Collaborative Streaming System Design as an Underlay to Cellular Networks

by

Mohammad Zulhasnine, MAsC

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering (OCIECE)
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada, K1S 5B6

April 2013

© 2013, Mohammad Zulhasnine

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-94540-7

Our file Notre référence

ISBN: 978-0-494-94540-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Media content sharing over a peer-to-peer (P2P) network is promising because of its distributive nature and scalability to network size. Recently device-to-device (D2D) communication as an underlay to cellular networks has been introduced as a technology component to the Long Term Evolution Advanced. The short range of D2D transmission exploits the reuse of frequency within a cell. In this work, we propose a collaborative streaming system where geographically close peers form groups and then share the content with others through a D2D multicast mechanism. We present the optimal grouping problem as the stable admission assignment problem. We also propose a distributive algorithm that forms groups of suitable sizes to tackle the retransmission problem. We present a mathematical model for peer grouping process that finds the relationship between two important measurements, the number of groups and group sizes, affecting the bandwidth savings and streaming rate. However, D2D transmitters may cause a significant amount of interference with the primary cellular network when radio resources are shared between them. We propose a heuristic radio resource allocation algorithm that can lessen interference to the primary cellular network with coordination from an LTE access point. We apply the theory of epidemiology to derive an analytical model for chunk distribution of live streaming content. The model finds buffer occupancy probability and provides insight on relationships among peer population size, buffer length and playback smoothness. We also propose a hybrid chunk selection method that is more scalable to the peer population in buffer space and also achieves better playback continuity by distributing urgent chunks first. We conduct extensive simulations to support our claims.

Acknowledgements

I feel the pleasure to express my thanks to those who made this thesis possible. I wish to express my deepest gratitude to my supervisor Professor Changcheng Huang, an extraordinary mentor, for his constant guidance and support throughout this thesis work. He sparked my interests to conduct innovative research and challenged me to reach my very best. I never felt I waited for my research work's feedback, he always provided quick and constructive feedback. I would also like to thank my examining committee members, Professor Hossam Hassanein, Professor Gregor Bochmann, Professor Yiqiang Zhao, Professor F. Richard Yu and Professor Halim Yanikomeroglu, for their careful reading, and valuable insights. I wish to extend my warmest thanks to Dr. Anand Srinivasan for his support and collaboration during my PhD study. I owe thanks to colleagues of our research group for many scientific discussions and their encouragements. I would also like to thank all of my friends in Ottawa for their encouragements and companionship. Finally I wish to thank my parents and other family members for their continuous encouragement and belief in me. I am also lucky to have a wonderful daughter who is still so little to comprehend how she could be a source of joy and inspiration in my work and in my life. And thanks to my beloved wife whose presence transformed my life into a beautiful and meaningful one. Her sacrifice to provide me with writing space is beyond my expectations.

Contents

Abstract	iii
Acknowledgement	iv
List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Motivation	1
1.2 Problem statement	3
1.3 Summary of contributions	5
1.4 Publication list	7
1.5 Outline of the thesis	8
2 Literature Review	10
2.1 Steaming over cellular network: issues and challenges	10
2.1.1 Designing an efficient overlay network	11
2.1.2 Wireless P2P streaming system	17
2.1.3 Peer selection strategy	23
2.1.4 Other issues on realizing cellular P2P streaming system	28
2.2 D2D communication underlaying a cellular network	31
2.2.1 Device discovery and connection setup	31
2.2.2 Interference management scheme for D2D enable network . . .	33

2.2.3	D2D multicast mechansim	39
2.2.4	Other coexisting networks	40
2.3	Streaming chunk selection strategy	41
2.4	Some related theories	43
2.4.1	Stable marriage	43
2.4.2	Spatial point process	45
2.4.3	Epidemiology	46
2.5	Summary	47
3	Resource Allocation for Device-to-Device Communication	50
3.1	Channel modeling	51
3.1.1	Radio resource and access technology	51
3.1.2	Channel model	52
3.2	Interference problem due shared spectrum	54
3.3	Problem formulation to mitigate interference	56
3.3.1	Downlink phase interference coordination	57
3.3.2	Uplink phase interference coordination	58
3.4	Penalty based genetic algorithm	60
3.5	Greedy heuristic RB selection algorithm	63
3.6	Simulation methodology	64
3.7	Results	70
3.7.1	Performance of heuristic D2D RB assignment algorithm for dif- ferent scheduling algorithms	70
3.7.2	Comparisons between random and heuristic D2D RB assign- ment algorithms	73

3.7.3	The impact of number of D2D connections on network throughput	75
3.8	Summary	76
4	Collaborative streaming employing device-to-device multicast	78
4.1	Agent and assisted peer selection as stable admission problem	79
4.2	The proposed collaborative streaming system	82
4.2.1	Agent selection and group formation	82
4.2.2	The content dissemination technique	87
4.2.3	Group maintenance	89
4.3	Stochastic properties of groups	89
4.3.1	Expected number of groups	92
4.3.2	Agent to farthest assisted peer expected distance	94
4.4	Performance evaluation	95
4.4.1	Parameters that affect the number of groups	97
4.4.2	eNB traffic off-loading and bandwidth saving	100
4.4.3	The impact of group size	102
4.4.4	Peer's experience of streaming rate	104
4.4.5	Streaming cost and fairness issue	105
4.5	Summary	108
5	Streaming chunk distribution: An epidemic spreading approach	110
5.1	System Model and Assumptions	110
5.2	The spread of chunks: an epidemic analogy	112
5.2.1	The deterministic formulation	114
5.2.2	The stochastic formulation	125
5.3	Simulation	128

5.3.1	The accuracy of the model	129
5.3.2	Different chunk selection algorithms	132
5.3.3	Influence of peer population and bandwidth to playback speed ratio on buffer length	133
5.4	Summary	135
6	Conclusion and future research	140
6.1	Summary of the work	140
6.2	Thesis contributions	143
6.3	Open problem and future Work	144
	References	148

List of Tables

3.1	Simulation parameters and values	77
-----	--	----

List of Figures

2.1	The relationship between physical and logical networks in the proposed C-Chord system. For the clarity of the picture, only the special key and part of the network are shown.	16
2.2	The ISP-controlled <i>Oracle</i> provides assistance on peering decision. For clarity, we only show logical connections between peers and <i>Oracles</i> , and among <i>Oracle</i> themselves.	27
2.3	Media-aware network elements for decoding and image resolution compatibility.	30
2.4	System level overview of LTE architecture.	32
3.1	The LTE downlink physical resource.	51
3.2	Interference problem when the primary cellular network shares channel with the D2D network.	55
3.3	Distance dependent macroscale path-loss map (in dB, including antenna gain) with 7 eNBs (in white circle).	67
3.4	Uncorrelated and space-correlated lognormal shadow fading path-loss map (in dB).	67
3.5	Simulation region (of size $1500 \times 1500 \text{ m}^2$) showing path-loss (dB) that includes antenna gain, distant dependent macroscale path-loss, and lognormal shadow fading path-loss.	68

3.6	Normalized network throughput for different scheduling algorithms with the proposed greedy heuristic RB assignment algorithm [scenario 1: the number of D2D connection is 10% (of total UEs)].	71
3.7	Normalized network throughput for different scheduling algorithms with the proposed greedy heuristic RB assignment algorithm [scenario 2: the number of D2D connection is 20% (of total UEs)].	72
3.8	Comparisons between the GA based solution and the greedy heuristic RB assignment algorithm in terms of normalized aggregate throughput for the PF scheduling approach.	73
3.9	Comparisons between random and heuristic D2D RB assignment algorithms in terms of normalized aggregate throughput.	74
3.10	Impact of the number of D2D connections on normalized aggregate throughput.	75
4.1	Cellular peers under an eNB from groups. Agents download contents from the Internet peers, and then share with multicast receivers employing D2D multicast. The D2D maximum distance $d_M = 30\text{ m}$. The maximum number of assisted peers an agent assists, $S = 3$. For clarity, only a part of the network is shown.	84
4.2	Node distribution in a compact set W in a tow-dimensional region $\mathbb{R}^2$. C the set of a perfect disk with center at o and radius r	90
4.3	The effect of D2D maximum distance and peer density on number of groups. The maximum number of allowable assisted peers is set to '4'.	97
4.4	The effect of maximum number of assisted peers on number of groups. The D2D maximum distance is set to 20 meter.	99

4.5	The comparison between simulation and analytical model in terms of number of groups. The D2D maximum distance and maximum allowable assisted peers in a group are set to 20 m and ‘3’ respectively. . .	100
4.6	The variation of eNB load with peer density and D2D maximum distance. The maximum allowable assisted peer is set to 4.	101
4.7	The variation of eNB load with peer density and maximum allowable assisted peer. The D2D maximum distance is 20 meter.	102
4.8	The variation of average distances from an agent to its farthest assisted peer with restriction on group size. There are 100 peers per eNB in the system. The D2D maximum distances are 15 and 25 m	103
4.9	The variation of agent to farthest peer distance with peer density. The maximum allowable assisted peers in a group and D2D maximum distance are set to ‘1’ and 25 m	104
4.10	Performance measurement in terms of streaming rate experienced by peer with varying network density. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30 m$ and 4 respectively.	105
4.11	Per peer streaming cost averaged over time with varying number of cellular peers. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30 m$ and 4 respectively. . .	106
4.12	Comparison of streaming cost fairness for different P2P systems. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30 m$ and 4 respectively.	107
5.1	The sliding window mechanism of a peer’s buffer B	111

5.2	The spread of chunks among peers in terms of buffer position.	123
5.3	The transition probability from state i to all possible states in Markov process.	126
5.4	The comparison among deterministic and stochastic versions of epidemic models, and simulation model with RF and MP algorithms. The stochastic version is repeated over ‘50’ times. $N = 1000$, $M = 40$, and $b/v = 1$	130
5.5	The gap between different models as a function of peer population size for RF algorithm. The stochastic version is repeated over ‘50’ times. $M = 40$, and $b/v = 1$	131
5.6	The gap between different models as a function of peer population size for MP algorithm. The stochastic version is repeated over ‘50’ times. $M = 40$, and $b/v = 1$	132
5.7	Comparison between various chunk selection strategies. $N = 1000$, $M = 40$, and $b/v = 1$	133
5.8	The effect of peer population size on RF and MP chunk selection strategy when the bandwidth to playback ratio remains constant ($b/v = 1$) as derived from the deterministic model.	134
5.9	The effect of peer population size on RF and MP chunk selection strategy when the bandwidth to playback ratio remains constant ($b/v = 1$) as measured in simulation model.	135
5.10	The effect of peer population size on HS when the bandwidth to playback speed ratio remains constant ($b/v = 1$) as derived from the deterministic model.	136

5.11	The effect of peer population size on HS when the bandwidth to playback speed ratio remains constant ($b/v = 1$) as measured in simulation model.	137
5.12	The effect of bandwidth to playback speed on RF and MP chunk selection strategy with peer population remains constant ($N = 1000$) as derived from the deterministic model.	138
5.13	The effect of bandwidth to playback speed on HS with peer population remains constant ($N = 1000$) as derived from the deterministic model.	138
5.14	The effect of bandwidth to playback speed on RF and MP chunk selection strategy with peer population remains constant ($N = 1000$) as measured in simulation model.	139
5.15	The effect of bandwidth to playback speed on HS with peer population remains constant ($N = 1000$) as measured in simulation model. . . .	139

List of Symbols

$>_u$	Preference orderings of peer u
Δy_m	Additional number of infectious peers within dt time for location m
Φ	Point Process
α	A constant that captures MP algorithm's procrastination behavior
β	Infection factor
γ	SINR
γ_d^{DL}	Downlink phase received SINR
γ_{eNB}^c	SINR at the base station due to CUE transmitter c
γ_{max}	Maximum SINR
γ^{MP}	The proportionality constant of MP algorithm
γ^{RF}	The proportionality constant of RF algorithm
η	Spectral efficiency
θ_{3dB}	3 dB beam width
Φ'	Thinned point process
λ	The density of peers
ξ	Constant path-loss exponent

ρ	Probability
$\mathcal{A}$	A set of agents
$\mathcal{M}$	A set of assisted peers
$\mathcal{O}$	A set of outliers
$\mathbb{R}^2$	A planar space
A	The area of the cellular coverage
$A_{dB}(\cdot)$	Attenuation in dB
A_m	Maximum attenuation
b	Peer's upload bandwidth
$B(1)$	The buffer space for the newest chunk
$B(M)$	The buffer space of the oldest chunk and peer's playback side
$c(o, r)$	A perfect disk of center o and of radius r
d_{corr}	Decorrelation distance
d_G	Random Euclidean distance of an agent to its farthest assisted peer
d_M	D2D maximum distance
d_{uv}	Distance between transmitter u and receiver v
d_u^{eNB}	The distance of peer u from the eNB
$E(d_G)$	Expected distance between an agent to its farthest assisted peer

$E(N_G)_{Matérn}$	Expected number of groups (Matérn hard-core process)
$E(N_G)_{prop}$	Expected number of groups (proposed method)
f_c	Carrier frequency
$f_{R_n}(\cdot)$	<i>pdf</i> of Euclidean distance
$F_{R_n}(\cdot)$	<i>cdf</i> of R_n
$f_{MP}(p_m)$	The function of chunk occupancy probability at m -th buffer position for MP algorithm
$f_{RF}(p_m)$	The function of chunk occupancy probability at m -th buffer position for RF algorithm
G_{uv}	The linear gain between u and v
h	Minimum distance for in hard-core operation
h_b	Base station antenna height
K_{uv}	Normalization constant
r	Distance
r_c^{DL}	Rate of communication corresponding to the SINR γ_c^{DL}
R_n	Random variable indicating distance of the n -th nearest peer
$L_{dB}(\cdot)$	Macroscopic path loss
m_c	Number of resources allocated to CE c

m_{sw}	Chunk position where RF algorithm switches to MP algorithm
M	Number of chunks in a buffer
p	Probability that any peer lies in $c(o, r)$
P_d	Transmission power of a device d
$PL_{dB,u,v}(\cdot)$	The total path-loss between u and the user v
p_m	Probability of finding the chunk at buffer position ' m ' at any instant of time
$p_{m,t}$	Probability of finding the chunk at buffer position ' m ' at time t
p_M	Peers playback continuity
p_r^m	Probability of a susceptible to convert to an infective peer
$p_r^{m,MP}$	Probability of a susceptible to convert to an infective peer for MP algorithm
$p_r^{m,RF}$	Probability of a susceptible to convert to an infective peer for RF algorithm
P_{sw}	Buffer occupancy probability at the switching chunk position
S	Maximum number of assisted peers
X_u	Lognormal shadow fading path-loss
N	Number of peers
N_0	Receiver's noise figure and thermal noise density

$\mathcal{N}_u$	A set representing a neighborhood of peer u
v	Playback speed
W	A compact set
$ W $	The area of $\mathbb{R}^2$
W_{eff}	Link level implementation overhead
$x^{(p)}$	The rising Pochhammer symbol
x_m	The number of peers that are susceptible to be infected due to disease m
y_m	The number of peers that have a chunk at m -th buffer position or number of infective peers due to disease m
y_{sw}	The number of infective peers at the time of switch
y'_{m+1}	The number of newly infected peers at buffer position $m + 1$

Abbreviations

3GPP	Third Generation Partnership Project
4G	Fourth-generation
<i>a</i> -Chord	Auxiliary Chord
ACK	Acknowledgement
ARQ	Automatic Repeat re-Quest
BGP	Border Gateway Protocol
BPP	Binomial Point Process
CAN	Content Based Network
CCCH	Common Control Channel
<i>cdf</i>	Cumulative Distributive Function
CDS	Connected Dominating Set
CR	Cognitive radio
CSMA	Carrier Sense Multiple Access
CQI	Channel Quality Indicator
C-Chord	Cellular Chord
Cell-ID	Base Station's Unique Numeric Identification Number
CDN	Content Distribution Network
CUE	C-Mode User Equipment
D2D	Device-to-Device
dB	decibel
DHT	Distributed Hash Table
DUE	D-mode User Equipment

E-PC	Evolved Packet Core
E-UTRAN	UMTS Terrestrial Radio Access Network
eNB	eNodeB
GGSN	Gateway General Packet Radio Service Support Node
GPS	Global Positioning System
HB	Heart Beating
HARQ	Hybrid Automatic Repeat Request
<i>id</i>	identifier
<i>i.i.d</i>	Independent and Identically Distributed
IP	Internet Protocol
ILP	Integer Linear Programming
ILA	Interference Limited Area
ISP	Internet Service Providers
kHz	Kilohertz
LC	Layered Coding
LTE	Long Term Evolution
LTE-Adv	Long Term Evolution Advanced
<i>m</i>	meter
<i>m</i> -Chord	Main Chord
MAC	Media Access Control
Max C/I	Maximum Carrier to Interference ratio
MANE	Media-Aware Network Element
MCS	Modulation and Coding Schemes
MHz	Megahertz

MDC	Media Description Coding
MINLP	Mixed Integer Nonlinear Programming
MBMS	Multimedia Broadcast/Multicast Service
MNO	Mobile Network Operator
NACK	Negative Acknowledgement
NAT	Network Address Translator
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	OFDM Access
<i>pdf</i>	Probability Distribution Function
P-GW	Packet Data Network Gateway
P2P	Peer-to-Peer
PAPR	Peak-to-Average Power Ratio
PDN	Packet Data Network
PF	Proportional Fair
PPP	Poisson Point Process
QoE	Quality of Experience
RB	Resource Block
RR	Round Robin
RTT	Round Trip Time
S-GW	Serving Gateway
SC-FDMA	Single Carrier Frequency Division Multiple Access
<i>std</i>	Standard Deviation
SINR	Signal to Interference and Noise Ratio
sHMG	Seamless Home Media Gateway

SI	Susceptible-Infective
SIS	Susceptible-Infective-Susceptible
SIR	Susceptible-Infective-Recovered
SM	Stable Marriage
SMS	Short Message Service
sNMG	Seamless Network Media Gateway
SNR	Signal to Noise Ratio
TDD	Time Division Duplex
TTL	Time-to-Leave
UDP	User Datagram Protocol
UE	User Equipment
VoD	Video on Demand
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WWAN	Wireless Wide Area Network

Chapter 1

Introduction

In this thesis, we propose a collaborative streaming system where some cellular peers download contents from the Internet peers, and then share the contents with geographically closed peers by employing operator controlled user-level multicast. The proposed system avoids bottleneck at the base station, and saves wireless bandwidth as much as possible but still solves the feedback implosion problem and achieves an effective multicast rate. We also propose an efficient streaming segment distribution strategy that ensures segment diversity and streaming chunks availability to improve the streaming quality.

1.1 Motivation

Contrary to increasing offers of appealing applications and services, we are being less wired everyday. Fourth-generation (4G) mobile technology aims at meeting the growing demands of supporting data applications at a faster speed. Mobile video traffic would exceed Exabyte per month by 2012-2013 [1], 20 years after the first short-message-service (SMS) was sent, whereas the online web reached the same milestone in 2004, more than 30 years after the first email was sent. This figure indicates the tremendous impact of wireless video contents in the near future. Peer-to-peer (P2P) live streaming is becoming extremely popular to the users who enjoy the Internet via wired broadband access or WiFi access that is connected to a wired backbone network.

This feature is still unavailable to mobile users even though they have access to the Internet.

Recently, device-to-device (D2D) communication as an underlay to cellular networks has been introduced as a technology component to the Long Term Evolution advanced (LTE-Adv) [2]. Cellular users in close proximity, with higher mutual signal to interference and noise ratio (SINR), may communicate directly rather than through the LTE base station. It would be highly advantageous, if some cellular users download content through the LTE base station while other cellular users retrieve it through D2D communication and thereby avoid congestion at the base station. Realizing this impact, industries are working on proximity-aware P2P applications. Qualcomm's recent innovation 'Flashlinq' technology allows devices to communicate directly in an operator spectrum [3]. 'Third Generation Partnership Project' (3GPP) is working on standardizing D2D communication that would provide a breakthrough to the collaborative streaming approach. The WINNER+ project [4] has also acknowledged the D2D communication as one of the solutions to overall throughput improvement. The next step is towards a live P2P streaming service on handheld devices. SopCast [5], a P2P streaming system over the Internet, already launched its mobile counterpart for Android devices. Other similar online streaming services will follow the suit.

The influence of mobile streaming systems, from a social network point of view, would be enormous. One can be a broadcaster, or even build one's own TV station without the help of a commercial server, using minimal resources. Imagine that you are enjoying a movie online on your tablet at home and you would like to share it with your family. Rather than huddling around the small screen, you can share contents on other D2D enabled devices by employing a P2P streaming protocol instantly. In

the world we live in today, people want entertainment anytime, anywhere. Social and community networks exhibit high locality characteristics. Whether you are in airports, stadiums, shopping centers, subway stations, traffic jams, or in any other crowded area, you would not miss any on-air TV shows. On a long distance train, a good number of people might be interested in watching the same live content (e.g., the coverage of an Olympic event). Multiplayer gaming and proximity video advertising could also benefit from the mobile P2P streaming concept. The rise of tablet usage has also boosted consumer demand for the cellular P2P streaming service.

1.2 Problem statement

Streaming over a P2P network is popular because of its scalability to network size and self-organizing nature. The paradoxical reality is that peers scatter randomly, which leads to needless traffic traversal through multiple links within a provider's network. Cellular channels are limited in number and are expensive. Besides, for cellular peers, all streaming content comes through the base station, raising the issue of congestion, and the system is limited by the capacity of the base station.

D2D communication as an underlying cellular network may empower user-driven rich multimedia applications. The short range of D2D transmission exploits a frequency reuse technique within a cell. However, D2D transmitters may cause a significant amount of interference to the primary cellular network when they share radio resources. The interference from D2D communications to the cellular network needs to be restricted to maintain a target performance level of the cellular network.

Geographically closed peers may collaborate to share content among them through a D2D multicast mechanism. Multicast over the wireless network may exploit the

inherent broadcast nature of the shared wireless channel. The time-varying nature of a wireless link requires a feedback mechanism to recover any missing segments of the content. Without a proper feedback method, it is also difficult to estimate the optimum transmission rate. Ironically, the same physical property of wireless links pose a challenge for incorporating a feedback mechanism while ensuring quality. The feedback mechanism may overwhelm the sender especially in a dense network.

P2P streaming applications are entirely user-driven, thus cooperation from Internet service providers (ISPs) or mobile network operators (MNOs) is not feasible. The LTE base station does not participate in the streaming process and is completely oblivious to user applications. Also, D2D communication can show better performance compared to the cellular network by serving the cell edge users or the users in poor channel conditions. A self-organizing and decentralized grouping algorithm is required. Designing such collaborative streaming system is extremely challenging as a smaller number of groups is desirable to save wireless bandwidth; at the same time proximity is required to increase the multicast rate. Choosing neighbors of close proximity is even more critical for video streaming due to a high volume of data transmission.

For streaming applications, users need to download a certain range of segments (also known as chunks) in a timely fashion. Chunks arriving after their scheduled playback time are useless; playback has to skip such content compromising the user-perceived visual quality. Meeting time constraint is therefore critical for the streaming service. The chunk selection strategy should facilitate chunk distribution by improving chunk diversity and ensuring the availability of chunks closer to the playback region. Thus an effective chunk selection strategy is required to improve playback smoothness.

In a P2P system, the overlay network is an application-level network on top of the physical network formed by the users (peers). In the proposed collaborative streaming system, cellular peers and Internet peers form overlay to share content among themselves. The cellular peers form topology-aware groups employing D2D communications. The D2D network lays beneath the cellular network keeping its interference to the cellular network below a tolerance level. Therefore we use the term ‘underlay’ referring to the part of the P2P network formed by the cellular users only using the D2D technique.

1.3 Summary of contributions

- We address the spectrum sharing problem between the 3GPP-LTE cellular network and underlaying the D2D communication network. We identify and analyze the interference problem of the primary cellular network caused by the D2D transmitter during the uplink and downlink phases separately.
- To the best of our knowledge, we are the first to formulate the D2D shared radio resource allocation problem as a mixed integer nonlinear programming (MINLP) problem where the LTE base station (also known as eNodeB) controls D2D connections. We propose a penalty-based heuristic genetic algorithm (GA) to solve the MINLP. We also propose an alternative greedy heuristic algorithm which can diminish interference to the primary cellular network utilizing channel gain information.
- The proposed greedy heuristic D2D radio resource allocation algorithm does not require any modification to the scheduling of the primary cellular network;

rather, it utilizes information provided by the primary scheduler. We also perform extensive simulations to prove the efficacy of the proposed algorithm over the random selection of radio resources.

- We infer the broadcaster/agent selection problem as the stable admission assignment problem and formulate the problem as an integer linear programming (ILP). The proposed system avoids bottleneck at the base station, and saves wireless bandwidth as much as possible, while still solving the feedback implosion problem and achieving an effective multicast rate.
- We also propose and implement a novel distributed algorithm to select the broadcasters where peers need information of only one-hop away neighbors. We show an application of D2D communication by designing a streaming system where some cellular peers download from the eNodeB (eNB), and then share with neighboring peers. We also provide the details of a content dissemination technique.
- The eNB loads, due to streaming content, largely depend on the number of groups. On the other hand, the assisted peer streaming rate critically relies on its distance from the agent. We use inter-peer distance distributions in spatial processes to find the tradeoff between eNB traffic reduction and peer's experience in terms of the streaming rate.
- We derive an analytical model that captures the chunk distribution process from a single source to the entire network under the influence of different chunk selection strategies. We apply the theory of epidemiology to derive the analytical model of the chunk distribution of live streaming content.

- The model finds the buffer occupancy probability and provides insight on relationships among peer population size, buffer length and playback smoothness. We demonstrate the accuracy of the analytical model by comparing it to the results obtained through simulations.
- We propose a hybrid chunk selection method that is scalable to peer population in terms of buffer space by exploiting chunk diversity. The hybrid method also achieves a better playback continuity by distributing urgent chunks first. Utilizing the model, we evaluate the performance of different chunk selection strategies: rarest first, most popular chunk and hybrid scheme.

1.4 Publication list

The work presented in the thesis has been published partially in a number of peer-reviewed conferences and journal as listed below:

- J1. M. Zulhasnine, C. Huang, and A. Srinivasan, "Towards an effective integration of cellular users to the structured Peer-to-Peer network," *J. Peer-to-Peer Netw. Appl.*, Springer, vol. 5, no. 2, pp. 178-192, 2012.
- B1. M. Zulhasnine, C. Huang, and A. Srinivasan, "P2P Streaming over Cellular Network: Issues, Challenges, and Opportunities", book chapter in *Building Next-Generation Converged Networks: Theory and Practice*, CRC press, USA, pp. 97-117, 2013.
- C1. M. Zulhasnine, C. Huang, and A. Srinivasan, "Exploiting cluster multicast for p2p streaming application in cellular system", in *Proc. Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, pp. 1-6, 2013.

- C2. M. Zulhasnine, C. Huang, and A. Srinivasan, "Topology-aware integration of cellular users into the P2P system" in *Proc. IEEE Vehicular Technology Conference Fall (VTC)*, San Francisco, USA, pp. 1-5, 2011.
- C3. M. Zulhasnine, C. Huang, and A. Srinivasan, "Penalty Function Method for Peer Selection over Wireless Mesh Network," in *Proc. IEEE Vehicular Technology Conference Fall (VTC)*, Canada, pp. 1-5, 2010.
- C4. M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient Resource Allocation for Device-to-Device Communication Underlying LTE Network," in *Proc. IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Canada, pp. 368-375, 2010.
- C5. M. Zulhasnine, C. Huang, and A. Srinivasan, "Favorable Peer Supported Throughput Optimization in Wireless Mesh Network," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, USA, pp. 1-5, 2010.

1.5 Outline of the thesis

We organize the remaining part of this thesis as follows. Chapter 2 reviews literature on P2P streaming in the context of wireless networks, D2D communications, and streaming chunk distribution strategies. This chapter also provides reviews on stable marriage and admission assignments, spatial point processes and epidemic theory as these are relevant to our work. Chapter 3 presents the resource allocation algorithm of the LTE network with underlying D2D communication. It presents a heuristic algorithm that enables D2D communication while limiting interference to primary cellular users. Chapter 4 presents the proposed collaborative streaming system. This

chapter illustrates the details of group formation employing the stable admission assignment and also presents an analytical approach to find the relationships among various factors affecting the streaming quality. Chapter 5 presents the chunk distribution model and also provides an analysis of different chunk selection strategies followed by the proposed hybrid chunk selection method. This chapter also discusses an alternate stochastic epidemic approach. Finally, chapter 6 concludes our work with future directions toward enabling P2P streaming over the cellular network.

Chapter 2

Literature Review

With an exploding number of handheld devices, from smart-phones, ePad, gaming consoles, to portable media players, the design of a cellular P2P streaming system is an important consideration, especially in the context of a next generation converged network. In this chapter, we first provide a comprehensive overview of the recent advancement in the area of wireless P2P steaming. We present architectures, protocols and issues towards enabling P2P streaming on wireless devices. We try to design a streaming system that exploits the D2D technique. Therefore we will address various issues related to a D2D network. We try to realize the chunk distribution process in the P2P system as it defines the P2P network capacity. Our literature review therefore also covers chunk distribution strategies. Fianlly we provide literature reviews on stable marriage, spatial point processes, and epidemic theory. We try to solve our streaming system related problems by applying these theories.

2.1 Steaming over cellular network: issues and challenges

We first describe the mechanism of constructing efficient overlay in detail. Then we illustrate a few existing wireless P2P streaming systems with the essence and strategy of peer-selection. We also discuss various issues on realizing a wireless streaming

system.

2.1.1 Designing an efficient overlay network

In the P2P system, users (peers) form an application-level overlay network on top of the physical network. Based on the overlay formation technique, P2P systems are classified into unstructured P2P systems such as Gnutella [6], KaZaA [7] and structured P2P systems such as Chord [8], Pastry [9], and content based network (CAN) [10]. Unstructured lookup protocols, in [6, 7], flood the queries with restrictions and therefore accrue considerable message overhead and lessen search accuracy. Structured lookup protocols in [8–10], construct the overlay network using a distributed hash table (DHT), thereby improving search efficiency and accuracy.

Streaming over P2P networks relies on an overlay network, and yields some penalties in terms of bandwidth usage due to the mismatch between the physical and the overlay network. The P2P streaming system needs to feed peers with current content continuously. Because of the magnitude of contents per unit time and the nature of playing same content throughout the entire system, an efficient overlay construction is the key to realize the cellular streaming system. Here, we will discuss a few efficient overlay systems.

Proximity-aware topology construction

Several researchers presented topology-aware clustering schemes where peers partition into groups depending on proximity-based network metric (e.g., latency, or round trip time) [11, 12]. This way, peers belonging to the same group are close to one another.

Network metric based clustering: The method proposed in [11] used information

derived from border gateway protocol (BGP) routing table snapshots, while the other [12] exploited landmark nodes to measure the proximity. The authors in [13] presented a topology-aware overlay named ‘mOverlay’. In mOverlay, geographically close peers form a group, using a dynamic landmark procedure. Unlike [12], mOverlay does not require any extra deployment of landmark nodes, rather groups within the mOverlay serve as the landmark nodes for a joining peer. The joining peer first contacts a peer of an arbitrary group. The peer then measures distance from that given group. The given group also provides their proximity information from other groups. The peer then joins the group with the closest distance. There is overhead to guide a joining node in finding its closest existing cluster.

Network prefix based clustering: Huang et al. [14] proposed a P2P scheme where peers form clusters according to their network prefix divisions. This divides the network into multiple network-aware clusters, and helps peers to search files in nearby peers first. One super-peer from each cluster maintains indexes of shared files and peers’ location. The bootstrap peer maintains an up-to-date cluster routing table to direct joining peers to the appropriate clusters. The bootstrap peer requires routing information from a BGP router to update the cluster routing table. The authors conducted simulations on WiFi-connected peers. However, peers, belonging to the same cluster and having the same network prefix division, could be far away from each other, let alone within a WiFi communication range. The presence of anonymous routers, whose existence can be detected (but not their addresses), also leads to distorted and inflated inferred topology.

Base station aware clustering: The work in [15] divided wireless users according to their Worldwide Interoperability for Microwave Access (WiMAX) base stations. Peers from the same base station form a local mesh network, exploiting WiFi-connectivity,

and the base stations construct a DHT-based overlay network. Peers first search content within the local mesh, and then on the DHT based system if required. The results show that this method can greatly reduce the number of lookup messages in physical networks. However, base stations' participation to the P2P system is not realistic. The method also requires knowledge of the physical layer and therefore is unsuitable for an application driven P2P system.

Physical layer dependent proximation: Wang and Ji [16] proposed a topology-aware P2P protocol based on a Chord for the mobile ad hoc network where physically close nodes get adjacent identifiers (*ids*) exploiting nodes' position information. Entering node broadcasts a request to join the existing wireless P2P network. Any peer, upon receiving the entering node's request, calculates the relative distance from the energy level of the received signal. From the received energy strength and *id* space margin information, the entering node gets a node *id* adjacent to the physically close nodes. This method is only suitable for the infrastructure-less wireless network and does not consider the Internet gateway in the design. The cost of device discovery should not outweigh the benefits of proximity-aware communication.

Geographical location based proximation: Canali et al. [17] proposed a location-aware overlay network, named MeshChord, for a wireless mesh network. To acquire location-awareness, the algorithm assigns adjacent *ids* to physically close peers. However, this method requires location information possibly by using global positioning system (GPS) enabled receivers. The users also may not feel comfortable to release their geographical location due to privacy and security reasons. The DHT-based P2P system 'GeoRoy' [18] is a location-aware variant of Viceroy [19]. GeoRoy is based on a two-tier architecture where mesh routers (super peers) form the DHT-based upper tier, and the mobile mesh clients (leaf peers) provide content to the system. The

DHT, in the upper tier, helps lower tier users to index the content. Thus, the super peers only provide a distributed resource catalog held by the leaf peers, and leaf peers publish and request resources with the aid of the super peers. The peers' *ids* are computed from geographical information in both GeoRoy [18], and MeshChord [16]. However, MeshChord is built on top of a Chord, while GeoRoy is built on top of Viceroy. The use, of a geographically aware hash function, reduces the mismatch between the overlay network and the underlying network and thereby makes the resource lookup more efficient in terms of hop distance.

Hierarchical overlay construction

A hierarchical DHT system improves the lookup performance by utilizing super peers. However, load balancing among super peers is a crucial problem of the hierarchical DHT system. In addition, such a system must avoid isolation of regular peers if a super peer fails.

Super peer based hierarchy: The P2P system by nature exhibits heterogeneity, and 'super peers', which are more powerful and stable, do exist. The authors in [20] proposed a two-level hierarchical Chord, named Chord², where super and regular peers join in two separate tiers. The motivation therein is to reduce the cost of maintaining the DHT-based P2P system. Regular peers, which form the outer tier, perform periodical maintenance of their successor links only. They inform the super peer if a change is detected due to a peer joining/leaving process. The super peer sends finger table update notifications to the responsible regular peers only. This way Chord² avoids periodical refreshing of the finger table by the regular peers and reduces maintenance costs. In [21], the network is divided into several groups. Each group consists of one stable and several unstable peers (who join/leave very frequently).

A stable peer gets higher bits on the left side of the *id* space and plays a more important role of routing and replication. The unstable peer gets a node *id* with lower bits on the left side that makes its *id* region as small as possible. The unstable peers only assist the reliable peers. The authors claimed that such a system minimizes maintenance overhead. Super-peers should be uniformly distributed throughout the overlay network and meet some other properties such as load balancing, adaptability to peer churn, and leveraging peer heterogeneity.

DHT based hierarchy: Zoels et al. [22] proposed a load-balancing algorithm for the two-tier DHT system with disjoint groups. To achieve an equal load for every super peer, a distributed load balancing algorithm determines which super peer the joining peer would connect to. Such a load balancing approach may create a mismatch between the physical layer and the logical layer of the overlay network.

DHT-based proximity-aware overlay

In ‘eQuus’ [23], a DHT based P2P system, network metric based close peers are grouped into the same cliques. CliqueStream [24], a P2P live streaming system, is built on top of ‘eQuus’ [23] to exploit proximity in terms of distant metric. In this system, nodes within a clique form a mesh structure and more stable nodes (in terms of available bandwidth) interconnect cliques in a tree structure. Cliques are forced to split and merge to maintain an upper and lower bound for the number of nodes. The authors in [25] proposed a topology-aware DHT system based on a Chord. In this scheme, the *id* assignment is relaxed, and logical successors are chosen according to their physical proximity in the ad hoc network. However, it is expensive to discover physically close peers in a larger network and often appears impractical.

Example of an application-driven efficient overlay system

In [26], we proposed an efficient overlay network named cellular Chord (C-Chord), which integrates cellular users into the P2P network in a topology-aware fashion. The hierarchical overlay system ‘C-Chord’ conceptually separates stable wired peers and wireless peers. Whereas the stable Internet peers form the large main Chord (m -

FIGURE 2.1: The relationship between physical and logical networks in the proposed C-Chord system. For the clarity of the picture, only the special key and part of the network are shown.

Chord), contrary cellular users form several (equal to the number of base stations) auxiliary Chords (a -Chords). The cellular peers, under the same base station, belong to a particular a -Chord. The base station does not participate in any of the Chords, and cellular users only utilize the base station’s unique numeric identification number (Cell-ID) to form the a -Chord. The a -Chord organizes the physically close cellular peers together. The Cell-ID serves as a rendezvous point for the peers under the same access point. Fig. 2.1 shows the relationship between physical and logical networks in the proposed C-Chord system. Here, the m -Chord consists of 4-wired

peers and each α -Chord consists of two cellular peers. Cellular users have the option of choosing between Internet peers and cellular peers. The cellular peers can either download diverse content from the stable Internet peers or download social content from the peers within the same base station. While the former option provides content diversity, the latter option avoids the Internet data penalty. Our proposed system reduces the high management cost of routing information with fewer entries in the routing table for cellular peers.

2.1.2 Wireless P2P streaming system

Recently several authors proposed wireless P2P streaming systems such as COSMOS [27], LocalTree [28], COMBINE [29], WiMA [30], and MOVi [31]. In COSMOS [27], one peer in a group randomly selects one video description and still it enjoys full quality of video through sharing. This feature saves valuable cellular bandwidth. However that peer requires peers with remaining video descriptions in its communication range. The peers interchange their roles alternatively to ensure the cost fairness. The density of peers always plays a crucial role in determining the rebroadcast scope. To avoid duplicate packets and the contention of the wireless medium, peers also exchange neighboring information and streaming buffer maps periodically. COSMOS employs a dynamic broadcast instead of a fixed broadcast where peers determine the broadcasting scope depending on their local density, thereby reduce flooding and channel redundancy. However the scheme does not mention how Internet peers and close-proximity wireless peers are organized. Also there is no feedback mechanism to recover any missing segments of the content. In LocalTree [28], a few cellular peers download streaming content from the Internet and then redistribute

the content at a multi-hop distance over the WiFi-connected unstructured mesh network. Each peer acquires received states of packets from its neighbors. To avoid collision, a peer waits in proportion to the number of received packets per neighbor before rebroadcasting. The peer, that acquires the least number of received packets per neighbor, is the most preferred to rebroadcast. Each time a peer rebroadcasts, it also decreases time-to-leave (TTL) by '1'. Peers in such a system may calculate the distance from the puller in terms of hop count from the TTL field. Any peer experiencing a stable hop count may join to construct a tree to further reduce battery consumption. The TTL determines the depth of the tree. The base mesh structure provides resilience against peer's connection failure. The proposed method minimizes battery consumption by reducing the rebroadcasting scope and constructing a stable tree on top of mesh-based overlay. Peers with better connections to the base station may fetch content from Internet peers in a more energy optimal way. The feedback problem would be severe in LocalTree. Ensuring playback smoothness is extremely difficult otherwise impossible because of the multi-hop distance. COMBINE [29] differs from other wireless streaming systems on how it forms the shared groups. In COMBINE, a requester (termed as initiator) identifies a set of partners in its vicinity and forms a shared group in an energy efficient way. To join the group, each potential member wakes up periodically to broadcast a short message, and keeps their wireless local area network (WLAN) cards in a low energy mode for the remaining time. The initiator always keeps its WLAN card switched on to discover such a message. The message contains the cost of the download (in terms of the monetary and the battery energy) and expected wireless wide area network (WWAN) speeds of the corresponding potential member. The initiator then chooses a set of partners based on received messages. WiMA [30] suggests that wireless peers with greater abilities

serve as broadcasters. However, WiMA does not describe how the comparison is made to select broadcasters from wireless peers. To find arriving peers, each broadcaster sends a heart beating (HB) packet periodically. The joining peer, upon receiving the HB packet, requests streaming content from the agent. In the stable condition, the agent simply broadcasts content to assisted peers. The assisted peer downloads missing contents from some other peers. WiMA also does not describe how an agent delivers content without a collision in the presence of another agent. The streaming quality of the wireless peers depends heavily on the availability of streaming contents in the agent. The wireless streaming system MOVi [31] deploys a centralized MOVi server located at the back-haul network. The server manages the connectivity maps of shared devices, tracks media segments and schedules the broadcasters to disseminate streaming content to the served-peers. The shared users serve as the cache for the media contents, and also responsible for data transfer under the control of the MOVi server. All of the above-mentioned streaming systems either lack an efficient group formation technique or do not have any feedback mechanism. The wireless service operators also do not have incentive as all of them use free broadcast channels such as the WiFi or the Bluetooth. Also none of them considered an efficient chunk distribution strategy which is the key to achieve playback smoothness and to ensure streaming quality.

Another way to save cellular bandwidth is to deploy cache. The authors suggested deploying cache at the mesh access point (MAP) that has inherent gateway to the Internet [32]. Clients form a P2P network within themselves and with the cached-MAPs and download/share content aided by a P2P tracker. However, this scenario requires ISP cooperation which is most unlikely in the context of the P2P network. A femto base station usually has large storage capacity and high-bandwidth

communication capability with its local users. However, a femto base station may not have enough high-speed backhaul to feed the aggregate data rate of all its served users. The authors in [33] proposed to cache the popular video content in femto base stations. Femto base stations may deliver the content to their local users upon requests by taking advantage of the short range communication. The cache placement problem is an active research area. The idea of video caching will not work for the live streaming case as the source generates content at real time and all users play the content at the same time.

Feedback mechanism to recover missing chunks

Multicasting in a collaborative streaming system exploits the inherent broadcast nature of the shared wireless channel. Ironically the same physical property poses a challenge in incorporating a feedback mechanism while ensuring quality. The feedback mechanism may overwhelm the sender, especially in a dense network. Without a proper feedback method, it is also difficult to estimate the optimum transmission rate.

We just open Pandora's box by stating that a feedback mechanism is essential to ensure quality while multicasting over the wireless channel. In traditional Automatic Repeat re-Quest (ARQ) error control method, an acknowledgement (ACK) is sent by the receiver to the sender to inform them that a correct frame has been received. The sender retransmits the corresponding frame if no ACK is received. When the traditional ARQ method is implemented in multicasting over the wireless channel, high bandwidth and extreme coordination are required to process ACKs by each receiver. This leads to a feedback implosion problem. An alternative solution to this problem would be to implement negative ACKs (NACKs) where the receiver(s) only

send(s) a NACK feedback to the sender upon receiving an erroneous frame. Even with the NACK based protocol, significant bandwidth is consumed and overhead is accrued if the sender needs to identify receivers that did not receive a correct frame. One way to circumvent the disadvantages and provide scalability is to combine the NACK messages. Still, the sender has to compromise the enumeration of the number of receivers with NACK messages, and is unable to determine the suitable rate of transmission.

In the wireless multicast method, the sender has to transmit at the lowest rate sustainable to the receiver at the worst condition environment. Valuable capacity is wasted because of the prolonged channel occupancy. The authors in [34] proposed a dynamic rate-adaptation mechanism where the sender adapts the multicast rate based on the quality of experience (QoE) feedbacks by the receivers. If the receiver experiences a lower QoE, the sender reduces the multicast transmission rate. On the other hand, if the receiver experiences a higher QoE, the sender increases the multicast transmission rate. This way the sender transmits at a rate higher than the lowest instant sustainable rate of the receiver with the worst channel condition.

For realtime traffic in user datagram protocol (UDP)-based streaming, in-time delivery is more crucial than providing a hard-core guarantee of the correct packet arrival. It is more acceptable to allow a few erroneous packets than waiting for re-transmission as long as the target level of user satisfaction is achieved. Selecting multiple senders is a viable alternative to the feedback mechanism and quality assurance solution.

Leveraging P2P concept to operator's multimedia system

The Multimedia Broadcast/Multicast Service (MBMS) in today's cellular networks is capable of broadcasting rich multimedia content to a large number of mobile peers. However, packet loss is inevitable as the optimal rate selection is challenging due to the time-varying nature of wireless links. The individual feedback mechanism for a retransmission request is not scalable and may overwhelm the multimedia server. Mobile users may form a P2P network based on secondary wireless interfaces such as 802.11 and repair packet losses. In PatchPeer [35], the base station, with multicast capability, delivers video on demand (VoD) to a set of mobile users. Any user with missing parts downloads the remaining parts from the neighboring user employing a unicast method. Selecting an appropriate neighboring peer is challenging, downloading from a peer at a multi-hop distance may cause delay. In [36], the base station broadcasts multimedia contents in blocks to mobile users. To repair the packet losses, mobile users are divided into several clusters. The connected dominating set (CDS) algorithm determines the dominator among the cellular users. The dominator is selected as the cluster-head and the remaining users join one of their neighboring dominators. The normal users wait for a random time between '0, 0.005' seconds, and then send packets to their dominators. This scheme assumed that one dominator identifies the lost packets on the whole network and coordinates with the other dominators to trigger the repair process. Each dominator then encodes all of its packets with a random linear network coding technique and then broadcasts that coded packet around its neighborhood. The proposed method thus reduces the burden of base station's downlink channels. Mobile video broadcast services may employ multiple transmissions from the base station to ensure video quality to the scattered

mobile users of variable channel quality. The authors in [37] proposed broadcasting different layers of video using different channels with different coverage range. The base layer is broadcasted throughout the entire cell to maintain the minimum quality for all viewers. More enhanced layers of video are broadcasted with decreasing covering range. To further improve the perceived video quality, mobile users that received enhancement layers forward those layers to users with a weaker channel condition through ad hoc links. Local broadcasters are selected from the average cellular data rate.

2.1.3 Peer selection strategy

The P2P streaming system is dynamic by nature, and exhibits heterogeneity in terms of bandwidth. A receiving peer in such a system relies on multiple sending peers to attain the best possible streaming quality. The receiver maintains a contact list of potential senders from which it chooses an active set for streaming. This process is known as the peer selection mechanism. The receiver first ranks potential senders based on some application level properties such as the contribution level, partnership willingness, upload bandwidth and energy levels. Appropriate peer selection from potential senders plays a vital role especially for wireless networks.

Related peer selection modules

Topology-aware peer selection: In [38], the receiver finds an optimal set of active peers by one of the three following approaches: random, end-to-end, or topology-aware. In a random approach, the receiver hardly achieves a minimal rate requirement from the aggregate rate of all connected peers because of the peers' low availability and

the congestion at the shared path. In an end-to-end approach, a receiver selects peers based on the individual end-to-end expected bandwidth and the availability information; this approach does not consider the shared segments. In a topology-aware approach, the receiver selects peers to maximize the expected aggregated rate at the receiver. The topology aware peer selection technique performs the best as it considers shared paths. This scheme applies to large-scale P2P networks and does not address wireless related issues. In [39], the authors addressed the issue of increased contention on the shared wireless channel when multiple nodes try to access the same file simultaneously. They also proposed a cooperative P2P file transfer protocol which increases the aggregate throughput by selecting potential download peers with a minimum interference on the download paths. However, this protocol selects potential download peers based on the current load and interference on the download paths for a single receiver at a time. The authors in [40] also emphasized the importance of appropriate peer selection from the discovered potential sender peers. In their proposed algorithm, the candidates for the senders send evaluation scores based on several factors such as energy, link quality, movement, lingering time and security. The receiver peer waits a period for evaluation scores from several candidates before ranking them. However, the evaluation scores change repeatedly in the time-varying wireless environment, and the price paid to acquire evaluation scores should be an important factor for designing an efficient streaming system.

Load-aware peer selection: In [41], the authors claimed that locality-aware peer selection might degrade the performance of the mobile P2P network over 3G cellular networks due to load mismatch. When a cellular peer selects multiple peers from the same base station as its partners, the underlying traffic still traverses through the gateway general packet radio service support node (GGSN), costing two-hop

expensive cellular channels. Choosing such a local peer also may cause load variation among the cell base stations. The authors suggested that a peer should select partner from the cell that has the least traffic load and better upload bandwidth. The authors did not mention how the peer selection module collects the base station's traffic load. Peer selection methods, proposed in [38–40], select potential senders that are optimal for a single receiving peer. Peering decisions (only at the receiver end) is always suboptimal, as that particular receiver does not consider other receivers' optimality. An efficient peer selection module should therefore should consider the scenario where multiple requesting peers download content from multiple potential senders. Both stable marriage (SM) algorithm and game theory are strong candidates to model such a problem where multiple peers involve their own selfish interests.

Matching algorithmic peer selection: Our work [42] presented a peer selection module for cellular peers based on the SM problem that chooses the appropriate candidate from the discovered potential senders. In that peer selection problem, a man is analogous to the requesting peer, and a woman is analogous to the potential sender. As a man proposes to a woman, the requesting peer sends a request to potential senders. We provided an alternate heuristic algorithm to solve the problem. The results showed that the intelligent peering decision improved the streaming quality.

Cooperative peer selection: A multi-armed bandit can be thought of a slot machine with multiple levers. When pulled, each lever earns a reward. A gambler's objective is to maximize the sum of the rewards earned from sequential lever pulls. The gambler needs to explore all levers to acquire a reward value associated with each lever; At the same time, the gambler should not lose too much rewards through explorations. The restless bandit algorithm is an extension of the multi-armed bandit problem. The work in [43] proposed a multiple senders selection scheme to improve the data

rate and energy saving in a multi-hop wireless network. The sender selection process is based on the restless-bandit algorithm where the potential senders' states (active or passive) are determined in a distributive and cooperative manner. The lever is equivalent to a sender, and pulling the lever is equivalent to setting the sender's status as active. The formulation is *restless*, as the reward, associated with each sender, changes independent of the sender selection process due to the time-varying nature of the wireless channel. The peer selection has a crucial influence on the overall network throughput and traffic distribution as well. Gürses and Kim [44] proposed a peer selection scheme that maximizes the overall throughput of the wireless ad hoc network. The scheme also allocates an optimal rate to receiving peers by controlling the MAC layer parameters. Such a control of the underlying parameters is not possible for real world P2P applications.

Game theoretic peer selection: Game theory analyzes the attitude of the decision makers ('players') whose rational decisions affect each other. Players interact among themselves, and form coalitions to gain (often-conflicting) rewards through their decisions. In [45], the authors proposed a peer selection process where peers play a strategic game to determine the parent-child relationship. Each parent forms a coalition with its child(ren). Each combination of coalitions offers a specific value. When a peer joins a coalition, it brings additional value to the existing coalition and gets its share of utility. Every player tries to maximize its share of utility by attempting to join a coalition that offers higher utility. A stable coalition is formed when no member of the coalition has an incentive to leave. A child usually accepts more than one parent until it meets its incoming traffic demand. The peer that has higher outgoing bandwidth gets less utility from a coalition. This drives that peer to form coalition with more parents. This way, peers of high outgoing bandwidth are resilient

to peer-failure and the system is less susceptible to peer dynamics.

ISP cooperation on peering process and the associated risks

Recently, a few researchers suggested that Internet service providers (ISPs) and P2P application providers could collaborate for peering decisions [46–50]. In [46], Aggarwal et al. proposed installing an ISP-owned facility - ‘*Oracle*’ that would assist a peer client on the locality-aware peering decision. P2P users provide a list of possible peers to the *Oracle*. The *Oracle* ranks candidate peers for the client based on BGP information. Such guidance on peering decisions reduces cross-domain traffic. *Oracles* from different ISPs may exchange information to make peer ranking more efficient [47]. Fig. 2.2 shows how an ISP-controlled *Oracle* assists peers selecting appropriate

FIGURE 2.2: The ISP-controlled *Oracle* provides assistance on peering decision. For clarity, we only show logical connections between peers and *Oracles*, and among *Oracle* themselves.

partners. Xie and Yang [48] proposed to introduce ‘*iTracker*’ to the ISP that performs a similar function as the *Oracle* [46, 47]. *iTracker* collects information from both P2P providers (application trackers) and ISPs, and provides peering decisions based on distance, load or peer’s bandwidth heterogeneity. In [49], the application

tracker provides peering instructions to clients from end-to-end path utilization information reported by ISPs. The objective of the application tracker is to balance the P2P traffic by minimizing the maximum link utilization through optimum partner selection. The content distribution network (CDN) connects clients with low-latency replica servers (located in ISPs). Peers utilize information collected by a CDN to select partners. For example, peers close to the same server select each other as partners to reduce unwanted traffic flows [50]. The CDN-assisted peer selection eliminates the need for additional entities such as *Oracle*. In [51], Kaya et al. proclaimed the potential risks coming out of cooperation between P2P providers and ISPs. The authors emphasized that the ISP owned *Oracle* should be avoided for inter-ISP peering decisions as ISPs have conflicting preferences for inter-ISP links. Thus another entity is required to aggregate the information from multiple trusted parties. A centralized *Oracle* monitored by a third party is needed to ensure fairness among multiple ISPs. However within each ISP, replicated servers are required to realize *Oracle*. There is also overhead to maintain a semi-static database.

2.1.4 Other issues on realizing cellular P2P streaming system

Copyright and legal issues

P2P services and applications are exposed to legal challenges due to frequent sharing of copyrighted material by the end users. The early P2P system Napster was built over a single-server tracker system that provided content lookup services to peers. Eventually, ‘Napster’ faced accusations of violating copyright, and ceased its operations [52]. Today’s real world P2P streaming systems, such as PPLive [53], SopCast [5], and Coolstreaming/DONet [54] employ ‘trackers’ only to connect peers sharing

common media content. The peers themselves do the streaming chunk discovery, utilizing gossip or DHT-based queries. It is controversial whether to hold such ‘trackers’ to account for copyright infringement.

Opportunistic use of WiFi hotspots

Smart-phones usually have multiple access interfaces and may access the Internet at WiFi hotspots, and thereby off-load base station’s traffic load. In [55], the authors presented experimental results on 3G mobile traffic off-loading through a WiFi hotspot. From statistical analysis, the authors claimed that WiFi hotspot connections could off-load about 65% of the total mobile data and save 55% of battery power. Collins et al. [56] reiterated the WLAN based universal coverage for mobile devices ranging from home to public places such as universities and business campuses, sport complexes and entertainment parks. Despite the rapid growth of WiFi hotspots, their coverage is still quite limited when it is compared to the large footprint of the cellular network. Moreover, many WiFi hotspot providers only allow a limited number applications to their subscribers, and prohibit them from getting rich multimedia content. Connection timeouts and frequent reconnection authentication may also hinder playing live streaming media.

Scalable video coding

Two major scalable video codings that cope with heterogeneity and jitter are media description coding (MDC) and layered coding (LC). Each description in MDC contributes a different amount of enhancement towards the final reconstructed video quality. With MDC, each peer adjusts the number of descriptions for playback according to its bandwidth and device capability. In LC, video is encoded into one base

layer and multiple enhancement layers. The base layer ensures the minimum quality of video as perceived by viewers. Viewers experience improved quality of video as the number of enhanced layers increases. The implementation of MDC or LC gives reality to live streaming on cellular devices.

Decoder compatibility and screen resolution heterogeneity

P2P users from different networks (for example, WiFi, Asymmetric Digital Subscriber Line, WiMAX, and cellular networks) may have decoding and image resolution limitations. The work in [57] proposed a content management system that offers adaptation functionalities to the peers from diverse networks. The proposed architecture introduces two media-aware network elements (MANEs): a seamless home media gateway (sHMG) at the edge of the extended home environment, and a seamless network media gateway (sNMG) at the edge of the cellular access network. Terminal-aware content adaptation and content protection are now distributed to the deployed media gateways. Fig. 2.3 shows the proposed architecture with MANE.

FIGURE 2.3: Media-aware network elements for decoding and image resolution compatibility.

2.2 D2D communication underlaying a cellular network

The D2D feature is a promising add-on component to the LTE-Adv network that may offer spectrum gain from sharing and a timely solution to locally aware services. A few industries are also eyeing on making a profit from the early deployment of proximity aware services such as FlashlinQ by Qualcomm [58], Nokia Instant Community [59], and AirDrop by Apple [60]. With Qualcomm's FlashLinQ technology, a user equipment (UE) can automatically discover thousands of other FlashLinQ enabled devices within one kilometer within a few seconds. Nokia's instant community system offers group text chatting and photo and video sharing by creating ad hoc WiFi connections between nearby devices. For example, audiences at sporting or concert events can share photos and videos in real time. Nokia's system does not require a WiFi hotspot; rather it relies on device's adhoc WiFi capability. However, the system does not scale well to high bandwidth services like video. Apple's "AirDrop" technology discovers nearby Macs and enables wireless P2P connections to synchronize and share files, also employing the WiFi radio interface. D2D communication mode may employ Internet Protocol (IP) header compression technique, and thereby may save bandwidth, and reduce packet error probability and energy consumption [61].

2.2.1 Device discovery and connection setup

The D2D communication requires an automated connectivity to sensors, machines, and other users to realize a wide variety of applications. Device and service discovery are crucial steps on achieving a seamless user experience. The criteria for designing

an device discovery mechanism are as the following:

- Energy-efficient in terms of battery uses;
- Low overhead on connection setup;
- Low maintenance overhead of the network, and;
- Robust with high device density.

FIGURE 2.4: System level overview of LTE architecture.

We note that these criteria are similar to that of the cognitive network. However unlike cognitive users, D2D users do not require to sense free spectrums and have less hardware limitations. Therefore, meeting these criteria is less challenging for the D2D network than that for the cognitive network. Two viable options for D2D session setup and management along with interference coordination to protect

the primary LTE cellular network are potential D2D traffic detection and dedicated radio bearer setup [2]. In 3GPP-LTE IP-based systems, the packet data network (PDN) gateway (P-GW) performs the routing from/to the Internet and also is aware of which eNB serves the UE (See Fig. 2.4). Therefore, the P-GW is able to detect potential D2D traffic from the source and the destination IP addresses of the UEs. Alternatively a new type of dedicated radio bearer may enable D2D communications and keeping control of the session setup and the radio resources. The work in [62] presented another device discovery scheme that uses beacons with cellular network assistance. Devices transmit beacons in parallel employing orthogonal frequency division multiple access (OFDMA) technology. The beaconing scheme complies with the LTE beacon structure. Synchronization is an important issue for a successful connection setup employing beacons. The authors suggested splitting the devices in two groups where the first group listens to the beacons of other groups. The problem with this solution is that the devices multiplexed together in the same OFDMA symbols cannot synchronize to each other. When devices are connected to different cells, synchronization for a successful beacon transmission is not possible. Our suggestion is that the device already in the network will beacon occasionally, and the joining device should scan for a larger period of time to scan the beacon.

2.2.2 Interference management scheme for D2D enable network

If the D2D users are assigned dedicated LTE resource blocks (RBs), there is no interference between cellular users and D2D users. This is only possible when a sufficient amount of resources is available. However, when the D2D communication shares the

same resources with the cellular communication, interference of D2D communication to the cellular network needs to be restricted to maintain a target performance level of the cellular network. The eNB controls the transmit power of D2D connections in order to restrict interference of D2D transmitters to the cellular network. Interference management schemes are inevitable to ensure the proper utilization of the D2D technology and to avoid harmful interference to cellular users. Two major techniques to avoid interference are: cellular-mode UE (CUE) and D2D-mode UE (DUE) pairing scheme and power control scheme.

Cellular and D2D user pairing technique

A CUE and A DUE are a match if the mutual channel gain is minimal and they are eligible for an Intra-cell spatial reuse of radio resources. In [63], the eNB uses information on channel gains between CUEs and D2D terminals, and allocates the shared resources by selecting appropriate CUEs and D2D transmitters such that a target performance level of the cellular network is maintained. In [64], the authors proposed an interference management scheme to share the download spectrum between CUEs and DUEs. The eNB broadcasts the ‘D2D common control channel (CCCH)’ including time and frequency to find the optimum match of the shared resources. All UEs estimate the interference level by listening the D2D CCCH, and report back to the eNB. In [65], CUEs report either their location information (using built-in GPS) or the detected DUEs’ position. The eNB broadcasts this information to the D2D transmitters. D2D transmitters then utilize this information to choose appropriate RBs. However, broadcasting CUEs’ location information to DUEs raises a security concern. Performing radio resource management autonomously by DUEs may also lead to further interference.

The authors in [66] proposed an interference management scheme by sharing uplink resources between CUEs and DUEs. The scheme limits the maximum transmitting power of a D2D transmitter to control the interference from the D2D transmitter to the cellular network. To manage interference from CUEs to DUEs, the scheme defines a region named δ_D -interference limited area (ILA). The D2D transmitter is only allowed to reuse the radio resources used by the CUEs outside the region δ_D -ILA. The reuse of radio resources of CUEs residing in the δ_D -ILA would cause interference and the SINR at the D2D receiver would fall short of the targeted threshold. The authors did not discuss the case of downlink mode. In [67], the authors proposed an uplink resource sharing strategy. The work derived an optimal minimum distance between the D2D receiver and its paired CUE to reduce the outage probability of D2D communications. The outage probability analysis and numerical evaluation showed the accuracy of the strategy. The work assumed that the power required for the D2D communication is lower enough to neglect the interference from the D2D transmitter to CUEs. The work in [68] proposed a shared resource allocation strategy between CUEs and DUEs in the downlink period of the cellular network. The cellular network employs a fractional frequency reuse technique. Any two interior CUEs use the shared spectrum while cell-edge CUEs use the partitioned spectrum. Any interior D2D pair is allowed to share the spectrum used by the different cell-edge CUEs. As the interior pair have a significant distance from the different cell-edge CUEs, the interferences generated between them are negligible. Any cell-edge D2D pair is allowed to share spectrum with the different cell-edge CUEs. Again, the scheme assumes that the power of the D2D transmitter is lower enough to maintain interference at a tolerable level.

Interference management through power control

The power control mechanism improves the performance of the D2D enabled cellular network both in terms of throughput and interference management. One simple way to prevent a D2D transmitter from causing too much interference is to reduce its power. With full channel state information (CSI), the eNB can coordinate the transmit powers of CUEs and DUEs to limit interference between them. The work in [69] aimed at maximizing the sum-rate under both maximum power and cellular priority constraints. The method in [70] utilizes cell statistics for an optimum power control when the instantaneous CSI is not available. The SINR behavior is formulated based on the position of the D2D pair. Then the power control method is applied to the D2D communication such that the SINR degradation of the cellular link does not fall behind a certain level. The results showed that the SINR distribution of the D2D users is comparable to that of the cellular user in most of the cell area. The authors also observed that resource sharing is more beneficial for the system when the D2D pair is farther away from the eNB. When the D2D pair is closer to the eNB, the downlink resource sharing performs better. The power control mechanism presented in [70] cannot reflect dynamic channel variations. In order to minimize the interference between CUEs and DUEs, the proposed scheme in [71] periodically applies a power control mechanism only to the D2D transmitter. To estimate the optimum power of the D2D transmitter, the method computes channel gains between a DUE and a CUE by listening to the sounding signal of the CUE. However, different CUEs have different channel gains with the particular D2D transmitter, and the scheme does not consider optimum matchings between CUEs and D2D transmitters for sharing same resources. Xing and Hakola [72] presented a power control scheme that utilizes

path-loss information. The scheme periodically updates the D2D transmitter's power to maintain a target level of interference level.

A faded channel may cause cellular outage due to channel information uncertainty at the eNB. The authors in [73] proposed that the eNB coordinates the interference between CUEs and DUEs. This work is similar to the work in [69] with the exception of meeting an intentionally increased level of cellular service guarantee. The cost paid to control cellular outage probability decreases the sum-rate of the hybrid LTE system; however the system still achieves significant gain over the single cellular case in most of the cell area. The work in [74] proposed a protocol to share radio resources between CUEs and DUEs during the uplink. In the first step of the protocol, a D2D transmitter sets its own power at a certain level such that it causes minimal interference to the base station. In the second step, the D2D transmitter floods the network to discover the route to the destination DUE. The route discovery packet utilizes transmission and interference information to lower the number of transmissions. The aforementioned work also analyzed the performance of the system in terms of the probability of outage.

Resource sharing mode adaptation

D2D enabled devices may switch between cellular and D2D modes to ensure a targeted rate. The eNB remains in control of selecting an appropriate transmission mode. A handful of research works [75–78] applied optimization techniques to ensure QoS requirements of D2D communication by selecting an appropriate transmission mode. In [75], the authors formulated a joint subcarrier allocation and transmission mode selection problem as an optimization framework that minimized the total transmission power of D2D transmitters in the downlink period of the cellular system.

The framework also ensures the rate of D2D communication. If a D2D transmitter requires more power than the permitted level to meet the rate constraint, the framework switches the transmission mode to the cellular mode. The authors also proposed a heuristic scheme that loosely follows the optimized framework. The simulation results showed that the proposed scheme can save total downlink transmission power. The interference posed by the D2D transmitter is different depending on its location; limiting all D2D transmission powers below a global maximum should not provide an optimal result. The stochastic optimization framework presented in [76] provided an optimum switching mechanism between cellular and D2D mode depending on the system environments and QoS requirements. The mechanism aimed at maximizing the average sum-rate of the system while ensuring the QoS requirements of all users. The work in [79] proposed a joint resource sharing mode selection and transmit power allocation algorithm that meets a target SINR for each link. The method requires the knowledge of channel gain information among CUEs and DUEs and between all CUEs and the eNB. The work in [77] also jointly optimized resource sharing modes and power allocation, however the target therein was to maximize the throughput rather than meeting SINR thresholds. The authors in [78] proposed a joint mode selection and scheduling task problem as an optimization problem that minimized the total transmission power during the uplink period of the cellular system. The framework also includes constraints to achieve a targeted rate of D2D communication. It is extremely difficult or otherwise impossible to solve the optimization problem. The authors also proposed a distributed scheme.

2.2.3 D2D multicast mechanism

The authors in [80] proposed a D2D clustering technique with an option to switch between cellular and D2D modes. The authors suggested that the network remains in control of the mode switching, the cluster head assignment and the cluster membership acceptance. The network may be unaware of a specific application running on cellular peers and may also fail to identify the source of freshly available content. Therefore the clustering process should be self-organizing. The authors also suggested that the cluster head transmits the content at a power determined by the weakest D2D link to meet the required service (data rate). Such a steadfast condition may drive into high power transmission and make interference management impossible. The authors in [81] proposed a reliable multicast technique with network-controlled D2D communications. The authors suggested a NACK based hybrid automatic repeat request (HARQ) level feedback mechanism where a receiver who does not receive content correctly, sends a NACK message. Due to the synchronous operation of HARQ, other receivers also hear retransmission requests. In a cooperative retransmission feature, members may also retransmit. However it is undetermined which members would retransmit. Identifying the exact receiver that sent a NACK message does not seem scalable. In [82], the authors proposed a two-step interference control scheme between CUEs and one D2D multicast group during the cellular uplink state. In the first step, the scheme derives an upper bound of the D2D transmit power to limit interference at the base station. In second step, the scheme assigned a matched resource block (RB) to the D2D transmitter such that a targeted SINR is achieved even under the presence of the interference from DUEs to CUEs.

2.2.4 Other coexisting networks

Spectrum sharing between cellular and adhoc network

There has been considerable research on spectrum sharing between cellular networks and infrastructure-less wireless networks [83–87]. Several authors emphasized the advantages of cellular and wireless local area network (WLAN) interworking systems such as a two-tier overlaying structure [83, 84]. In this structure, disjoint WLANs provide local coverage in hotspot areas, while cellular networks provide a wide area coverage. Mutual interference and vertical handoffs (switching between two access networks) are inevitable problems with this type of coexisting network. Due to heavier download traffic, the uplink spectrum is under-utilized in a frequency division duplex based cellular system with equal bandwidth allocated for uplink and downlink transmissions. The authors in [85] suggested that network capacity can be improved significantly when ad hoc users make use of unoccupied uplink sub-channels. The authors in [86] also proposed a spectrum reuse protocol where D2D users are only allowed to communicate with each other during the uplink frame of the network. This is due to the fact that during the uplink period only the base station is exposed to interference by the D2D users. In this scheme, the D2D user measures its pathloss from the known base station power and the received power during the downlink control frame. The D2D user then calculates its transmit power such that the SINR ratio of the base station does not fall below a threshold level. The authors in [88] advocated to enable direct communication between mobile devices. In their proposed architecture, mobile devices discover partners using cognitive radio and constitute cooperative clusters with neighboring devices. The embedded femtocells in a macrocell may use

either a shared spectrum or a partitioned spectrum. The authors in [89] proposed a flexible hybrid spectrum allocation scheme for the LTE network where inner femto-cells operate in a partitioned spectrum usage and outer femtocells operate in a shared spectrum usage. In [90], the authors studied D2D communication using a WiFi ad hoc mode in the license exempt bands. The authors concluded that mobile devices (in close proximity), when communicating directly instead of infrastructure modes, can reduce contentions/collisions and also avoid bottleneck on an access point. However, cellular operators are more likely to nourish D2D communications on a licensed band as a controlled underlay to a cellular network. With controlled D2D communication, they can design the network with a guaranteed lower level of interference while making an additional profit out of it.

Opportunistic spectrum sharing in cognitive network

The cognitive radio (CR) technology, a very active research area for the past two decades, also reuses the under-utilized spectrum in an opportunistic manner [91–93]. CRs sense the radio frequency environment to detect temporal and spatial “holes” in the spectrum and thereby avoid interference with the primary users. However, the sensing ability of secondary users demands a high level of complexity to detect weak primary signals. Even CR system requires some degree of coordinations among secondary users to ensure fairness of spectrum usages [92].

2.3 Streaming chunk selection strategy

The chunk selection strategy is an important factor that influences the P2P network capacity. The authors in [94] defined chunk’s utility as a function of its rarity

and urgency. They transformed the scheduling problem to a min-cost flow problem and proposed a chunk scheduling algorithm to improve the system-wide utility under stringent buffer and heterogeneous bandwidth constraints. The algorithm is computationally expensive and peers cannot afford to solve the optimization problem due to limited resources. The authors in [95] proposed a chunk scheduling algorithm where chunks are forwarded according to peers' upload bandwidth. Each chunk is tagged with a pseudo random number, and each parent-and-child pair is associated with a subrange of random numbers. The subrange's size is proportional to the upload bandwidth of the corresponding parent. Chunks are forwarded through the link associated with a subrange containing the random number. The peer's buffer occupancy probability is in tune with the chunk propagation process in the P2P network. The work [96] presented a P2P streaming network measurement method based on a peer's buffer occupancy probability. However, the propagation of a chunk in that model is split into two phases. The rates of chunk distribution are different at different phases. Therefore, there could be an abrupt change from one phase to another, and the model lacks reality. In [97], the authors analyzed different chunk distribution strategies to random peers by using stochastic theory. They claimed that the rarest first (RF) algorithm is more scalable to peer population, and most popular (MP) algorithm improves playback continuity by distributing urgent chunks first. The authors presented a simple probabilistic model of the buffer occupancy, and also proposed a mixed strategy for chunk distribution that ensures both scalability and playback continuity. The proposed model is too simple; the probability distribution of the buffer state is the same for all chunks. The model does not consider the network homogeneity in terms of bandwidth constraints, and also does not account for a peer selection strategy. The work in [98] proposed a chunk selection algorithm that

targets improving playback continuity and balancing the load among peers. In the proposed method, a receiving peer first computes the expected delivery time of a rare chunk corresponding to all potential sending peers from the chunk size and estimated bandwidth. The chunk is assigned to a sending peer that delivers it earlier and is not overloaded itself. The basic difference from the RF algorithm is that the scheme does not overload the fast peers holding rare chunks. However, it is not understood how the scheme maintains tradeoffs between deadline constraint and load balancing. Bracciale *et al.* [99] proposed a sender-optimal chunk scheduling algorithm for a large scale P2P system. The sender enumerates the number of neighbors to which it would upload content depending on its chunk transfer time. To avoid its upload bandwidth waste, the sender also prefers peers with higher downlink bandwidth. The work [100] evaluates the performance of a multi-tree based P2P system where each peer delivers chunks in a store-and-forward manner. They derived the theoretical bounds for the number of peers reachable in a given interval of time.

2.4 Some related theories

2.4.1 Stable marriage

The stable matching algorithm can provide practical solutions to many real-world problems such as college admissions, hospital interns' assignments, and allocations of human organs for transplant recipients. Gale and Shapley used marriage as an illustrative example of pairwise matching. The stable marriage (SM) algorithm finds a stable matching between men and women through a sequence of proposals from men to women. Both men and women have preference lists. A matching is called

stable when there is no man/woman pair of which both have stimulus to elope. Gale and Shapley [101] first introduced this algorithm and showed that for two disjoint sets, there always exists at least a single stable matching regardless of individual preferences. Gusfield and Irving [102] later extended the algorithm to many variations. McVitie and Wilson [103] proved that the classical SM algorithm, presented by Gale and Shapley, can only produce a single extreme outcome, that is either a male- or a female-optimal solution. The authors also provided an algorithm to compute the full set of stable matchings. Irving et al. [104] proposed an algorithm that finds the optimal matching under some equitable or egalitarian criteria. Here a solution achieves optimality when the total ‘satisfaction’ of all individuals is maximized.

Many real life problems, such as matchings between students and colleges, interns and hospitals, and workers and firms, give rise to the polygamous version of the classical SM algorithm. Baïou and Balinski [105] followed a graph-theoretic approach to investigate the many-to-many matching - a generalized form of the classical SM algorithm. The stable solution does not require total satisfaction to be maximum, nor it requires to be a fair one. When individuals’ stable partners are ordered in terms of satisfaction, the assignment of median choices provide a compromise mechanism. Teo and Sethuraman [106] and later Sethuraman et al. [107] showed that the perfect compromise solution may achieve fairness for a classical SM and many-to-one matching problems respectively. Notably, when each individual pairs with an individual of his or her median choice, the assignment is a matching and is stable as well. Therefore the outcome of a stable matching problem solution is an equitable or socially optimal stable match. Vande Vate [108] formulated the stable matching problem as a system of linear inequalities to find the optimal solution. The author characterized stable matchings as the extreme point of a certain polytope. Rothblum [109] presented an

extension to the matching problem to many-to-one matchings that is either a male or a female is allowed to be married to more than one individual of the opposite sex.

2.4.2 Spatial point process

The spatial point process studies patterns of random points in a topological space [110]. This branch of stochastic geometry is useful to analyze patterns of points with statistical measures of important properties. Some applications of point process are the location study of animals in a forest, bird nests, and crime cases. The strength of the received signal power and interference is distant dependent. The performance of the network relies intensely on the spatial configuration of users and the network access point. The spatial point process is also an important tool to measure network performances in terms of interference distributions and link outages. The work in [111] analyzed the outage probability of ALOHA and carrier sense multiple access (CSMA) protocols at the media access control (MAC) layer using point process. Spatial distributions critically affect the performance of many emerging network technologies, such as Femtocells, hotspots, relays, and white space harvesters as they are often overlaid with traditional cellular networks [112]. In [113], Srinivasa and Haenggi presented nodal distance distributions in uniformly random networks. The authors also applied the knowledge of internodal distances to characterize wireless networks in terms of energy consumption, interference, outage, and connectivity. A more detailed work on internodal distance distributions is provided in [114]. Sometimes it is necessary to obtain a new point process by applying thinning, clustering or super-position operation on the basic point process to solve many pragmatic problems. For example, the work in [115] applies Matérn hard-core process, a dependent thinning operation, to

determine the set of concurrent transmitters.

2.4.3 Epidemiology

Throughout history infectious diseases cost lives and prosperity. The mathematical model of epidemics had been studied to analyze the spread and control of infectious diseases over the past centuries. Hamer [116] first introduced ‘mass action’ principle to develop a deterministic epidemic model. Ross [117, 118], Kermack and McKendrick [119], and Soper [120] also investigated the deterministic epidemiological model in the early 20th century. Reed and Frost introduced the probabilistic form of epidemic model delivered first as class lectures at the Johns Hopkins University during 1928. Although they never published their work, other authors had cited their work [121–123]. The lectures are now available in article form in [124]. Abbey [121] and Maia [122] provided the basic equation of the Reed-Frost model with a few ramifications and adaptations. Fine [123] gave a comprehensive description of the Reed-Frost model. Sartwell [125] summarized the sequel of work on the Reed-Frost model. Jacquez [126] claimed that the Reed-Frost model lacks generality in generating a function model and special assumptions made in this model are unreasonable. Longini Jr. [127] compared the stochastic and deterministic models using a numerical example, and showed that the two models are equivalent for a large population. The stochastic model is more appropriate for a small population. Liu *et al.* [128, 129] investigated the dynamics of epidemiological models with the nonlinear incident rate of the form $\beta I^p S^q$. Here β is the rate of contact, and I, S denote the number of infectious and susceptible individuals respectively. The authors claimed that the epidemiological model exhibits a qualitatively different dynamic behavior with $p > 1$, whereas it

does not exhibit altered dynamic behavior with $q > 1$. With $p > 1$, there may exist multiple equilibria or periodic solutions emanated from Hopf bifurcation depending on some parameter values. Wilson and Worcester [130, 131] studied the epidemiological model with general infection rate of the form S^q .

2.5 Summary

We have conducted extensive reviews related to our thesis work.

- In this chapter we first provided reviews on architecture, protocols and issues towards P2P streaming in the context of wireless networks.
- As the streaming system relies on the overlay network, designing an efficient overlay is the key to combat against the bandwidth wastage. A few important aspects of an efficient overlay network are scalability, proximity aware, and low maintenance overhead. We found that the topology-aware overlay system is the most efficient one. In designing our collaborative streaming system, we try to form groups among geographically closed peers.
- There are many variants of wireless P2P streaming systems. Topology construction processes depend on the users' access technology. The D2D enabled network is a new concept. There is potential to build a wireless P2P streaming system on top of that and we aim to design a streaming system that utilizes D2D multicast.
- Multicast techniques could be exploited in P2P streaming systems by utilizing broadcast nature of the shared wireless channel. However, a feedback mechanism needs to be addressed, especially in a dense network. Also without a proper

feedback method, it is also difficult to estimate the optimum transmission rate.

- Peering decisions play an important role, especially for wireless networks. Peer selection should leverage the underlying network conditions. However, the price paid to infer topology should not exceed the reward achieved from it. The peer selection technique should sustain to peer churn rate, with peer joining/leaving the system all the time. The receiver should experience improved streaming quality by choosing potential senders with better load distribution. P2P streaming is always user driven, ISP cooperation is only possible for a CDN-based streaming system.
- There are manifold benefits of enabling D2D communication in a cellular network under the control of the LTE system. With control over D2D connectivity, the operators can incorporate the D2D service exploiting existing cellular architecture, offer new services with new revenue opportunities. The eNB can employ power control mechanism for D2D connections and guarantee limited interference to the cellular network. However, an efficient resource allocation is required to decide whether to allocate dedicated resources for D2D services when ample amount of resources are available.
- The chunk distribution strategy is vital as it influences P2P network capacity. The existing RF algorithm may overload fast peers holding rare chunks; whereas the MP algorithm cannot provide chunk diversity. An efficient chunk selection algorithm is required for an effective streaming system.
- The SM algorithm offers many real-life problems. Usually peers have conflicting interests, therefore the SM algorithm is suitable for designing a peer selection

module. The spatial point process is useful to measure statistical performance in the context of wireless network. We try to use the epidemic model to describe the chunk distribution process.

Chapter 3

Resource Allocation for Device-to-Device Communication

The basic building block of the collaborative streaming system is the D2D enabled component. The D2D communication, as an underlying cellular network, empowers user-driven rich multimedia applications and has also proven to be network efficient, off-loading eNB traffic. However, D2D transmitters may cause a significant amount of interference to the primary cellular network when radio resources are shared between them. During the downlink phase, the primary CUE may suffer from interference by the D2D transmitter. On the other hand, the eNB is the victim of interference by the D2D transmitter during the uplink phase when radio resources are allocated randomly. Such interference can be avoided or otherwise diminished, if radio resources are allocated intelligently with coordination from the eNB. In this chapter, we formulate the problem of radio resource allocation to D2D communications as an MINLP. We propose a penalty-based GA to solve the MINLP. It is not possible to implement the penalty-based GA within the fast scheduling period of the LTE network. In this chapter, we therefore propose an alternative greedy heuristic algorithm that can lessen interference to the primary cellular network by utilizing channel gain information. We also perform extensive simulations to prove the efficacy of the proposed algorithm.

FIGURE 3.1: The LTE downlink physical resource.

3.1 Channel modeling

3.1.1 Radio resource and access technology

In the LTE system, bandwidth is divided into equal size physical RBs. Each RB physically occupies one slot in the time domain (0.5 millisecond) and 180 kilohertz (kHz) in the frequency domain with a subcarrier spacing of 15 kHz. Fig. 3.1 illustrates the LTE downlink physical resource. The LTE has adopted OFDM based radio interface due to its higher spectral efficiency and resilience against a multi-path delay spread. There is, however, one important difference between the feasible assignments on the uplink and downlink shared channels. In LTE, the multiple access scheme for downlink (from eNB to UEs) is OFDMA. The radio access technology in uplink is a

single carrier frequency division multiple access (SC-FDMA) due to its characteristics of low peak-to-average power ratio (PAPR) that enables higher transmit power efficiency for the UEs. The physical properties of SC-FDMA require that RBs allocated to a single user be contiguous in frequency. The minimum scheduling period in the time domain is one millisecond; therefore the smallest unit of resource that can be assigned is two RBs.

3.1.2 Channel model

In order to measure two important parameters, SINR and channel gain, we consider both distant dependent macroscopic path-loss and shadow fading path-loss. For urban and suburban areas, the macroscopic path-loss in decibel (dB) between an eNB and a UE at a distance of r meter (m) is:

$$L_{dB}(r) = 40(1 - 4 \times 10^{-3}h_b) \log_{10}(r/1000) - 18 \log_{10}(h_b) + 21 \log_{10}(f_c) + 80 \quad (3.1)$$

Here, f_c is the carrier frequency in megahertz (MHz), and h_b is the base station antenna height (in m). For different scenarios and details, the readers are encouraged to read [132].

The slow fading or shadowing on a wireless channel is caused by obstacles (such as buildings, hills, trees) in the propagation path of the radio waves and obviously is location-dependent [133]. Typically, the shadow fading (X_s) is measured by a Gaussian (normal) distribution in the logarithmic scale with mean μ_s and standard deviation (*std*) σ_s [134]. The probability distribution function (pdf) of shadowing in

logarithmic scale is given by:

$$p(X_{s,dB}) = \frac{1}{\sigma_{s,dB}\sqrt{2\pi}} e^{-\frac{(X_{s,dB}-\mu_{s,dB})^2}{2\sigma_{s,dB}^2}} \quad (3.2)$$

The distribution in Eq. (3.2) follows a log-normal distribution in the linear scale. Therefore shadow fading is also known as log-normal fading. Typical values of the mean and the *std* of shadow fading variables are 0 dB and 8 dB for an urban environment [135]. Without doubt, the terrain and the morphology structures do not change abruptly; the slow fading values should possess a spatial correlation. When shadowing is modeled only by log-normally distributed random variables, the variables do not meet the spatial correlation properties. Finding an accurate fading model is a myth. However, we generate realistic looking patterns as described in [135]. Uncorrelated log-normal fading fields are generated first and then convoluted two-dimensionally along with the following exponential correlation function.

$$\tilde{X}_{s,dB}(x, y) = \sum_{i=-a}^a \sum_{j=-b}^b X_{s,dB}(x+i, y+j) C_s(\Delta d) \quad (3.3)$$

Here $C_s(\Delta d)$ is an exponential correlation function as given by

$$C_s(\Delta d) = e^{-\frac{\Delta d}{d_{corr}} \ln 2} \quad (3.4)$$

Δd is the Euclidean distance between the points (x, y) and $(x+i, y+j)$. a, b define the dimension of the mask whose area is very small than the cell area. The decorrelation distance d_{corr} depends on the environment. For the urban environment, $d_{corr} = 20 \text{ m}$ works satisfactory for an distance up to 500 *m* [135]. We assume that the impact of

fast fading is averaged out over a certain time period.

Each eNB consists of three sectoral antenna with the following radiation pattern as described in [136]:

$$A_{dB}(\theta) = -\min \left[(\theta/\theta_{3dB})^2, A_m \right]; \quad -180^\circ \leq \theta \leq 180^\circ$$

θ_{3dB} is 3 dB beam width, and A_m is the maximum attenuation. The total path-loss, which includes the antenna gain, between eNB B and the user u , is:

$$PL_{dB,B,u}(\cdot) = L_{dB}(r) + \log_{10}(\tilde{X}_{s,u}) - A_{dB}(\theta) \quad (3.5)$$

where $X_{s,u}$ is the modified log-normal shadow fading path-loss of user u . The linear gain between the eNB and a user u is

$$G_{Bu} = 10^{-PL_{dB,B,u}/10} \quad (3.6)$$

For D2D communication, the gain between two UEs u and v is $G_{uv} = K_{uv}d_{uv}^{-\xi}$ [137]. Here, d_{uv} is the distance between transmitter u and receiver v , ξ is a constant path-loss exponent and K_{uv} is a normalization constant. The normalization constant depends on the radio propagation properties of the environment, and also accounts for the effects of coding gain, spreading gain, beamforming, etc.

3.2 Interference problem due shared spectrum

The D2D commutation takes place underlaying the primary cellular network. We acknowledge that the radio resources are valuable and DUEs rather share resources

with CUEs than use dedicated resources. Hence, the interference of D2D transmitters to the cellular network needs to be restricted to maintain a target performance level of the cellular network. If the distance between D2D is too long, the D2D will require high power. If that is the case, the primary cellular network would suffer from interference. Fig. 3.2 illustrates the interference problem when the primary cellular

FIGURE 3.2: Interference problem when the primary cellular network shares channel with the D2D network.

network shares a channel with the D2D network. During the download period of the cellular network (see Fig. 3.2(a)), cellular UE_0 is exposed to interference when either of the D2D transmitter (UE_1 , or UE_3) transmits using the same allocated subband. Also, the D2D receiver (UE_2 , or UE_4) will suffer from interference caused by the cellular UE_0 , whichever shares the same subband with the cellular UE_0 . One can easily speculate that the amount of interference will depend not only on D2D transmit power, but also on spatial distance between the D2D transmitter and the CUE. On the other hand, during the uplink period of the cellular network (see Fig. 3.2(b)), the

D2D transmitter (UE₁ or UE₄) causes interference only to the immobile eNB. The D2D receiver (UE₂ or UE₃) also suffers from interference from the UE₀. Anyone can whisper that the amount of interference depends on the transmitter power and the spatial distance as well. Hence, intelligent selection of the shared RB would result in better performance in terms of network throughput. According to the suggestion [2], we deny D2D connections if the maximum distance between DUEs is higher than $d_M = 25$ m. Since the eNB coordinates the D2D session setup, we assume that both UEs need to be in the same cell for a possible D2D connection.

3.3 Problem formulation to mitigate interference

The eNB schedules the primary cellular network in the standard way, and our solution does not require any modification for D2D resource allocation. However, we want to allocate the same assigned RBs of any CUE c to one of the D2D connections for which total network throughput is increased. We formulate the problem of assigning appropriate RBs for underlying D2D communication as an optimization problem that achieves higher throughput without impairing the existing cellular network. Consider a time division duplex (TDD) cellular network with identical splitting of uplink and downlink resources. The total number of available RBs during the downlink and uplink is equal. This is due to the fact that because of TDD, there is only a single-carrier frequency and the uplink and downlink transmissions are separated in time, and also on cell site basis; however this does not affect our proposed scheme. We assume an infinitely backlogged model where data is always available from each UE. We also assume that a more advanced inter-cell interference mitigation scheme works on top of our scheme. The eNB serves a set $\mathcal{C} = \{1, \dots, C\}$ of CUEs and also coordinates a

set $\mathcal{D} = \{1, \dots, D\}$ of D2D pairs. According to our assumptions, $C \gg D$.

3.3.1 Downlink phase interference coordination

During the download phase of the cellular network, any CUE is exposed to interference when any DUE is allowed to transmit using the same allocated subband. The amount of interference will depend not only on D2D transmit power but also the channel gain between the D2D transmitter and the CUE. G_{cd} denotes the channel gain between UEs c and d . G_{Bc} denotes the channel gain between the eNB and UE c . P_B and P_d are the transmission powers of the base station and the device d respectively. When D2D devices utilize downlink RBs, any D2D transmitter d causes interference to CUE c while using RBs assigned to it. The downlink SINR of user c is

$$\gamma_c^{DL} = \frac{P_B G_{Bc}}{N_0 + I + \sum_d x_c^d P_d G_{cd}} \quad (3.7)$$

The SINR of the D2D pair d is

$$\gamma_d^{DL} = \frac{\sum_c x_c^d P_d G_{dd}}{N_0 + I + \sum_c x_c^d P_B G_{dB}} \quad (3.8)$$

The variable x_c^d satisfies $x_c^d = 1$, if D2D pair d uses RB(s) assigned to CUE c . N_0 accounts for the receiver's noise figure and thermal noise density; I accounts for intercell interference. Let us define rates r_c^{DL} and r_d^{DL} corresponding to the SINR γ_c^{DL} and γ_d^{DL} as determined by the Shannon capacity model. We want to maximize the sum rate of the primary CUEs and secondary DUEs. The problem is formulated

as an MINLP problem as below:

$$\text{Maximize} \quad \sum_c^C m_c r_c^{DL} + \sum_d^D \sum_c^C x_c^d m_c r_d^{DL} \quad (3.9)$$

Subject to:

$$P_B G_{Bc} \geq \gamma_{c,tgt}^{DL} \left(N_0 + I + \sum_d x_c^d P_d G_{cd} \right), \quad \forall c \in \mathcal{C} \quad (3.10)$$

$$\sum_c x_c^d P_d G_{dd} \geq \gamma_{d,tgt}^{DL} \left(N_0 + I + \sum_c x_c^d P_B G_{dB} \right), \quad \forall d \in \mathcal{D} \quad (3.11)$$

$$\sum_c x_c^d \leq 1, \quad \forall d \in \mathcal{D} \quad (3.12)$$

$$\sum_d x_c^d \leq 1, \quad \forall c \in \mathcal{C} \quad (3.13)$$

Here, m_c denotes the number of RBs allocated to the CUE c at each time slot during the downlink period. x_c^d and m_c are binary and integer decision variables, respectively. P_d is a fractional decision variable. The value of r_d^{DL} depends on the variable P_d . Constraints in (3.10) and (3.11) guarantee the target SINR of the CUE and the DUE, respectively. Constraints in (3.12) ensure that each device shares one user's RBs at most, whereas constraints in (3.13) ensure that, at most, one D2D pair shares any of the user's RBs.

3.3.2 Uplink phase interference coordination

During the uplink phase of the cellular network, the D2D transmitter causes interference only to the immobile eNB. A D2D receiver is also exposed to interference from the UE whose subband is being shared. The SINR of the D2D receiver d during the

uplink phase is γ_d^{UL} and is calculated as:

$$\gamma_d^{UL} = \frac{\sum_c y_c^d P_d G_{dd}}{N_0 + I + \sum_c y_c^d P_c G_{cd}} \quad (3.14)$$

$$\gamma_{eNB}^c = \frac{P_c G_{cB}}{N_0 + I + \sum_d y_c^d P_d G_{dB}} \quad (3.15)$$

The variable y_c^d satisfies $y_c^d = 1$, if a D2D pair d uses RB(s) assigned to the CUE c . Let us define rates r_d^{UL} and r_{eNB}^c corresponding to the SINR γ_d^{UL} and γ_{eNB}^c as determined by the Shannon capacity model. We want to maximize the sum rate of the primary CUEs and secondary DUEs. The problem is formulated as a MINLP problem as below:

$$\text{Maximize} \quad \sum_c n_c r_{eNB}^c + \sum_d \sum_c y_c^d n_c r_d^{UL} \quad (3.16)$$

Subject to:

$$P_c G_{Bc} \geq \gamma_{eNB}^{c,tgt} \left(N_0 + I + \sum_d y_c^d P_d G_{dB} \right); \forall c \in \mathcal{C} \quad (3.17)$$

$$\sum_c y_c^d P_d G_{dd} \geq \gamma_{d,tgt}^{UL} \left(N_0 + I + \sum_c y_c^d P_c G_{cd} \right); \forall d \in \mathcal{D} \quad (3.18)$$

$$\sum_c y_c^d \leq 1; \forall d \in \mathcal{D} \quad (3.19)$$

$$\sum_d y_c^d \leq 1; \forall c \in \mathcal{C} \quad (3.20)$$

Here, n_c denotes the number of RBs allocated to the CUE c at each time slot during the uplink period. y_c^d and n_c are binary and integer decision variables, respectively.

P_d is a fractional decision variable. The value of r_d^{UL} depends on the variable P_d . Constraints in (3.17) and (3.18) guarantee the target SINR of the CUE and the DUE, respectively. Constraints in (3.19) ensure that each device shares one user's RBs at most, whereas constraints in (3.20) ensure that, at most, one D2D pair shares any of the user's RBs.

3.4 Penalty based genetic algorithm

The optimal solution to an MINLP problem is in general hard to achieve, both in theory and in practice, due to the intractability caused by mixed-integer and nonlinear characteristics of decision variables. To achieve near-optimal result of the MINLP problem, we propose a GA-based solution. The GA is a class of stochastic optimization search methods that mimics the process of Darwinian natural selection. The GA always maintains a large candidate solutions rather than generates a single potential solution.

Since the GA is suitable for unconstrained optimization problems, constraint handling is a critical concern when the GA applies to constrained optimization problems. Penalty and barrier methods are among the most common techniques that transform the problem from a constrained optimization to an unconstrained problem for the GA implementation. While the former belongs to an exterior-point method, the latter belongs to an interior-point method. Penalty method adds a penalty function to the objective function for any violation of the constraints and forces the solution to the feasible region and a subsequent optimum. The barrier method also adds a penalty-like term to the objective function that acts as a barrier. This ensures that

Algorithm 1: The penalty based GA for solving the MINLP during downlink period

Input : Network parameters:

$\mathcal{C}, \mathcal{D}, P_B, G_{Bc}, G_{dB}, G_{cd}, G_{dd}, \gamma_{c,tgt}^{DL}, \gamma_{d,tgt}^{DL}, N_0, I$

GA parameters: $pop_size, p_C, p_M, N_G, R_{l,i}$

Output : m_c, x_c^d, P_d

begin

Set up and initialize network and GA parameters;

Generate initial population $\mathcal{P}^0$;

for $i \leftarrow 1$ **to** N_G **do**

for $j \leftarrow 1$ **to** pop_size **do**

if $\text{mod}(j, 2) == 0$ **then** *Crossover* with p_C ;

Mutate with p_M ;

end

foreach *individual of the current generation* **do**

 Evaluate fitness with penalty using Eq. (3.21);

end

 Perform *selection* using roulette wheel sampling scheme;

end

return m_c, x_c^d, P_d

end

the search always remains to the interior-point of the feasible region. The interior-point barrier method requires a feasible starting point. Also it is computationally inefficient to check that the search never leaves the feasible domain. Therefore, we prefer an exterior-point penalty method. Unlike [138], our approach does not require to eliminate integer variables. The penalty based GA to solving the MINLP for the downlink period is outlined in Algorithm 1 and illustrated as follows:

Representation and initialization: The algorithm starts with an initial population $\mathcal{P}^0$ of size pop_size . Each individual in the population is a string entity of an artificial chromosome, representing a solution to the problem at hand. The candidate solutions for P_d are derived from the chromosomes of length D . Each of these chromosomes is a binary string and directly coded as a real value within its corresponding bound. The candidate solutions for x_c^d are derived from the corresponding elements of a chromosome comprising two-dimensional binary string of length $C \times D$. And the candidate solutions for m_c are derived from C chromosomes. Each of these chromosomes is coded as an integer value. To generate $\mathcal{P}^0$, we assign random numbers to the binary string. The chromosomes then evolve through successive iterations called generations.

Crossover and Mutation: During each generation, new chromosomes (offspring) are produced using two genetic operations, crossover and mutation. Crossover operator generates offspring by recombining two chromosomes from the current generation and thereby provides diversity to the population. The crossover point is chosen with a crossover probability p_C . We perform the mutation operation with a probability p_M that prevents premature loss of hereditary information by randomly modifying alleles within a chromosome.

Selection: The performance of the penalty method depends on the magnitude of the penalty factor. If the penalty factors are too large, the GA based method

converges to premature solution inside the feasible region. On the other hand, if the penalty factors are too small, this method may divert towards the infeasible region due to the occurrence of excessive violation. Therefore, it is important to choose appropriate penalty factors to maintain the diversity and check the violations at the same time. We propose a GA based penalty function method for solving nonlinear constrained optimization problems. This approach utilizes user defined multi-stage penalty factors for each constraint in the objective function to account for violation. Multi-stage penalty factors depend on the magnitude of the violation of the constraints. The fitness function is evaluated as the following:

$$Eval(\cdot) = \sum_c^C m_c r_c^{DL} + \sum_d^D \sum_c^C x_c^d m_c r_d^{DL} - \sum_i (R_{l,i} \times \max[0, g_i(\cdot)]) \quad (3.21)$$

Where, $R_{l,i}$ is the penalty factor of the l -th violation level corresponding to i -th constraint. Constraints are rearranged of the form $g_i(\cdot) \leq 0$. Constraints are normalized to avoid any sort of bias toward any of them. We use roulette-wheel selection, that selects individuals with a probability proportional to their fitness values and thus keeps the population size constant. After N_G iterations (generations), the algorithm converges to the best (fittest) chromosomes that represent a near-optimal or suboptimal solution to the problem. We follow a similar procedure to find a near-optimal solution to the MINLP for uplink period.

3.5 Greedy heuristic RB selection algorithm

Although the GA-based heuristic algorithm in Section 3.4 provides a suboptimal solution to the MINLP, the technique still requires a considerable computational effort.

The LTE scheduling period is 1 millisecond and the GA based solution remains far from being suitable for real applications. Here, we present an alternative greedy heuristic RB selection algorithm to solve the MINLP optimization problem that requires much less computation time while maintaining the overall performance close to the optimal. During the downlink phase, higher values of γ_c^{DL} and γ_d^{DL} would aid in increased cell and D2D throughput, respectively. Through careful observation of Eq. (3.7), it is obvious that lower channel gain (G_{cd}) between CUE c and D2D transmitter d will result in higher γ_c^{DL} ; i.e. it will cause less interference to CUE c . Therefore, any UE with a higher channel quality indicator (CQI) can share RBs assigned to them with the D2D transmitter with lower channel gain between them. Here, we assume that the eNB receives channel gain (between CUE and DUE) information through control as it remains in control of D2D connections.

For the uplink phase, by observing Eq. (3.14) and Eq. (3.15), we intuitively decide that a lower G_{cd} or increased spatial distance between CUE and the D2D receiver d would result in higher D2D throughput. However, of paramount importance is to limit the interference at the eNB caused by the D2D connections. Both algorithms are summarized in Algorithms 2 and 3.

3.6 Simulation methodology

We perform extensive simulations to evaluate the efficacy of the proposed RB sharing method. We developed a simulator using *C++* programming language that realistically modeled the LTE system. Table 3.1 summarizes a list of simulation parameters and their default values. We generated distant dependent macroscopic path-loss including antenna gain with a 7-cell, 3 sector hexagonal layout. Fig. 3.3 shows the

Algorithm 2: Downlink D2D RB allocation scheme

 $\mathcal{C}$: Sorted list of CQIs for all downlink UEs in decreasing order G_{cd} : Channel gain matrix $\mathcal{D}$: List of D2D connections yet to be assigned RBs m_c : Number of RBs assigned to CUE c **begin** $c \leftarrow 1$;**while** $\mathcal{D} \neq \emptyset$ **or** $c == C$ **do** Pick RBs with c^{th} largest value; Find the D2D transmitter d for which channel gain is minimum;

$$\gamma_c^{DL} \leftarrow \frac{P_B G_{Bc}}{N_0 + I + P_d G_{cd}};$$

$$\gamma_d^{DL} \leftarrow \frac{P_d G_{dd}}{N_0 + I + P_B G_{dB}};$$

if $\gamma_c^{DL} \geq \gamma_{c,tgt}^{DL}$ **and** $\gamma_d^{DL} \geq \gamma_{d,tgt}^{DL}$ **then** Share all RBs of the CUE c with D2D connection d ; $\mathcal{D} = \mathcal{D} - \{d\}$; $c \leftarrow c + 1$ **else** Do not assign RBs to D2D connection d **end****end****end**

Algorithm 3: Uplink D2D RB allocation scheme

 $\mathcal{C}$: Sorted list of CQIs for all uplink UEs in decreasing order G_{cd} : Channel gain matrix $\mathcal{D}$: List of D2D connections yet to be assigned RBs m_c : Number of RBs assigned to CUE c **begin** $c \leftarrow 1$;**while** $\mathcal{D} \neq \emptyset$ **or** $c == C$ **do** Pick RBs with c^{th} largest value; Find the D2D transmitter d for which channel gain is minimum;

$$\gamma_d^{UL} \leftarrow \frac{P_d G_{dd}}{N_0 + I + P_c G_{cd}};$$

$$\gamma_{eNB}^c \leftarrow \frac{P_c G_{Bc}}{N_0 + I + P_d G_{dB}};$$

if $\gamma_d^{UL} \geq \gamma_{d,tgt}^{DL}$ **and** $\gamma_{eNB}^c \geq \gamma_{eNB}^{c,tgt}$ **then** Share all RBs of the UE c with D2D connection d ; $\mathcal{D} = \mathcal{D} - \{d\}$; $c \leftarrow c + 1$ **else** Do not assign RBs to D2D connection d **end****end****end**

FIGURE 3.3: Distance dependent macroscale path-loss map (in dB, including antenna gain) with 7 eNBs (in white circle).

FIGURE 3.4: Uncorrelated and space-correlated lognormal shadow fading path-loss map (in dB).

FIGURE 3.5: Simulation region (of size $1500 \times 1500 \text{ m}^2$) showing path-loss (dB) that includes antenna gain, distant dependent macroscale path-loss, and lognormal shadow fading path-loss.

aforementioned path-loss map. Fig. 3.4 shows the shadow fading map with random (uncorrelated) shadow fading variables (left) and with convolutional correlated shadow fading variables (right). Fig. 3.5 shows the path-loss map (dB) that includes antenna gain, distant dependent macroscale path-loss, and lognormal shadow fading loss. Users are distributed uniformly inside the scenario. The SINR is calculated from the received signal power and interference power level. The CQI is calculated from the SINRs at the UEs and then fed back to the eNB. Link adaptation is performed based on the CQI at the eNB. Various modulation and coding schemes (MCS, QPSK, 16QAM, and 64QAM) are considered with varying coding rates ranging from 1/8 to 4/5 (see [136, 139] for details). The rate corresponds to a certain block error rate (BLER, %10) are also estimated from the ACK/NACK of past transmissions. The

number of RBs is 100 for 20MHz bandwidth. Uplink and downlink bandwidth is divided equally into m RBs. The uplink/downlink bandwidth ratio does not effect the proposed scheme. Three different scheduling algorithms are employed for the primary cellular network: Round Robin (RR), Maximum Carrier to Interference ratio (Max C/I) and Proportional Fair (PF) with CQI feedback. In the time domain context, at each time slot, the PF scheduler opportunistically serves a user that would achieve the highest instant data rate scaled by the long term data rate in the past window [140]. That is at each time slot, the PF scheduler selects user

$$j = \arg \max_i \frac{r_i(t)}{R_i(t)} \quad (3.22)$$

where $r_i(t)$ is the data rate reported through uplink CQI, and $R_i(t)$ is the long term data rate of the user i at time t . The value of $R_i(t)$ is updated by an exponential filter as the following:

$$R_i(t+1) = \begin{cases} (1-\psi)R_i(t) + \psi r_i(t) & \text{if } i = j; \\ (1-\psi)R_i(t) & \text{if } i \neq j. \end{cases} \quad (3.23)$$

where ψ is a time constant. The PF scheduling achieves a trade-off between throughput and fairness among all users by giving priority to the user with a higher achievable data rate and a lower long term data rate. It has been shown that maximizing the sum of logarithmic long term data rate ($\max \sum_i \ln R_i$) also achieves the PF criteria as defined in Eq. (3.22) [141, 142]. In [143], the authors proposed a frequency domain

resource allocation algorithm with the following PF metric:

$$\max \sum_c \sum_i \frac{x_i^c(t) r_i^c(t)}{R_i(t)} \quad (3.24)$$

where $x_i^c(t)$ is a binary variable to indicate whether or not RB c is allocated to user i at time slot t . The throughput is calculated by mapping the SINR to the ideal link-adaptation based on the LTE link-level capacity as follows:

$$\eta = \begin{cases} 0, & \text{if } \gamma < \gamma_{min} \\ W_{eff} \times \log_2(1 + \gamma), & \text{if } \gamma_{min} \leq \gamma < \gamma_{max} \\ 4.7, & \text{if } \gamma \geq \gamma_{max}; \end{cases}$$

where η is the estimated spectral efficiency in bps/Hz, γ is the SINR and W_{eff} (0.57) is the attenuation factor applied to the Shannon bound. W_{eff} accounts for link level implementation overhead and also for overhead due to pilot and control channel signals. We apply hard spectral efficiency, which gives $\eta_{max} = 4.7$ bps/Hz at $\gamma_{max} = 25$ dB or higher. Also, $\eta = 0$ at $\gamma_{min} = -10$ dB or lower.

3.7 Results

In this section, we analyze and discuss the simulation results.

3.7.1 Performance of heuristic D2D RB assignment algorithm for different scheduling algorithms

The D2D commutation takes place underlying the primary cellular network. We limit the number of possible D2D connections as a percentage of total UEs in the

FIGURE 3.6: Normalized network throughput for different scheduling algorithms with the proposed greedy heuristic RB assignment algorithm [scenario 1: the number of D2D connection is 10% (of total UEs)].

network. We perform the simulations for a period of 500 seconds on two scenarios where the number of D2D connections are restricted to 10% and 20% of the total UEs in the network. As the patterns remain the same, we present the results for the first 100 seconds to show the initial variations. To show how the proposed D2D RB assignment algorithm performs, we plot the sum rate of the cell and D2D throughput. We also plot the cell throughput to investigate whether the primary cellular network does not degrade due to interference from the D2D connections. Fig. 3.6 and Fig. 3.7 show the normalized throughput variation with time for RR, Max C/I, PF with CQI scheduling algorithms in the case of scenario 1 and 2, respectively. For both scenarios, RR performs poorer in terms of cell and sum of cell and D2D throughput than that of the other two scheduling algorithms. This is due to the fact that RR

FIGURE 3.7: Normalized network throughput for different scheduling algorithms with the proposed greedy heuristic RB assignment algorithm [scenario 2: the number of D2D connection is 20% (of total UEs)].

does not consider current channel conditions and allocates equal resources to all users. Max C/I performs the best in terms of cell and sum of cell and D2D throughput by exploiting channel conditions. Moreover, for all algorithms, the network throughput is higher when D2D communication is integrated underlaying the cellular network.

We use of free *GAlib* (version 2.4.7) tool for our GA based solution [144]. We also compare the performance of the GA and the proposed greedy RB assignment algorithm for the PF scheduling approach as shown in Fig. 3.8. The GA algorithm performs better both in terms of cell and the sum of cell and D2D throughput than that of the greedy heuristic algorithm. However, the GA algorithm requires a significant amount of time making it an impractical solution.

FIGURE 3.8: Comparisons between the GA based solution and the greedy heuristic RB assignment algorithm in terms of normalized aggregate throughput for the PF scheduling approach.

3.7.2 Comparisons between random and heuristic D2D RB assignment algorithms

We also compare different RB assignment algorithms (GA based solution, greedy heuristic and random algorithms). We plot the normalized throughput with a varying number of active UEs in the cell (see Fig. 3.9). We kept the number of D2D connections constant (20% of the CUEs). The cell throughput is the highest for the GA based solution (marked as ‘Cell, MINLP-GA’) for all active numbers of UEs in the cell. And the cell throughput is higher for the greedy heuristic algorithm than that of the random algorithm for all active numbers of UEs in the cell. For instance with 30 active users per site, the cell throughput are 10% and 3% higher for GA based solution than that of the random (marked as ‘Cell, Random’) and greedy heuristic (marked as ‘Cell, Heuristic’) solutions, respectively. This is due to the fact that the

FIGURE 3.9: Comparisons between random and heuristic D2D RB assignment algorithms in terms of normalized aggregate throughput.

random RB assignment causes a significant amount of interference on the primary cellular network. As the GA based solution is the most efficient to avoid interference to CUEs than the other two algorithms, the sum of the cell and D2D throughput is also the highest when compared to that of the other two approaches. However, the GA based approach consumed a significant amount of time. The greedy heuristic algorithm performs better in the sum of cell and D2D throughput than that of the random algorithm. For example, with 30 active UEs per site, normalized cell throughput are 2% and 11% higher in the case of the GA based solution than that of the heuristic algorithm and the random approach, respectively.

FIGURE 3.10: Impact of the number of D2D connections on normalized aggregate throughput.

3.7.3 The impact of number of D2D connections on network throughput

Finally, we present the impact of the number of D2D connections on the network throughput with varying numbers of active UEs for the proposed greedy heuristic algorithm. The number of D2D connections is either 10% or 20% of the number of active UEs. For any specific number of active UEs, the cell throughput is lower with 20% D2D connections than that with 10% D2D connections. For example, with 40 active UEs per eNB, the cell throughput is 5% lower with 20% D2D connections than that with 10% D2D connections. With more D2D connections, the interference is higher and the cell throughput is lower. However, for any specific number of active UEs, the sum throughput of CUEs and DUEs is higher with 20% D2D connections than

that with 10% D2D connections. With higher D2D connections the sum throughput increases from the contribution of D2D communications.

3.8 Summary

The D2D feature is a promising add-on component to the LTE-Adv network. However, the offered spectrum gain requires an efficient interference management scheme. During the download phase of the cellular network, the UE is exposed to interference from the D2D transmitter, thus sharing its spectrum, and the DUE is exposed to interference from the base station; whereas during the upload phase of the cellular network, the D2D transmitter causes interference only to the eNB. The D2D receiver is also exposed to interference from the UE whose allocated RBs are being shared. In this chapter, we formulated the shared spectrum allocation problem between CUEs and DUEs as an optimization problem that maximizes throughput without impairing the existing cellular network. We also proposed an alternative greedy heuristic RBs selection algorithm. The notion of the heuristic algorithm is that when CUE and D2D transmitters share spectrum, mutual lower channel gains increase the sum throughput of CUEs and DUEs. A higher spatial distance between DUEs and the eNB also increases the network throughput. We also perform extensive simulations to evaluate the efficacy of the proposed RB sharing methods.

TABLE 3.1: Simulation parameters and values

Parameter	Values
Spectrum allocation(uplink/downlink)	20 MHz
Carrier frequency	2 GHz
Number of subcarriers per RB	12
Neighboring subcarrier spacing	15 kHz
RB bandwidth	180 kHz
Number of available RBs	100
Max eNB Tx power	20 W
Max antenna gain	15 dBi
Max UE Tx Power	250 mW [24 dBm]
MCS	QPSK: 1/6, 1/3, 1/2, 2/3 16QAM: 1/2 , 2/3, 3/4 64QAM: 1/2 , 2/3, 3/4, 4/5
Slot duration	0.5 ms
Number of symbols per slot	7(1 Pilot+6 Data)
3dB beam width, θ_{3dB}	65°
Maximum attenuation, A_m	20 dB
Cell-level user distribution	Uniform
Number of active users per cell	10, 20, 30, 40, 50
User speed	5 km/h
Distance attenuation	$L = 35.3 + 37.6 \times \log(r)$
Log-normal shadow fading	8 dB <i>std</i>
Shadowing decorrelation distance	$d_{corr} = 20 \text{ m}$
Cell layout (Hexagonal)	3-sector sites, 21 sectors in total
Cell radius	167 m (500 m inter-site distance)
UE noise figure	9 dB
UE thermal noise density	-174 dBm/Hz

Chapter 4

Collaborative streaming employing device-to-device multicast

A topology-aware peering decision is the key to an efficient P2P streaming system, which could avoid needless traffic traversal through multiple links due to the mismatch between the physical and the overlay network. In this chapter we propose a collaborative streaming system where geographically closed peers form groups, one peer from each group downloads content from Internet peers and then shares the content within groups through D2D multicast mechanism. We present the broadcaster/agents and their optimal assisted peer selection problem as the stable admission assignment problem and formulate the problem as an ILP problem. We also present a distributed algorithm to select agents among the cellular peers with a suitable number of assisted peers for each agent to tackle the retransmission problem. The cellular peers change roles to be a broadcaster or a multicast receiver to ensure fairness. We present a mathematical model for the peer grouping process. The model finds the relationship between two important measurements, namely number of agents and number of assisted peers in a group, affecting the bandwidth savings and streaming rate. We also perform extensive simulations to show the efficacy of our design and to verify our claims.

4.1 Agent and assisted peer selection as stable admission problem

Our proposed P2P streaming system comprises of both Internet and cellular peers. We name it as a collaborative streaming system because of the collaboration involved among peers within each group. The streaming source is located on the Internet. We need to design a scheme to disseminate streaming content among peers efficiently. To save the expensive wireless bandwidth, and to avoid congestion at the base station, we need to select broadcasters among cellular peers that download content from Internet peers, and then share with other neighboring cellular peers through a multicasting mechanism. The peers that broadcast (we name them agents) and the peers that receive content from the agents are assisted peers. We consider the following issues to design an efficient streaming system:

- Minimizing the number of agents would save wireless bandwidth by delivering content to assisted peers with a minimized number of transmissions.
- There should be enough agents so that none of them has to feed more than a specific number of assisted peers due to limitations of retransmission
- An agent should prefer a closer one among two assisted peers and vice versa. This would increase the multicast rate.

The cluster based broadcasting approach has been around for a long time. The cluster heads (agents in our case) selection criteria proposed in the past are based on node's *id* [145, 146], node's connectivity [147–149], or node's mobility [150]. However, we need to find agents that serve as many assisted peers as possible while not

exceeding a certain limit to tackle the retransmission problem. Furthermore, agents should choose closer peers to the assisted ones to maximize the multicast rate. The problem is rather close to many-to-one matching or the stable admission problem [105, 107, 151], which finds a stable assignment between universities and students under a strict order of individual preference. The assignment is stable when there are no university or student pairs who prefer each other over the outcome they get from the given assignments. In our problem, universities are analogous to agents and students are analogous to assisted peers. A university, on one hand, can offer admission to as many students as it may afford, while a student accept admission to only one university. Likewise, one agent may assist a limited number of peers through broadcasting, and an assisted peer may only get assistance from one agent. Preparing a preference list is simple and straightforward. Optimum performance is achieved from the stable admission algorithm by utilizing individual preference. To increase the multicast rate within a group, an agent prefers a closer peer over a distant peer and vice versa. In addition to solving the admission problem, we need to minimize the number of agents/universities as well. Also, the roles of peers have not been determined beforehand. Therefore we formulate the stable admission or the agent selection problem as an ILP framework.

In the steady state, cellular peers under an arbitrary eNB station are represented by an undirected graph $G = (\mathcal{N}, \mathcal{L})$. Peers are labeled through 1 to N . The presence of a link, represented by a pair (u, v) of distinct nodes, means that peer u and v are in direct communication range. The subset $\mathcal{A} \subseteq \mathcal{N}$ represents a set of agents. $\mathcal{N}_u$ is a set representing a neighborhood of peer u . One agent can handle maximum S assisted peers due to limitations of retransmissions. For all $u \in \mathcal{N}$, d_u^{eNB} represents the distance of peer u from the eNB, whereas d_{uv} represents the distance between

two peers, u and v . We define a binary variable x_u which satisfies $x_u = 1$, if $u \in \mathcal{A}$, and '0' otherwise. Let y_{uv} represent a binary variable such that $y_{uv} = 1$, if peer u assists peer v , and '0' otherwise. The symbol $>_u$ indicates the preference orderings of peer u . For example, $j >_u v$ indicates that agent u prefers peer j over peer v as an assisted peer. Similarly $i >_v u$ denotes that an assisted peer v prefers peer i over peer u as an agent. Agents and assisted peers always prefer closer peers to maximize the multicast rate. Therefore if $d_{uj} \leq d_{uv}$ then $j >_u v$, else $v >_u j$. The following ILP model describes the problem of selecting appropriate agents:

$$\text{Minimize} \quad \sum_{u=1}^N x_u d_u^{eNB} \quad (4.1)$$

Subject to:

$$\sum_{\forall u \in \mathcal{N}_v} y_{uv} + x_v = 1, \quad \forall v \in \mathcal{N} \quad (4.2)$$

$$\sum_{\forall v \in \mathcal{N}_u} y_{uv} \leq S, \quad \forall u \in \mathcal{N} \quad (4.3)$$

$$x_u \geq y_{uv}, \quad \forall u \in \mathcal{N}, \forall v \in \mathcal{N}_u \quad (4.4)$$

$$S y_{uv} + S x_v + S \sum_{i >_v u} y_{iv} + \sum_{j >_u v} y_{uj} \geq S x_u, \quad \forall (u, v) \in \mathcal{L} \quad (4.5)$$

The objective function in Eq. (4.1) minimizes the number of agents, which is analogous to minimizing the number of broadcast transmissions. The constraints in Eq. (4.2) ensure that each peer is either serving as an agent or getting assistance from only one agent. The constraints in Eq. (4.3) guarantee that an agent assists S peers at most. The constraints in Eq. (4.4) meet the requirements that an assisted peer is only served by one agent. The constraints in Eq. (4.5) ensure that a sufficient

number of agents are selected and the selected pair is a match. The formulated ILP problem ensures that there are enough agents to avoid a feedback implosion problem and therefore it differs from the problem of finding a minimum dominating set.

4.2 The proposed collaborative streaming system

The ILP-based agent selection problem, presented in Section 4.1, requires the knowledge of the global topology. The optimal solution becomes unsolvable or at least intractable in polynomial time due to the increase in the size of the solution space. A distributive grouping algorithm is required not only due to the enormous size of the solution space, but more importantly because of the self-organizing and decentralized nature of the P2P system. In this section, we first describe distributed agent selection problem that captures the objective and the constraints presented in the ILP formulation, and then provide a detailed description of the content dissemination technique. Our main focus is on the cellular part of the streaming network, and the air interface of the cellular network is the LTE.

4.2.1 Agent selection and group formation

In our design, agents which have a typically good connection to the eNB, first download content from the eNB through the Internet Gateway and then share content with peers closer to them in terms of mutual distance. We assume that there is no cooperation from either the ISP or the MNO on peering and grouping decisions. The MNO only allocates D2D resources whenever there is a request from the UEs. CUEs may employ network controlled beacons to discover peers around its neighborhood

[62]. Alternatively, users may utilize the topology-aware C-Chord [26] system to communicate with a neighboring device directly instead of connecting through the eNB. We now briefly present the agent selection algorithm. A peer only communicates with one-hop neighbors to determine its role as an agent or as an assisted peer in a distributive manner. Peers who do not have neighbors within the D2D maximum distance, download content through the eNB.

- At the beginning of the agent selection, peers exchange information of their eNB distance with their one-hop neighbors. If the distance information is unavailable, peers exchange an exponential moving average of their signal-to-noise ratio (SNR) corresponding to the eNB with their one-hop neighbors. Any peer who does not have any neighbors within the D2D maximum distance remains as a stand-alone peer and downloads content from Internet peers through the eNB. We named them as outliers.
- Each peer, whose eNB distance is greater compared to any of its neighboring peer's eNB distance, contacts the nearest agent for assistance. If the request for assistance is rejected by one agent, it contacts the next agent in order of mutual distance. If peer's proposals to be served are rejected by all possible agents and the peer itself has potential neighbors to serve, it declares itself as an agent.
- Each agent accepts requests from at most S subordinates with the preference to the closest one first. Due to peer dynamics, if an agent does not have enough assisted peers, it contacts the closest agent in order of mutual distance, and invites the agent to become its sensing agent.

Algorithm 4 describes how a cellular peer u determines its role as an agent or an assisted peer. The value of the integer variable s_u is the final outcome of the

FIGURE 4.1: Cellular peers under an eNB from groups. Agents download contents from the Internet peers, and then share with multicast receivers employing D2D multicast. The D2D maximum distance $d_M = 30$ m. The maximum number of assisted peers an agent assists, $S = 3$. For clarity, only a part of the network is shown.

algorithm that defines the role of the peer u . Figure 4.1 illustrates the outcome of the distributed solution to the problem of selecting agents at a particular instant in time. For clarity, only a part of the network is shown. The agents (marked by the square) download streaming content from the eNB, and then disseminate content to their one-hop assisted peers (marked by a filled circle), exploiting the broadcast nature of the wireless medium. The maximum number of assisted peers (S) that an agent can assist, is set to ‘3’. For example, two agents i, j were selected with a maximum number of ‘3’ assisted peers per agent even though all of them are close. The agents employ a D2D multicast mechanism; there is no collision as the eNB allocates channels appropriately. Therefore, agents i, j concurrently disseminate content to their assisted peers. Peers only request D2D resources for multicast application; the eNB is not required to support P2P streaming system.

Algorithm 4: The functionality of peer u to determine the broadcasting role in a distributed manner.

Symbol definition:

S : maximum number of assisted peers
 T : an integer larger than S
 s_u : status of peer u ; undetermined (-1), agent (0 to S), assisted ($T > S$)
 $\mathcal{N}_u$: set consisting one-hop neighbor of peer u
 d_u^{eNB} : distance of peer u from the eNB
 req_u : number of request for assistance to u

Initial assignment:

Send d_u^{eNB} to all $v \in \mathcal{N}_u$
 Receive d_v^{eNB} from all $v \in \mathcal{N}_u$
 $s_u = -1$ (status undetermined);
 $req_u = 0$;
begin
 while $s_u == -1$ **do**
 if $d_u^{eNB} \leq \min(d_v^{eNB}, \forall v \in \mathcal{N}_u \ \& \ s_v < 0)$ **then**
 $s_u = 0$; (set itself as agent)
 while wait for assistance request **do**
 if Receive request for assistance **then**
 $req_u = req_u + 1$;
 end
 end
 while $s_u < S \ \& \ req_u > 0$ **do**
 Accept request from the closest one;
 $s_u = s_u + 1$;
 $req_u = req_u - 1$;
 end
 else
 Find agents for assistance;
 if found & $s_{agent} < S$ **then**
 Send request to the nearest agent for assistance
 if accept **then**
 $s_u = T$; (status is assisted)
 end
 end
 end
 if $s_u > S; \forall v \in \mathcal{N}_u$ **then**
 $s_u = -2$; force termination; no potential neighbor left
 end
 end
end

Algorithm complexity: We measure the performance by communication complexity. Assuming for an arbitrary eNB, $\Delta = \max(|\mathcal{N}(u)|), \forall u \in \mathcal{N}$, denotes the maximum degree among N cellular peers. Then the total number of messages to disseminate neighboring information among peers at the beginning of the algorithm is $O(N\Delta)$. If a peer determines its state to be an agent, it needs to send at most S communication messages of ‘acceptance’ to the assisted peers. Therefore total $O(AS)$ messages are required for A number of agents. Any peer seeking assistance requires $\lceil \frac{\Delta}{S} \rceil$ messages in the worst case scenario to send a request for assistance. So the total number of messages in this case is $O((N - A) \lceil \frac{\Delta}{S} \rceil)$. The number of messages required is scalable, and the price paid for it is negligible when compared to the amount of cellular bandwidth saving to play the bandwidth hungry media.

Extension to the algorithm (Incentive mechanism): An incentive mechanism is essential to enforce collaboration and exchange data among peers. This would also ensure the fairness. We incorporate a distributed reciprocal incentive mechanism into the proposed streaming system. Each cellular peer maintains additional information on pair-wise shared amount with its neighboring peers, named as ‘reward’. For any peer, a *reward* with respect to its neighbor is the ratio of content it has delivered to the neighbor to the content it has received from that particular neighbor. Therefore a high value of *reward* indicates that its neighbor’s contribution is negligible, whereas a low value of *reward* signifies that a peer receives more content. Periodically, peers apply a *tit-for-tat* strategy when serving as an agent. It denies multicast service to its neighbor if the *reward* with respect to that giving neighbor is below a threshold value. Therefore a peer will serve more to other peers who have contributed as agents. A free rider which does not play a role as an agent for any other peers would not get streaming services from others and would download contents through the eNB. Such

an incentive mechanism among peers for getting qualitative service translates into overall improved system performance. However a joining peer must get assistance without considering the *reward* it may offer. Also in the real world such a hard and fast incentive method may be relaxed so that heterogeneous peers with lower upload bandwidth get streaming services from the contribution of other altruistic peers.

4.2.2 The content dissemination technique

In the proposed collaborative streaming system, a source in the Internet progressively generates new chunks with consecutive time-stamps. We now discuss how to synchronize content download and broadcast sessions. Let us assume agents download M chunks within τ_1 seconds. The broadcasting session completes within τ_2 seconds. Then for τ_3 seconds, cellular peers download missing chunks from neighboring agents or from the Internet. During these τ_3 seconds, peers also exchange information to determine the agents on the next cycle. Here we describe how to disseminate M chunks within $\tau_1 + \tau_2 + \tau_3$ seconds beginning from a time-stamp t .

1. From t to $t + \tau_1$ seconds, selected agents download M chunks from Internet peers. An agent might miss a few chunks from the M chunks, even if the agent selection algorithm always favors cellular peers with target SNR as the candidates. If a cellular peer does not have a neighbor, it simply downloads M chunks from the Internet within time-stamps $t, t + \tau_1 + \tau_2 + \tau_3$.
2. From $t + \tau_1$ to $t + \tau_1 + \tau_2$ seconds, the agents share M chunks (excluding any missing chunks) with neighboring peers employing D2D multicast applications. The eNB is not aware of the application and only remains in control of mode switching between cellular to D2D roles and vice versa. Selecting a multicast

transmission rate is critical to avoid packet errors. We illustrate the multicast mechanism at the end of this discussion.

3. Peers in a streaming system maintain lists of multiple senders. From $t + \tau_1 + \tau_2$ to $t + \tau_1 + \tau_2 + \tau_3$ time, multicast receivers (assisted peers) prepare a list of missing chunks, and send requests for missing chunks to the agents. If time allows, peers continuously download missing chunks, or otherwise play media at a degraded rate. During this period, each cellular peer runs the distributed agent selection algorithm for the next cycle.

If the streaming rate is R_s kilobyte per second (kbps), then each cellular peer needs $R_s \times (\tau_1 + \tau_2 + \tau_3)$ kb content in $\tau_1 + \tau_2 + \tau_3$ seconds. Each broadcaster has to download at the rate of $R_s \times (\tau_1 + \tau_2 + \tau_3)/\tau_1$ kbps, and then transmit at the rate $R_s \times (\tau_1 + \tau_2 + \tau_3)/\tau_2$ kbps employing a multicast scheme.

D2D multicast reliability mechanism: In our agent selection algorithms, we allow agents to support a limited number of assisted peers. By limiting the multicast group to a smaller size we avoid limiting the multicast transmission rate to the lowest instant sustainable rate for the weakest peers around the neighborhood. Limiting the number of assisted peers for each agent also makes retransmission possible. The agent adapts the multicast rate based on the QoE feedback by assisted peers [34] to avoid huge errors or to grab a higher rate if achievable. As the agents prefer closest peers as assisted peers, the adapted multicast transmission rate is also high. For real-time traffic like UDP-based streaming, in-time delivery is more crucial than providing a hard core guarantee of the correct packet arrival. It is more acceptable to allow a few erroneous packets than waiting for retransmission as long as the target level of user satisfaction is achieved. Our P2P architecture offers multiple-peer selection

capability. In case any assisted peer misses a few pieces of video chunks, it simply finds an alternate Internet peer to download the missing chunks from, typically at a lower rate without affecting other peers in its group.

4.2.3 Group maintenance

A peer network is dynamic in nature; a peer joins or leaves the system at any time due to mobility or other reasons. A peer may leave the group without any notification. Group maintenance is therefore essential. The agent sends multicast beacons periodically to its assisted peers at an interval T_b . If an assisted peer does not receive two consecutive multicast beacons from the agent, it considers that the agent is no longer serving. If an agent leaves the system, the assisted peer with the lowest distance from the eNB among group members sets itself as an agent. Assisted peers send acknowledgements to their corresponding agent. If an agent does not receive two consecutive acknowledgements from the same assisted peer, it considers that that assisted peer has left the system.

4.3 Stochastic properties of groups

In our proposed collaborative streaming system, we presented a mechanism where peers form groups/clusters themselves to share content. We also presented an ILP framework that finds an optimum formation of clusters in terms of eNB off-load amounts with guaranteed streaming quality. In this section, we present an analytical approach to derive the expected number of groups and group sizes.

We first derive distance distribution among peers. Homogeneous Poisson point process (PPP) and binomial point process (BPP) are two most common ways to

FIGURE 4.2: Node distribution in a compact set W in a tow-dimensional region $\mathbb{R}^2$. C the set of a perfect disk with center at o and radius r .

represent node distribution of a random network. When the number of nodes is known, the PPP may not be accurate enough as the ‘Poisson’ distribution permits an unbounded number of nodes. The assumption, that the number of nodes in disjoint areas is independent in a PPP, also affects the accuracy [113]. For example, if most of the nodes are located in a particular region, the remaining area is inevitably almost empty. The PPP fails to capture this fact. The number of peers in the steady state is fixed and known for the purpose of stochastic analysis, therefore we assume that peers are distributed as a BPP (denoted as Φ). Let W be a compact set in a planar space $\mathbb{R}^2$ that contains exactly N peers. The area of $\mathbb{R}^2$ is denoted as $|W|$. Let R_n be a random variable that indicates the Euclidean distance of the n -th nearest peer of BPP Φ . Let $c(o, r)$ be a perfect disk of center o and radius of r . See Fig. 4.2. If $c(o, r)$ contains exactly $n - 1$ peers, and n -th peer is located at the circumference of the disc, then the radius r is the Euclidean distance of the n -th nearest peer. As N peers are independent and identically distributed (*i.i.d*), the cumulative distributive function (*cdf*) of R_n is the probability that there are at least n points in $c(o, r)$ as the

following:

$$\begin{aligned}
F_{R_n}(r) &= \mathbb{P}\{\text{at least } n \text{ peers in the disk}\} \\
&= \sum_{k=n}^N \binom{N}{k} p^k (1-p)^{N-k}, \quad 0 \leq r \leq R \\
&= I_p(n, N - n + 1)
\end{aligned} \tag{4.6}$$

provided the center of the disk does not coincide with any of the peer location¹. Here $p = |c(o, r) \cap W|/|W|$ is the probability that any peer lies in $c(o, r)$ when uniformly distributed over $|W|$. The cdf $F_{R_n}(r)$ is expressed with a regularized incomplete beta function with the following definition [152, p. 231, 1473]:

$$I_x(a, b) = \frac{\int_0^x t^{a-1} (1-t)^{b-1} dt}{B(a, b)}, \quad (a, b) > 0 \tag{4.7}$$

The probability distribution function (*pdf*) of Euclidean distance r is :

$$\begin{aligned}
f_{R_n}(r) &= \frac{d}{dr} F_{R_n}(r) \\
&= \frac{d}{dr} I_p(n, N - n + 1) \\
&= \frac{p^{n-1} (1-p)^{N-n}}{B(n, N - n + 1)} \frac{dp}{dr}
\end{aligned} \tag{4.8}$$

See [152, p. 204] for an explanation of the derivative. $f_{R_n}(r)$ in Eq. (4.8) is an analytical expression of the Euclidean distance of the n -th peer from a reference point.

We want to find the expected number of groups and expected distance from the

¹When o coincides on a peer location, N is replaced by $N - 1$ in BPP Φ .

D2D transmitter to the farthest multicast receiver within a group. While the former measure affects the eNB load amounts, the latter measure affects the streaming rate within a group. In the steady state, there are N users under an arbitrary eNB station labeled 1 to N that will participate in a collaborative streaming system. The area of the cellular coverage is A that is $|W| = A$. The intensity of the system is $\lambda = \frac{N}{A}$. Each peer either serves as an agent, is assisted, or stands as an outlier. $\mathcal{M}$, and $\mathcal{O}$ denote the sets of assisted peers and outliers respectively. Therefore, $\mathcal{A} \cap \mathcal{M} \cap \mathcal{O} = \phi$ and $\mathcal{A} \cup \mathcal{M} \cup \mathcal{O} = \mathcal{N}$.

4.3.1 Expected number of groups

In the collaborative streaming system, agents use orthogonal spectrum resources, and may have overlapping transmission regions in a dense network where they serve a restricted number of peers. As mentioned earlier, we restricted the maximum number of assisted peers within S to avoid the feedback implosion problem. We consider aforementioned conditions to find the expected number of groups in the system. We take a reference point that coincides with the location of a peer u . There are $N - 1$ peers left instead of N peers. Thus the probability that peer u has k neighbours is:

$$\mathbb{P}\{\text{peer } u \text{ has } k \text{ neighbors}\} = \binom{N-1}{k} \rho^k (1-\rho)^{N-1-k} \quad (4.9)$$

where the probability ρ is defined as the following

$$\rho = \frac{\pi(d_M)^2}{A} \quad (4.10)$$

Now in a group of size k , all members are equally likely of being an agent with probability $\frac{1}{k}$. Therefore the probability that peer u is an agent is

$$\mathbb{P}\{u \in \mathcal{A}\} = \binom{N-1}{k} \sum_{k=1}^S \frac{1}{k+1} \rho^k (1-\rho)^{N-1-k} \quad (4.11)$$

Similarly we can find the probability that peer u is an outlier as the following:

$$\mathbb{P}\{u \in \mathcal{O}\} = (1-\rho)^{N-1} \quad (4.12)$$

Let $\mathcal{C}$ be the set of peers which download content from the eNB using cellular links. Like an agent, an outlier has to download from the eNB. Therefore $\mathcal{C} = \mathcal{A} \cup \mathcal{O}$. From Eqs. (4.11) and (4.12) we get the probability that peer u download from the eNB:

$$\mathbb{P}\{u \in \mathcal{C}\} = \sum_{k=0}^S \frac{1}{k+1} \binom{N-1}{k} \rho^k (1-\rho)^{N-1-k} \quad (4.13)$$

As there are N peers in the system, the expected number of groups in the system is:

$$E[N_G]_{prop} = N \mathbb{P}\{u \in \mathcal{C}\} = N \sum_{k=0}^S \frac{1}{k+1} \binom{N-1}{k} \rho^k (1-\rho)^{N-1-k} \quad (4.14)$$

As a cross-check of the number of groups in the system, we apply the Matérn hard-core operation [153] to get an alternative expression. The Matérn hard-core point process does not allow points to lie closer together than a certain minimum distance h . The resulting process Φ' from the hard-core operation is a subset of the basic process Φ

$$\Phi' \subset \Phi \quad (4.15)$$

Here each point is assigned a random mark. A node is retained if it does not have neighbors with marks higher than itself within distance h ; any other nodes are thinned out from the basic process. The retained nodes after the removal of thinned out nodes possess the intensity λ' from the initial intensity λ [110, p. 162] as:

$$\lambda' = \frac{1 - \exp(-\lambda\pi h^2)}{\pi h^2} \quad (4.16)$$

There is a similarity between the behaviors of agent selection algorithm and the Matérn hard-core process. The retained points produce non-overlapping disks of radius $h/2$ with centers coinciding with their locations. We assume that our proposed agent selection algorithm also tries to get non-overlapping disk of a radius equal to the median of the D2D maximum distance. Utilizing $\lambda = N/A$ and $h = d_M$, the expected number of groups in the system with Matérn hard-core process

$$E[N_G]_{\text{Matérn}} = \frac{N}{\lambda} \frac{1 - \exp(-\lambda\pi(d_M)^2)}{\pi(d_M)^2} \quad (4.17)$$

The Matérn hard-core process is accurate enough when N is large and the area of the disk ($\pi(d_M)^2$) is small enough compared to cell area A . In the real world, the accuracy could be smaller as we neglect temporal and spatial correlations. Eqs. (4.17) and (4.14) provide the expected number of groups with the Matérn hard-core process and proposed method respectively.

4.3.2 Agent to farthest assisted peer expected distance

Now we will derive the expected distance between an agent to its farthest assisted peer. The agent coincides with a peer location, and there are $N - 1$ remaining

peers. d_G denotes the random Euclidean distance of an agent to its farthest assisted peer when the agent is serving S assisted peers. From Eq. (4.8) and by utilizing $r = (A\rho/\pi)^{1/2}$, we get the expected distance between an agent to its farthest assisted peer from the following expression:

$$\begin{aligned}
E[d_G] &= E[R_S | R_S \leq d_M] \\
&= \frac{\int_0^{d_M} r f_{R_S}(r) dr}{F_{R_S}(d_M)} \\
&= \left(\frac{A}{\pi}\right)^{\frac{1}{2}} \int_0^\rho \frac{\rho^{S-1+\frac{1}{2}}(1-\rho)^{N-1-S}}{I_\rho(S, N-S)B(S, N-S)} d\rho \\
&= \left(\frac{A}{\pi}\right)^{\frac{1}{2}} \frac{I_\rho(S+1/2, N-S)B(S+1/2, N-S)}{B(S, N-S)} \\
&= \left(\frac{A}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma(S+\frac{1}{2})\Gamma(N)}{\Gamma(N+\frac{1}{2})\Gamma(S)} \frac{I_\rho(S+1/2, N-S)}{I_\rho(S, N-S)} \\
&= \left(\frac{A}{\pi}\right)^{\frac{1}{2}} \frac{S^{(\frac{1}{2})}}{N^{(\frac{1}{2})}} \frac{I_\rho(S+1/2, N-S)}{I_\rho(S, N-S)} \tag{4.18}
\end{aligned}$$

Here $x^{(p)} = \frac{\Gamma(x)\Gamma(x+p)}{\Gamma(x)}$ denotes the rising Pochhammer symbol that obeys the transformation due to Euler [152, p. 2267]. When an agent serves less than S assisted peers, the expected distance between an agent to its farthest assisted peer would be lower than $E[d_G]$.

4.4 Performance evaluation

We perform extensive simulations to evaluate the performance of our proposed collaborative streaming system. We developed a simulator using $C++$ programming language that realistically modeled the LTE system [154]. We run an event-driven

program on top this simulator. Our distributed scheme is solvable for any size of network. Finding an optimal solution of the ILP based stable admission approach is NP-hard, and such a global formulation is not realizable in the context of the P2P network. We solve the ILP based stable admission assignment formulation with the IBM ILOG Optimizer [155] for the purposes of comparison.

Topology: We implement the streaming system on several different networks that include peers from the Internet and also from cellular networks. Peers join the system following the Poisson arrival process. Peers stay in the system unless the streaming ends or the peers fail. The physical links between Internet peers are generated by using a stochastic loss model. The available bandwidth is set randomly in the range $[0.75R_0, 1.25R_0]$, where R_0 is the base bandwidth. We also set the maximum bandwidth of cellular peers to $0.1R_0$, 10 times smaller than the wired network. We set $R_0 = 300 \text{ kbps}$, which is realistic as sources are tapped while uploading. The simulator in [154] captures the time varying capacity of the wireless channel. There are 7-cells in a 3-sector hexagonal layout.

Workload: Each peer participates in downloading a 97.5 MB video that is divided into 390 chunks of size 256 KB.

Parameter setting: We choose $\tau_1 = 5$ seconds which is sufficiently higher than the round-trip-time (RTT). τ_2 does not have any impact on the system performance as long as it remains lower than the RTT. The values of τ_2 and τ_3 are 4 and 5 seconds respectively. The run time of the simulation is 6 minutes excluding the initial setup time. In all cases where we use a confidence interval, we repeat our simulation 50 times and plot the curve with the error bar showing 95% confidence.

First we evaluate the streaming system in terms of the number of groups and the group sizes. We also verify the impact of these two measures using our derived closed

form mathematical expressions. We compare the proposed distributive collaborative streaming system with ILP based optimal and traditional P2P streaming systems in terms of eNB loading, peer's achievable streaming rate, and streaming cost.

4.4.1 Parameters that affect the number of groups

The performance of the streaming system depends on the number of groups and group sizes. The peer density, D2D maximum distance, and maximum number of assisted peers in a group affect these two measures. Smaller number of groups leads to a reduction in eNB loads. On the other hand, the larger the group size is the higher the distance of an agent to its farthest assisted peer is, *i.e.* the lower the multicast rate is.

FIGURE 4.3: The effect of D2D maximum distance and peer density on number of groups. The maximum number of allowable assisted peers is set to '4'.

To investigate how the peer density and D2D maximum distance affects the number of groups, we compute the number of groups for three different values of D2D maximum distances 10, 15, and 25 m . For each set, we vary the number of peers per eNB. The maximum number of assisted peers in a group is set to ‘4’. Figure 4.3 illustrates the variation in the number of groups with peer density and different sets of D2D maximum distances. First, there is negligible difference between the ILP based stable admission approach (marked as ‘CS-ILPSA’) and the proposed distributed scheme (marked as ‘CS-DISTR’). For example, with ‘120’ peers per eNB and a 15 meter D2D maximum distance, the number of groups with an ILP based stable admission approach and proposed distributed scheme are ‘60’ and ‘68’ respectively. However, the gap between the ILP and the distributed schemes is higher with high peer density than that with the low peer density. When the peer density is low, choices for selecting suitable group members are limited. As the peer density increases, combinations of groups with different members are many. Therefore the ILP scheme performs better than the distributed scheme when the peer density is higher. Second, as the number of peers increases, the number of groups increases. However, the number of groups to total peers ratio decreases as the number of peers increases. In a dense network, eNB off-load amounts would be higher than that of a sparse network. Also, with a larger D2D maximum distance, the number of groups is smaller. For example, when the number of peers per eNB is ‘120’ in the distributive scheme, the number of groups is ‘86’ with $d_M = 10$, whereas the number of groups with $d_M = 25$ is only ‘46’. Therefore, the number of groups is smaller for a larger D2D maximum distance and in a denser network, .

Figure 4.4 depicts the variation of the number of groups with peer density and different maximum number of assisted peers. As before, the proposed distributive

FIGURE 4.4: The effect of maximum number of assisted peers on number of groups. The D2D maximum distance is set to 20 meter.

scheme closely follows the trend of the ILP based stable admission approach. For example, when the maximum allowable assisted peers is ‘2’ and the number of peers is ‘120’, the number groups with ILP based stable admission assignment algorithm and proposed distributed scheme are ‘55’ and ‘61’ respectively. The number of groups decreases as the maximum allowable assisted peers in a group increases. For example, with ‘180’ peers per eNB in the proposed distributive scheme, the number of groups is ‘102’ with a maximum of ‘1’ assisted peers, whereas the number of groups is only ‘67’ with a maximum of ‘6’ assisted peers. The result indicates that with more allowable assisted peers, the group size increases and the number of groups decreases.

We compare the number of groups measured from the simulation method with our derived mathematical expressions [Eqs. (4.17) and (4.14)] as shown in Fig 4.5. The Matérn model is not accurate enough as it does not have the dependency on group size restriction. Also, the assumption on the separation ($h = d_M$) is partly

FIGURE 4.5: The comparison between simulation and analytical model in terms of number of groups. The D2D maximum distance and maximum allowable assisted peers in a group are set to 20 m and ‘3’ respectively.

accountable for the error. Our proposed model on the number of groups calculation is more accurate than that of the Matérn model.

4.4.2 eNB traffic off-loading and bandwidth saving

To investigate the eNB congestion and wireless bandwidth saving, we monitor the traffic through a particular eNB averaged over time. As the patterns remain the same, we only present the simulation result for a particular eNB. Figures 4.6 and 4.7 depict the eNB traffic averaged over time for the proposed collaborative and traditional P2P streaming system. In all cases, streaming contents are normalized to the highest traffic for the purpose of comparison. In the traditional P2P streaming system, the amount of eNB traffic due to streaming (marked as ‘Trad-P2P-Stream-Syst’) grows almost linearly. The amount of eNB traffic due to streaming in the proposed system and

FIGURE 4.6: The variation of eNB load with peer density and D2D maximum distance. The maximum allowable assisted peer is set to 4.

ILP based approach are less than that of traditional P2P system. This is due to the fact that in the traditional P2P system, all cellular peers download content through the eNB, and do not share content between them. Also, being oblivious to network architecture, cellular peers (in a traditional system), often make peering decisions with other cellular peers through the eNB and accrue two-way wireless bandwidth, whereas in our proposed collaborative streaming system with an increase in the number of cellular peers, more and more peers join the D2D multicast application and off-load the traffic through the eNB. Also, the difference between the proposed collaborative and ILP based streaming systems in terms eNB off-loads is negligible. Hence, the proposed collaborative streaming system reduces eNB traffic by taking the advantage of the D2D multicasting technique. For the proposed distributive streaming system (as shown in Fig. 4.6), when the number of peers per eNB is 120, the normalized load at the eNB is ‘0.48’ with $d_M = 10$, whereas the same distance is ‘0.26’ with $d_M = 25$.

FIGURE 4.7: The variation of eNB load with peer density and maximum allowable assisted peer. The D2D maximum distance is 20 meter.

Therefore with a larger D2D maximum distance, fewer agents are required which leads to a reduction of eNB load. Again, in Fig. 4.7 for the proposed distributive streaming system, when the number of peers per eNB is 120, the normalized load at the eNB is ‘0.38’ with $S = 1$, whereas the same distance is ‘0.31’ with $S = 6$. Therefore if the streaming system allows more assisted peers for each agent, fewer agents are required and the load on eNB decreases.

4.4.3 The impact of group size

There is a limit on how much the streaming system off-loads content from eNB without affecting its performance. A smaller number of groups leads to a larger distance between agent to farthest peer distance and thereby lowers the streaming rate in a group. Figure 4.8 shows how the restriction on group size and D2D maximum

FIGURE 4.8: The variation of average distances from an agent to its farthest assisted peer with restriction on group size. There are 100 peers per eNB in the system. The D2D maximum distances are 15 and 25 m .

distance affect the average distance from an agent to its farthest assisted peer. The more the maximum number of allowable assisted peers is, the higher the distance is from the agent to the farthest peer. A lower number of maximum allowable assisted peers forces the agent to choose closer peers only. For example, with 100 peers per eNB and $d_M = 25$ m in the proposed distributive scheme, the agent to farthest assisted peer is 12.2 m with 1 assisted peer's maximum, whereas the same distance is 17.9 m with 6 assisted peers' maximum. Therefore by allowing more peers in a group, the agent to farthest assisted peer distance increases. The differences among the proposed collaborative streaming system, ILP based approach and mathematical model ($E[d_G]$) are negligible, and they all follow a similar pattern.

Figure 4.9 depicts the variation of agent to farthest peer distance with peer density. As the network density is increased, agents have more chances to fulfill their quota

FIGURE 4.9: The variation of agent to farthest peer distance with peer density. The maximum allowable assisted peers in a group and D2D maximum distance are set to ‘1’ and 25 m .

of the maximum number of assisted peers. They choose peers who are geographically closer to them with a restriction on the maximum number of assisted peers. That is in a dense network, agent to farthest peer distance decreases as the total peers increases. For example, when the number of peers per eNB is 60 in the distributive scheme, the agent to farthest peer distance is 12 m , whereas the same distance is 12.3 m with 180 peers per eNB. Also the performance of the proposed distributive scheme is close enough to the ILP based stable admission approach and mathematical model in terms of agent to the farthest peer distance.

4.4.4 Peer’s experience of streaming rate

We evaluate the performance of the streaming system in terms of peer’s experience on achievable streaming rate. We monitor the content received by the cellular peers,

FIGURE 4.10: Performance measurement in terms of streaming rate experienced by peer with varying network density. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30$ m and 4 respectively.

and take the average of the content received over time. We repeat the procedure for the varying number of cellular peers. All content is normalized to the highest average content among all systems. Due to the mismatch between the physical and overlay network, the average content received by the cellular peers is the lowest for the traditional P2P streaming system for all network sizes. See Fig. 4.10. The average content received with the proposed collaborative system is comparable with that of the ILP based streaming system. Hence, our proposed distributive agent selection algorithm performs close to the optimal solution.

4.4.5 Streaming cost and fairness issue

The streaming cost is defined by the amount of content downloaded through the eNB by the cellular peers. The cost is per user averaged over time, and then normalized

FIGURE 4.11: Per peer streaming cost averaged over time with varying number of cellular peers. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30\text{ m}$ and 4 respectively.

to the highest cost of all systems for the purpose of comparison. We do not count the streaming cost for those D2D multicast applications where the eNB allocates shared resources between the D2D transmitter and cellular users by some interference avoidance technique [154]. Figure 4.11 depicts the per peer streaming cost averaged over time with a varying number of cellular peers. The streaming cost is the highest for the traditional P2P streaming system for all network sizes. The per peer streaming cost is high for both the distributed and optimal algorithm, however it decreases with the increase in peer density. This is due to the fact that with the higher peer density, a single broadcaster is able to share streaming with more cellular peers. The per peer streaming cost in the proposed streaming system is always comparable to that of optimal solution for all network sizes. Hence the proposed streaming system effectively reduces the streaming cost.

FIGURE 4.12: Comparison of streaming cost fairness for different P2P systems. The D2D maximum distance and maximum allowable assisted peers in a group are set to $d_M = 30\text{ m}$ and 4 respectively.

We measure the cost fairness using Jain's fairness [156] index between 0 and 1. The larger the value of index, the more fair it is, and the cost is distributed among the cellular peers. Figure 4.12 illustrates the fairness of the streaming cost for different P2P systems. Although the traditional P2P system shows the highest cost fairness, that aspect is meaningless as peers do not share streams in such a system. The variation is due to the difference in the number of missing chunks among the peers. Streaming cost fairness is lower for both the distributed and optimal solution, however it increases with peer density. As more and more peers share the streaming cost with the increased network size, the fairness of streaming cost increases. Streaming cost fairness for the distributed algorithm closely follows that of the optimal solution.

4.5 Summary

As expected, the amount of eNB traffic due to streaming in the proposed collaborative streaming system is less than that of the traditional P2P system. This is due to the fact that in the traditional P2P system, all cellular peers download content through the eNB, and do not share content between them. Also, being oblivious to the network architecture, cellular peers in a traditional system often make peering decisions with other cellular peers through the eNB and accrue two-way wireless bandwidth, whereas the proposed collaborative streaming system reduces eNB traffic by taking the advantage of D2D multicasting technique. The proposed distributive collaborative streaming system is efficient as it closely follows the performance of ILP based streaming systems in terms eNB off-load.

The performance of the streaming system depends on the number of groups and group sizes. Peer density, D2D maximum distance, and maximum number of assisted peers in a group affect these two measures. For both larger D2D maximum distance and denser network, the number of groups is smaller. And smaller number of groups leads to much reduction in eNB loads. With an increased number of cellular peers, more and more peers join the D2D multicast application and off-load the traffic through the eNB. If the streaming system allows more assisted peers for each agent, fewer agents are required and the load on eNB decreases. There is a limit on how much the streaming system could off-load content from the eNB without affecting the performance. A smaller number of groups leads to larger distance between agent to farthest peer and thereby lowers the streaming rate in a group. However, if the network density is increased, agents have a higher chances to fulfill their quota of the maximum number of assisted peers. And they choose peers who are geographically

closer to them with a restriction on the maximum number of assisted peers. That is in a dense network, agent to farthest peer distance decreases and the multicast rate increases as the total peers increase.

The per peer streaming cost in the proposed streaming system is lower than that of the traditional P2P streaming system for all network sizes. The per peer streaming cost decreases with an increase in peer density. This is due to the fact that with higher peer density, a single broadcaster is able to share streaming with more cellular peers. Therefore the proposed streaming system effectively reduces the streaming cost. Streaming cost fairness increases with peer density. As more and more peers share the streaming cost with the increased network size, the fairness of streaming cost increases. Streaming cost fairness for the distributed algorithm closely follows that of the optimal solution.

Chapter 5

Streaming chunk distribution: An epidemic spreading approach

Apart from peering decisions, the chunk selection strategy is an important factor that influences the user-perceived visual quality directly. Live streaming is delay sensitive and chunks are required to meet time-constraints. In this chapter, we derive an analytical model for chunk distribution of live streaming content by drawing an analogy with the theory of epidemiology. The model finds buffer occupancy probability and provides insight on the relationship between peer population size, buffer length and playback smoothness. Utilizing the model, we evaluate the performance of different chunk selection strategies: RF and MP chunks. The former is more scalable to peer population in terms of buffer space through improving chunk diversity, whereas the latter achieves better playback continuity by distributing urgent chunks first. We also propose a hybrid chunk selection method that exploits the best of both strategies. We verify our claims with extensive simulations.

5.1 System Model and Assumptions

We model the chunk distribution under the influence of chunk selection strategy. Utilizing the model, we determine the buffer occupancy and compare the chunk selection strategies in terms of scalability and playback smoothness. At any given instance,

each peer caches up to a few minutes worth of chunks in the buffer. Let us assume that buffer B can hold up to M chunks. We specify the buffer positions according to the age of the chunks stored. That is the greater the chunk position m is, the older the chunk is. The peer plays back the oldest chunk at $B(M)$ immediately and uses $B(1)$ to store the newest chunk at the current time slot. In the next time slot, the peer removes the chunk at $B(M)$ that is already played, and shifts all the chunks by '1'. The emptied buffer position $B(1)$ can be used to store the newest chunk. Therefore, the buffer acts as a sliding window into the stream of chunks distributed by the streaming source. Figure 5.1 illustrates a peer's sliding window of length M .

FIGURE 5.1: The sliding window mechanism of a peer's buffer B .

We do the following assumptions to derive the model:

1. Peer population N is large enough not to be affected by peer arrival and departure during a short period of time.
2. All peers maintain an equal length of sliding windows.
3. The playback rate does not change.
4. The playback windows of all peers are synchronized.

It is worth mentioning that for real-time live streaming applications, this assumption on synchronisation is reasonable because all peers in a streaming session will try to synchronize with the source. Buffer length is very small compared the to the total length of the video. There could be instances where a few peers may not be fully synchronised. However this instance doesn't effect the overall behavior of the chunk distribution; our simulation results confirms that. We avoid complexity due to loose synchronization and keep the problem tractable. Buffer length is directly proportional to the delay of the playback start time. Let us assume at time t , there are total N peers. $p_{m,t}$ is the probability of finding the chunk at m -th buffer position ' m ' at time t of an arbitrary peer. The probability reaches a steady state for sufficiently large values of t . At that time t , $p_{m,t}$ is independent of t , is denoted as p_m . One important characteristic of the P2P network is that all peers in the same P2P network exhibit similar buffer occupancy distribution because the probability p_m captures the average behavior of all peers in a specific P2P system. As p_m distribution is extendable to the whole P2P system network, we have

$$p_m = \frac{y_m}{N} \quad (5.1)$$

Here, y_m denotes the number of peers that have a chunk at $B(m)$.

5.2 The spread of chunks: an epidemic analogy

The epidemic model describes the spread of diseases and rumors in a population. In general, the population is divided into two or more subgroups: infective, susceptible and removed/isolated individuals. When an infected individual invades a susceptible

individual, the susceptible individual becomes infected. An infected individual may recover from disease through vaccination and attain immunity from further infections. Now if the immunity is temporary, then the Susceptible-Infective-Susceptible (SIS) model characterizes the spread of the disease. Whereas if the immunity is long lasting or permanent, the Susceptible-Infective-Recovered (SIR) model characterizes the epidemic process. If there is no immunity once infected, the epidemic process falls into the category of the Susceptible-Infective (SI) case.

We will model the chunk distribution behavior by applying the SI epidemic process. Just like an infectious disease passes through susceptible and infective individuals, a chunk is distributed amongst peers from those who already have that chunk to those who are still missing the same. That is, a peer is either willing to download a missing chunk (i.e. a peer is susceptible to the disease), or it already has the chunk (i.e. it is already infected by the disease).

Recall that a single peer's buffer $B(m)$ caches different chunks for different values of m . However, all chunks have to face the same fate by moving from buffer $B(1)$ to buffer $B(M)$. Therefore, how a single chunk evolves from buffer $B(1)$ to $B(M)$ is analogous to view M different chunks at buffer B at any instance of time. For our analogy with the epidemic process, we say that a chunk is analogous to a infectious disease that transmits to peers. That is, a peer is susceptible to be affected by M different diseases. From Eq. (5.1) we infer that the number of peers holding a chunk at buffer $B(m)$ is the number of infectious individuals. The population is then analogous to the peers in the system. We assume that during this time the population remains constant.

Let x_m, y_m be the number of susceptible and infective respectively at a discrete

position of buffer $m = 1, 2, \dots, M$. As the population remains closed

$$x_m + y_m = N; \quad \forall m \geq 1 \quad (5.2)$$

5.2.1 The deterministic formulation

Now we will derive the chunk distribution model based on the deterministic version of the epidemic model. From this model we will analyse the prospects and constraints of two existing chunk selection algorithms: RF and MP. This analysis provides us with a framework to design an efficient hybrid chunk distribution strategy. In the following analysis, we let m , x_m , and y_m relax to be continuous variables to make it more tractable. Let Δy_m be the additional number of infectious individuals/peers within dt time for location m , *i.e.*

$$\Delta y_m = y_{m,dt} - y_m \quad (5.3)$$

The disease/chunk passes from an infectious individual to a susceptible individual. The more the infective individuals are, the higher their chances of passing diseases. Similarly, the more susceptible individuals are, the higher their chances of getting infected. On the other hand, the higher the peer's upload bandwidth is, the higher the possibility to disseminate a larger number of chunks/diseases. Therefore no matter what chunk selection algorithm is used and considering dt is small, we can assume with first order accuracy that the additional number of infectious individuals is proportional jointly to peer's upload bandwidth (b), the product of the number of infectious individuals (y_m), the number of susceptible individuals (x_m) and the duration

of infection.

$$\Delta y_m \propto by_mx_m dt \quad (5.4)$$

Let β be the ratio of Eq. (5.4). We have

$$\Delta y_m = \beta by_mx_m dt \quad (5.5)$$

Assume the number of chunks played per unit of time or the playback speed is v , we have

$$dm = v dt \quad (5.6)$$

Let us define dy_m as

$$dy_m = y_{m+dm} - y_m \quad (5.7)$$

Due to the sliding window structure, we have

$$dy_m = y_{m,dt} - y_m = \Delta y_m \quad (5.8)$$

From Eqs. (5.5, 5.6, 5.8) we get

$$dy_m = \frac{\beta by_mx_m}{v} dm \quad (5.9)$$

The infection factor β depends on the chunk selection algorithm. First we analyze the behavior of the two known algorithms: RF and MP chunk selection strategies.

Rarest chunk first selection

For the RF algorithm, the rarer a chunk is, the larger the infection factor is. As x_m is the number of peers that still missing the chunk at buffer position ' m ', x_m characterizes the chunk's rareness at m -th buffer position. The larger the value of x_m is, the larger the number of peers missing the chunks (at m -th buffer position) are. That is for RF algorithm, we assume

$$\beta = \gamma^{RF} x_m \quad (5.10)$$

Here γ^{RF} is the proportionality constant. Therefore for RF algorithm, using Eq. (5.9) we get

$$dy_m = \frac{\gamma^{RF} b y_m x_m^2}{v} dm \quad (5.11)$$

From Eqs. (5.2, 5.11)

$$dy_m = \frac{\gamma^{RF} b y_m (N - y_m)^2}{v} dm \quad (5.12)$$

We can find the value of γ^{RF} from the buffer occupancy characteristic closer to the newest chunk buffer $B(1)$, i.e. the boundary condition. From Eq. (5.12) we get

$$dy_m|_{m \rightarrow 1} \approx \frac{\gamma^{RF} b y_m N^2}{v} dm \quad (5.13)$$

Here as $m \rightarrow 1$, $N \gg y_m$. Only a few peers have newer chunks and the chunk distribution is limited by the number of peers holding new chunks. Therefore from this point of view with $m \rightarrow 1$, the additional number of infectious individuals/peers is equal to the product of the number of peers that have the chunk and peer's average

bandwidth b scaled by v , dm .

$$dy_m|_{m \rightarrow 1} \approx \frac{by_m}{v} dm \quad (5.14)$$

Comparing Eqs. (5.13, 5.14) we get

$$\gamma^{RF} = \frac{1}{N^2} \quad (5.15)$$

From Eqs. (5.12, 5.15) we get

$$dy_m = \frac{by_m(N - y_m)^2}{N^2 v} dm \quad (5.16)$$

The differential equation of Eq. (5.16) can be solved by variable separation technique.

$$\frac{v}{b} \left[\ln \frac{y_m}{N - y_m} + \frac{N}{N - y_m} - \ln \frac{y_1}{N - y_1} - \frac{N}{N - y_1} \right] = m$$

To get the average behavior, utilizing Eq. (5.1) we replace y_m with p_m as follows:

$$\frac{v}{b} \left[\ln \frac{p_m}{1 - p_m} + \frac{1}{1 - p_m} \right] = \frac{v}{b} \left[\ln \frac{p_1}{1 - p_1} + \frac{1}{1 - p_1} \right] + m \quad (5.17)$$

Therefore for RF algorithm we get

$$f_{RF}(p_m) = C_{RF} + m_{RF} \quad (5.18)$$

where,

$$f_{RF}(p_m) = \frac{v}{b} \left[\ln \frac{p_m}{1-p_m} + \frac{1}{1-p_m} \right] \quad (5.19)$$

$$C_{RF} = \frac{v}{b} \left[\ln \frac{p_1}{1-p_1} + \frac{1}{1-p_1} \right] \quad (5.20)$$

$f_{RF}(p_m)$ is a function of chunk occupancy probability at m -th buffer position for RF algorithm.

Most popular chunk first selection

The closer to the playback a chunk position is, the more popular that chunk is because of urgency. Therefore, for the MP algorithm, peer prefers chunk close to the playback position. We capture the phenomenon with the following Eq. (5.21)

$$\beta = \gamma^{MP} \frac{y_m}{1 + \alpha y_m} \quad (5.21)$$

Here γ^{MP} is the proportionality constant. α is a constant that captures MP algorithm's procrastination factor. Considering $\alpha > 1$, the value of β in Eq. (5.21) is less when $m \rightarrow 1$ or y_m is small. For chunks closer to the playback, β increases slowly. This characterizes the procrastination of the MP algorithm which emphasizes downloading chunks closer to the playback first. From the order of x_m and y_m in (5.9) we get $\gamma^{MP} = \frac{1}{N}$. From Eq. (5.2, 5.9, 5.21)

$$dy_m = \frac{by_m^2(N - y_m)}{N(1 + \alpha y_m)v} dm \quad (5.22)$$

The differential equation of Eq. (5.22) can be solved by variable separation technique.

$$\frac{v}{b} \frac{1 + \alpha N}{N} \left[\ln \frac{y_m}{N - y_m} - \frac{1}{y_m} - \ln \frac{y_1}{N - y_1} + \frac{1}{y_1} \right] = m$$

Again by using Eq. (5.1), we replace y_m with p_m to get the following

$$\frac{v}{b} \frac{1 + \alpha N}{N} \left[\ln \frac{p_m}{1 - p_m} - \frac{1}{N p_m} \right] = \frac{v}{b} \frac{1 + \alpha N}{N} \left[\ln \frac{p_1}{1 - p_1} - \frac{1}{N p_1} \right] + m \quad (5.23)$$

Therefore for MP algorithm we get

$$f_{MP}(p_m) = C_{MP} + m_{MP} \quad (5.24)$$

where,

$$f_{MP}(p_m) = \frac{v}{b} \frac{1 + \alpha N}{N} \left[\ln \frac{p_m}{1 - p_m} - \frac{1}{N p_m} \right] \quad (5.25)$$

$$C_{MP} = \frac{v}{b} \frac{1 + \alpha N}{N} \left[\ln \frac{p_1}{1 - p_1} - \frac{1}{N p_1} \right] \quad (5.26)$$

$f_{MP}(p_m)$ is a function of chunk occupancy probability at m -th buffer position for MP algorithm.

Impact of peer population and playback continuity on buffer length

We would like to find the relationships among the following important parameters under the influences of RF and MP algorithms.

- The buffer length (M)
- The size of peer population (N)

- The playback continuity (p_M). It should be noted that p_M reflects the continuity of the playback process. The higher p_M is, the better the playback continuity is.

$\frac{\partial M}{\partial N}$ represents the reactivity of buffer length to population when other parameters such as continuity p_M , bandwidth b , and playback rate v are held constant.

$$\begin{aligned}
 \frac{\partial M_{RF}}{\partial N} &= -\frac{\partial C_{RF}}{\partial N} \\
 &= -\frac{v}{b} \frac{\partial}{\partial N} \left[\ln \frac{1}{N-1} + \frac{N}{N-1} \right] \\
 &\approx \frac{v}{b} \frac{1}{N}; \quad N \gg 1
 \end{aligned} \tag{5.27}$$

$$\begin{aligned}
 \frac{\partial M_{MP}}{\partial N} &= -\frac{\partial C_{MP}}{\partial N} \\
 &= -\frac{v}{b} \alpha \frac{\partial}{\partial N} \left[\ln \frac{1}{N-1} - 1 \right]; \quad \alpha N \gg 1 \\
 &\approx \frac{v}{b} \frac{\alpha}{N}; \quad N \gg 1
 \end{aligned} \tag{5.28}$$

The gradient of M with respect to N is positive for both algorithms, which indicates that as the peer population increases, the system needs more buffer space to achieve the same level of continuity.

$\frac{\partial M}{\partial p_M}$ reflects the reactivity of buffer length to playback continuity when other parameters such as peer population N , bandwidth b , and playback rate v are held

constant. For RF algorithm the gradient of M relative to p_M is

$$\begin{aligned}
 \frac{\partial M_{RF}}{\partial p_M} &= \frac{\partial f_{RF}}{\partial p_M} \\
 &= \frac{v}{b} \frac{\partial}{\partial p_M} \left[\ln \frac{p_M}{1-p_M} + \frac{1}{1-p_M} \right] \\
 &= \frac{v}{b} \left[\frac{1}{p_M(1-p_M)} + \frac{1}{(1-p_M)^2} \right] \\
 &= \frac{v}{b} \frac{1}{p_M(1-p_M)^2}
 \end{aligned} \tag{5.29}$$

For MP algorithm the gradient of M relative to p_M is

$$\begin{aligned}
 \frac{\partial M_{MP}}{\partial p_M} &= \frac{\partial f_{MP}}{\partial p_M} \\
 &= \frac{v}{b} \frac{1+\alpha N}{N} \frac{\partial}{\partial p_M} \left[\ln \frac{p_M}{1-p_M} - \frac{1}{Np_M} \right] \\
 &\approx \frac{v}{b} \alpha \left[\frac{1}{p_M(1-p_M)} + \frac{1}{Np_M^2} \right]; \quad \alpha N \gg 1 \\
 &\approx \frac{v}{b} \alpha \left[\frac{Np_M + 1 - p_M}{Np_M^2(1-p_M)} \right] \\
 &\approx \frac{v}{b} \alpha \left[\frac{Np_M + 1}{Np_M^2(1-p_M)} \right]; \quad N \gg 1 \\
 &\approx \frac{v}{b} \frac{\alpha}{p_M(1-p_M)}
 \end{aligned} \tag{5.30}$$

From Eqs. (5.29, 5.30), it is apparent that the gradient of M relative to p_M is positive. This indicates that the system needs more buffer space to achieve a higher level of continuity while the population remains constant.

Comparison between RF and MP algorithms

We would like to investigate the effects of N , M , and p_M on RF and MP algorithms. From Eqs. (5.27, 5.28), it is evident that the MP algorithm needs α times larger

buffer than the RF algorithm when only the population increases. Therefore the buffer length in the RF strategy is more scalable than that of the MP algorithm to the peer population.

Eqs. (5.29, 5.30) also portray that the MP strategy needs $\alpha(1 - p_M)$ times more or less buffer (depending on actual value of M) than that of the RF strategy as the playback continuity (p_M) increases. For lower values of p_M , as long as $\alpha(1 - p_M) > 1$, the MP strategy requires $\alpha(1 - p_M)$ more buffer than that of the RF strategy. However when playback continuity is larger, $\alpha(1 - p_M)$ becomes less than '1'. Then the MP strategy requires $\alpha(1 - p_M)$ time less buffer than that of the RF algorithm. So intuitively for a given peer population, the MP strategy attains better continuity than the RF strategy for a sufficiently larger buffer length. Now we derive the condition under which the MP algorithm performs better in terms of continuity over the RF algorithm. Let $p_{M'}$ denotes the playback continuity for which $M_{MP} < M_{RF}$. Therefore using Eqs. (5.17, 5.23) we get

$$\alpha \left[\ln \frac{p_{M'}}{1 - p_{M'}} - \frac{1}{N p_{M'}} - \ln \frac{1}{N - 1} + 1 \right] < \left[\ln \frac{p_{M'}}{1 - p_{M'}} + \frac{p_{M'}}{1 - p_{M'}} - \ln \frac{1}{N - 1} - 1 \right]; \quad \alpha N \gg 1$$

As $N \gg 1$ we get

$$(\alpha - 1) \ln \frac{p_{M'}}{1 - p_{M'}} - \frac{1}{1 - p_{M'}} + (\alpha - 1) \ln N + 2 < 0 \quad (5.31)$$

where $p_{M'}$ is the playback continuity for which the buffer length of MP is less than that of the RF. Eq. (5.31) imposes the condition for which MP algorithm would have better continuity than the RF algorithm.

Hybrid strategy

As newer chunks are rarer, the RF algorithm prefers newer chunks. Chunks closer to the playback position are urgent and therefore popular. The MP algorithm prefers older chunks or chunks closer to the playback position. Therefore the RF strategy performs better in terms of chunk distribution closer to the newer chunk position while the MP performs better closer to the playback side. We may employ mix strategies that download newer chunks at the beginning, and the most popular chunks closer to the playback side.

FIGURE 5.2: The spread of chunks among peers in terms of buffer position.

In the quest of optimum switching from RF to MP chunk selection strategy, we first inspect the buffer occupancy probability p_m from the curve derived from the deterministic model. We set peer population $N = 1000$, buffer length $M = 40$, and the bandwidth to playback speed (b/v) ratio to '1'. We compute buffer position m for every infected peers (y_m) for RF and MP strategies from Eqs. (5.17, 5.23)

respectively. We also find additional infected peers per chunk position dy_m/dm from Eqs. (5.16, 5.22). Figure 5.2 depicts how chunks spread among peers with the aging of buffer position for RF and MP chunk selection strategies. As expected, the number of infected peers for the RF strategy (RF-Inf-Peers) is larger than that of the MP strategy (MP-Inf-Peers) for newer chunk positions. Also the additional number of infected peers per chunk position for the RF strategy (RF-Add-Inf-Peers) is larger than that of the MP strategy (MP-Add-Inf-Peers) for newer chunk positions. The maximum number of additional infected peers per chunk position for the RF strategy occurs at a lower chunk position when compared to the MP strategy.

The hybrid strategy (HS) takes advantage of both strategies. The HS employs the RF strategy to download newer chunks and then switches to the MP strategy to download urgent chunks. The switching chunk position m_{sw} is the position up to which the additional number of infected peers for RF strategy is just larger than that of the MP strategy. We may find the position m_{sw} from the following condition

$$\frac{dy_m}{dm}|_{RF} = \frac{dy_m}{dm}|_{MP} \quad (5.32)$$

From Eqs. (5.32, 5.16, 5.22, 5.2) we derive the following two Eqs. (5.33, 5.34).

$$y_{sw} \approx \frac{\alpha - 1}{\alpha} N \quad (5.33)$$

$$p_{sw} \approx \frac{\alpha - 1}{\alpha} \quad (5.34)$$

that leads to Eq. (5.35) using Eq. (5.18)

$$m_{sw} = f_{RF}(p_{sw}) - C_{RF} \quad (5.35)$$

From Eqs. (5.18, 5.19, 5.20, 5.24, 5.25, 5.26), the following Eqs. (5.36, 5.37) define the HS

$$\frac{v}{b} \left[\ln \frac{p_m}{1-p_m} + \frac{1}{1-p_m} \right] = \frac{v}{b} \left[\ln \frac{p_1}{1-p_1} + \frac{1}{1-p_1} \right] + m; \quad \forall m \leq m_{sw} \quad (5.36)$$

$$\begin{aligned} \frac{v}{b} \frac{1+\alpha N}{N} \left[\ln \frac{p_m}{1-p_m} - \frac{1}{Np_m} \right] &= \frac{v}{b} \frac{1+\alpha N}{N} \times \left[\ln \frac{p_{sw}}{1-p_{sw}} - \frac{1}{Np_{sw}} \right] + m - m_{sw}; \\ &\forall m > m_{sw} \end{aligned} \quad (5.37)$$

That is for buffer position $m \leq m_{sw}$ it employs the RF strategy, and for $m > m_{sw}$ it employs the MP strategy.

5.2.2 The stochastic formulation

The chunk distribution analysis through a deterministic epidemic model is suitable when the peer population is large. The stochastic model captures inherent fluctuations in the buffer occupancy distribution. In this section, we analyze the chunk distribution from the Markov chain based epidemic model. If the number of newly infected peers at buffer position $m+1$ is y'_{m+1} then

$$y_{m+1} = y_m + y'_{m+1} \quad (5.38)$$

The discrete distributions ‘Poisson’ and ‘geometric’ would allow for an unbounded number of infectious individuals. On the contrary, the binomial discrete distribution restricts the number of possible infectious individuals in one period to be bounded within zero to a fix positive integer. Therefore we assume that the new number of infected peers from x_m number of susceptible peers on the next buffer position follows the binomial distribution with parameters x_m, p_r^m . Here p_r^m is the probability of the susceptibility to convert to an infective peer. Therefore, a one step transition probability from state i to state j is

$$P_{ij}[y_{m+1} = j + i | y_m = i] = \binom{x_m}{j} (p_r^m)^j (1 - p_r^m)^{x_m - j} \quad (5.39)$$

Here, the number of infected peers in state i is i , and the number of newly infective

FIGURE 5.3: The transition probability from state i to all possible states in Markov process.

peers at state j is j . State j depends only on the current state of i , therefore the system follows the Markov process. Figure 5.3 illustrates the transition probability from state i to all possible states. The encircled state number represents the total number of infected peers in the current state. The expected number of newly infected peers on state j is

$$E[y_{m+1} = j + i | y_m = i] = \sum_{k=0}^{x_m} \binom{x_m}{k} (p_r^m)^k (1 - p_r^m)^{x_m - k} \times k \quad (5.40)$$

Irrespective of the peer selection method, the value of p_r^m is proportional to bandwidth b , the number of infectious peers y_m and is inversely proportional to playback speed v . The number of susceptible peers (x_m) characterizes the chunk's rareness at m -th buffer position. Therefore for the RF algorithm

$$p_r^{m,RF} = \frac{by_mx_m}{N^2v} \quad (5.41)$$

On the other hand, $\frac{y_m}{1+\alpha y_m}$ delineates the MP chunk selection mechanism. Therefore for the MP algorithm

$$p_r^{m,MP} = \frac{by_m^2}{N(1+\alpha y_m)v} \quad (5.42)$$

For comparison, we express Eqs. (5.41, 5.42) as the following

$$p_r^{m,RF} = \kappa \frac{x_m}{N} \quad (5.43)$$

$$p_r^{m,MP} = \kappa \frac{y_m}{1+\alpha y_m} \quad (5.44)$$

where $\kappa = \frac{by_m}{Nv}$. For lower values of m , the part x_m/N of RF algorithm is close to '1', and the part $\frac{y_m}{1+\alpha y_m}$ of MP strategy is much smaller than '1' (as $\alpha > 1$). However, as m increases, more susceptible individuals become infected and x_m/N becomes smaller than $\frac{y_m}{1+\alpha y_m}$. Therefore for lower values of m , $p_r^{m,RF} > p_r^{m,MP}$ which recapitulates the fact that the RF strategy performs better in terms of chunk distribution closer to the newer chunk position while the MP performs better closer to the playback side. Therefore, at optimum switching from RF to MP algorithm

$$p_r^{m_{sw},RF} = p_r^{m_{sw},MP} \quad (5.45)$$

which leads to the same switching position as in the deterministic model defined by Eq. (5.33).

5.3 Simulation

We perform extensive simulations to validate our claims on the behavior and performance of the different chunk selection algorithms. First we describe how we attain results from the different models.

Deterministic model

To simulate the deterministic version of the epidemic model, we vary the buffer occupancy probability (p_m) within '0' to '1' and compute the corresponding buffer position (m) from Eqs. (5.17, 5.23, 5.36, 5.37). This way we avoid deriving an explicit form of the variable y_m . The obvious assumption in this model is that buffer length is continuous. The deterministic model portrays the average fluctuating behavior. We derive the value of $\alpha(= 3)$ by comparing the performance of the deterministic model to that of the simulation model with MP algorithm. Once derived, we keep the same value in all scenarios of the simulation.

Stochastic model

To simulate the stochastic model in Eq. (5.39), a random number $r \in [0, 1]$ is generated at every chunk position m and for each susceptible peer exposed to y_m number of infectious peers. If $r \leq p_r^m$, that susceptible peer becomes infected, otherwise it escapes from being infected. Then we calculate the buffer occupancy distribution p_m

from Eq. (5.1). The value of p_m derived from stochastic model depends on the generated series of random numbers. Average results are taken from multiple stochastic simulations.

Simulation model

We developed a simulator using C++ programming language that realistically modeled the P2P streaming system [154]. We run an event-driven program on top of this simulator. Peers form an overlay network on top of the physical network. A single source progressively generates new chunks with consecutive time-stamps to stream a total of 97.5 MB of video. The video is divided into 390 chunks of 256 KB each. At every time slot, each peer makes an attempt to download a single chunk from one of the potential senders. A sending peer fulfills one or more requests on each time slot depending on its bandwidth capacity. We randomly select a few peers and monitor its buffer over a period of time. Then we enumerate the probabilities of occupancy at each position of the buffer space from the events of ‘buffer position is fill’ over that time. In the following and unless otherwise stated, the population size, $N = 1000$, and the buffer length $M = 40$, and the bandwidth to playback speed ratio, $b/v = 1$.

5.3.1 The accuracy of the model

To investigate the accuracy of the model, we plot the curves representing probabilities of buffer occupancy for RF and MP chunk selection strategies in Fig. 5.4. The stochastic model uses a two-sided 95% confidence interval with the error bar centered on the mean. The region inside the error bar portrays a measure of fluctuation in stochastic model. From the curves in 5.4, it is evident that both the deterministic

FIGURE 5.4: The comparison among deterministic and stochastic versions of epidemic models, and simulation model with RF and MP algorithms. The stochastic version is repeated over ‘50’ times. $N = 1000$, $M = 40$, and $b/v = 1$.

and stochastic version of the epidemic models closely follow the results from the simulation. For example, RF curves from the deterministic and stochastic models (marked by ‘RF-DET’, and ‘RF-STO’ respectively) are next to the RF curve from the simulation model (marked by ‘RF-SIM’). Similar to the MP algorithm, the curves from the deterministic, stochastic and simulation models (marked by ‘MP-DET’, ‘MP-STO’, and ‘MP-SIM’ respectively) are close to one another. The differences between the deterministic epidemic model and simulation model are partly due to the assumption of continuity of chunk positions in the deterministic model.

Next we find the dependency of model accuracy on peer population size. If the variables p_m^d and p_m^s define the buffer occupancy probabilities at m -th buffer position for the deterministic and simulation models respectively, then the expression $\sum_{m=0}^{M-1} |p_m^d - p_m^s|/M$ defines the overall gaps between two models. We compute the

FIGURE 5.5: The gap between different models as a function of peer population size for RF algorithm. The stochastic version is repeated over ‘50’ times. $M = 40$, and $b/v = 1$.

gaps for a different combination of the models with the identical chunk selection algorithm as shown in Figs. 5.5, 5.6. For example, the curve in Fig. 5.5 marked by ‘SIM-DET-Gap-RF’ represents the overall gaps of the deterministic model from the simulation model with RF chunk selection. As expected, the overall gaps decrease as the peer population increases. The overall gaps of the deterministic model from the simulation model with the MP algorithm (marked as ‘SIM-DET-Gap-MP’ in Fig. 5.6) also decrease with an increase in the size of the population. The overall gaps of the stochastic model from the simulation model with both RF and MP algorithms exhibit a similar trend (marked as ‘SIM-STO-Gap-RF’ in Fig. 5.5 and as ‘SIM-STO-Gap-MP’ in Fig. 5.6). The overall gaps of the stochastic model from the deterministic model (marked as ‘DET-STO-Gap-RF’ in Fig. 5.5 and as ‘DET-STO-Gap-MP’ in Fig. 5.6) also decrease with an increased peer population size. Therefore the question of whether the chunk distribution process is best described by the stochastic and

FIGURE 5.6: The gap between different models as a function of peer population size for MP algorithm. The stochastic version is repeated over ‘50’ times. $M = 40$, and $b/v = 1$.

deterministic mechanisms largely depends on the peer population. When the peer population is large, both models exhibit almost equivalent results. The stochastic model is more appropriate for a smaller number of peer population to capture the small variation.

5.3.2 Different chunk selection algorithms

Fig. 5.7 shows the buffer occupancy probabilities with different chunk selection algorithms computed from the simulation model. From the curves it is obvious that the buffer occupancy probabilities are higher for the RF algorithm (marked as ‘RF-SIM’) when compared to the MP algorithm (marked ‘MP-SIM’) at buffer positions holding the new chunks (lower values of m). Therefore the RF chunk selection algorithm prefers newer chunks as they are rarer, whereas the MP chunk selection algorithm prefers older chunks as those are more urgent to be played therefore popular. The

FIGURE 5.7: Comparison between various chunk selection strategies. $N = 1000$, $M = 40$, and $b/v = 1$.

HS (mark as ‘HS-SIM’) attains the highest probability of occupancy at all buffer positions among all chunk selection strategies. The HS clearly enjoys the benefits of both strategies.

5.3.3 Influence of peer population and bandwidth to playback speed ratio on buffer length

Figures 5.8 and 5.9 show the effects of peer population on RF and MP algorithms in the deterministic and simulation models respectively. In the deterministic epidemic model, if the peer population increases from ‘100’ to ‘1000’, the RF algorithm requires $(12.3 - 9.98 =) 2.3$ units and the MP algorithm requires $(25.9 - 19 =) 6.9$ units of additional buffer space to achieve the same level of buffer occupancy probability ($p_m = 0.8$). Therefore the MP algorithm needs $2.94 (\approx \alpha)$ times more buffer space than that of the RF algorithm. In the simulation model, the MP strategy requires 2.67

FIGURE 5.8: The effect of peer population size on RF and MP chunk selection strategy when the bandwidth to playback ratio remains constant ($b/v = 1$) as derived from the deterministic model.

($\approx \alpha$) times more buffer space than that of the RF strategy as the peer population increases from ‘100’ to ‘1000’ (See Fig. 5.9). The results validate our claim in Section 5.2.1. With HS, the additional buffer spaces required are $(11.7 - 9.4 =) 2.3$ and $(12.4 - 9 =) 3.4$ in the deterministic (in Fig. 5.10) and simulation (in Fig. 5.11) models respectively. These amounts are comparable with the RF algorithm and are much less than that of the MP strategy. Therefore HS attains the scalable property of the RF algorithm. Even with huge increases in peer population, the additional buffer requires less amount of increase. The results reassure the scalability of the P2P system.

Figures 5.12 and 5.13 show the effects of bandwidth to playback speed on chunk selection algorithms in the deterministic model. As expected in the deterministic epidemic model with RF algorithm, high $b/v (= 2)$ requires less buffer space (6.18)

FIGURE 5.9: The effect of peer population size on RF and MP chunk selection strategy when the bandwidth to playback ratio remains constant ($b/v = 1$) as measured in simulation model.

and less $b/v (= 1)$ needs more buffer space (12.3) to achieve the same level of buffer occupancy probability ($p_m = 0.8$). With the MP algorithm, if the bandwidth to playback speed ratio decreases (from $b/v = 2$ to $b/v = 1$), the buffer space requirement increases (from 13 to 25.9). Similarly with the HS, if the buffer space requirement increases, the bandwidth to playback speed decreases (as shown in Fig. 5.13). Congruent results are found with the simulation model (See Figs. 5.14, 5.15). Note that if $b/v < 1$, the playback continuity may reside well below the satisfaction level, we reckon that peers with $b/v < 1$ would just leave system.

5.4 Summary

The chunk selection strategy is an important factor that influences the P2P network capacity and the user-perceived visual quality. Utilizing our chunk distribution model,

FIGURE 5.10: The effect of peer population size on HS when the bandwidth to playback speed ratio remains constant ($b/v = 1$) as derived from the deterministic model.

we found the relationships among buffer length, peer population, and playback continuity under the influences of RF and MP algorithms. If the population remains constant, the system needs more buffer space to achieve a higher level of continuity. Therefore the buffer length in the RF strategy is more scalable than that of the MP algorithm to the peer population. The RF chunk selection scheme may overload peers holding rare chunks and peers holding popular chunks could be under-utilized. In case of the MP chunk selection algorithm, a sufficient number of peers held popular chunks (i.e. chunks that are closer to the playback side) and an overloading problem is not a constraint. For a given peer population, the MP strategy attains better continuity than the RF strategy with a sufficiently larger buffer length. The proposed HS takes advantage of both strategies. The HS employs the RF strategy to download newer chunks and then switches to the MP strategy to download the urgent chunks. Finally we performed extensive simulations to validate our claims on the behavior

FIGURE 5.11: The effect of peer population size on HS when the bandwidth to playback speed ratio remains constant ($b/v = 1$) as measured in simulation model.

and performance of the different chunk selection algorithms.

FIGURE 5.12: The effect of bandwidth to playback speed on RF and MP chunk selection strategy with peer population remains constant ($N = 1000$) as derived from the deterministic model.

FIGURE 5.13: The effect of bandwidth to playback speed on HS with peer population remains constant ($N = 1000$) as derived from the deterministic model.

FIGURE 5.14: The effect of bandwidth to playback speed on RF and MP chunk selection strategy with peer population remains constant ($N = 1000$) as measured in simulation model.

FIGURE 5.15: The effect of bandwidth to playback speed on HS with peer population remains constant ($N = 1000$) as measured in simulation model.

Chapter 6

Conclusion and future research

In this chapter, we conclude our thesis work, recap our contributions and also provide future research directions.

6.1 Summary of the work

In this section we list the most important findings of our thesis work. The insights gained are useful towards designing a user driven effective streaming system.

- Operator-controlled D2D communication is a promising integration for LTE-Adv network. When resources are shared among CUEs and DUEs, D2D transmitters cause interference to the primary CUEs during the downlink period and to the eNB during the uplink period. With an efficient shared resource allocation scheme, this feature has offered improved network performance by keeping the interference limited.
- Simulation results showed that the proposed greedy heuristic algorithm can improve network performance in terms of the sum of cell and D2D throughput without causing any significant harm to the primary cellular network for all standard scheduling algorithms of the primary network. The results also showed that the cell, and sum of cell and D2D throughput is higher for the proposed heuristic algorithm than the random algorithm. In a dense D2D network

scenario, although the cell throughput has decreased, the sum of cell and D2D throughput has increased significantly with the proposed algorithm.

- In Chapter 4, we have designed a collaborative streaming system for a cellular system where peers with close proximity form groups and share content amongst them through operator-controlled direct communication. We have proposed and implemented a distributed algorithm to select an appropriate number of broadcasters with limited number of assisted peers to ensure the quality of the streaming content. We have presented the details of the content dissemination technique that schedules the direct download from the Internet and the broadcasting session (Section 4.1).
- Using inter-peer distance distribution in the spatial process, we have derived analytical expressions for the expected number of groups and the expected distance of D2D transmission, two important measures for streaming quality (Section 4.3).
- Through performance evaluation (Section 4.4), we have shown that the proposed distributive agent selection algorithm has performed close to the optimal solution. The proposed system has off-loaded eNB traffic and saved cellular channels by taking advantage of the D2D multicast technique. The proposed streaming system has effectively reduced more and more streaming costs, and showed better and better cost fairness with an increase in peer density. If the agent supports more peers, the number of groups decrease, which results in a higher reduction of eNB loads. However, this may reduce the streaming rate due to increases in the distance from the agent to farthest peer. Therefore there is a trade off opportunity between the amount of off-loaded content and multicast

rate by tuning the maximum number of allowable assisted peers in a group.

- In Chapter 5, we derived an analytical model that captures the chunk distribution process from a single source to the entire network under the influence of different chunk selection strategies. Using epidemic theory, we have analyzed the features and performance of the RF and MP chunk selection algorithms.
- The model captures essential relationships among buffer size, peer population and playback continuity. If the peer population increases, the system requires additional buffer space for both RF and MP strategies to maintain the same level of buffer occupancy probability. However, the additional buffer length required in the MP algorithm case is larger than that of the RF algorithm case. Therefore the buffer length in the RF strategy is more scalable than that of the MP algorithm in terms of the size of the peer population. However for a given peer population, the MP strategy attains a better continuity than the RF strategy for a sufficiently larger buffer length. We have also proposed the HS chunk selection strategy that exploits and takes advantage of both methods. We have demonstrated the accuracy of the analytical models by comparing them to the results obtained through simulation.
- To simulate realistic mobile radio networks, we have carefully chosen simulation parameters. The simulation parameters we used are taken from referred journals and technical reports. The parameters are either supported by real data, or are derived from empirical measurements. In all cases, we provide appropriate citations.

6.2 Thesis contributions

In this section, we address the contributions of our thesis work.

- We have formulated the shared spectrum allocation problem between the cellular network and the underlaying D2D communication network as an MINLP problem. We also solved the spectrum allocation problem by employing a greedy heuristic algorithm that has assigned appropriate resources utilizing channel gain information. Our proposed solution has diminished interference to the primary cellular network and has also showed improved performance in terms of the sum of cell and D2D throughput. The proposed greedy heuristic D2D radio resource allocation algorithm has not required any modification to the scheduling of the primary cellular network; rather, it has utilized information provided by the primary scheduler.
- As a primary goal of our thesis, we have designed a collaborative streaming system for a cellular network. The proposed streaming system has saved wireless bandwidth through D2D multicast mechanism and off-loaded high bandwidth streaming traffic.
- We have formulated the grouping process for collaboration as the stable admission assignment problem. The formulation has captured both topology-aware peer selection and appropriate agent selection problems. Our proposed distributive agent selection algorithm has avoided bottleneck at the base station, achieved an effective multicast rate, and also solved the feedback implosion problem as well. We have also provided the details of the content dissemination technique.

- Through the spatial point process, we have derived the expected number of groups and the expected distance from an agent to the farthest multicast receiver. Using this information, we have shown the tradeoff between the eNB off-load amount and the achievable multicast rate. We have also proved the accuracy of the model through simulations.
- We have derived an analytical model of the chunk distribution process by applying epidemic theory and have evaluated the performance of the streaming system under the influence of different chunk selection strategies. The model has measured the buffer occupancy probability and has shown a relationship among peer population size, buffer length, playback smoothness, playback rate and link bandwidth. Through extensive simulation, we have validated the accuracy of the analytical models. To the best of our knowledge, we are the first to apply the epidemic theory to model the chunk distribution process.
- From the insights we gained from the chunk distribution model, we have proposed a hybrid chunk selection method that has shown scalable property to peer population in terms of buffer space and has achieved better playback continuity by optimal switching between rarest and urgent chunk first strategies.

6.3 Open problem and future Work

There are still a few open problems that need to be addressed to realize a successful collaborative streaming system for cellular users.

- The proposed resource allocation strategy for a D2D enabled network, presented in Chapter 3, has shown throughput improvement as the CUE and DUE sum

rate with a guaranteed target SINR. It would be valuable to perform some analysis and numerical evaluation in terms of outage probability to ensure the effectiveness of the scheme.

- Although we have scheduled a few peers leaving the system and few others joining the system at random times in our simulation, we have not measured the degree of fault tolerance. Measuring the degree of fault tolerance even with concurrent join/failure would be interesting.
- A prolonged viewing time demands that the cellular streaming system be energy efficient. The scope of our proposed streaming system did not include energy constraint in terms of peer's battery life. We acknowledge that peers' collaboration in favor of energy saving is worth investigation.
- In our simulation, we have considered a few users randomly walk at a speed close to 5km/h. D2D communication takes the advantage of short range communication. If the DUEs move very fast, they would not be able to form effective multicast groups. It is worth investigating the impact of mobility in terms of user's QoE.
- Depending on the mobility, carrier frequency and Doppler spread time, the time scale of fast fading is on order of milliseconds [157]. The scheduler may not be capable of tracking the instantaneous CQI which changes very frequently. Therefore, it is impractical, otherwise impossible to implement PF scheduling with instantaneous CQI information even for the LTE fast scheduler. Few works suggested that the scheduler might use the stochastic behavior of the channel instead of the instantaneous CQI [157, 158]. In our proposed shared spectrum

allocation algorithm between cellular and D2D networks, we assumed that the impact of fast fading is averaged out over a certain time period. We did consider the slow fading, the time scale order of which is several seconds and the LTE scheduler is capable of tracking the expected CQI.

- In our design, we have only considered regular cellular users to share their spectrums with D2D users. We did not explore the opportunity of sharing micro- or femtocell's spectrums with the D2D network. Typically, a femtocell provides coverage to a location where cellular coverage is limited, or otherwise unavailable. A microcell provides coverage to a limited area such as a hotel or a mall with high density cellular users. Therefore, it would be more challenging to share spectrums between the D2D network and micro- and femtocell networks.
- We assessed our streaming system with *tit-for-tat* strategy. It would be interesting to see how the streaming system would behave with other incentive mechanisms such as reputation, game theory, and taxation [159].
- We presented results to evaluate the performance of the streaming chunk selection strategy in terms of peer's experience averaged over time. An individual peer may experience less quality over a short period of time than an expected quality averaged over time. This issue is worth further investigation.
- The chunk distribution model, presented in Chapter 5, evaluated streaming quality in terms of playback smoothness. We acknowledge that other parameters, such as delay and jitter, are also important to investigate the impact of chunk allocation strategies. In our design, we assume peer selection to be random. However, our model shows the variation due to peer bandwidth and

playback rate. There is a strong intuitive relationship between the bandwidth and peering decision. More efforts can be made to include the peer selection component in the proposed model.

- The collaborative streaming system relies on D2D communication that is agnostic to the traditional cellular communication. However, we have shown that the system can save a significant amount of cellular bandwidth and gain an effective streaming rate over the traditional system. Operators have to come forward with an open mind to implement the streaming system and thereby make a profit from the early deployment. With exploding number of handheld devices and innovative user-driven applications, we believe that the cellular P2P streaming application would be successful in near future.

References

- [1] Cisco, “Visual networking index: Forecast and methodology: Global mobile data traffic forecast update, 2011-2016,” *White Paper*, 2012.
- [2] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, “Device-to-device communication as an underlay to LTE-advanced networks,” *IEEE Commun. Mag.*, vol. 47, no. 12, pp. 42–49, Dec. 2009.
- [3] Qualcomm, “Flashlinq,” *GSMA Mobile World Congress*, 2012.
- [4] A. Osseiran and A. Gouraud, “Wireless World Initiative New Radio - WINNER+ D2.1: Preliminary WINNER+ system concept,” Tech. Rep., May 2009.
- [5] “SopCast.” [Online]. Available: <http://www.sopcast.com/>
- [6] “Gnutella protocol development.” [Online]. Available: <http://rfe-gnutella.sourceforge.net/index.html/>
- [7] “KaZaA.” [Online]. Available: <http://www.kazaa.com/>
- [8] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A scalable peer-to-peer lookup service for internet applications,” in *Proc. ACM SIGCOMM Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2001, pp. 149–160.
- [9] A. I. T. Rowstron and P. Druschel, “Pastry: Scalable, decentralized object location, and routing for large-scale Peer-to-Peer systems,” in *Proc. IFIP/ACM Int. Conf. Distributed Systems Platforms Heidelberg*, 2001, pp. 329–350.

-
- [10] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in *Proc. ACM SIGCOMM Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2001, pp. 161–172.
 - [11] B. Krishnamurthy and J. Wang, "On network-aware clustering of web clients," in *Proc. ACM SIGCOMM Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2000, pp. 97–110.
 - [12] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Topologically-aware overlay construction and server selection," in *IEEE Int. Conf. Computer Communications (INFOCOM)*, 2002, pp. 1190–1199.
 - [13] X. Y. Zhang, Q. Zhang, Z. Zhang, G. Song, and W. Zhu, "A construction of locality-aware overlay network: mOverlay and its performance," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 1, pp. 18–28, Jan. 2004.
 - [14] C.-M. Huang, T.-H. Hsu, and M.-F. Hsu, "Network-aware P2P file sharing over the wireless mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 1, pp. 204–210, Jan. 2007.
 - [15] J. Li, L. Huang, W. Jia, M. Xiao, and P. Du, "An efficient implementation of file sharing systems on the basis of WiMAX and Wi-Fi," in *Proc. IEEE Int. Conf. Mobile Adhoc and Sensor Systems*, 2006, pp. 819–824.
 - [16] S. Wang and H. Ji, "Realization of topology awareness in peer-to-peer wireless network," in *Int. Conf. Wireless Communications, Networking and Mobile Computing*, 2009, pp. 1–4.

-
- [17] C. Canali, M. E. Renda, P. Santi, and S. Burresi, "Enabling efficient peer-to-peer resource sharing in wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 3, pp. 333–347, Mar. 2010.
 - [18] L. Galluccio, G. Morabito, S. Palazzo, M. Pellegrini, M. E. Renda, and P. Santi, "Georoy: A location-aware enhancement to Viceroy peer-to-peer algorithm," *J. Comput. Netw.*, vol. 51, no. 8, pp. 1998–2014, Jun. 2007.
 - [19] D. Malkhi, M. Naor, and D. Ratajczak, "Viceroy: A scalable and dynamic emulation of the butterfly," in *Proc. Symp. Principles of Distributed Computing*, 2002, pp. 183–192.
 - [20] Y.-J. Joung and J.-C. Wang, "Chord²: A two-layer chord for reducing maintenance overhead via heterogeneity," *Computer Networks*, vol. 51, no. 3, pp. 712–731, 2007.
 - [21] K. Kim and D. Park, "Mobile NodeID based P2P algorithm for the heterogeneous network," in *Int. Conf. Embedded Software and Systems*, 2005, pp. 1–8.
 - [22] S. Zoels, Z. Despotovic, and W. Kellerer, "Load balancing in a hierarchical DHT-based P2P system," in *Int. Conf. Collaborative Computing: Networking, Applications and Worksharing*, 2007, pp. 353–361.
 - [23] T. Locher, S. Schmid, and R. Wattenhofer, "eQuus: A provably robust and locality-aware peer-to-peer system," in *IEEE Int. Conf. Peer-to-Peer Computing*, 2006, pp. 3–11.
 - [24] S. Asaduzzaman, Y. Qiao, and G. Bochmann, "CliqueStream: Creating an

- efficient and resilient transport overlay for peer-to-peer live streaming using a clustered DHT,” *J. Peer-to-Peer Netw. Appl.*, vol. 3, no. 2, pp. 100–114, 2010.
- [25] C. Cramer and T. Fuhrmann, “Proximity neighbor selection for a DHT in wireless multi-hop networks,” in *IEEE Int. Conf. Peer-to-Peer Computing*, 2005, pp. 3–10.
- [26] M. Zulhasnine, C. Huang, and A. Srinivasan, “Topology-aware integration of cellular users into the P2P system,” in *IEEE Vehicular Technology Conf. (VTC), Fall*, 2011, pp. 1–5.
- [27] M.-F. Leung and S. H. G. Chan, “Broadcast-based peer-to-peer collaborative video streaming among mobiles,” *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 350–361, Mar. 2007.
- [28] B. Zhang, S. G. Chan, G. Cheung, and E. Y. Chang, “LocalTree: An efficient algorithm for mobile peer-to-peer live streaming,” in *IEEE Int. Conf. Communications (ICC)*, 2011, pp. 1–5.
- [29] G. Ananthanarayanan, V. N. Padmanabhan, L. Ravindranath, and C. A. Thekkath, “COMBINE: leveraging the power of wireless peers through collaborative downloading,” in *Proc. Int. Conf. Mobile Systems, Applications and Services (MobiSys)*, 2007, pp. 286–298.
- [30] W. Jiang, X. Liao, H. Jin, and Z. Yuan, “WiMA: A novel wireless multicast agent mechanism for live streaming system,” in *Int. Conf. Convergence Information Technology*, 2007, pp. 2467–2472.
- [31] H. Yoon, J. Kim, F. Tan, and R. Hsieh, “On-demand video streaming in mobile

- opportunistic networks,” in *IEEE Int. Conf. Pervasive Computing and Communications (PerCom)*, 2008, pp. 80–89.
- [32] Y. Zhu, W. Zeng, H. Liu, Y. Guo, and S. Mathur, “Supporting video streaming services in infrastructure wireless mesh networks: Architecture and protocols,” in *IEEE Int. Conf. Communications (ICC)*, 2008, pp. 1850–1855.
- [33] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, “Femtocaching: Wireless video content delivery through distributed caching helpers,” in *IEEE Int. Conf. Computer Communications (INFOCOM)*, 2012, pp. 1107–1115.
- [34] K. Piamrat, A. Ksentini, J. M. Bonnin, and C. Viho, “Q-DRAM: QoE-based dynamic rate adaptation mechanism for multicast in wireless networks,” in *IEEE Global Telecommunications Conf. (GLOBECOM)*, 2009, pp. 1–6.
- [35] T. Do, K. Hua, N. Jiang, and F. Liu, “PatchPeer: A scalable video-on-demand streaming system in hybrid wireless mobile peer-to-peer networks,” *J. Peer-to-Peer Netw. Appl.*, vol. 2, no. 3, pp. 182–201, 2009.
- [36] Y. Liu, B. Guo, C. Zhou, and Y. Cheng, “A CDS based cooperative information repair protocol with network coding in wireless networks,” in *IEEE Global Telecommunications Conf. (GLOBECOM)*, 2010, pp. 1–5.
- [37] S. Hua, Y. Guo, Y. Liu, H. Liu, and S. S. Panwar, “Scalable video multicast in hybrid 3g/ad-hoc networks,” *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 402–413, Apr. 2011.

-
- [38] M. Hefeeda, A. Habib, B. Botev, D. Xu, and B. Bhargava, "PROMISE: peer-to-peer media streaming using CollectCast," in *Proc. ACM Int. Conf. on Multimedia*, 2003, pp. 45–54.
- [39] S. M. EIRakabawy and C. Lindemann, "Peer-to-peer file transfer in wireless mesh networks," in *Proc. Wireless on Demand Network System and Services (WONS)*, 2007, pp. 114–121.
- [40] X. Li, H. Ji, R. Zheng, Y. Li, and F. R. Yu, "A novel team-centric peer selection scheme for distributed wireless P2P networks," in *IEEE Wireless Communication and Networking Conf. (WCNC)*, 2009, pp. 1–5.
- [41] Y. Zhang, X. Zhou, F. Bai, and J. Song, "Peer selection for load balancing in 3G cellular networks," in *Int. Conf. Information Science and Engineering (ICISE)*, 2010, pp. 2227–2230.
- [42] M. Zulhasnine, C. Huang, and A. Srinivasan, "Towards an effective integration of cellular users to the structured peer-to-peer network," *J. Peer-to-Peer Netw. Appl.*, vol. 5, no. 2, pp. 178–192, 2012.
- [43] P. Si, F. R. Yu, H. Ji, and V. C. M. Leung, "Distributed multisource transmission in wireless mobile peer-to-peer networks: A restless-bandit approach," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 420–430, Jan. 2010.
- [44] E. Gurses and A. N. Kim, "Maximum utility peer selection for P2P streaming in wireless ad hoc networks," in *IEEE Global Telecommunications Conf. (GLOBECOM)*, 2008, pp. 1–5.
- [45] M. K. H. Yeung and Y.-K. Kwok, "On game theoretic peer selection for resilient

- peer-to-peer media streaming,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 10, pp. 1512–1525, Oct. 2009.
- [46] V. Aggarwal, A. Feldmann, and C. Scheideler, “Can ISPs and P2P users cooperate for improved performance?” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 3, Jul. 2007.
- [47] Z. Duliński, M. Kantor, W. Krzysztofek, R. Stankiewicz, and P. Cholda, “Optimal choice of peers based on BGP information,” in *IEEE Int. Conf. Communications (ICC)*, 2010, pp. 1–6.
- [48] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, “P4P: provider portal for applications,” in *ACM SIGCOMM Comput. Commun. Rev.*, 2008, pp. 351–362.
- [49] C. Wang, N. Wang, M. Howarth, and G. Pavlou, “An adaptive peer selection scheme with dynamic network condition awareness,” in *IEEE Int. Conf. Communications (ICC)*, 2009, pp. 1–5.
- [50] D. R. Choffnes and F. E. Bustamante, “Taming the torrent: A practical approach to reducing cross-ISP traffic in peer-to-peer systems,” *ACM SIGCOMM Comput. Commun. Rev.*, pp. 363–374, 2008.
- [51] A. O. Kaya, M. Chiang, and W. Trappe, “P2P-ISP cooperation: Risks and mitigation in multiple-ISP networks,” in *IEEE Global Telecommunications Conf. (GLOBECOM)*, 2009, pp. 1–8.
- [52] “A & M Records, Inc. v. Napster,” vol. 239 F.3d 1004 (9th Cir.), 2001.
- [53] “PPLive.” [Online]. Available: <http://www.pplive.com/>

-
- [54] X. Zhang, J. Liu, B. Li, and Y. S. P. Yum, "CoolStreaming/DONet: a data-driven overlay network for peer-to-peer live media streaming," in *IEEE Int. Conf. Computer Communications (INFOCOM)*, 2005, pp. 2102–2111.
- [55] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: how much can WiFi deliver?" in *Int. Conf. Emerging Networking EXperiments and Technologies (CoNEXT)*, 2010, pp. 1–12.
- [56] K. Collins, S. Mangold, and G. M. Muntean, "Supporting mobile devices with wireless LAN/MAN in large controlled environments," *IEEE Commun. Mag.*, vol. 48, no. 12, pp. 36–43, Dec. 2010.
- [57] L. Garcia, L. Arnaiz, F. Alvarez, J. M. Menendez, and K. Gruneberg, "Protected seamless content delivery in P2P wireless and wired networks," *IEEE Wireless Commun.*, vol. 16, no. 5, pp. 50–57, Oct. 2009.
- [58] M. S. Corson, R. Laroia, J. Li, V. Park, T. Richardson, and G. Tsirtsis, "Toward proximity-aware internetworking," *IEEE Wireless Comm.*, vol. 17, no. 6, pp. 26–33, Dec. 2010.
- [59] "Nokia research center: Nokia instant community," Tech. Rep., 2010. [Online]. Available: <http://research.nokia.com/news/9391>
- [60] "Apple: Airdrop software," Tech. Rep., 2010. [Online]. Available: <http://www.apple.com>
- [61] T. K. Madsen, Q. Zhang, F. H. P. Fitzek, and M. Katz, "Design and evaluation of IP header compression for cellular-controlled P2P networks," in *IEEE Int. Conf. Communications (ICC)*, 2007, pp. 4046–4051.

-
- [62] K. Doppler, C. B. Ribeiro, and J. Knecht, "Advances in D2D communications: Energy efficient service and device discovery radio," in *Int. Conf. Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE)*, 2011, pp. 1–6.
- [63] 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 Conf. (VTC), Spring*, 2009, pp. 1–5.
- [64] 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 Conf. (VTC), Fall*, 2010, pp. 1–5.
- [65] S. Xu, H. Wang, T. Peng, and Q. Huang, "Effective labeled time slots based D2D transmission in cellular downlink spectrums," in *IEEE Vehicular Technology Conf. (VTC), Spring*, 2010, pp. 1–5.
- [66] 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 Trans. Wireless Commun.*, vol. 10, no. 12, pp. 3995–4000, Dec. 2011.
- [67] H. Wang and X. Chu, "Distance-constrained resource-sharing criteria for device-to-device communications underlaying cellular networks," *Electron. Lett.*, vol. 48, no. 9, pp. 528–530, 2012.
- [68] H. S. Chae, J. Gu, B.-G. Choi, and M. Y. Chung, "Radio resource allocation scheme for device-to-device communication in cellular networks using fractional

- frequency reuse,” in *Asia-Pacific Conference on Communications (APCC)*, 2011, pp. 58–62.
- [69] C.-H. Yu, O. Tirkkonen, K. Doppler, and C. Ribeiro, “Power optimization of device-to-device communication underlaying cellular communication,” in *IEEE Int. Conf. Communications (ICC)*, 2009, pp. 1–5.
- [70] —, “On the performance of device-to-device underlay communication with simple power control,” in *IEEE Vehicular Technology Conf. (VTC), Spring*, 2009, pp. 1–5.
- [71] J. Gu, S. J. Bae, B.-G. Choi, and M. Y. Chung, “Dynamic power control mechanism for interference coordination of device-to-device communication in cellular networks,” in *Int. Conf. Ubiquitous and Future Networks (ICUFN)*, 2011, pp. 71–75.
- [72] H. Xing and S. Hakola, “The investigation of power control schemes for a device-to-device communication integrated into OFDMA cellular system,” in *Int. Symp. Personal Indoor and Mobile Radio Communications (PIMRC)*, 2010, pp. 1775–1780.
- [73] C.-H. Yu, K. Doppler, C. Ribeiro, and O. Tirkkonen, “Performance impact of fading interference to device-to-device communication underlaying cellular networks,” in *IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, 2009, pp. 858–862.
- [74] B. Kaufman, J. Lilleberg, and B. Aazhang, “Spectrum sharing scheme between cellular users and ad-hoc device-to-device users,” *IEEE Trans. Wireless Commun.*, vol. preprint, pp. 1–12, 2013.

-
- [75] 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 Conf. (VTC), Fall*, 2011, pp. 1–5.
- [76] 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 Conf. (VTC), Fall*, 2012, pp. 1–5.
- [77] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.
- [78] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for D2D communications," in *IEEE Global Telecommunications Conf. (GLOBECOM) Workshops*, 2011, pp. 358–362.
- [79] S. Hakola, T. Chen, J. L. andki, and T. Koskela, "Device-to-device (D2D) communication in cellular network - performance analysis of optimum and practical communication mode selection," in *IEEE Wireless Communication and Networking Conf. (WCNC)*, 2010, pp. 1–6.
- [80] T. Koskela, S. Hakola, T. Chen, and J. Lehtomaki, "Clustering concept using device-to-device communication in cellular system," in *IEEE Wireless Communication and Networking Conf. (WCNC)*, 2010, pp. 1–6.
- [81] J. Seppala, T. Koskela, T. Chen, and S. Hakola, "Network controlled device-to-device (D2D) and cluster multicast concept for LTE and LTE-A networks," in *IEEE Wireless Communication and Networking Conf. (WCNC)*, 2011, pp. 986–991.

-
- [82] D. Wang, X. Wang, and Y. Zhao, "An interference coordination scheme for device-to-device multicast in cellular networks," in *IEEE Vehicular Technology Conf. (VTC), Fall*, 2012, pp. 1–5.
- [83] E. Stevens-Navarro and V. W. S. Wong, "Virtual partitioning for connection admission control in cellular/WLAN interworking," in *IEEE Wireless Communication and Networking Conf. (WCNC)*, 2008, pp. 2039–2044.
- [84] W. Song and W. Zhuang, "Multi-service load sharing for resource management in the cellular/WLAN integrated network," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 725–735, Feb. 2009.
- [85] K. Huang, Y. Chen, B. Chen, X. Yang, and V. Lau, "Overlaid cellular and mobile ad hoc networks," in *IEEE Int. Conf. Communication Systems (ICCS)*, 2008, pp. 1560–1564.
- [86] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *Asilomar Conf. Signals, Systems and Computers*, 2008, pp. 1537–1541.
- [87] K. Huang, V. K. N. Lau, and Y. Chen, "Spectrum sharing between cellular and mobile ad hoc networks: transmission-capacity trade-off," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1256–1267, Sep. 2009.
- [88] F. Fitzek, M. Katz, and Q. Zhang, "Cellular controlled short-range communication for cooperative P2P networking," *Wireless Personal Communications*, vol. 48, no. 1, pp. 141–155, 2009.
- [89] Y. Bai, J. Zhou, and L. Chen, "Hybrid spectrum usage for overlaying LTE

- macrocell and femtocell,” in *IEEE Global Telecommunications Conf. (GLOBE-COM)*, 2009, pp. 1642–1647.
- [90] B. Xing, K. Seada, and N. Venkatasubramanian, “An experimental study on Wi-Fi Ad-Hoc mode for mobile device-to-device video delivery,” in *IEEE Int. Conf. Computer Communications (INFOCOM) Workshops*, 2009, pp. 1–6.
- [91] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [92] A. Sahai, R. Tandra, S. M. Mishra, and N. Hoven, “Fundamental design trade-offs in cognitive radio systems,” in *Proc. Int. Workshop Technology and Policy for Accessing Spectrum*, 2006, pp. 1–9.
- [93] T. Yucek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications,” *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 2009.
- [94] M. Zhang, Y. Xiong, Q. Zhang, L. Sun, and S. Yang, “Optimizing the throughput of data-driven peer-to-peer streaming,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 1, pp. 97–110, Jan. 2009.
- [95] B. A. Alghazawy and S. Fujita, “Probabilistic packet scheduling scheme for hybrid pull-push P2P live streaming protocols,” in *Int. Conf. Networking and Computing (ICNC)*, 2011, pp. 248–251.
- [96] Y. Chen, C. Chen, and C. Li, “Measure and model P2P streaming system by buffer bitmap,” in *IEEE Int. Conf. High Performance Computing and Communications (HPCC)*, 2008, pp. 242–249.

-
- [97] Y. Zhou, D.-M. Chiu, and J. C. S. Lui, "A simple model for chunk-scheduling strategies in P2P streaming," *IEEE/ACM Trans. Netw.*, vol. 19, no. 1, pp. 42–54, Feb. 2011.
- [98] G. Kowalski and M. Hefeeda, "Empirical analysis of multi-sender segment transmission algorithms in peer-to-peer streaming," in *IEEE Int. Symp. Multimedia (ISM)*, 2009, pp. 243–250.
- [99] L. Bracciale, F. L. Piccolo, D. Luzzi, S. Salsano, G. Bianchi, and N. Blefari-Melazzi, "A push-based scheduling algorithm for large scale P2P live streaming," in *Int. Telecommunication Networking Workshop on QoS in Multiservice IP Networks (IT-NEWS)*, 2008, pp. 1–7.
- [100] G. Bianchi, N. B. Melazzi, L. Bracciale, F. L. Piccolo, and S. Salsano, "Stream-line: An optimal distribution algorithm for peer-to-peer real-time streaming," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 6, pp. 857–871, Jun. 2010.
- [101] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.
- [102] D. Gusfield and R. W. Irving, *The stable marriage problem : structure and algorithms*. Cambridge, Mass. :: MIT Press, 1989.
- [103] D. G. McVitie and L. B. Wilson, "The stable marriage problem," *ACM Commun.*, vol. 14, no. 7, pp. 486–490, Jul. 1971.
- [104] R. W. Irving, P. Leather, and D. Gusfield, "An efficient algorithm for the "optimal" stable marriage," *J. ACM*, vol. 34, no. 3, pp. 532–543, Jul. 1987.

-
- [105] M. Baïou and M. Balinski, “Many-to-many matching: stable polyandrous polygamy (or polygamous polyandry),” *Discrete Appl. Math.*, vol. 101, no. 1-3, pp. 1–12, Apr. 2000.
 - [106] C.-P. Teo and J. Sethuraman, “The geometry of fractional stable matchings and its applications,” *J. Math. Oper. Res.*, vol. 23, no. 4, pp. 874–891, Nov. 1998.
 - [107] J. Sethuraman, C.-P. Teo, and L. Qian, “Many-to-one stable matching: Geometry and fairness,” *J. Math. Oper. Res.*, vol. 31, no. 3, pp. 581–596, Aug. 2006.
 - [108] J. H. V. Vate, “Linear programming brings marital bliss,” *J. Oper. Res. Lett.*, vol. 8, no. 3, pp. 147–153, 1989.
 - [109] U. G. Rothblum, “Characterization of stable matchings as extreme points of a polytope,” *Math. Program.*, vol. 54, no. 1-3, pp. 57–67, 1992.
 - [110] D. Stoyan, W. Kendall, and J. Mecke, *Stochastic geometry and its applications*. Wiley, New York, 1995.
 - [111] R. Giacomelli, R. K. Ganti, and M. Haenggi, “Outage probability of general ad hoc networks in the high-reliability regime,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 1151–1163, Aug. 2011.
 - [112] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, “A primer on spatial modeling and analysis in wireless networks,” *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 156–163, Nov. 2010.
 - [113] S. Srinivasa and M. Haenggi, “Distance distributions in finite uniformly random

- networks: Theory and applications,” *IEEE Trans. Veh. Technol.*, vol. 59, no. 2, pp. 940–949, Feb. 2010.
- [114] D. Moltchanov, “Distance distributions in random networks,” *J. Ad Hoc Netw.*, vol. 10, no. 6, pp. 1146–1166, 2012.
- [115] M. Haenggi, “Mean interference in hard-core wireless networks,” *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 792–794, Aug. 2011.
- [116] W. H. Hamer, “The Milroy lectures on epidemic disease in England - The evidence of variability and of persistency of type.” *The Lancet*, vol. 167, no. 4307, pp. 733–739, 1906.
- [117] R. Ross, *The prevention of malaria*. London: John Murray, 1911.
- [118] ———, “Some a priori pathometric equations,” *J. Brit. Med.*, pp. 546–547, 1915.
- [119] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proc. Roy. Soc. of London, Series A*, vol. 115, no. 772, pp. 700–721, 1927.
- [120] H. E. Soper, “The interpretation of periodicity in disease prevalence,” *J. Roy. Stat. Soc.*, vol. 92, no. 1, pp. 34–73, 1929.
- [121] H. Abbey, “An examination of the Reed-Frost theory of epidemics,” *J. Human Biol.*, vol. 24, no. 3, pp. 201–233, Sep. 1952.
- [122] J. Maia, “Some mathematical developments on the epidemic theory formulated by Reed and Frost,” *Human Biol.*, vol. 24, no. 3, pp. 167–200, Sep. 1952.

-
- [123] P. E. M. Fine, "A commentary on the mechanical analogue to the Reed-Frost epidemic model," *Am. J. Epidemiol.*, vol. 106, no. 2, pp. 87–100, 1977.
- [124] W. H. Frost, "Some conceptions of epidemics in general," *Am. J. Epidemiol.*, vol. 103, no. 2, pp. 141–151, 1976.
- [125] P. E. Sartwell, "Memoir on the Reed-Frost epidemic theory." *Am. J. Epidemiol.*, vol. 103, no. 2, pp. 138–140, Feb. 1976.
- [126] J. A. Jacquez, "A note on chain-binomial models of epidemic spread: What is wrong with the Reed-Frost formulation?" *Math. Biosci.*, vol. 87, no. 1, pp. 73–82, 1987.
- [127] I. M. L. Jr., "A chain binomial model of endemicity," *Math. Biosci.*, vol. 50, no. 1-2, pp. 85–93, 1980.
- [128] W.-M. Liu, S. A. Levin, and Y. Iwasa, "Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models," *J. Math. Biol.*, vol. 23, no. 2, pp. 187–204, 1986.
- [129] W.-M. Liu, H. W. Hethcote, and S. A. Levin, "Dynamical behavior of epidemiological models with nonlinear incidence rates," *J. Math. Biol.*, vol. 25, no. 4, pp. 359–380, 1987.
- [130] E. B. Wilson and J. Worcester, "The law of mass action in epidemiology - part I," *Proc. Nat. Acad. Sci.*, vol. 31, no. 1, pp. 24–34, 1945.
- [131] ———, "The law of mass action in epidemiology - part II," *Proc. Nat. Acad. Sci.*, vol. 31, no. 1, pp. 109–116, 1945.

-
- [132] 3GPP TR 25.814, "Physical layer aspects for E-UTRA," 2006.
- [133] F. Graziosi and F. Santucci, "A general correlation model for shadow fading in mobile radio systems," *IEEE Commun. Lett.*, vol. 6, no. 3, pp. 102–104, Mar. 2002.
- [134] R. Madan, S. Cui, S. Lal, and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3142–3152, Nov. 2006.
- [135] I. Forkel, M. Schinnenburg, and M. Ang, "Generation of two-dimensional correlated shadowing for mobile radio network simulation," in *Int. Symp. Wireless Personal Multimedia Communication*, 2004, pp. 314–319.
- [136] 3GPP TR 36.942 v1.2.0, "E-UTRA radio frequency (RF) system scenarios (release 8)," 2007.
- [137] M. Zulhasnine, C. Huang, and A. Srinivasan, "Penalty function method for peer selection over wireless mesh network," in *IEEE Vehicular Technology Conf. (VTC), Fall*, 2010, pp. 1–5.
- [138] M. Mehrjoo, S. Moazeni, and X. Shen, "An interior point penalty method for utility maximization problems in OFDMA networks," in *IEEE Int. Conf. Communications (ICC)*, 2009, pp. 1–5.
- [139] P. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the shannon bound," in *IEEE Vehicular Technology Conf. (VTC), Spring*, 2007, pp. 1234–1238.

-
- [140] M. Andrews, *A Survey of Scheduling Theory in Wireless Data Networks*, ser. Wireless Communications. Springer New York, 2007, vol. 143, pp. 1–17.
- [141] H. Kushner and P. Whiting, “Asymptotic properties of proportional-fair sharing algorithms,” in *Allerton Conference*, vol. 3, 2002, pp. 1–10.
- [142] D. Tse, “Multiuser diversity in wireless networks,” 2013. [Online]. Available: <http://www.eecs.berkeley.edu/~dtse/ima810.pdf>
- [143] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, “Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink,” in *IEEE Int. Conf. Computer Communications (INFOCOM)*, 2009, pp. 2611–2615.
- [144] M. Wall, “GAlib: A C++ library of genetic algorithm components,” *Mechanical Engg. Dept., Massachusetts Institute of Technol.*, 1996.
- [145] A. Ephremides, J. E. Wieselthier, and D. J. Baker, “A design concept for reliable mobile radio networks with frequency hopping signaling,” *Proc. IEEE*, vol. 75, no. 1, pp. 56–73, Jan. 1987.
- [146] C. R. Lin and M. Gerla, “Adaptive clustering for mobile wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 15, no. 7, pp. 1265–1275, Sep. 1997.
- [147] S. S. Yau and W. Gao, “Multi-hop clustering based on neighborhood benchmark in mobile ad-hoc networks,” *J. Mob. Netw. Appl.*, vol. 12, no. 5, pp. 381–391, Dec. 2007.
- [148] L. Ramaswamy, B. Gedik, and L. Liu, “A distributed approach to node clustering in decentralized peer-to-peer networks,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 16, no. 9, pp. 814–829, 2005.

-
- [149] G. Chen, F. G. Nocetti, J. S. Gonzalez, and I. Stojmenovic, "Connectivity based k-hop clustering in wireless networks," in *Proc. Annual Hawaii Int. Conf. System Sciences*, 2002, pp. 2450–2459.
- [150] P. Basu, N. Khan, and T. D. C. Little, "A mobility based metric for clustering in mobile ad hoc networks," in *Int. Conf. Distributed Computing Systems Workshop*, 2001, pp. 413–418.
- [151] M. Baïou and M. Balinski, "The stable admissions polytope," *J. Math. Program.*, vol. 87, no. 3, pp. 427–439, 2000.
- [152] E. W. Weisstein, *Concise encyclopedia of mathematics*, 2nd ed. CRC Press, 2003.
- [153] B. Matérn, *Spatial variation*, 2nd ed. New York: Springer Verlag, 1980, vol. 36.
- [154] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," in *IEEE Int. Conf. Wireless And Mobile Computing, Networking and Communications (WiMob)*, 2010, pp. 368–375.
- [155] I. Corporation, "IBM ILOG CPLEX Optimizer," 2012. [Online]. Available: <http://www.ibm.com>
- [156] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Tech. Rep. DEC-TR-301, 1984.
- [157] J. Zhang, S.-C. Liew, and L. Fu, "On fast optimal stdma scheduling over fading wireless channels," in *IEEE INFOCOM 2009*, 2009, pp. 1710–1718.

-
- [158] Q. Sun, Y. Zhang, S. Jin, and X. Gao, "Statistical channel state information aided proportional fair scheduling scheme for highly transmit correlated channels," *Wireless Personal Communications*, pp. 1–23, 2012.
- [159] X. Su and S. K. Dhaliwal, "Incentive mechanisms in P2P media streaming systems," *IEEE Internet Comput.*, vol. 14, no. 5, pp. 74–81, Oct. 2010.