing a uniform service quality regardless of high mobility. Thus, mobility prediction plays an important role in designing context aware procedures aiming to provide uniform service quality to users. The user mobility is not random but is direction oriented and many users take similar paths on daily basis. Section.4.3 used information of such user's origin and destination to enhance accuracy of mobility prediction in an urban scenario. The enhanced mobility prediction is then used to initiate context aware resource allocation. This ensures a uniform service quality for specific service classes (e.g. non real-time streaming) in day-to-day scenarios like coverage holes. In case of a non-guaranteed bit rate service (e.g., file transfer) around 35% improvement is observed in user throughput (in routes with coverage hole) with PRB = 2 and improvement of 65% is observed with PRB = 3. For a non-real time streaming service use of the origin and destination information to trigger resource allocation leads to throughput improvement of around 80% (in roads with coverage hole) and service interruption is reduced by circa 90%. Simulation results showed substantial improvements in average user throughput and reduced service interruption times, indicating a better service continuity.

Further, section. 4.6 investigated commonalities among various context awareness schemes with user mobility as key focus, outlined common inputs/information, signaling and interfaces required to support mobility context awareness. The new interfaces required in vehicular infrastructure were also discussed and overall architecture to provide support for mobility context awareness was presented.

Dynamic Crowd Formation Prediction and Resource Management

In real world scenarios, such as stadium, convention halls, public events, etc., large number of users assemble to form a crowd. When high number of users in such a crowded situation avail cellular services, then blocking and dropping of users are very high. In the forthcoming generation of mobile communications (5G), users expect a better quality of experience (QoE) even in the presence of a crowd. Section. 4.4 proposed a framework to predict formation of crowds based on the mobility pattern of users in neighboring cells. The methods of future cell prediction, cluster detection and trajectory prediction are used in coalition to deduce severity of a crowd formation and predict the probable location of crowd formation. Subsequently, load balancing is proactively triggered at the site of crowd formation or a small cell is switched on to serve users in a crowd. Simulation results demonstrate significant improvements in KPIs. Due to LB, dropping of users is reduced by $\approx$23%, blocking of users by $\approx$10% and blocked handover attempts by $\approx$15%. Further, activation of a small cell at site of frequent crowd formation exhibited reductions in dropping of users by $\approx$60%, blocking of users by $\approx$56% and blocked handover attempts by $\approx$59%.

Radio Resource Management in Vehicular Traffic Jams

Popularity of broadband multimedia services has been resulting in an increase of vehicular users availing such services. In day-to-day scenarios, large number of such vehicular users travel in and out of a cell, changing its data traffic load. At certain sites, traffic jams occur frequently leading to a hotspot situation in the serving cell. This leads to a higher dropping/blocking of users, thereby hindering their QoE. Section. 4.5 proposed a framework to indicate traffic status of a cell and to predict the occurrence of traffic jams at specific sites. The mobility behavior of vehicular users in and around the site of interest were investigated and a traffic status indicator was designed from a cellular network perspective. Further, RRM strategies namely traffic status aware load balancing and small cell activation/deactivation were presented, which work in tandem with traffic prediction framework. The traffic status aware LB, reduced dropping of users by $\approx$18%, blocking of new access attempts by $\approx$10% and blocked handover attempts by $\approx$18%. Further, the small cell activation resulted in decrease of dropped user connections by $\approx$82%, blocked access attempts by $\approx$42% and blocked handovers are reduced

by ≈81%. These proactive RRM schemes resulted in substantial reductions in blocking/dropping of users and blocked handover attempts, demonstrating a better QoE for users even in traffic jam situations.

7.2 Context Aware D2D Communications

Contemporary mobile systems are aimed at satiating broadband demand from users who are sparsely distributed in public areas e.g., parks, bus-stops etc. However, with increase in user density, the quality of user's experience deteriorates. Typically, data traffic of the UE is routed via base station, which handles connection establishment between the UEs. Thus, overload is experienced at the base station due to such large number of users. However, when UEs that desire to communicate are in close proximity, it is possible to assign a direct link between them for communication and relieve the base station from traffic routing. This thesis proposed solutions for such D2D communication in a crowded situation, especially regarding resource and power allocation enabling efficient D2D communications.

Location aware D2D Resource Allocation

Section. 5.2 introduced the concept of sectoring enabling reuse of radio resources among D2D links and cellular users, with a motive to achieve optimal interference aware resource allocation. The proposed scheme is compared with existing RA techniques and is shown to be performing on par with the optimal distance based scheme that requires precise geo-location information. In scenario with high target SINR for cellular-UEs and D2D links (8 and 10 dB respectively), it was observed that sectoring scheme outperforms Random RA with ≈ 20% more established connections and ≈ 15% more system capacity.

Further, section. 5.3 evaluated the performance of several location based resource allocation schemes against position estimation errors based on Monte Carlo simulations, for real world deployment. The evaluation showed that positioning errors show non linear effects on throughput of D2D users, throughput at base station and overall throughput. However, the degradations in throughput are in order of 11 Kbps (low SINR targets) and 30 Kbps (high SINR targets), which are not much compared to the overall throughput which is in order of Mbps. Thus, in real world scenario where perfect knowledge of user locations are not available, location based D2D resource allocation

schemes can still be deployed with minimum deviations in their performance.

Post-RA Power Allocation in Cellular Networks

D2D underlay results in better spectrum utilization but at the cost of mutual interference. Severe interferences due to D2D transmissions on the same PRBs as uplink transmissions of cellular users(CUE), suppresses the cellular user performance and could eventually lead to connection dropping. Possible countermeasures include regulating D2D transmissions to certain power levels, regions etc. However, this results in a sub-optimal exploitation of D2D transmission opportunities. In this regard, section. 5.4 presented post-resource allocation-power control (PC) at CUEs as a solution to overcome high interference from D2D transmissions, without hindering D2D performance. Three discrete schemes namely, interference aware PC, blind PC and threshold based PC were presented. Simulation results demonstrated, a reduction in dropping of CUEs due to interference and an improvement in throughput at base station, while not effecting the D2D performance. In specific, the post-RA interference aware PC maintained cellular dropping below 5% for wide range of considered D2D SINR targets.

7.3 Future Work

The following points could be considered for the future work :

- *Mobility context awareness*: Several practical scenarios were considered in this thesis such as moving networks, dynamic crowds, traffic jams etc that leads to high load situations. RRM were triggered based on the built mobility context awareness. Two main RRM methods used were load balancing and small cell activation. There are several new emerging techniques such as beam forming [OMM16] that can be used as alternatives to RRM techniques used in this thesis. Beam forming can be proactively managed based on built context awareness (e.g., crowd formation, traffic jams etc) to alleviate high load situations. In addition, study of different spectral bands for small cells can be a future research direction. Further, the presented framework can be extended to include context awareness schemes focusing on device types, applications, service/traffic types etc., along with mobility.

- *Context Aware D2D Communications*: D2D communication is explored as a viable solution to provide service in crowded scenarios and solutions for resource allocation and power control are provided in this thesis. One of the key directions for future work is to investigate further on efficient interference estimation techniques to obtain interference estimate from D2D transmitters, to facilitate real world deployment of interference aware PC. Further, application of the proposed solutions for vehicle-to-vehicle (V2V) communications and service types such as mMTC is an interesting research direction as well.

Bibliography

[3GP04] 3GPP2. 3gpp2 and cdma development group (cdg): cdma2000 evaluation methodology, 2004. Available online at http://www.3gpp2.org/public_html/specs/C.R1002-0_v1.0_041221.pdf.

[3GP06] 3GPP. Physical layer aspects for evolved universal terrestrial radio access (e-utra) (release 7), 2006. Available online at http://www.3gpp.org/ftp/specs/html-INFO/25814.htm.

[3GP09] 3GPP. E-utra physical layer procedures, 2009. Available online at http://www.etsi.org/deliver/etsi_ts/136200_136299/136213/08.08.00_60/ts_136213v080800p.pdf.

[3GP10] 3GPP. Further advancements for e-utra physical layer aspects (release 9), 2010. Available online at http://www.3gpp.org/ftp/specs/html-INFO/36814.htm.

[3GP11] 3GPP. Self-configuring and self-optimizing network use cases and solutions (release 9), 2011. Available online at http://www.etsi.org/deliver/etsi_tr/136900_136999/136902/09.03.01_60/tr_136902v090301p.pdf.

[3GP12a] 3GPP. E-utra physical channels and modulation (release 11), 2012. Available online at http://www.3gpp.org/ftp/specs/html-INFO/36211.htm.

[3GP12b] 3GPP. E-utra physical layer procedures (release 11), 2012. Available online at www.3gpp.org/dynareport/36213.htm.

[3GP13] 3GPP. Discussion on ue-ue channel model for d2d studies, 2013. Available online at http://www.3gpp.org/DynaReport/TDocExMtg\T1\textendashR1-72b\T1\textendash30042.htm.

[Ach00] S.B. Achelis. Technical analysis from a to z. *McGraw-Hill, Second edition*, 2000.

[AL09] Alcatel-Lucent. The lte network architecture-a comprehensive tutorial, 2009.

[ASN+14] Z. Altman, S. Sallem, R. Nasri, B. Sayrac, and M. Clerc. Particle Swarm Optimization for Mobility Load Balancing SON in LTE Networks. In *International Workshop on Self Organizing Networks at IEEE Wireless Communications and Networking Conference*, 2014.

[AWM14] A. Asadi, Q. Wang, and V. Mancuso. A survey on device-to-device communication in cellular networks. *IEEE Communication Surveys and Tutorials, Vol. 16*, pages 1801–1819, 2014.

[AWVK10] A. Awada, B. Wegmann, I. Viering, and A. Klein. A game-theoretic approach to load balancing in cellular radio networks. In *International Symposium on Personal, Indoor and Mobile Radio Communications*, 2010.

[AzHV13] H. Abou-zeid, H.S. Hassanein, and S. Valentin. Optimal Predictive Resource Allocation: Exploiting Mobility Patterns and Radio Maps. In *IEEE Globecom - Wireless Networking Symposium*, 2013.

[BBC97] P.J. Brown, J.D. Bovey, and X. Chen. Context-aware applications: From the laboratory to the marketplace. *IEEE Personal Communications, Vol.4*, pages 58–64, 1997.

[BK10] S. Bellahsene and L. Kloul. A new Markov-based mobility prediction algorithm for mobile networks. In *European performance engineering conference on Computer performance engineering*, 2010.

[BLRC10] X. Bao, U. Lee, I. Rimac, and R.R Choudhury. Dataspotting: Offloading cellular traffic via managed device-to-device data transfer at data spots. *ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 14*, pages 37–39, 2010.

[BY12] P. Bao and G. Yu. An interference management strategy for device-to-device underlaying cellular networks with partial location information. In *International Symposium on Personal, Indoor and Mobile Radio Communications(PIMRC)*, 2012.

[CCZ+12] X. Chen, L. Chen, M. Zeng, X. Zhang, and D. Yang. Downlink resource allocation for device-to-device communication underlaying cellular networks. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2012.

[CDY+12] P. Cheng, L. Deng, H. Yu, Y. Xu, and H. Wang. Resource allocation for cognitive networks with D2D communication: An evolutionary approach. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.

[CFF+16] W. Chen, Z. Feng, Z. Feng, Q. Zhang, and B. Liu. Discrete location-aware power control for d2d underlaid cellular networks. In *IEEE Wireless Communications and Networking Conference*, 2016.

[CJC17] P. Cerwall, P. Jonsson, and S. Carson. Ericsson mobility report, November 2017. Available online at `https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017.pdf`.

[CKS17] P. Chakraborty, N.P. Kuruvatti, and H.D. Schotten. A Novel Serious Game Engineering Based Interactive Visualization and Evaluation Platform For Cellular Technologies. In *International Symposium on Networks, Computers and Communications*, Marrakech, Morocco, 2017.

[DBC03] L. Du, J. Bigham, and L. Cuthbert. Towards intelligent geographic load balancing for mobile cellular networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol.33*, pages 480–491, 2003.

[DDF+06] S. Dobson, S. Denazis, A. Fernández, D. Gaïti, et al. A survey of autonomic communications. *ACM Transactions on Autonomous and Adaptive Systems, Vol.1*, pages 223–259, 2006.

[Dey98] A.K. Dey. Context-aware computing: The cyberdesk project. In *Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium on Intelligent Environments*, Palo Alto, USA, 1998.

[DRJ+09] K. Doppler, M.P. Rinne, P. Janis, et al. Device-to-Device communications; functional prospects for LTE-Advanced networks. In *IEEE Workshop at International Conference on Communications*, 2009.

[DRW+09] K. Doppler, M.P. Rinne, C. Wijting, et al. Device-to-device communication as an underlay to lte-advanced networks. *IEEE Communications Magazine, Vol. 47*, pages 42–49, 2009.

[DSJ98] S.K. Das, S.K. Sen, and R. Jayaram. A novel load balancing scheme for the tele-traffic hot spot problem in cellular networks. *Wireless Networks, Vol.4*, pages 325–340, 1998.

[DSS13] Q. Duong, Y. Shin, and O.S. Shin. Resource allocation scheme for device-to-device communications underlaying cellular networks. In *International Conference on Computing, Management and Telecommunications (ComManTel)*, 2013.

[DY95] S.V.A Davies and J. Yin. Crowd monitoring using image processing. *IEEE Electronic and Communications Engineering Journal, Vol. 7*, pages 37–47, 1995.

[DYRJ10] K. Doppler, C.H. Yu, C.B. Ribeiro, and P. Janis. Mode selection for device-to-device communication underlaying an LTE-advanced network. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[DZX+12] J. Du, W. Zhu, J. Xu, Z. Li, and H. Wang. A compressed HARQ feedback for device-to-device multicast communications. In *IEEE Vehicular Technology Conference-Fall*, 2012.

[FDM+12] G. Fodor, E. Dahlman, G. Mildh, et al. Design aspects of network assisted device-to-device communications. *IEEE Communications Magazine, Vol. 50*, pages 170–177, 2012.

[FM12] P. Fazio and S. Marano. Mobility prediction and resource reservation in cellular networks with distributed Markov chains. In *International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2012.

[FN73] Z. Fluhr and E. Nussbaum. Switching plan for a cellular mobile telephone system. *IEEE Transactions on Communications, Vol. 21*, 1973.

[Fre12] A. Freedman. Velocity estimation of cellular terminals in urban environment using ray tracing. In *Workshop on Positioning, Navigation and Communication (WPNC)*, 2012.

[Fri95] B. Fritzke. A growing neural gas network learns topologies. *Advances in Neural Information Processing Systems, Vol.7*, pages 625–632, 1995.

[FS08] G. Frahling and C. Sohler. A fast k-means implementation using coresets. *International Journal of Computational Geometry and Applications, Vol.18*, pages 605–625, 2008.

[GCL+15] Z. Gao, C. Chen, Y. Li, B. Wen, L. Huang, and Y. Zhao. A Mobility Load Balancing Algorithm based on Handover Optimization in LTE Network. In *International Conference on Computer Science and Education*, 2015.

[GDM12] N. Golrezaei, A.G. Dimakis, and A.F. Molisch. Device-to-device collaboration through distributed storage. In *IEEE Global Communications Conference (GLOBECOM)*, 2012.

[GMD12] N. Golrezaei, A.F. Molisch, and A.G. Dimakis. Base-station assisted device-to-device communications for high-throughput wireless video networks. In *IEEE International Conference on Communications*, 2012.

[GP15a] 5G-PPP. 5g-ppp phase-1 projects, 2015. Available online at `https://5g-ppp.eu/5g-ppp-phase-1-projects/`.

[GP15b] 5G-PPP. 5g vision, 2015. Available online at `https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf`.

[GP17] 5G-PPP. 5g-ppp phase-2 projects, 2017. Available online at `https://5g-ppp.eu/5g-ppp-phase-2-projects/`.

[Gud91] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronics Letters, Vol. 27*, pages 2145–2146, 1991.

[HCLK10] S. Hakola, T. Chen, J. Lehtomaki, and T. Koskela. Device-to-device (D2D) communication in cellular network-performance analysis of optimum and practical communication mode selection. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[Hei05] J.L. Heilbron. The oxford guide to the history of physics and astronomy. *Oxford University Press, ISBN:0195171985*, 2005.

[HK81] T. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory, Vol. 27*, pages 49–60, 1981.

[HKL12] M.H. Han, B.G. Kim, and J.W. Lee. Subchannel and transmission mode scheduling for D2D communication in OFDMA networks. In *IEEE Vehicular Technology Conference-Fall*, 2012.

[HT05] H. Holma and A. Toskala. Wcdma for umts: Radio access for third generation mobile communications. *John Wiley and Sons Ltd. ISBN:9780470870969*, 2005.

[HT09] H. Holma and A. Toskala. Lte for umts - ofdma and sc-fdma based radio access. *John Wiley and Sons, ISBN: 9780470745472*, 2009.

[HYW14] W. Hu, L. Yan, and H. Wang. Traffic jams prediction method based on two-dimension cellular automata model. In *International conference on Intelligent Transportation Systems*, 2014.

[HYXY12] T. Han, R. Yin, Y. Xu, and G. Yu. Uplink channel reusing selection optimization for device-to-device communication underlaying cellular networks. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2012.

[IEE05] IEEE. Ieee 802.20 project group: 802.20 evaluation criteria, 2005. Available online at `http://ieee802.org/20/P_Docs/IEEE_802.20-PD-09.doc`.

[IEE08] IEEE. Ieee 802.16m project group: Evaluation methodology document (emd), 2008. Available online at `http://ieee802.org/16/tgm/docs/80216m-08_004r2.pdf`.

[II07] WINNER II. Winner ii channel models, 2007.

[IKY13] S. Ikeda, N. Kami, and T. Yoshikawa. Adaptive mobility management in cellular networks with multiple model-based prediction. In *International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2013.

[IR97] ITU-R. Guidelines for the evaluation of radio transmission technologies for imt-2000, 1997. Available online at `http://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1225-0-199702-I!!PDF-E.pdf`.

[IR08] ITU-R. Requirements related to technical performance for imt-advanced radio interface(s), 2008.

[JBT+10] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kurner. Handover Parameter Optimization in LTE Self-Organizing Networks. In *IEEE-Vehicular Technology Conference*, 2010.

[JF12] M. Jar and G. Fettweis. Throughput Maximization for LTE Uplink via Resource Allocation. In *International Symposium on Wireless Communication Systems*, 2012.

[JFTK04] J. Jobin, M. Faloutsos, S.K. Tripathi, and S.V. Krishnamurthy. Understanding the effects of hotspots in wireless cellular networks. In *IEEE International Conference on Computer Communications, INFOCOM*, Hong Kong, China, 2004.

[JHC12] M. Jung, K. Hwang, and S. Choi. Joint mode selection and power allocation scheme for power-efficient Device-to-Device (D2D) communication. In *IEEE Vehicular Technology Conference-Spring*, 2012.

[JKR+09] P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl. Interference-aware resource allocation for device-to-device radio underlaying cellular networks. In *IEEE Vehicular Technology Conference-Spring*, 2009.

[Joe70] A.E. Jr. Joel. Mobile communication system. *Patent No. US3663762 A*, 1970.

[KA08] B. Kaufman and B. Aazhang. Cellular networks with an overlaid device to device network. In *Asilomar Conference on Signals, Systems and Computers*, 2008.

[KE89] J. Karlsson and B. Eklundh. A cellular mobile telephone system with load sharing-an enhancement of directed retry. *IEEE Transactions on Communications, Vol.47*, pages 530–535, 1989.

[Ken10] J. Kennedy. Particle swarm optimization. *Encyclopedia of Machine Learning, Springer-Verlag*, pages 760–766, 2010.

[KER+17] C. Kilinic, M. Ericson, P. Rugeland, A. Zaidi, O. Aydin, V. Venkatkumar, M.C. Filippou, M. Mezzavilla, N.P. Kuruvatti, and J.F. Monserrat. 5G Multi-RAT integration evaluations using a common PDCP layer. In *IEEE-Vehicular Technology Conference*, Sydney, Australia, 2017.

[KFP10] P. Kolios, V. Friderikos, and K.P. Papadaki. Load balancing via store-carry and forward relaying in cellular networks. In *IEEE Global Telecommunications Conference*, 2010.

[KHS19] N.P. Kuruvatti, R. Hernandez, and H.D. Schotten. Interference Aware Power Management in D2D Underlay Cellular Networks. In *IEEE-AFRICON Conference*, Accra, Ghana, 2019.

[KKL+15] N.P. Kuruvatti, A. Klein, J. Lianghai, R. Sattiraju, et al. Robustness of Location Based D2D Resource Allocation Against Positioning Errors. In *IEEE-Vehicular Technology Conference*, Glasgow, Scotland, 2015.

[KKS15] N.P. Kuruvatti, A. Klein, and H.D. Schotten. Prediction of Dynamic Crowd Formation in Cellular Networks for Activating Small Cells. In *International Workshop on 5G Architecture at IEEE-Vehicular Technology Conference*, Glasgow, Scotland, 2015.

[KKSS13a] A. Klein, N.P. Kuruvatti, J. Schneider, and H.D. Schotten. Fuzzy Q-Learning for Mobility Robustness Optimization in Wireless Networks. In *International Workshop on Emerging Technologies*

for LTE-Advanced and Beyond 4G at IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, USA, 2013.

[KKSS13b] N.P. Kuruvatti, A. Klein, J. Schneider, and H.D. Schotten. Exploiting Diurnal User Mobility for Predicting Cell Transitions. In *International Workshop on Broadband Wireless Access at IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, USA, 2013.

[Kle08] A. Klein. Analyse und vergleich verschiedener mobilfunkstandards. *Diploma thesis at University of Kaiserslautern, Chair for Wireless Communication and Navigation*, 2008.

[KLH99] J. Kim, T. Lee, and C.S. Hwang. A dynamic channel assignment scheme with two thresholds for load balancing in cellular networks. In *IEEE Radio and Wireless Conference*, 1999.

[KMF+16] C. Kilinic, J.F. Monserrat, M.C. Filippou, N.P. Kuruvatti, A. Zaidi, I. Silva, and M. Mezzavilla. New Radio 5G User Plane Design Alternatives. In *International Workshop on 5G RAN Design at IEEE GLOBECOM*, Washinton DC, USA, 2016.

[KMS17a] N.P. Kuruvatti, J.F.S. Molano, and H.D. Schotten. Mobility Context Awareness to Improve Quality of Experience in Traffic Dense Cellular Networks. In *International Conference on Telecommunications*, Limassol, Cyprus, 2017.

[KMS17b] N.P. Kuruvatti, J.F.S. Molano, and H.D. Schotten. Monitoring Vehicular User Mobility to Predict Traffic Status and Manage Radio Resources. In *IEEE-Vehicular Technology Conference*, Sydney, Australia, 2017.

[KMS18] N.P. Kuruvatti, H.M. Mutabazi, and H.D. Schotten. Exploiting Mobility Context Awareness in Cellular Networks for Assisting Vehicular Use Cases. In *IEEE-Vehicular Technology Conference*, Chicago, USA, 2018.

[KS13] W. Kim and Y. Suh. Enhanced Adaptive Periodic Mobility Load Balancing Algorithm for LTE Femtocell Networks. In *IEEE International Conference on Network of the Future*, 2013.

[KS16a] N.P. Kuruvatti and H.D. Schotten. Framework to Support Mobility Context Awareness in Cellular Networks. In *IEEE-Vehicular Technology Conference*, Nanjing, China, 2016.

[KS16b] N.P. Kuruvatti and H.D. Schotten. Post-Resource Sharing Power Allocation in Cellular Networks to Coexist with D2D Underlay.

In *International Conference on Networks of the Future*, Buzios, Brazil, 2016.

[KS19] N.P. Kuruvatti and H.D. Schotten. Mobility Context Awareness in Heterogeneous Networks to Enhance Multipath Communications. In *IEEE-Vehicular Technology Conference (VTC-Spring)*, Kuala Lumpur, Malaysia, 2019.

[Kuh56] H.W. Kuhn. Variants of the hungarian method for assignment problems. *Naval Research Logistics Quarterly, Vol. 3*, pages 253–258, 1956.

[KV03] S. Kim and P.K. Varshney. Adaptive Load Balancing with Preemption for Multimedia Cellular Networks. In *IEEE Wireless Communications and Networking Conference*, 2003.

[KZS16] N.P. Kuruvatti, W. Zhou, and H.D. Schotten. Mobility Prediction of Diurnal Users for Enabling Context Aware Resource Allocation. In *IEEE-Vehicular Technology Conference*, Nanjing, China, 2016.

[Laa05] K. Laasonen. Clustering and Prediction of Mobile User Routes from Cellular Data. In *European conference on Principles and Practice of Knowledge Discovery in Databases*, 2005.

[LH00] Y.D. Lin and Y.C. Hsu. Multihop cellular: A new architecture for wireless communications. In *IEEE International Conference on Computer Communications, INFOCOM*, 2000.

[LH05] S. Liou and Y. Huang. Trajectory predictions in mobile networks. *International Journal of Information Technology, Vol. 11*, pages 109–122, 2005.

[LH16] H. Li and D. Hu. Mobility Prediction based Seamless RAN-Cache Handover in HetNet. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016.

[LKK+14] J. Lianghai, A. Klein, N.P. Kuruvatti, R. Sattiraju, and H.D. Schotten. Dynamic Context-aware Optimization of D2D Communications. In *International Workshop on 5G Mobile and Wireless Communication System for 2020 and Beyond at IEEE-Vehicular Technology Conference*, Seoul, Republic of Korea, 2014.

[LKKS14] J. Lianghai, A. Klein, N.P. Kuruvatti, and H.D. Schotten. System Capacity Optimization Algorithm for D2D Underlay Operation. In *International Workshop on 5G Technologies at IEEE In-*

ternational Conference on Communications, Sydney, Australia, 2014.

[LLAH15] N. Lee, X. Lin, J.G. Andrews, and R.W. Heath. Power control for d2d underlaid cellular networks: Modeling, algorithms, and analysis. *IEEE Journal on Selected Areas in Communications*, pages 1–13, 2015.

[LLG12] J.C. Li, M. Lei, and F. Gao. Device-to-Device (D2D) communication in MU-MIMO cellular networks. In *IEEE Global Communications Conference (GLOBECOM)*, 2012.

[LMQ+14] Q. Lv, Z. Mei, Y. Qiao, Y. Zhong, and Z. Lei. Hidden Markov Model Based User Mobility Analysis in LTE Network. In *International Symposium on Wireless Personal Multimedia Communications*, 2014.

[LSSK14] Y.V. Leshchik, D.M. Sonkin, M.A. Sonkin, and S.A. Khrul. Applying of combined algorithm for prediction urban road traffic based on fuzzy search theory. In *International Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, 2014.

[LZLS12] L. Lei, Z. Zhong, C. Lin, and X. Shen. Operator controlled device-to-device communications in lte-advanced networks. *IEEE Wireless Communications, Vol. 19*, pages 96–104, 2012.

[MAA17] A.M. Mustafa, O.M. Abubakr, and O. Ahmadien. Mobility Prediction for Efficient Resources Management in Vehicular Cloud Computing. In *International Conference on Mobile Cloud Computing, Services, and Engineering*, 2017.

[Max65] J.C. Maxwell. A dynamical theory of the electromagnetic field. *Philosophical Transactions: The Royal Society Publishing*, pages 459–512, 1865.

[MBM+14] Y.V.L Melo, R.L. Batista, T.F Maciel, et al. Power Control with Variable Target SINR for D2D Communications Underlying Cellular Networks. In *European Wireless Conference*, 2014.

[MLLZ12] Y. Meng, J. Li, H. Li, and P. Zhang. An Adaptive Network Selection Scheme Based on the Combination of User Mobility and Traffic Load. In *International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, 2012.

[MLPH11] H. Min, J. Lee, S. Park, and D. Hong. Capacity enhancement using an interference limited area for device-to-device uplink underlaying cellular networks. *IEEE Transactions on Wireless Communications, Vol. 10*, pages 3995–4000, 2011.

[MLS08] R. Mangharam, I. Lee, and O. Sokolsky. Real-Time Traffic Congestion Prediction. In *NSF-NCO/NITRD National Workshop on High Confidence Transportation Cyber-Physical Systems*, 2008.

[MMT13a] S.S. Mwanje and A. Mitschele-Thiel. A Q-Learning Strategy for LTE Mobility Load Balancing. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications: Mobile and Wireless Networks*, 2013.

[MMT13b] S.S. Mwanje and A. Mitschele-Thiel. Minimizing Handover Performance Degradation due to LTE Self Organized Mobility Load Balancing. In *IEEE Vehicular Technology Conference-Spring*, 2013.

[MNK+07] P. Mogensen, W. Na, I.Z. Kovács, F. Frederiksen, A. Pokhariyal, et al. LTE Capacity Compared to the Shannon Bound. In *IEEE Vehicular Technology Conference-Spring*, 2007.

[MOH+15] A. Mohamed, O. Onireti, S.A. Hoseinitabatabaei, et al. Mobility Prediction for Handover Management in Cellular Networks with Control/Data Separation. In *IEEE International Conference on Communications: Mobile and Wireless Networking Symposium*, 2015.

[Mol11] A.F. Molisch. Wireless communications. *John Wiley and Sons Ltd. ISBN:9780470741870*, 2011.

[MP12] METIS-Project. Mobile and wireless communications enablers for the 2020 information society, 2012. Available online at `https://www.metis2020.com`.

[MP13a] METIS-Project. D1.1 - future radio access scenarios, requirements and kpis, 2013. Available online at `https://www.metis2020.com`.

[MP13b] METIS-Project. D4.1 - summary on preliminary trade-off investigations and first set of potential network-level solutions, 2013. Available online at `https://www.metis2020.com`.

[MP13c] METIS-Project. D6.1 - simulation guidelines, 2013. Available online at `https://www.metis2020.com`.

[MPM12] S. Michaelis, N. Piatkowski, and K. Morik. Predicting next network cell IDs for moving users with Discriminative and Generative Models. In *Mobile Data Challenge Workshop (by Nokia)*, 2012.

[MSA+14] R. Margolies, A. Sridharan, V. Aggarwal, et al. Exploiting Mobility in Proportional Fair Cellular Scheduling: Measurements and Algorithms. In *IEEE International Conference on Computer Communications, INFOCOM*, 2014.

[Net11] Nokia Siemens Networks. 2020: Beyond 4g radio evolution for the gigabit experience, 2011.

[Net12] 3GPP Technical Specification Group Radio Access Network. E-utra radio resource control (rrc) (release 11), 2012. Available online at `http://www.3gpp.org/ftp/Specs/html-info/36331.htm`.

[NGM07] NGMN. Radio access performance evaluation methodology, 2007.

[OMM16] A. Osseiran, J.F. Monserrat, and P. Marsch. 5g mobile and wireless communications technology. *Cambridge University Press, ISBN:9781107130098*, 2016.

[PGM+17] E. Pateromichelakis, J. Gebert, T. Mach, J. Belschner, W. Guo, N.P. Kuruvatti, V. Venkatkumar, and C. Kilinic. Service-Tailored User-Plane Design Framework and Architecture Considerations in 5G Radio Access Networks. *IEEE Open Access Journal*, 2017.

[PLW+09] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang. Interference avoidance mechanisms in the hybrid cellular and device-to-device systems. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2009.

[PP14] N. K. Pratas and P. Popovski. Underlay of low-rate machine-type D2D links on downlink cellular links. In *IEEE Workshop at International Conference on Communications*, 2014.

[Pro10] C-CAST Project. Context casting (c-cast), 2010. Available online at `https://cordis.europa.eu/project/rcn/85341_en.html`.

[Pro15] METIS-II Project. Mobile and wireless communications enablers for the 2020 information society-ii, 2015. Available online at `https://metis-ii.5g-ppp.eu/`.

[Pro16] METIS-II Project. D2.1-performance evaluation framework, 2016. Available online at `https://metis-ii.5g-ppp.eu/`.

[Pro17] 5G-NetMobil Project. 5g lösungen für die vernetzte mobilität der zukunft, 2017. Available online at `https://5g-netmobil.de/`.

[PW12] T. Poegel and L. Wolf. Prediction of 3G Network Characteristics for Adaptive Vehicular Connectivity Maps. In *IEEE Vehicular Networking Conference (VNC)*, 2012.

[PW14] O. Perkasa and D. Widyantoro. Video-based system development for automatic traffic monitoring. In *International Conference on Electrical Engineering and Computer Science (ICEECS)*, 2014.

[RLJ00] L.K. Rasmussen, T.J. Lim, and A.L. Johansson. A matrix-algebraic approach to successive interference cancellation in cdma. *IEEE Transactions on Communications, Vol. 48*, pages 145–151, 2000.

[RPM97] N.S. Ryan, J. Pascoe, and D.R. Morse. Enhanced reality fieldwork: the context-aware archaeological assistant. *Computer Applications and Quantitative Methods in Archaeology*, pages 182–196, 1997.

[Sch03] J.H. Schiller. Mobile communications. *Pearson Education Limited ISBN:032112381 6*, 2003.

[Sch05] M. Schwartz. Mobile wireless communications. *Cambridge University Press, ISBN:139780511264238*, 2005.

[Sid83] J.N. Siddall. Probabilistic engineering design: Principles and applications. *CRC Press, ISBN:9780824770228*, 1983.

[SJWL13] L. Su, Y. Ji, P. Wang, and F. Liu. Resource allocation using particle swarm optimization for D2D communication underlay of cellular networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[SK04] W. Soh and H.S. Kim. Dynamic bandwidth reservation in cellular networks using road topology based mobility predictions. In *IEEE International Conference on Computer Communications, INFOCOM*, 2004.

[SKL+15] R. Sattiraju, A. Klein, J. Lianghai, N.P. Kuruvatti, H.D. Schotten, et al. Virtual Cell Sectoring for Enhancing Resource Allocation and Reuse in Network Controlled D2D Communication. In *IEEE-Vehicular Technology Conference*, Glasgow, Scotland, 2015.

[SMBS10] S.K Sadhukhan, S. Mandal, P. Bhaumik, and D. Saha. A Novel Direction-Based Diurnal Mobility Model for Handoff Estimation in Cellular Networks. In *IEEE India conference (INDICON)*, 2010.

[SP10] SOCRATES-Project. D5.9 - final report on self-organisation and its implications in wireless access networks, 2010. Available online at `http://www.fp7-socrates.eu`.

[SSS09] I. Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. *IEEE Transactions on Patterns Recognition and Machine Intelligence, Vol. 31*, pages 1472–1485, 2009.

[ST94] B.N. Schilit and M.M. Theimer. Disseminating Active Map Information to Mobile. *IEEE Network, Vol.8*, pages 22–32, 1994.

[SVV13] P. Szilagyi, Z. Vincze, and C. Vulkan. Integrated Mobility Load Balancing and Traffic Steering Mechanism in LTE. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications: Mobile and Wireless Networks*, 2013.

[SWYS10] H. Si, Y. Wang, J. Yuan, and X. Shan. Mobility Prediction in Cellular Network Using Hidden Markov Model. In *IEEE Consumer Communications and Networking Conference (CCNC)*, 2010.

[Tec15] Unity Technologies. Unity - game engine, tools and multiplatform, 2015. Available online at `http://unity3d.com`.

[TK15] N. Tuncel and M. Koca. Joint Mobility Load Balancing and Inter-cell Interference Coordination for Self-Organizing OFDMA Networks. In *IEEE Vehicular Technology Conference-Spring*, 2015.

[TMC04] Y.K. Tang, J. Mišic, and I. Chlamtac. An On-line Hot-spot Detection Scheme in DS-CDMA Networks - Single Traffic Type. In *IEEE Globecom 2004 - Wireless Communications, Networks, and Systems*, Dallas, Texas, USA, 2004.

[TV05] D. Tse and P. Viswanath. Fundamentals of wireless communication. *Cambridge University Press. ISBN:9780521845274*, 2005.

[WC12] H. Wang and X. Chu. Distance-constrained resource-sharing criteria for device-to-device communications underlaying cellular networks. *Electronics Letters, Vol. 48*, pages 528–530, 2012.

[WCW98] J. Wu, J. Chung, and C. Wen. Hot-spot traffic relief with a tilted antenna in cdma cellular networks. *IEEE Transactions on Vehicular Technology, Vol.47*, pages 1–9, 1998.

[Wei97] J.W. Weibull. Evolutionary game theory. *MIT Press, ISBN: 9780262731218*, 1997.

[XCG10] L. Xu, Y. Chen, and Y. Gao. Self-organizing Load Balancing for Relay Based Cellular Networks. In *IEEE International Conference on Computer and Information Technology*, 2010.

[XCL+15] L. Xu, X. Cheng, Y. Liu, W. Chen, Y. Luan, K. Chao, M. Yuan, and B. Xu. Mobility Load Balancing Aware Radio Resource Allocation Scheme for LTE-Advanced Cellular Networks. In *International Conference on Communication Technology*, 2015.

[XH10] H. Xing and S. Hakola. The Investigation of Power Control Schemes for a Device–to-Device Communication integrated into OFDMA Cellular System. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2010.

[XjHF+10] W. Xi-jun, T. Hui, J. Fan, L. Xiang-yan, H. Xuan-ji, and L. Tai-ri. Cell-Cluster Based Traffic Load Balancing in Cooperative Cellular Networks. In *IEEE Consumer Communications and Networking Conference*, 2010.

[XSH+12a] C. Xu, L. Song, Z. Han, D. Li, and B. Jiao. Resource allocation using a reverse iterative combinatorial auction for device-to-device underlay cellular networks. In *IEEE Global Communications Conference (GLOBECOM)*, 2012.

[XSH+12b] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, and B. Jiao. Interference-aware resource allocation for device-to-device communications as an underlay using sequential second price auction. In *IEEE International Conference on Communications (ICC)*, 2012.

[XTL11] X. Xiao, X. Tao, and J. Lu. A QoS-aware power optimization scheme in OFDMA systems with integrated Device-to-Device (D2D) communications. In *IEEE Vehicular Technology Conference-Fall*, 2011.

[XWC+10] S. Xu, H. Wang, T. Chen, Q. Huang, and T. Peng. Effective interference cancellation scheme for device-to-device communication underlaying cellular networks. In *IEEE Vehicular Technology Conference-Fall*, 2010.

[XZZG12] H. Xiong, D. Zhang, D. Zhang, and V. Gauthier. Predicting Mobile Phone User Locations by Exploiting Collective Behavioral Patterns. In *International Conference on Ubiquitous Intelligence and Computing*, 2012.

[YB00] D.J. Young and N.C. Beaulieu. The generation of correlated rayleigh random variates by inverse discrete fourier transform. *IEEE Transactions on Communications, Vol. 48*, pages 1114–1127, 2000.

[YDRT11] C.H. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen. Resource sharing optimization for device-to-device communication underlaying cellular networks. *IEEE Transactions on Wireless Communications, Vol. 10*, pages 2752–2763, 2011.

[YHPH06] G. Youjun, T. Hui, Z. Ping, and X. Haibo. A QoS-Guaranteed adaptive resource allocation algorithm with low complexity in OFDMA system. In *International Conference on Wireless Communications, Networking and Mobile Computing*, 2006.

[YMT+02] E. Yanmaz, S. Mishra, O. K. Tonguz, H. Wu, and C. Qiao. Efficient dynamic load balancing algorithms using iCAR systems: A generalized framework. In *IEEE Vehicular Technology Conference, VTC-Fall*, 2002.

[YN10] Z. Yang and Z. Niu. A New Relay Based Dynamic Load Balancing Scheme in Cellular Networks. In *IEEE Vehicular Technology Conference-Fall*, 2010.

[YSTL14] C. Yang, M. Sheng, H. Tian, and J. Li. Double Threshold Design for Mobility Load Balancing in Self-Optimizing Networks. In *IEEE Vehicular Technology Conference-Spring*, 2014.

[YT12] C.H. Yu and O. Tirkkonen. Device-to-device underlay cellular network based on rate splitting. In *IEEE Wireless Communications and Networking Conference*, 2012.

[YTDR09] C.H. Yu, O. Tirkkonen, K. Doppler, and C. Ribeiro. Power optimization of device-to-device communication underlaying cellular communication. In *IEEE International Conference on Communications*, 2009.

[YV09] E.S. Yami and V.T. Vakili. Adaptive Bandwidth Allocation and Mobility Prediction in Mobile Cellular Networks. In *International Conference on Next Generation Mobile Applications, Services and Technologies*, 2009.

[YW93] T. Yum and W. Wong. Hot spot traffic relief in cellular systems. *IEEE Journal on Selected Areas in Communications, Vol.11*, pages 934–940, 1993.

[ZCYJ13] R. Zhang, X. Cheng, L. Yang, and B. Jiao. Interference-aware graph based resource sharing for device-to-device communications underlaying cellular networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013.

[ZGFW12] L. Zimmermann, A. Goetz, G. Fischer, and R. Weigel. GSM mobile phone localization using time difference of arrival and angle of arrival estimation. In *International Multi-Conference on Systems,Signals and Devices*, 2012.

[Zha04] G. Zhang. Subcarrier and bit allocation for real-time services in multiuser OFDM systems. In *IEEE International Conference on Communications*, 2004.

[ZHHC13] B. Zhou, H. Hu, S.Q. Huang, and H.H Chen. Intracluster device-to-device relay algorithm with optimal resource utilization. *IEEE Transactions on Vehicular Technology, Vol. 62*, pages 2315–2326, 2013.

[ZJLZ14] H. Zhou, Y. Ji, J. Li, and B. Zhao. Joint mode selection, mcs assignment, resource allocation and power control for d2d communication underlaying cellular networks. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2014.

[ZKAQ01] J. Zander, S.-L. Kim, M. Almgren, and O. Queseth. Radio resource management for wireless networks. *Artech House Publishers, ISBN: 9781580531467*, 2001.

[ZMT13] N. Zia and A. Mitschele-Thiel. Self-Organized Neighborhood Mobility Load Balancing for LTE Networks. In *IEEE International Wireless Days Conference*, 2013.

Curriculum Vitae

Personal Details

Name:	Nandish P. Kuruvatti
Place of birth:	Davanagere, India

Professional Experience

04/2012 - present	Technische Universität Kaiserslautern Wireless Communications & Navigation research group Research associate & teaching assistant
07/2008 - 06/2009	Nokia Siemens Networks R&D Centre Bangalore, India Software developer

Education

04/2012 - present	Technische Universität Kaiserslautern PhD thesis: "Incorporating Context Awareness in Cellular Networks to Enhance System Performance and User Mobility Support"
04/2010 - 01/2012	Technische Universität Kaiserslautern Master degree in Electrical & Computer Engineering with focus on Communication Technology Thesis: "Extracting Secret Keys from Variations in Wireless Channels"
09/2004 - 06/2008	Visvesvaraya Technological University Bachelor degree in Electronics & Communications Engineering Thesis: "Design and Simulation of Phase Locked Loop using VLSI"