OULU 2012
C 422

ACTA UNIVERSITATIS OULUENSIS

Beatriz Lorenzo Veiga

NEW NETWORK PARADIGMS FOR FUTURE MULTIHOP CELLULAR SYSTEMS

UNIVERSITY OF OULU GRADUATE SCHOOL;
UNIVERSITY OF OULU, FACULTY OF TECHNOLOGY,
DEPARTMENT OF COMMUNICATIONS ENGINEERING;
CENTRE FOR WIRELESS COMMUNICATIONS;
INFOTECH OULU

ACTA UNIVERSITATIS OULUENSIS
C Technica 422

BEATRIZ LORENZO VEIGA

NEW NETWORK PARADIGMS FOR FUTURE MULTIHOP CELLULAR SYSTEMS

Academic dissertation to be presented, with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu, for public defence in OP -sali (Auditorium L10), Linnanmaa, on 28 June 2012, at 12 noon

UNIVERSITY OF OULU, OULU 2012

Acta Univ. Oul. C 422, 2012

Supervised by
Professor Savo Glisic

Reviewed by
Professor Allen B. MacKenzie
Professor Luis M. Correia

ISBN 978-951-42-9854-7 (Paperback)
ISBN 978-951-42-9855-4 (PDF)

ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)

Cover Design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2012

Lorenzo Veiga, Beatriz, New network paradigms for future multihop cellular systems.
University of Oulu Graduate School; University of Oulu, Faculty of Technology, Department of Communications Engineering; Centre for Wireless Communications; Infotech Oulu, P.O. Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. C 422, 2012
Oulu, Finland

Abstract

The high increase in traffic and data rate for future generations of mobile communication systems, with simultaneous requirement for reduced power consumption, makes Multihop Cellular Networks (MCNs) an attractive technology. To exploit the potentials of MCNs a number of new network paradigms are proposed in this thesis.

First, a new algorithm for efficient relaying topology control is presented to jointly optimize the relaying topology, routing and scheduling resulting in a two dimensional or space time routing protocol. The algorithm is aware of intercell interference (ICI), and requires coordinated action between the cells to jointly choose the relaying topology and scheduling to minimize the system performance degradation due to ICI. This framework is extended to include the optimization of power control. Both conventional and cooperative relaying schemes are considered.

In addition, a novel sequential genetic algorithm (SGA) is proposed as a heuristic approximation to reconfigure the optimum relaying topology as the network traffic changes. Network coding is used to combine the uplink and downlink transmissions, and incorporate it into the optimum bidirectional relaying with ICI awareness.

Seeking for a more tractable network model to effectively use context awareness and relying on the latest results on network information theory, we apply a hexagonal tessellation for inner partition of the cell into smaller subcells of radius r. By using only one single topology control parameter (r), we jointly optimize routing, scheduling and power control to obtain the optimum trade-off between throughput, delay and power consumption in multicast MCNs. This model enables high resolution optimization and motivates the further study of network protocols for MCNs. A new concept for route discovery protocols is developed and the trade-off between cooperative diversity and spatial reuse is analyzed by using this model.

Finally, a new architecture for MCN is considered where multihop transmissions are performed by a Delay Tolerant Network, and new solutions to enhance the performance of multicast applications for multimedia content delivery are presented.

Numerical results have shown that the algorithms suggested in this thesis provide significant improvement with respect to the existing results, and are expected to have significant impact in the analysis and design of future cellular networks.

Keywords: cooperative diversity, dynamic traffic distribution, intercell interference, multicast, multihop cellular network, network optimization, reuse factor, routing, scheduling, topology control

Lorenzo Veiga, Beatriz, Tulevaisuuden monihyppyisten matkapuhelinjärjestelmien uudet paradigmat.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Teknillinen tiedekunta, Tietoliikennetekniikan osasto; Centre for Wireless Communications, PL 4500, 90014 Oulun yliopisto; Infotech Oulu, PL 4500, 90014 Oulun yliopisto
Acta Univ. Oul. C 422, 2012
Oulu

Tiivistelmä

Tiedonsiirron ja tiedonsiirtonopeuksien suuri kasvu sekä tehonkulutuksen pieneneminen tulevien sukupolvien matkapuhelinjärjestelmissä tekevät monihyppyiset matkapuhelinverkot houkuttteleviksi vaihtoehdoiksi. Tässä työssä esitetään uusia tiedonsiirtoverkkojen paradigmoja monihyppyisten matkapuhelinverkkojen hyödyntämiseksi.

Työssä esitellään uusi algoritmi tehokkaaseen releointitopologian hallintaan, joka optimoi yhtäaikaisesti topologian, reitityksen sekä lähetyshetkien ajoituksen ja mahdollistaa tila-aika-reititysprotokollan toteutuksen. Esitetty algoritmi huomioi solujen keskinäishäiriön ja vaaditulla solujen välisellä koordinoidulla hallinnalla saadaan yhdessä valittua topologia ja ajoitus, jotka minimoivat solujen keskinäisistä häiriöistä johtuvan suorituskyvyn heikentymisen. Myöhemmin tätä viitekehystä on laajennettu lisäämällä siihen tehonsäädön optimointi. Työssä on tutkittu sekä perinteisiä että kooperatiivisia releointimenetelmiä.

Lisäksi työssä esitetään uusi geneettinen algoritmi heuristiseksi approksimaatioksi verkon liikenteen muutoksen vaatimaan releointitopologian uudelleen järjestelyyn. Työssä tarkastellaan lisäksi verkkokoodausta ylä- ja alasuuntaan tapahtuvan tiedonsiirron yhdistämiseksi sisällyttämällä se solujen keskinäishäiriön huomioivaan kahdensuuntaiseen releointiin.

Etsittäessä paremmin mukautuvaa ja kontekstitietoisuutta hyödyntävää verkkomallia, joka käyttää hyväkseen viimeisimpiä verkkojen informaatioteoreettisia tuloksia, voidaan verkon solut pilkkoa pienempiin kuusikulmaisiin alisoluihin. Käyttämällä ainoastaan näiden alisolujen sädettä r voidaan puolestaan verkon reititys, ajoitus ja tehon säätö optimoida yhtäaikaisesti saavuttaen paras mahdollinen kompromissi verkon läpäisyn, viiveen ja tehonkulutuksen välillä. Kehitetty malli mahdollistaa korkean resoluution optimoinnin ja motivoi uusien verkkoprotokollien kehitystä monihyppyisissä matkapuhelinverkoissa. Tätä mallia käyttäen esitellään myös uusi konsepti reitinetsintäprotokollille sekä analysoidaan kooperatiivisen diversiteetin ja tila-avaruudessa tapahtuvan uudelleenkäytön välistä kompromissiratkaisua.

Lopuksi työssä tarkastellaan monihyppyisen matkapuhelinverkon uutta arkkitehtuuria, jossa monihyppylähetykset suoritetaan viivesietoisella verkolla ja esitetään uusia ratkaisuja multimediasisällön monilähetysten tehokkuuden parantamiseksi.

Työssä saadut tulokset osoittavat, että ehdotetut algoritmit parantavat järjestelmien suorituskykyä verrattuna aiemmin tiedossa olleisiin tuloksiin. Työn tuloksilla voidaan olettaa myös olevan suuri vaikutus tulevaisuuden matkapuhelinverkkojen analysointiin ja suunnitteluun.

Asiasanat: dynaaminen liikenteen jakauma, kooperatiivinen diversiteetti, monihyppyinen matkapuhelinverkko, monilähetys, reititys, topologiakontrolli, uudelleenkäyttökerroin, verkon optimointi

To my family

Preface

The research work presented in this thesis was developed at the Center for Wireless Communications (CWC), University of Oulu, Finland, during the years 2008-2012. First of all, I would like to thank my supervisor, Professor Savo Glisic, and the director of CWC, Lic. Tech. Ari Pouttu for giving me the opportunity to work in such inspiring and highly professional research unit.

The support, patience, encouragement and guidance from my supervisor Professor Savo Glisic over the years have been invaluable and I wish to thank him for that. This cooperation will have large impact on my future professional career.

I would like to express my gratitude to Professor Matti Latva-aho and Pentti Leppänen, Heads of the Department of Communication Engineering during these years, for creating efficient environment for my work in the department.

I am grateful to my reviewers Professor Allen B. MacKenzie (Virginia Tech, USA) and Professor Luis M. Correia (Technical University of Lisbon, Portugal) for their thorough examination of the dissertation as well as to my opponents Professor Luis DaSilva (Virginia Tech, USA) and Professor Alhussein Abouzeid (Rensselaer Polytechnic Institute, USA) in my public defense.

The finantial support for this work was provided by the Finnish Funding Agency for Technology and Innovation (Tekes), the Academy of Finland, Nokia, Nokia Siemens Networks, Elektrobit and Infotech Oulu Graduate School. I was privileged to receive personal grants for Doctoral studies from the following Finnish foundations: Tauno Tönning, Nokia, Finnish Cultural foundation and Riita ja Jorma J. Takasen foundation. These acknowlegments encourage me to go on with my research work and they are gratefully recognized.

I would also like to thank Professor Markku Juntti for his support within Infotech and Professor Jari Iinatti for his support within my participation in teaching.

I want to thank Juha-Pekka Mäkelä for preparing the Finnish translation of the thesis abstract and his help in many aspects of my work.

Many thanks to the whole CWC staff for providing me a friendly working environment. My special thanks go to my friends Emmi Kaivanto, Maria Kangas and Mariella Sarestoniemi for the good moments that I have shared with them.

I extend my special appreciation to the administrative personnel of the CWC. I am indeed grateful to Eija Pajunen, Kirsi Ojutkangas, Elina Komminaho, Haana Saarela, Timo Äikäs, Jari Sillanpää, Antero Kangas and Vaili Jamsa for their invaluable help.

I want to express my unreserved gratitude to my kummi family Kari, Ritva and Sonja Heljasvaara for their support, kindness and help during these years.

I am deeply indebted to my loved parents, Miguel and Puri, for their love and support through my life. To them I owe, among many other things, my deep interest for this profession. Special thanks to my sister Blanca for being always there for me. I would like to thank my grand-mother, grand-ants and cousins. Without the love and support of my family, I would never imagine what I have achieved.

Finally, I give a special thank you to all my friends from Spain.

Oulu, May 17, 2012 Beatriz Lorenzo Veiga

List of symbols and abbreviations

$\mathbf{A}$	Availability matrix
$\mathbf{a}$	Access vector
$\mathbf{a'}$	Differential access vector
$\mathbf{a}_{\mathrm{I}}$	Access vector corresponding to the intial traffic
$\mathbf{a}_{\mathrm{F}}$	Access vector after the traffic has changed
AP_i	Interfering Access Point
AP_r	Reference Access Point
b	Time slot index
B	Length of the scheduling cycle
B_1	Number of time slots for the initial optimum topology
B'	Number of time slots for the optimum topology associated to the differential access vector $\mathbf{a'}$
$\mathbf{c}$	Capacity vector
c_l	Capacity of link l
$c_{\mathfrak{R}}$	Capacity of the route $\mathfrak{R}$
$\mathbf{D}$	Destination matrix
D	Number of destinations
d_{CD}	Channel defading distance
d_i	Interfering distance
$d_{i,,j}$	Distance between user i and user j
d_0	Unit distance
d_r	Relaying distance
D^x	Infection rate of destination infected by packet x
$\mathcal{D}$	Set of all destination users
$E[T]$	Average of packet delivery delay
$\mathbf{f}$	Vector of flow rates
f_l	Flow on link l
$\mathcal{F}$	Set of contention based schemes
g_m	Multicast gain
$g(\boldsymbol{\mu})$	Lagrange dual function
G	Graph
$G_{i,j}$	Channel gain between user i and user j
G_{L^f}	Number of times a packet is copied in its entire lifetime
G_{T^f}	Number of times a packet is copied at the time of delivery
h	Hop index

h_l	Hop distance
H	Maximum number of hops
$\mathbf{I}$	Identity matrix
$\mathrm{I}_{i,j}$	Interference power at the position of the reference receiver j due to the cochannel interfering signal transmitted by i
irf	Intercell reuse factor
$\mathcal{I}$	Set of base stations
l	Link index
L	Number of links
L^f	Lifetime of packet f
$\mathcal{L}$	Set of links
$\mathcal{L}(b)$	Set of links actived in slot b
$\mathcal{L}(n)$	Set of links used by node n
M	Maximum number of sessions in the network
m	Mobile user
$m(h,\theta)$	User location in polar coordinates
m_d	Mobile destination
m_i	Mobile user located in reference cell i
m_r	Mobile user located in reference cell r
$m^{\Gamma}(h,\theta)$	Location in polar coordinates of the user belonging to cluster Γ
$\mathcal{M}$	Set of intermediate users
N	Number of users in the network
N_c	Number of cells
N_{es}	Number of topologies generated by exhaustive search
N_{ga}	Number of topologies generated by genetic algorithm
N_s	Number simultaneous transmissions
n	Node
n_m	Number of mutations
$\mathcal{N}$	Set of nodes
$\mathcal{N}(l)$	Set of nodes using link l
P	Transmission power
p	Protocol index
$P_{\max}$	Maximum power
$P_{\min}$	Minimum power
$P_{i,j}$	Transmision power from user i to user j
$p_r(t)$	Probability of recovery from infection
PT^b	Partial topology in slot b

$\mathscr{P}$	Set of multicast protocols
r	Subcell radius in the tessellation scheme
R	Cell radius
$\mathbf{R}$	Routing matrix
$\mathbf{R}(b)$	Partial routing matrix in slot b
R^x	Recovery rate from packet x
$\mathcal{R}$	Set of relaying users
$\Re^{(2)}$	Set of two dimensional relaying topology
S	Cell area
$\mathbf{T}$	Topology matrix
T_D	Packet delivery delay for D destinations
Thr	Throughput
$\mathcal{T}^b$	Set of candidate users available to transmit in slot b
$\Im^{(2)}$	Set of two dimensional topologies
$\mathfrak{T}$	Topology submatrix
U	Utility function
$U^{(down)}$	Utility function on the downlink
$U^{(up)}$	Utility function on the uplink
v	Terminal speed
w_1, w_2	Weights of the optimization
$\mathbf{x}$	Transmission rate vector
$\mathbf{x}^{(2)}$	Extended rate vector
x_n	Rate of source n
z_j	Resource allocation at network element j
$\mathcal{Z}$	Set of physical layer resource allocation schemes

α	Propagation constant
$\boldsymbol{\alpha}(\alpha_{h,\theta})$	Routing matrix in polar coordinates
$\boldsymbol{\beta}$	Spatial user distribution matrix
γ	Gene index
$\boldsymbol{\Gamma}$	Clustering factor
$\boldsymbol{\Delta}$	Variation of the traffic in the network
$\boldsymbol{\delta}$	Overall downlink network traffic
ε_e	Energy efficiency
ε_t	Time efficiency
θ	Location of the user in polar coordinates

$\Theta^{(h)}$	Set of angles for the users located in hop h
$\boldsymbol{\lambda}$	Overall uplink network traffic
λ_m	Arrival rate of user m
μ	Lagrange multiplier
ξ	Diversity order
Π	Scheduling set
φ	Location of the interfering user in polar coordinates
ρ_u	Density of users
$\boldsymbol{\Psi}^b$	Transmission matrix in slot b
υ	Timer
Ω_d	Broadcast directivity

2G	2nd Generation
3G	3rd Generation
4G	4th Generation
5G	5th Generation
AODV	Ad hoc On-Demand Distance Vector
AP	Access Point
BS	Base Station
CD	Channel Defading
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CONR	CONventional Relaying
COOR	COOperative Relaying
CPU	Central Processing Unit
DCM	Destination Cooperative Multicast
DFRP	Directed Flooding Routing Protocol
DNCM	Destination Non-Cooperative Multicast
DSL	Digital Subscriber Lines
DSR	Dynamic Source Routing
DTN	Delay Tolerant Network
DVBH	Digital Video Broadcast–Handheld
FFR	Fractional Frequency Reuse
GA	Genetic Algorithm
I^2M	Intercell Interference Management
ICFC	Intercell Floofing Coordination

ICI	Intercell Interference
IEEE	Institute of Electrical and Electronics Engineers
IFNC	Inter Flooding Network Coding
InSyNet	Inter System Networking
LAR	Load Aware Routing
LTE	Long Term Evolution
MAC	Medium Access Control
MANET	Mobile Ad hoc NETwork
MBMS	Multimedia Broadcast Multicast Services
MCN	Multihop Cellular Network
MIMO	Multiple Input Multiple Output
MRC	Maximum Ratio Combining
MSN	Mobile Sensor Network
NP-hard	Non-deterministic Polynomial-time hard
NRDP	Nano Route Discovery Protocol
NSCM	Nano Scale Channel Model
NSNM	Nano Scale Network Model
NUM	Network Utility Maximization
ODE	Ordinary Differential Equations
OFDMA	Orthogonal Frequency Division Multiple Access
PER	Polymorphic Epidemic Routing
PF	Proportional Fairness
QoS	Quality of Service
RREQ	Route REQuest
RRES	Route RESponse
RRM	Radio Resource Management
SAPR	Shortest Available Path Routing
SCF	Store-Carry and Forward
SCN	Single-hop Cellular Networks
SDMA	Space Division Multiple Access
SGA	Sequential Genetic Algorithm
SI	Spatial Interleaving
SINR	Signal-to-Interference-plus-Noise Ratio
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
$(TC)^2$	Traffic Cognitive Topology Control
TCN	Traffic Cognitive Network

TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TSL	Topology Search program
TTL	Time-To-Live
UMTS	Universal Mobile Telecommunications System
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

Contents

1 Introduction

1.1 Motivation

Multihop cellular networks (MCNs) are proposed in respond to the demand for next generation cellular systems to support high data rates with efficient power consumption, enlarge coverage area and provide good QoS for multimedia applications [1], [2]. Simultaneous need of increasing the capacity and reducing the power will require more spatial reuse. One technique under consideration to achieve this goal is the deployment of small cells. By scaling down the cell size and so increasing the total number of channels in space, the network capacity can be linearly increased, proportional to the number of new base stations (BSs) or the scaling factor. However, the deployment of more BSs and their interconnections to the wired backbone results in high network cost. This problem can be overcome by deploying wireless multihop routers instead of new BSs or allowing selected mobile terminals to act as routers, to establish a wireless MCN. In this way, by shortening the links, the required transmit power is reduced which is highly desirable in interference-limited networks and provides the opportunity for capacity increase when suitable techniques are applied.

MCNs are economically convenient due to the capability of providing faster deployment by using the existing infrastructure of cellular networks. Different architectures based on 2G, 3G and WiMAX can coexist and different types of networks such as femtocells, delay tolerant networks, WLANs might be used as an augmented technology.

The concept of adding ad hoc capabilities to cellular nodes is widely explored in MCNs [1], [2]. The advantages of this *hybrid* architecture include increasing the throughput of the network, enlarging the coverage area of the base station, decreasing the power consumption of the mobile users, and increasing the network scalability. In order to exploit those advantages, the selection of the most appropriate relays among the existing mobile terminals [3] should be jointly considered with routing and scheduling. The throughput on each hop and opportunity for spatial reuse increases with the number of hops, but the complexity of the system also increases. Consequently, a large number of possibilities results in a large scale optimization problem. Additionally, the delay from source to destination is increased with more hops, which may not be tolerated by delay-sensitive services. The above problem becomes more complex

in the multicell scenario where intercell interference (ICI) is present. Thus, to exploit the potentials of MCN, a systematic approach to network optimization is needed to study the gains and trade-offs associated with this type of networks.

A number of radio resource management (RRM) schemes, such as relay selection and radio resource partition, along with a number of routing and topology control algorithms have been proposed for ad hoc networks [4]. Some of the earliest information theoretic work by Cover and El Gamal [5] gave capacity bounds for the simple relay channel, while more recent work by Gupta and Kumar [6] expanded this work to give asymptotic results for general ad hoc relay networks. However the problems associated with this type of networks are different from those of cellular networks, so the results are not directly applicable to cellular multihop relay scenarios.

A number of potential opportunities and challenges are related to MCNs. To take advantage of such potentials, it is necessary to overcome important technological challenges, such as the design and joint optimization of robust, adaptive and context aware multihop routing protocols, as well as scheduling and energy efficient radio resource allocation. Different architectures, protocols, and analytical models for MCNs have been proposed in the literature where different system aspects were investigated. This chapter aims to provide a survey of the major research issues and challenges in MCNs.

1.2 Research challenges in multihop cellular networks

1.2.1 Overview

In this section, some of the most important research issues in MCNs are summarized, and in each subsection the main solutions for the introduced problems are presented.

The architecture of MCNs consists of cellular and ad hoc relaying components as shown in Fig. 1. In such *hybrid* network architecture, MCNs combines the benefits of single-hop cellular networks (SCNs) and ad hoc networks. The SCNs have reliable performance and mature technology support. However, their infrastructure is costly to build and suffers from some limitations on the channel data rate when the number of mobile users is high or there is heavy traffic during peak hours. They also have limitations on system capacity and network expansion. On the other hand, ad hoc networks are cheap to deploy but

channel contention and interference between nodes are more difficult to predict and control, and the end-to-end paths between source and destination are more vulnerable to node mobility and failure. To preserve the advantages and cope with the limitations of both networks when operating standalone, a number of factors should be taken into account for designing a MCN.

The most important factors are multihop routing, topology control, the design of RRM protocols, particularly for the management of the ICI, and load balancing schemes. These factors are closely inter-related and affect power consumption, capacity, coverage and QoS provisioning.

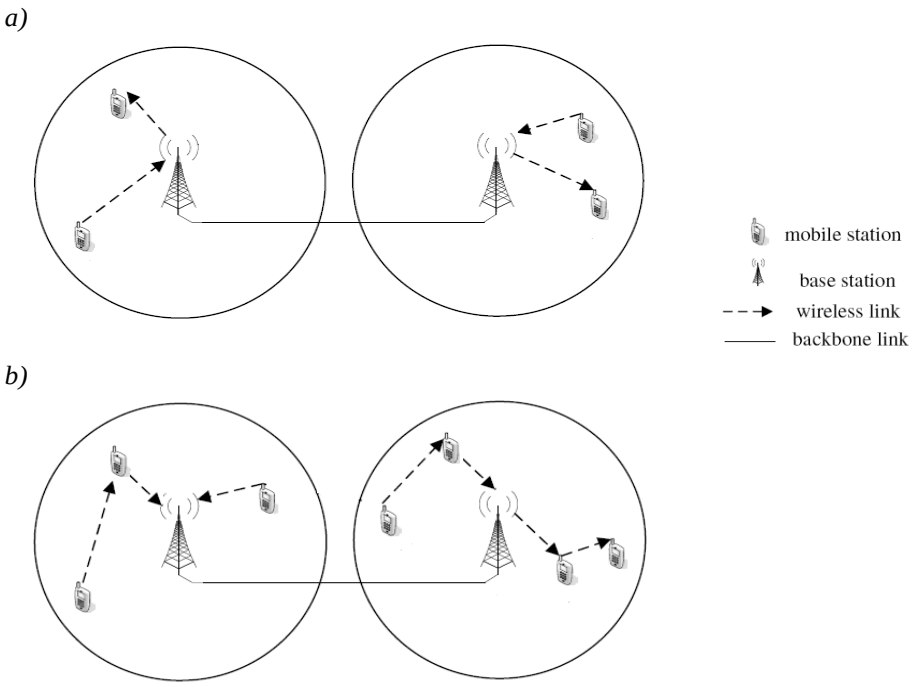

Fig. 1. a) Single hop cellular network; b) Multihop cellular network.

In Fig. 2 we represent the protocol stack indicating where each of those functions is located. A number of papers have shown that by exploiting useful interactions of protocols in different layers, the network performance can be improved significantly (cross-layer optimization). For example, the coordination between

routing and resource scheduling in MCNs is crucial and warrants careful investigation [7]. A cross-layer routing protocol with constrains in relay node selection and source to destination path selection is proposed in [8] for a single cell scenario. In [9] a cross-layer throughput analysis is presented for fixed topologies and without optimization of the power allocation. The jointly optimization of ICI avoidance and load balancing schemes in a multicell network is addressed in [10].

The above examples are just few, from a vast variety of issues addressed in the literature of MCNs. In the remainder of this section, each of those main problems is discussed in detail in separate subsections to bring more insight in their impact on overall characteristics of MCNs.

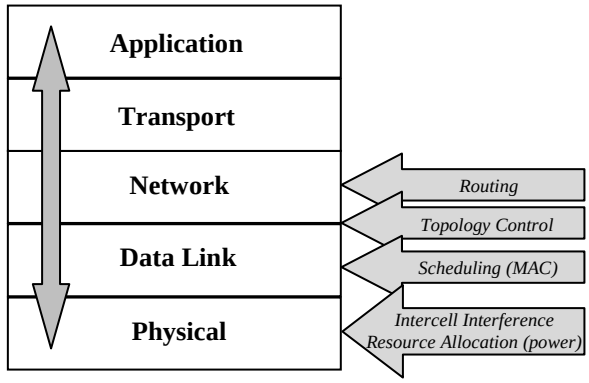

Fig. 2. Protocol stack illustrating different design decision factors in MCNs.

1.2.2 Multihop routing

The relaying technology has been studied intensively for applications in MCNs and is included in most third- and fourth-generation wireless system developments and standardizations [11]-[12]. The relay channel was introduced in [13] by assuming that there is a source that wants to transmit information to a single destination and a relay terminal that is able to help the destination (*relay-assisted transmission*). The relaying concept is the basis of multihop routing and cooperative transmission too [14].

Routing is a major issue in MCNs because it affects packet delay and system throughput. In mobile ad hoc network (MANET) many routing algorithms have been proposed [15]. These algorithms are designed with network infrastructure nonexistence in mind, and their main objective is to establish/maintain network connectivity, rather than to maximize system capacity. As a result, these algorithms are not suitable for MCNs.

A routing algorithm in MCN introduces extra signaling overhead when broadcasting route information which adds extra interference. The effect of the interference is normally ignored in MANETs but cannot be neglected in cellular networks. This is mainly because the transmission power of nodes in MCNs can be several orders of magnitude higher than that of nodes in MANETs. In both MANETs and MCNs, the amount of signaling overhead mainly depends on the chosen routing algorithm. The routing algorithms can generally be classified into two categories: a) *proactive* routing and b) *reactive* routing [16], [17]. Proactive routing mechanisms discover and calculate routes all the time. Each node periodically exchanges its routing information with its neighbors by continuously broadcasting hello/topology messages, and thus, its signaling overhead depends on the broadcasting interval and the number of nodes in the network. On the other hand, reactive routing schemes find and maintain routes only when needed. The signaling overhead of reactive routing increases with the increasing number of active communication pairs as well as with the number of nodes [16], [17]. In MCNs, the radio resources are centrally controlled, and thus, a mobile terminal has to establish a connection with the BS before data is transmitted. In such an environment, reactive routing offers several advantages over proactive routing.

First, reactive routing produces less signaling overhead, as there is no routing unless data transmission is required. Second, reactive routing only maintains necessary routing entries. Most of the routing entries maintained by proactive routing could be obsolete due to discontinuous reception (DRX) [18] or users' mobility. In reactive routing, a source node normally utilizes flooding to deliver a route request (RREQ) packet to the destination. Once an RREQ reaches its destination, the destination reports a route response (RRES) back to the source along the nodes that the RREQ has traversed. In the case when multiple RREQs are received, the route with the best performance metric would be reported.

During the route-discovery phase, the RREQ can be broadcast to the entire network (*i.e.*, complete flooding) or a certain part of it (*i.e.*, directed flooding). For example, dynamic source routing (DSR) [19] utilizes complete flooding to find a route to its destination if a source cannot reach the destination in a single

hop. In contrast, the Ad-hoc On-Demand Distance Vector (AODV) routing protocol [20] uses incremental scoped flooding to find a route. A source gradually enlarges the flooding diameter until it finds a destination or the search diameter reaches a predefined "time-to-live (TTL)" threshold (*i.e.*, the maximum number of relay nodes in the routing path). AODV should use complete flooding if no route is found when the search diameter hits the threshold. The drawback of the proactive routing is the delay in the data transmission.

It seems that for MCNs that enable DRX the reactive routing approach would be a better choice. Hence, the existing routing protocols proposed for MCNs normally adopt DSR to discover the best route. Some routing protocols utilize a scoped flooding approach to reduce the signaling overhead of DSR. For example, Choi and Cho [21] proposed an inhibit access control method that utilized the path loss (or, equivalently, distance) to eliminate useless forwarding participants. In [18] TTL threshold is used to limit the search diameter of each RREQ. Generally, the TTL threshold can be derived based on the given system level constraints of MCNs. For example, the TTL threshold may depend on the maximum intracell interference [22], the end-to-end delay requirement of the multihop transmission [23], the maximum route discovery time [21], or the performance metric of the routing protocol [18].

When designing a routing protocol, the control strategy and path selection metric (cost function) need to be defined. As MCNs contain coordinators (BSs or APs) and mobile users, routing control may be *centralized*, *de-centralized*, or *hybrid*. In centralized routing, BSs are responsible for route discovery and maintenance. BSs have unlimited power supply and high computational power which helps to avoid consuming the limited battery power of mobile nodes for route information exchange and route computation. In CAHAN [24], a central controller periodically receives the location information from each user in the cell to determine the route of the *ad hoc* subnet (cluster) heads with which mobile users communicate. However, when mobiles are outside of the maximum transmission range of a BS or an AP, a decentralized (distributed) routing scheme, such as DSR, is desirable. Some MCN proposals employ distributed routing schemes. For example, in mobile-assisted data forwarding (MADF) [25], mobile nodes may be willing to relay data packets based on their local traffic condition. If the traffic is less than a certain threshold, they broadcast a message to their neighboring mobile nodes indicating that they have available channels for relaying data packets. Then, a mobile node in a congested cell chooses a relaying

node to relay its data packets to a less congested neighboring cell based on the link quality between itself and the relaying node and estimated packet delay.

In MCNs, a hybrid routing approach is commonly used. Route control is shared by the BS and mobile users. For example, in cellular based routing (CBR) [26] and cellular based source routing (CBSR) [27], mobile nodes collect information about the neighborhood and send it to the BS for route computation. This helps reduce the route computation overhead at relaying nodes. In addition, not only source node can initiate a relaying request, a relaying node can also take the initiative by advertising their free channels (available capacity) for relaying [25, 26, 28]. Hence, routing overhead is shared amongst source nodes and relaying nodes.

Different routing protocols consider different path selection metrics. Metrics include BS reachability, hop count, path loss, link quality, signal strength, bit error rate (BER), carrier-to-interference ratio (C/I), delay-sensitivity, throughput, power, battery level, mobile speed, and energy consumption. If BS reachability information is available e.g., provided by relaying nodes, mobile nodes can select the best next hop relaying node to reach the BS. Limiting the number of hops helps bound the packet delay, but reduces the chance of obtaining relaying paths, and, hence, the reachability. This can be overcome by using topology control as it will be explained in the next subsection. Nevertheless, choosing paths based on the smallest number of hops also raises fairness and energy efficiency issues [29], [30].

Several routing algorithms have been proposed for MCNs based on e.g., location [31], path-loss [32], transmission-power [33], and congestion [34]. In [35] the relay station overload problem is considered in the route selection protocol. But in these approaches the selected routes are not necessarily optimal in terms of the system resource utilization and the signaling overhead was ignored. Link quality may be expressed as a function of path loss, BER, and C/I. Delay and throughput are common metrics because they reflect the network performance directly. Minimum power routing is important in CDMA-based MCNs to reduce interference and achieve high cell capacity. Battery level, mobile speed, energy consumption are useful for assuring the reliability of relaying paths. Other possible metrics include traffic load, mean queue length, and number of packets queued along the path.

Joint routing and resource management schemes

The coordination between routing and resource scheduling in MCNs is crucial due to the strong interdependency between the two functions. In MCNs, multihop transmissions normally consume less system capacity if routing is appropriately performed, and the radio resource scheduler should promptly capture the saved resources and assign them to others who suffer a deficit. Consequently, radio resource scheduling should be based on the results of routing. On the other hand, radio resource scheduling affects the system interference/loading pattern, which in turn might affect the decisions of user route selection. Performing joint routing and scheduling is known to produce superior performance results, as compared with decoupled scheduling and routing [36]. The optimal radio resource allocation problem in MCNs, with the objective of throughput maximization, is proven to be NP-hard [37]. So, it is quite challenging to devise efficient RRM schemes that tackle the joint problem.

The strategy for effective coordination of routing and packet scheduling in packet-based MCNs is addressed in [38], and a heuristic algorithm, named integrated radio resource allocation (IRRA) algorithm is proposed to find suboptimal solutions. In [39] the optimal placement of relay nodes and the time allocation were studied for the system employing one relay in a cell with uneven traffic distribution.

Several existing routing algorithms proposed in the literature aim to minimize total transmission power or maximize the transmission rate on each routing path while ignoring interference due to concurrent transmissions on different hops and among different routing paths [40]. When the effects of interference are not considered, the optimum routing path and/or optimum number of hops can usually be found given high node density. These achievable capacity gains are, however, very optimistic and much higher than what could be achieved in real networks. When both intracell and intercell interference as well as self-interference on each routing path are taken into account, there is a tight coupling between the interest in high spatial reuse for efficient radio resource consumption and the interference level in the network [39, 41]. In fact, the interference level of the network can be quantified through a Perron-Frobenius eigenvalue of the system path gain matrix [41]. Therefore, the design of a joint resource allocation and routing scheme should be done in such a way that the interference level is low enough and the desired QoS performance in terms of bit error rate (BER) or signal-to-interference-and noise ratio (SINR) can be achieved.

There are two popular approaches to modeling interference in an MCN. In the first approach interference is explicitly captured by SINR, and the feasibility of a QoS constraint can be checked through the Perron-Frobenius eigenvalue of the channel gain matrix [39, 41]. This approach was employed to develop an interference aware routing algorithm in [42]. In that paper the authors first obtained the minimum path loss routing solution. Then this initial routing solution was renavigated to find a routing path that improves the interference level in the network based on the Perron-Frobenius eigenvalue. Two-hop relaying schemes are the most commonly considered [43]. Limiting the number of hops to two degenerates the routing problem into a relay selection one [39], which can simplify the protocol design and minimize the communication overhead significantly, but this is a quite artificial model that may be far from a real network.

For the second approach, the joint resource allocation and routing problem is solved by using graph theory [44]. In this approach transmission links that interfere with each other are assumed to be known (based on interference range). Given this information, only links that do not interfere with one another are allowed to be active (*i.e.*, transmitting data) at the same time. Given a routing path for end-to-end data delivery (*i.e.*, from the source node to the destination node), there is an optimal transmission schedule of minimum length where in each time slot of the schedule only noninterfering links are allowed to transmit. Thus, the joint resource allocation and routing problem is equivalent to finding routing paths for all active mobile users and a transmission schedule such that the total number of time slots required to activate each link once on these routing paths is minimized. If all links in the network transmit at the same rate (*i.e.*, single-rate transmission), the end-to-end throughput for each active mobile user is equal to the ratio between this transmission rate and the length of the schedule (*i.e.*, the minimum number of time slots used in the schedule). If we map each time slot in the schedule to one color, the underlying problem is equivalent to a graph-coloring problem which is usually NP-hard [44]. Therefore, good polynomial-time heuristic algorithms with probable performance bounds are usually developed to solve the problem. The penalty of suboptimality is, however, quite high in many cases, which may ultimately result in very poor performance. For example, the algorithm proposed in [44] for the multicast problem achieves only a quarter of maximum throughput in the worst case, which may be unacceptable considering the potential gain due to multihop implementation.

The latest trend in this field, especially for multimedia applications, is based on matrix game theory and soft graph colouring [45].

Multipath routing

The “cooperative diversity” concept in multihop relaying networks is explored in [14, 46]. The main objective of the cooperative diversity is to improve the performance of cellular networks by using multiple nodes between the user and BS to simultaneously carry the same information. This idea resembles Multiple Input Multiple Output (MIMO) systems in a distributed manner. Since it is physically difficult to deploy multiple antennas on a single palm-sized mobile host, receiving multiple replicas of the main message from different relay nodes may improve the system performance due to its diversity nature. Multipath routing is one way for such cooperation by using multiple parallel paths between source and destination nodes, where the main data stream is split into streams of lower data rates and routed to the destination through the MCN. Multipath cellular networks are capable of supporting high data rate services with less transmission power consumption.

Several works explore the idea of multipath routing in MCNs [47- 49]. However, sufficient attention is not given for resource allocation and power conservation in these works. The key issues related to cooperation in multipath cellular networks are efficient relay selection and resource allocation. The aim is to find the best set of relays nodes that can cooperate with the user and the BS to establish a high data rate cellular connection and, a resource allocation algorithm that assigns appropriate transmission power and data rates to each of the selected relay nodes. Relay nodes will be selected among all idle nodes based on their willingness to cooperate, their channel quality, and their remaining battery resources.

1.2.3 Topology control

Topology control was originally developed for wireless sensor networks [50] to reduce energy consumption and interference. It works as a middle ware, connecting routing and lower layers as shown in Fig. 2. Topology control focuses on network connectivity with the link information provided by medium access control (MAC) and physical layers. When constructing network topology in MCNs, topology control takes care about the interference and link availability

prediction. The way the network topology is defined has a strong impact on routing. Topology control aims to simplify the routing process by providing: a) connectivity between nodes, b) energy efficient links, c) robustness against changes in location and removal of nodes, and d) maximization of link capacity.

From routing perspective, it is expected that data packets are routed via a stable and reliable path to avoid frequent rerouting problem, since frequent rerouting may induce broadcast storm to the network, waste scarce radio resources and degrade end-to-end network performance such as throughput and delay [51] which is especially critical in MCNs.

Previous work on interference avoidance topologies is based on one of these two assumptions: the network has power control or the network has channel control. Topology control through transmit power control generally utilizes a single shared channel and assumes a MAC for temporal separation of interfering transmissions. Burkhart [52] pioneered the power control based approach, assigning weights to connections that are equal to the number of radios the connection interferes with. This is used in the Min-Max Link Interference with a property *P* (MMLIP), Minimize the Average Interference Cost while Preserving Connectivity (MAICPC) and Interference Minimum Spanning Tree (IMST) [53] algorithms. Another power control based approach uses a radio interference function, in which the interference contribution of a radio is the maximum interference of all connections incident upon it. This is used in the Min-Max Node Interference with a property *P* (MMNIP) [53] and the Low Interference-Load Topology (LILT) [54] algorithms. Alternatively, the Average Path Interference (API) [55] approach trims high-interference, redundant edges from the Gabriel graph (GG)[1]. The channel control approach assumes the connectivity of the network is fixed and that two radios can only communicate if they share a common channel, of which there are fewer available than needed; this is illustrated by Connected Low Interference Channel Assignment (CLICA) [56], a heuristic approach, and Subramanian's [57] Tabu-search based algorithm. Several cooperation-based topology control algorithms have been proposed to create power efficient topologies (for a recent survey, see [50]). These algorithms assume total cooperation amongst radios, which collectively set their transmission power level so as to achieve a network-level goal.

[1] Graph with vertex set *S* in which any points *P* and *Q* in *S* are adjacent if they are distinct and, the *closed* disc of the line segment *PQ* is a diameter containing no other elements of *S*.

In cellular networks there are nodes of different classes (e.g., BS and mobile users), where BS is the origin/termination of all downlinks/uplinks of its service area. Topology control algorithms for ad hoc and sensor networks are designed with infrastructureless networks in mind and usually distributed, whereas in cellular networks it is possible to use centralized algorithms run within the base station. In the former, the aim is to maintain the connectivity between the nodes with minimum energy consumption. In contrast, in MCNs, energy consumption is less important than link throughput and delay. In [58] several classic topology control algorithms for ad hoc networks, such as Gabriel graph (GG), relative neighborhood graph (RNG), Yao graph (YG) and Delaunay graph (DG) are adapted to the cellular environment.

Limited work has been done in designing effective topology reconfiguration algorithms to offer optimal routing solutions in MCNs. The future data transmissions will have to face multiple radio access standards and complex spectrum allocation situations, and topology reconfiguration has emerged as a key technological enabler for supporting transmissions among heterogeneous networks, adapting to the time-varying environment and managing the joint radio resources across different spectrum bands. In [59], the authors provide an overview of the research in the field of topology control for cognitive radio networks, proposing Prediction-based Cognitive Topology Control (PCTC) to predict the duration of link availability. Based on this prediction, PCTC constructs a reliable topology which is aimed at improving network performance. Recently, some work has been done in applying bio-inspired algorithms for topology reconfiguration [60]-[62]. In [60] a particle swarm optimization is presented for minimum spanning tree (MST) problem for WSNs. In [61] a genetic algorithm is used for topology control in ad hoc networks to minimize the node degree, while preserving the network connectivity. A genetic algorithm with immigrants and memory schemes is presented in [62] to solve the dynamic shortest path (SP) problem in MANETs.

1.2.4 Intercell interference management

The exponential increase on data traffic demand in cellular networks requires a highly efficient exploitation of the available spectrum. 4G cellular standards are targeting aggressive spectrum reuse (frequency reuse 1) to achieve high system capacity and simplify radio network planning. The increase in system capacity comes at the expense of link SINR degradation due to increased ICI (*i.e.*

interference that two or more neighboring cells using the same frequency resource cause to each other), which severely impacts cell-edge user capacity and overall system throughput. Hence, advanced interference management schemes are critical to improve the performance of cell edge users. The main ICI management schemes for improving system performance can be classified in the following three groups:

a) Radio resource management through adaptive fractional frequency reuse:

Multicellular RRM efficiently partitions resources across cells in order to manage per resource interference experienced in each cell. Hybrid schemes that are a combination of universal reuse and higher reuse factors, so called fractional reuse (FFR) partitioning, were first introduced in [63] and are suggested for standards like LTE and WiMAX [64].

In particular, a mix of high and low reuse frequency resources (e.g., frequency reuse 1 and 3, respectively) are allowed in each cell as shown in Fig. 3. Resources governed by frequency reuse 1 can be assigned to users that are closer to the center of the cell and hence experience less interference from other cells, while the lower reuse resources are assigned to interference-limited users at the cell edge. Allowing a combination of frequency reuse patterns overcomes the capacity limitation inherent with lower frequency reuse, while also keeping a low interference environment to retain throughput and coverage for cell edge users. The definition of what constitutes cell center versus cell edge users is an important part of FFR design and is typically based on SINR metrics rather than actual user location within the cell.

Different FFR schemes are proposed for interference management in the downlink, while uplink is closely tied to power control mechanisms for interference management. From a link perspective the downlink allows for a more tractable analysis since if the desired mobile terminal location is known, the distances to all potential interfering BSs can be easily determined based on the network geometry, and hence a probabilistic estimate of the SINR can be calculated based on the channel fading conditions for the desired signal and the interfering signals.

Analysis of the uplink interference requires knowledge of not only the location of the desired mobile terminal under consideration, but also the relative locations of all potential interfering mobile terminals. This includes the locations of the interfering terminals, the number of potential terminals, and their speed.

Nevertheless, these a-priori hand-crafted schemes are still far from optimal in the sense that they do not adapt to dynamic network environments, e.g., time-varying user loads/locations. In addition, user scheduling working opportunistically based on perceived time-varying channels must be considered in conjunction with ICI management to achieve a high performance gain.

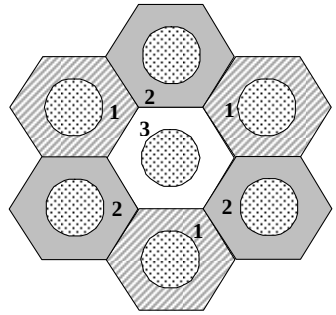

Fig. 3. FFR scheme.

More elaborate work on mitigating ICI has been done by [65]–[66]. Resource allocation management can prevent in-band concurrent transmissions to cause intra-cell and inter-cell interference by full time and frequency orthogonalization of resources. But such orthogonal allocations are not spectral efficient. Li *et al.* [65] formulated an optimization problem to maximize the system throughput in a multicell OFDMA system. In their solution, a Radio Network Controller (RNC) coordinates the interference among multiple cells so that each cell utilizes not all but around 80% of its subbands to avoid the dominant ICI. Bonald *et al.* [66] examined the capacity gains achievable by intercell time resource sharing in CDMA/HDR systems. They formulated an optimization problem which coordinates the activity phases of BSs so as to provide higher data rate for boundary users by mitigating ICI. In both [65] and [66], it is noteworthy that using only partial resources (frequency and time, respectively) is essential to obtain potential performance gains associated with mitigating ICI.

b) Power control:

Historically power control has been employed in cellular systems to minimize near-far dynamic range effects by constraining the uplink power to be received with a constant power level at the base station. Such an approach, while not

optimal from an aggregate throughput or spectral efficiency perspective, assures fairness to cell edge users.

Recently many works have been done on coordinated resource allocation in cellular wireless networks, including both *centralized* and *distributed* procedures. Centralized algorithms (e.g., [67], [68], [69], and many references in [70]) require global information to compute the transmit power. Due to the hardness of the problem, however, even centralized algorithms cannot guarantee that the globally optimal solution is found. Optimal binary power control (BPC)[2] for sum rate maximization was considered in [67]. They showed that BPC could provide reasonable performance compared with the multi-level power control in the multi-links system.

On the other hand, distributed algorithms (e.g., [71]-[75]) do not require a central controller and may demand less information exchange and computational complexity. Huang *et al.* [71] derive a distributed algorithm for interference mitigation by using a game-theoretic approach: here, the transmit power is considered as a continuous variable which is adjusted to maximize some network utility function. In [72], it is proposed to first identify the users whose power should be set to zero and then Huang's approach [71] is applied. Continuous power control requires more information exchange without significant benefits as shown in [68], [69]. Another drawback of the solutions in [71], [72] is that they require the knowledge of the channel gains from all other BSs to the scheduled users. Kiani *et al.* [73] propose a distributed binary power control algorithm for maximizing the total throughput which makes use of a simplified interference model. Stolyar *et al.* [74] propose a distributed algorithm aimed at minimizing the network power consumption while maintaining constant bit-rate for every user in each cell. In [75] a multicell power control optimization for interference management is presented to improve the spatial reuse factor.

Another approach to power control for multicell systems is Opportunistic power control (OPC) and has been shown to be throughput optimal for data traffic [76]. OPC exploits channel fluctuations such that it increases the transmission power when the channel is good and the transmission rate is adjusted according to the received SINR ratio. Although the OPC concept is attractive because it maximizes the multicell throughput and lends itself for distributed implementations, it can become extremely unfair. Indeed, previous works

[2] In the two-link case, BPC assigns full power to one link and minimum to the other, or full power on both links depending on the noise and channel gains.

proposed computationally efficient algorithms that deal with the fairness issue of OPC [76], [77], typically at the expense of some loss in the overall throughput.

c) Smart antenna techniques to null interference from other cells:

To attain the full potential gain of multicell networks, smart antenna techniques that exploit the spatial diversity for interference mitigation are proposed in the literature [78]. The key idea is to equip transmitters and receivers in the system with multiple antennas and then utilize the directivity and/or diversity properties of the multi-antenna processing. The main methods are

c1) Transmit beamforming:

Transmit beamforming is an efficient way to combat ICI, and in particular to protect the user at the cell edge. By transmitting in a narrow beam directed towards the desired user instead of a sector-wide beam, it is possible to reduce the interference spread to other cells in the system. In addition, the transmitted signal also gets a power boost from the resulting array gain. Beamforming can be carried out in different ways; at the highest level we distinguish *adaptive* and *fixed* beamforming. In adaptive beamforming the antenna weights are adaptively set in order to optimize the antenna pattern according to some optimization criteria. Fixed beamforming is a low-complexity approach, in which a finite set of antenna weights is used which generates a set of predefined beams. Hence, the beamforming problem reduces to beam selection, which requires less feedback information than adaptive approaches.

Research on downlink beamforming using antenna arrays at the BS can be categorized into two classes. The first class of research focuses on designing algorithms for computing the beamforming weights (*i.e.*, relative amplitudes and phase shifts of antenna elements) and transmission power for each user given a set of scheduled users [79]-[81]. This is often modeled as an optimization problem, where the objective is to minimize the total transmission power subject to the constraint that each user's SINR requirement is satisfied. The second class of research focuses on the MAC layer with physical layer user separability constraints. The goal of this class of research is to maximize the number of scheduled users while satisfying their SINR constraint. This problem is extended and combined with other multi-user access schemes such as TDMA, OFDM and CDMA in [82]. The performance of various beamforming techniques has also been studied in the context of UMTS [83]-[85].

c2) Spatial antenna techniques such as MIMO and SDMA:

Gains due to SDMA [86], [87] and MIMO [88], [89] implementations have been extensively researched, and a number of MIMO and SDMA techniques are included in the LTE standard. The notion of network MIMO involves the use of multiple antennas at both the transmitter and receiver side. Joint encoding over geographically distributed antennas renders the network into a super-cell, which is related to the MIMO broadcast scenario [90]. On the downlink, multiple base stations can transmit one or more MIMO paths to a mobile, whereas on the uplink, the transmission by a mobile can be received by one or more BSs. The gains of these techniques have been well established through simulation [91] as well as trial implementations. In case that full channel state information and all data are available at a central controller, network MIMO can efficiently exploit all spatial degrees of freedom to eliminate ICI. Although the network's performance is no longer limited by interference, there is a huge amount of additional complexity and coordination overhead compared to single cell signal processing.

c3) Decoding algorithms:

The use of multiple antennas at receivers facilitates establishment of spatial diversity branches, which can be used for implementation of receive diversity and/or interference rejection techniques in the receive processing. Since the radio channels from a transmit antenna to the receive antennas tend to fade differently, multi-antenna receivers provide diversity –both for the signal of interest and for the interference. With appropriate selection of the antenna combining weights, accounting for the radio channel, the interference power and the spatial coloring of the interference, such multi-antenna receivers may provide increased robustness to both fading and interference.

The most well-known method for receive diversity is traditional Maximum Ratio Combining (MRC) [92]. Other recent receiver decoding innovations include Sphere and dirty paper coding [93].

1.2.5 Load balancing

Another important issue in multicell networks is to resolve the load imbalance problem between cells. In order to balance the load among different cells, it is needed to transfer the over-loaded traffic from "hot" cells to neighboring "cooler"

ones. Various dynamic load balancing schemes to deal with the unbalanced traffic problem are proposed in the literature. We can broadly classify them into four groups: a) Strategies based on channel borrowing from cooler cells [94]; b) Strategies based on BS selection [95]; c) Strategies based on power control and cell breathing [96],[97],[99]; and d) Strategies based on relay-assisted traffic transfer [100]-[102].

The basic idea of channel borrowing is to borrow a set of channels from “cooler” cells (with less traffic load) to “hot” cells. However, this will change the pre-defined spectrum reuse pattern and introduce more cochannel interference. Also, as future cellular networks move towards to universal frequency reuse, there is little space for channel borrowing schemes. In BS selection schemes, mobile users in hot cells will try to associate with a BS in a neighboring cooler cell and get service, but the throughput is limited due to low signal strength. The cell breathing effects allow adjustment of transmit power to reduce the size of hot cells to release over-loaded traffic to neighboring cooler cells. Sang *et al.* [96] proposed an integrated framework consisting of a MAC layer cell breathing technique and load aware handover/cell-site selection to deal with load balancing. Bu *et al.* [97] were first to rigorously consider a mathematical formulation of proportional fairness PF [98] in a network-wide manner with users' associations to BSs. They showed that the general problem is NP-hard and proposed a heuristic algorithm to approximately solve the problem. [99] extends this network-wide PF to the multicell network with partial frequency reuse where each BS has limited resources based on ICI pre-coordination scheme and independently runs a PF scheduler. Therefore, in cell breathing schemes, close cooperation among adjacent cells is required to guarantee full coverage and mitigate ICI. The last strategy consists of taking the advantage of MCNs to relay over-loaded traffic from hot cells to cooler cells. Load balancing in MCNs not only involves balancing among cells, but also balancing among relaying nodes and the choice of relaying device. Compared with previously discussed dynamic load balancing schemes, relay-based load balancing schemes are more flexible and will introduce less interference.

In [100], a mobile-assisted call admission scheme is proposed to achieve load balancing in cellular networks, which requires an ad hoc overlay network on the cellular network. The authors divided the channels into two groups, one for the ad hoc overlay network and the other for the cellular network. The simulation results showed that a fixed division of channels is not efficient. In [101]-[102], the authors proposed dynamic load balancing schemes in the integrated cellular and

ad hoc relaying systems (iCAR) [101] and PARCelS [102]. The ad hoc relaying stations (ARS) compose an overlay ad hoc network, which can help relay traffic among different cells.

In iCAR, low cost limited mobility ARSs are placed in hot spot areas for traffic relaying. This strategy is not only costly, but also not flexible enough to handle the highly dynamic load scenarios in 4G networks. PARCelS uses mobile nodes for relaying. When a BS is congested, mobile nodes search best routes to other non-congested cells. Route information is forwarded to BSs for selection. This strategy requires considerable routing overhead and does not take advantage of the presence of powerful BSs. In addition, both schemes do not take into account the load balancing among mobile users. Balancing among the users is important to avoid the situation where over-loaded relaying nodes run out of battery. This affects the availability of routes and connectivity. Although this issue is more related to routing, balancing load among cells and mobile users is important to achieve good network performance.

ALBA [103] is a dynamic load balancing scheme for CDMA-based MCNs which considers the location and priority of mobile nodes for load migration. The basic idea is to shift traffic load from a hot cell to cooler cells in a best effort manner by checking periodically the load status of the cells in the network. Best effort is assumed because relaying routes for load migration may not exist especially in a highly dynamic loaded network. If cell load deviation is greater than a global load deviation threshold, then starts load migration planning. ALBA may be also applied to any heterogeneous load environment. Although simulation results show that this scheme has good performance in terms of throughput and lower call blocking ratio, like most load balancing schemes, ALBA is a heuristic.

A novel message forwarding mechanism for load balancing in relay based multicell topologies is presented in [104]. When considering mobile relay nodes, the actual mobility of nodes can be used to physically propagate information messages. Furthermore, while full connectivity can be provided by the supporting BS, *store-carry and forward* (SCF) paradigm is proposed to provide the target performance gain at the expense of message delivery delay. It is important to note that the SCF paradigm was originaly conceived as a way to provide communication in intermittently connected networks [105]. However in this case, message forwarding is deliberately delayed to allow for the physical propagation of information messages. With knowledge on data traffic load conditions of neighboring cells, a BS has the flexibility to redirect delay tolerant information messages to adjacent cells by utilizing the underlay SCF scheme. In this way,

delay tolerant traffic can be guided to neighbor cells which have lower utilization level to avoid resource stagnation at the targeted cell.

Load balancing for multicast applications

Network integration also represents an interesting way to support multimedia services in cellular systems as it allows increasing the efficiency of the whole system in terms of coverage area, resource capacity, and number of simultaneously active users. The network load in this type of applications is especially critical.

Cellular service providers have already had difficulties to keep up with the staggering increase in data traffic [106], [107], and will have to carefully engineer their networks to support the tremendous amount of mobile video traffic in the future. Today, cellular networks are unable to handle large scale live video distributions since existing cellular deployments do not natively support multicast and broadcast.

Cellular service providers may address the capacity issue by: a) deploying more base stations, b) upgrading their base stations, e.g., to support Multimedia Broadcast Multicast Services (MBMS) [108], or c) building dedicated broadcast networks, such as Digital Video Broadcast–Handheld (DVBH) [109]. However, these solutions incur high infrastructure costs and may not be compatible with current mobile devices. Hence, a better solution is needed. Since modern mobile devices are equipped with multiple network interfaces, cellular service providers may *offload* mobile video traffic to an *auxiliary* network. In MCNs mobile devices relay video data among each other using ad hoc links. Exploiting such a *free* mechanism of distribution alleviates bottlenecks and reduces cost for cellular service providers.

While MCNs have the potential to capitalize on the complementary features of both networks for low cost yet reliable massive live video distribution, transmission of video data must adhere to the timing needs inherent in the delivery and playback of video content.

Law *et al.* [110] evaluate a hybrid network in which some mobile devices act as gateways and relay data to mobile users outside the range via a multihop ad hoc network. Lao and Cui [111] propose a hybrid network, in which each multicast group is either in the cellular mode or in the ad hoc mode. Park and Kasera [112] consider the gateway node discovery problem, and model ad hoc interference as a graph coloring problem. Bhatia *et al.* [113] formulate a problem

of finding the relay users to maximize the overall data rate, and they propose an approximation algorithm to solve it. Qin and Zimmermann [114] present an adaptive strategy for live video distribution to determine the number of quality layers to be transmitted between two mobile devices. Hua *et al.* [115] formulate an optimization problem in a hybrid network to determine the cellular broadcast rate of each quality layer. In the ad hoc network, a flooding routing protocol is used to discover neighbors and a heuristic is employed to forward video data. A lot of work remains to be done in this area since most of the previous works are based on single-cell scenario.

1.3 Aims and outline of the thesis

The aim of the thesis is to present a number of new network paradigms for future MCNs. The contributions include solutions for relaying topology control optimization, network reconfiguration issues, scheduling, new multihop routing protocols and different proposals for multicast traffic optimization in cellular networks as well as the integration of different types of networks within the MCN. A novel approach to the optimization, control and analysis of MCNs is used to address those paradigms.

Physical layer issues, such as a new channel model for multihop networks, new interference management schemes and power control optimization, are also covered in this thesis.

The thesis is organized in 6 chapters:

- In Chapter 1, the introduction is presented as a literature review of the most important research results for MCNs. Open problems and future research directions are pointed out to highlight the motivation for the research. The main contributions of this thesis are included from Chapter 2 to 5 as follows:
- Chapter 2, the results of which have been presented in [116]-[118], presents an algorithm for efficient relaying topology control, which is aware of intercell interference. The algorithm jointly chooses the relaying topology and scheduling in the adjacent cells in such a way to minimize the system performance degradation due to the intercell interference. A new topology search (TSL) program is developed to find the best topology in accordance with a given objective function. The set of constraints in the optimization program includes relaying specific system parameters and temporal and spatial nonuniform traffic distribution.

This framework is also extended to include the optimization of the power allocation and, the network performance is compared by using cooperative diversity relaying scheme (COOR) and conventional relaying scheme (CONR), resulting in two intercell interference management protocols I^2M-COOR and I^2M-CONR, respectively. By including weights in the utility function we analyze the trade-off between throughput and power allocation.

Numerical results demonstrate that an adaptive relaying topology control provides the network utility improvements and presents the framework for quantifying these improvements for spatially and temporally varying traffic.

- Chapter 3, the results of which have been presented in [119]-[121], extends our previous results on relaying topology optimization and presents a novel sequential genetic algorithm (SGA) for dynamic reconfiguration of the relaying topology to the traffic variations in the network.

  Duplex transmission is considered and network coding is used to combine the uplink/downlink transmissions and, incorporate it into the *optimum bidirectional relaying with ICI awareness* resulting in a comprehensive solution for 4G/5G common air interface.

  Numerical results show that SGA-TSL provides both high performance improvements in the system, fast convergence (at least one order of magnitude faster than exhaustive search) in a dynamic network environment.
- Chapter 4, the results of which have been presented in [122]-[124], presents a new approach to optimization in MCNs. A nano scale network model (NSNM) is developed for high resolution optimization. By applying hexagonal tessellation, the cell is partitioned into smaller subcells of radius *r*. By adjusting the radius of the subcell *r*, different hopping ranges are obtained which directly affect the throughput, power consumption and interference. With *r* as the optimization parameter, we jointly optimize routing, scheduling and power control to obtain the optimum trade-off between throughput, delay and power consumption. By using only one single topology control parameter (*r*) for multi objective system optimization we minimize the control traffic overhead that makes the system feasible for practical implementation.

  A set of numerical results demonstrates that NSNM enables high resolution optimization of the system and an effective use of the context awareness. A special nano scale channel model (NSCM) for this application is also included together with a new concept for route discovery protocols for MCNs which is aware of the mutual impact of all routes in the cell.

Numerical results show that our proposed algorithm is superior when compared to other existing route discovery protocols adapted to this scenario.

In addition, the NSNM is used to analyze the trade-off between cooperative diversity and spatial reuse in multihop cellular networks.

- Chapter 5, the results of which have been presented in [125], provides a promising way to support multimedia services in cellular networks by integration of Delay Tolerant Network (DTN) and cellular network. Polymorphic Epidemic Routing (PER) is proposed for multicast DTN and new adaptive recovery schemes are developed to remove the delivered packets from the network (recovery from infection). Analytical model of this system is presented to study the effects of different recovery schemes on the performance of multicast DTN. Numerical results show that our adaptive schemes increase the efficiency of the whole system in terms of reducing the message delivery delay and the resource consumption compared to existing schemes.
- Finally, Chapter 6 concludes the thesis. The main results are summarized and future research directions are suggested.

1.4 Author´s contribution to the publications

This thesis is mainly based on four journal papers [116, 119, 122 and 125], and six conference papers [117, 118, 120, 121, 123 and 124]. Most of the results have been published or are under consideration for publication. I have coauthored all these papers with my supervisor Prof. Glisic. The author has had the main responsibility for performing the analysis, developing the simulation software, generating the numerical results, and writing all the aforementioned papers. Prof. Glisic provided opportunities, motivations, reviews and suggestions related to technical issues, editorial corrections and guidance in the study and publication process. All results and analysis presented in this thesis have been produced by the author.

2 Optimization of relaying topology in MCN

In this chapter, we propose an optimization framework for relaying topology control in MCNs aware of the intercell interference (ICI). Firstly, we define the relaying topology in MCNs to answer the questions who is transmitting to whom, and when, in such a way to insure the best system performance. The transmission scheduling is included in the optimization to reduce the effects of the ICI while providing channel reuse factor one. The algorithm jointly chooses the relaying topology and scheduling in the adjacent cells in such a way to maximize the system performance. This results in a multicell jointly optimal relaying topology. In case of temporally and spatially varying traffic distribution, the optimal topology will also vary in time and an efficient way of topology control is needed. The optimization problem is formulated by using NUM (Network Utility Maximization) formulation and the aim is to jointly optimize the relaying topology, routing and scheduling in MCNs with ICI awareness. The utility function includes data rate, power consumption and delay. The overall optimization is solved by using the combination of a new topology search program (TSL) developed for this application and CVX program [126].

Numerical results demonstrate that an adaptive relaying topology control provides the network utility improvements and presents the framework for quantifying these improvements for spatially and temporally varying traffic.

This framework is then extended to include the optimization of the power allocation and the utility function is modified by adding weights to model the trade-off between throughput and power consumption. We also compare the performance of conventional relaying (CONR) and cooperative relaying (COOR) schemes, resulting into two intercell interference management protocols I^2M-CONR and I^2M-COOR, respectively. Numerical results show that I^2M-COOR offers an improvement in the network throughput of at least 4 times and a reduction of power consumption up to 3 times compared to I^2M-CONR.

The remainder of this chapter is organized as follows. An overview and background of cross-layer optimization is given in Section 2.1. System model and assumptions are then given in Section 2.2. The description of the relaying topology and motivating example are presented in Section 2.3. Section 2.4 extends this model to cooperative relaying scheme (COOR). The traffic modeling is presented in Section 2.5. Jointly optimization of relaying topology control, routing and scheduling is shown in Section 2.6. In Section 2.7, the throughput-power trade-off is presented. Numerical results are shown in Section 2.8 to

illustrate the performance of the proposed algorithms. Finally, Section 2.9 concludes the chapter.

2.1 Overview and background

In the past, network protocols in layered architectures have been obtained on an ad hoc basis, and many of the recent cross-layer designs are also conducted through piecemeal approaches. Only recently, network protocol stacks are analyzed and designed as distributed solutions to some global optimization problems in the form of generalized Network Utility Maximization (NUM), providing insight on what they optimize and on the structures of the network protocol stack. Such a framework of "*layering as optimization decomposition*" provides a common framework for modularization, a way to deal with complex, networked interactions. It exposes the interconnection between protocol layers and can be used to study rigorously the performance trade-off in protocol layering, as different ways to modularize and distribute a centralized computation. Even though the design of a complex system will always be broken down into simpler modules, this theory will allow us to systematically carry out this layering process and explicitly trade off design objectives.

By this framework, the network utility function is maximized by its decomposition into components which are implemented in different layers of the protocol stack. Each layer corresponds to a decomposed subproblem, and the interfaces among layers are quantified as functions of the optimization variables coordinating the subproblems. Different layers iterate on different subsets of the decision variables using local information to achieve individual optimality. Taken together, these local algorithms attempt to achieve a global objective.

Intuitively, layered architectures enable a scalable, evolvable, and implementable network design. Each layer in the protocol stack hides the complexity of the layer below and provides a service to the layer above. It adopts a modularized and often distributed solution approach to network coordination and resource allocation. Such a design process of modularization can be quantitatively understood through the mathematical language of *decomposition theory* for constrained optimization [127]. Decomposition theory provides the analytical tool for the design of modularized and distributed control of networks.

The application of "*layering as optimization decomposition*" has been illustrated through many case studies. The congestion control functionality of TCP has been reverse engineered by implicitly solving the *Basic NUM* problem

[128]. Other results also show how to reverse engineer Border Gateway Protocol (BGP) as the solution to the Stable Path Problem [129], and contention based Medium Acces Control (MAC) protocols as a noncooperative selfish utility maximization game [130]. A number of papers have been published in this area with the focus on jointly optimized congestion control and routing [131], [132]; routing, scheduling and power control [133], [134]; congestion control, routing and scheduling [135], [136]; congestion control and physical recource allocation [137], [138]; and different combinations of those problems [139]-[142]. The feasibility of using the above framework in mobile communications with fading and mobility is described in [140].

Since the early 1990s, it has been recognized that for efficient solution of optimization problems we need *convexity*. Convex optimization has become a computational tool of central importance in engineering, thanks to its ability to solve very large, practical engineering problems reliably and efficiently. Many communication problems can either be cast as or be converted into convex optimization problems, which greatly facilitate their analytic and numerical solutions. Furthermore, powerful numerical algorithms exist to find the optimal solution of convex problems efficiently [143].

In the sequel we provide the formulation of NUM, a description to the most common decomposition methods and the decomposition of NUM. For the basics in the area of convex optimization, the basics of convexity, Lagrange duality, distributed subgradient methods and other solution methods for convex optimization, the reader is referred to [143].

NUM formulations

We consider a network modeled as a set $\mathcal{L}$ of links (scarce resources) with finite capacities $\mathbf{c} = (c_l, l \in \mathcal{L})$. They are shared by a set $\mathcal{N}$ of sources indexed by *n*. Each source n uses a set $\mathcal{L}(n) \subseteq \mathcal{L}$ of links. Let $\mathcal{N}(l) = \{n \in \mathcal{N} \mid l \in \mathcal{L}(n)\}$ be the set of sources using link l. The sets $\{\mathcal{L}(n)\}$ define a $L \times N$ routing matrix $\mathbf{R} = \{r_{ln}\}$ with $r_{ln} = 1$, if $l \in \mathcal{L}(n)$, *i.e.*, source *n* uses link *l* and 0, otherwise.

The *Basic NUM* problem is the following formulation, known as monotropic programming. TCP variants have recently been reverse engineered to show that they are implicitly solving this problem, where source rate vector $\mathbf{x} \geq 0$ is the only set of optimization variables, and routing matrix $\mathbf{R}$ and link capacity vector $\mathbf{c}$ are both constants:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \sum_n U_n(x_n) \\ \text{subject to} \quad & \mathbf{Rx} \le \mathbf{c}, \end{aligned} \tag{1}$$

where the utility function U_n is often assumed to be smooth, increasing, concave, and dependent on local rate only, although recent investigations have removed some of these assumptions for applications where they are invalid.

Many of the papers on "*layering as optimization decomposition*" are special cases of the following *Generalized NUM* for the entire protocol stack:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \sum_n U_n(x_n, P_{e,n}) + \sum_j V_j(z_j) \\ \text{subject to} \quad & \mathbf{Rx} \le \mathbf{c}(\mathbf{z}, \mathbf{P}_e), \\ & \mathbf{x} \in \mathcal{C}_1(\mathbf{P}_e),\ \mathbf{x} \in \mathcal{C}_2(\mathbf{F})\ or \in \mathbf{\Pi}(\mathbf{z}), \\ & \mathbf{R} \in \mathfrak{R},\ \mathbf{F} \in \mathcal{F},\ \mathbf{z} \in \mathcal{Z}, \end{aligned} \tag{2}$$

where the rate of source n is designated as x_n, and these rates are arranged in vector $\mathbf{x}$. Parameter z_j denotes the physical layer resource at network element j. The utility functions U_n and V_j are concave functions and in general may be any nonlinear, monotonic functions. $\mathbf{R}$ is the routing matrix, and $\mathbf{c}$ are the logical link capacities as functions of both, physical layer resources $\mathbf{z}$ and targeted decoding error probabilities $\mathbf{P}_e$ which capture for example, the functional dependency of power control and interference level in the network. Routing matrix $\mathbf{R} = [r_{ln}]$ has entries $r_{ln} = 1$ if source n (n = *1,2,..,N*) is using link l (l = *1,2,..,L*) and 0 otherwise. If in addition an ARQ protocol is used, a constraint set $\mathcal{C}_1(\mathbf{P}_e)$ is included. The issue of rate-reliability trade-off and coding is captured in this constraint. The rates are further constrained by the medium access success probability which is modeled by $\mathcal{C}_2(\mathbf{F})$, where $\mathbf{F}$ is the contention matrix, or the scheduling constraint set $\mathbf{\Pi}$. The sets of possible routing algorithms, scheduling or contention based MAC schemes and physical layer resource allocation schemes are represented by $\mathfrak{R}$, $\mathcal{F}$ and $\mathcal{Z}$, respectively. Within these algorithms the optimization variables are $\mathbf{x}$, $\mathbf{z}$, $\mathbf{P}_e$, $\mathbf{F}$ and $\mathbf{R}$.

Decomposition Methods

The basic idea of decomposition is to split the original large problem into smaller subproblems, which are then coordinated by a master problem by means of some kind of signalling. Most of the existing decomposition techniques can be classified into *primal decomposition* and *dual decomposition* methods. The

former is based on decomposing the original primal problem, whereas the latter is based on decomposing the Lagrange dual of the problem. Primal decomposition methods have the interpretation that the master problem directly gives each subproblem an amount of resources that it can use. The role of the master problem is then to properly allocate the existing resources. In dual decomposition methods, the master problem sets the price for the resources to each subproblem which has to decide the amount of resources to be used depending on the price. The role of the master problem is then to obtain the best pricing strategy. Primal decomposition and dual decomposition can in fact be inter-changed by introducing auxiliary variables [127]. Almost all the papers in the vast, recent literature on NUM use a standard dual-based distributed algorithm. Contrary to the apparent impression that such a decomposition is the only possibility, there are in fact many alternatives to solve a given network utility problem in different but all distributed manners [147], including multi-level and partial decompositions as shown in Fig. 4. Each of the alternatives provides possibly different network architectures with different engineering implications.

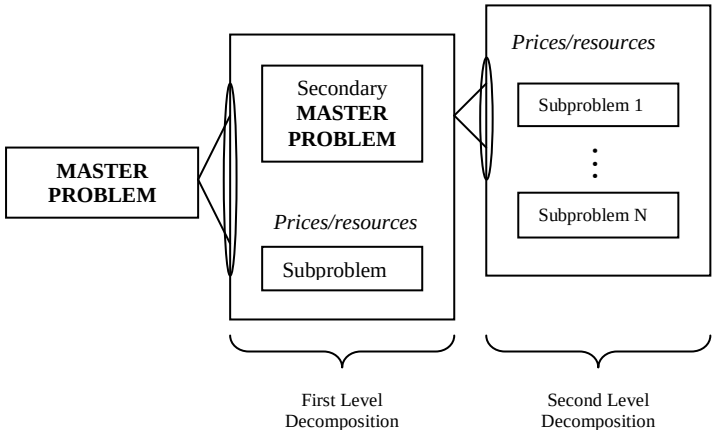

Fig. 4. Multilevel decomposition.

Dual decomposition of the Basic NUM

In this section, we illustrate how the dual decomposition approach can be applied to the *Basic NUM* problem to produce the standard dual-decomposition-based

distributed algorithm. Assume that the utility functions are concave, and possibly linear functions. The Lagrange dual problem of (1) is formed as

$$L(\mathbf{x},\boldsymbol{\mu}) = \sum_n U_n(x_n) + \sum_l \mu_l \left(c_l - \sum_{n \in \mathcal{N}(l)} x_n \right), \tag{3}$$

where $\mu_l \geq 0$ is the Lagrange multiplier (*i.e.*, link price) associated with the linear flow constraint on link *l*. Additivity of total utility and linearity of flow constraints lead to a Lagrangian dual decomposition into individual source terms

$$\begin{aligned} L(\mathbf{x},\boldsymbol{\mu}) &= \sum_n [U_n(x_n) - (\sum_{l \in \mathcal{L}(n)} \mu_l) x_n] + \sum_l c_l \mu_l \\ &= \sum_n L_n(x_n, q_n) + \sum_l c_l \mu_l, \end{aligned} \tag{4}$$

where $q_n = \sum_{l \in \mathcal{L}(n)} \mu_l$. For each source *n*, $L_n(x_n, q_n) = U_n(x_n) - q_n x_n$ only depends on local rate x_n and the path price q_n (*i.e.*, sum of μ_l on links used by source *n*).

The Lagrange dual function $g(\boldsymbol{\mu})$ is defined as the maximized $L(\mathbf{x}, \boldsymbol{\mu})$ over $\mathbf{x}$ for a given $\boldsymbol{\mu}$. This "net utility" maximization obviously can be conducted distributively by each source

$$x_n^*(q_n) = \arg\max_x \ [U_n(x_n) - q_n x_n], \forall n. \tag{5}$$

The Lagrangian maximizer $\mathbf{x}^*(\boldsymbol{\mu})$ will be referred to as price-based rate allocation (for a given price $\boldsymbol{\mu}$). The Lagrange dual problem of (1) is

$$\begin{aligned} \underset{\boldsymbol{\mu}}{\text{minimize}} \quad & g(\boldsymbol{\mu}) = L(x^*(\boldsymbol{\mu}), \boldsymbol{\mu}) \\ \text{subject to} \quad & \boldsymbol{\mu} \geq 0, \end{aligned} \tag{6}$$

where the optimization variable is $\boldsymbol{\mu}$. Since $g(\boldsymbol{\mu})$ is the point wise supremum of a family of affine functions in $\boldsymbol{\mu}$, it is convex and (6) is a convex minimization problem. Since $g(\boldsymbol{\mu})$ may be nondifferentiable, an iterative subgradient method can be used to update the dual variables $\boldsymbol{\mu}$ to solve the dual problem (6)

$$\mu_l(t+1) = [\mu_l(t) - \beta(t)(c_l - \sum_{n \in N(l)} x_n(q_n(t)))]^+, \forall l, \tag{7}$$

where $c_l - \sum_{n \in \mathcal{N}(l)} x_n(q_n(t))$ is the *l*th component of a subgradient vector of $g(\boldsymbol{\mu})$, *t* is the iteration index, and $\beta(t) > 0$ is the step size. Certain choices of step sizes, such as $\beta(t) = \beta_0/t$, $\beta > 0$, guarantee that the sequence of dual variables $\boldsymbol{\mu}(t)$ converges to the dual optimal $\boldsymbol{\mu}^*$ as $t \to \infty$. It can be shown that the primal variable $x^*(\boldsymbol{\mu}(t))$ also converges to the primal optimal variable $\mathbf{x}^*$. For a primal

problem that is a convex optimization the convergence is towards a global optimum.

In summary, the sequence of source and link algorithms (5), (7) forms a standard dual-decomposition-based distributed algorithm that globally solves NUM (1) and the dual problem (6), *i.e.*, computes an optimal rate vector $\mathbf{x}^*$ and optimal link price vector $\boldsymbol{\mu}^*$ without explicit need for signalling. This is because the subgradient is precisely the difference between the fixed link capacity and the varying traffic load on each link, and the subgradient update equation has the interpretation of weighted queuing delay update.

In the *Generalized NUM* coupling can happen in the constraints, and also in the objective function, where the utility of source n, $U_n(x_n, \{x_i\}_{i \in \mathcal{I}(n)})$, depends on both its local rate x_n and the rates of a set of other sources with indices in set $\mathcal{I}(n)$. If U_n is an increasing function of $\{x_i\}_{i \in \mathcal{I}(n)}$, this coupling models cooperation in a clustered system, otherwise it models competition such as power control in wireless network or spectrum management in DSL. Such coupling in the objective function can be decoupled [147] by first introducing auxiliary optimization variables and consistency equality constraints, thus shifting coupling in objective to coupling in constraints. This can be decoupled by applying dual decomposition and solved by introducing "consistency prices". These consistency prices are iteratively updated through local message passing. For more details in the area of decomposition theory, the reader is refered to [127].

In what follows, we present a new design of relaying topology control algorithms for MCNs based on Network Utility Maximization (NUM) formulation. The design jointly optimizes the relaying topology, routing and scheduling resulting in a two dimensional or space time routing protocol. This framework is also extended to include the optimization of the power allocation in cooperative and conventional relaying schemes.

2.2 System model and assumptions

In this section, we describe the system model adopted for relaying in MCNs and the model of ICI. We use generic notation for variable X as $X_{transmitter,receiver}$ where the first index refers to the position of the transmitter (terminal or base station) and the second one to the position of the receiver.

We consider the uplink transmission in a cellular network with a set $\mathcal{I} = \{i\}$ of base stations. For a systematic presentation of the problem, hexagonal cells are approximated by circles of radius R and the cell area is divided into concentric

rings with index m_i as shown in Fig. 5. We assume that the *mobile* users are located on those rings at distance $d_{m_i,i} = m_i d_0$ from the base station i where m_i is an integer and d_0 is a unit distance. We consider that there is one user per channel in each ring. Users are equipped with omnidirectional antennas and transmission rates depend on the instantaneous SINR at the receivers.

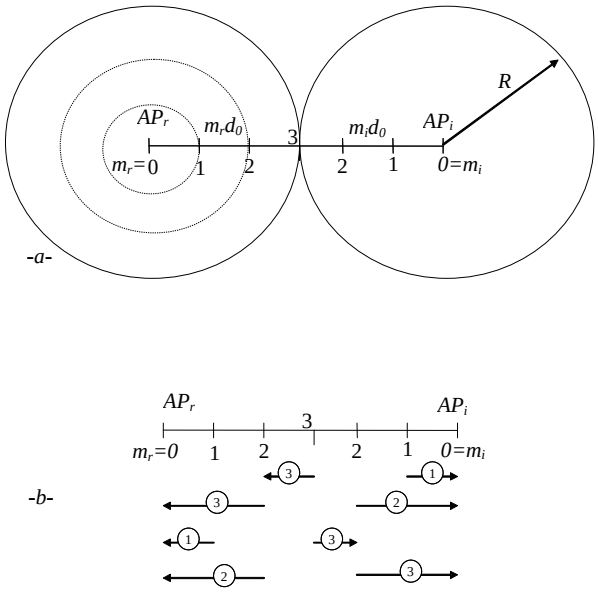

Fig. 5. a) Modeling interfering users positions for 2-cells; b) Possible transmission schedule.

Let us assume that in a reference cell with index $i = r$, there is a reference *mobile* user m_{r1} transmitting (relaying) to another *mobile* user m_{r2} with channel gain $G_{m_{r1},m_{r2}}$. At the same time, a cochannel *interfering mobile* user m_{i1} is transmitting to another *mobile* user m_{i2} in cell i, $i \in \mathcal{I}_{-r} = \{i \neq r\}$ with channel gain $G_{m_{i1},m_{i2}}$.

We will use notation $(m_{r_1}, m_{r_2}, \mathbf{m}_{i_1}, \mathbf{m}_{i_2})$, to denote *simultaneous* transmission (relaying) on reference route from mobile user m_{r_1} *to* m_{r_2}, $r \in \mathcal{I}$ and cochannel interfering mobile users from m_{i_1} *to* m_{i_2}, $i \in \mathcal{I}_{-r}$ position in all interfering cells.

Assuming that $P_{m_{r_1},m_{r_2}}$ and $P_{m_{i_1},m_{i_2}}$ are the transmission power of the useful and interference signal respectively, the received power at users m_{r2} and m_{i2} are given by

$$S_{m_{r_1},m_{r_2}} = P_{m_{r_1},m_{r_2}} G_{m_{r1},m_{r2}}, \; r \in \mathcal{I} \tag{8}$$

$$S_{m_{i1},m_{i2}} = P_{m_{i1},m_{i2}} G_{m_{i1},m_{i2}}, \; i \in \mathcal{I}_{-r}. \tag{9}$$

Let $I_{m_{i1},m_{r2}}$ represents the interference power at the position of the reference receiver m_{r2} due to the interfering cochannel signal transmitted by m_{i1}. This can be presented as

$$I_{m_{i1},m_{r2}} = P_{m_{i1},m_{i2}} G_{m_{i1},m_{r2}} = S_{m_{i1},m_{i2}} G_{m_{i1},m_{r2}} / G_{m_{i1},m_{i2}}, \tag{10}$$

where $G_{m_{i1},m_{r2}}$ is the gain of the channel between the interfering user m_{i1} and the reference user m_{r2}.

The signal to interference plus noise ratio $SINR_{m_{r1},m_{r2}}$ at m_{r2} in the presence of all interfering users is

$$SINR_{m_{r1},m_{r2}}\left(\mathbf{m}_{i_1},\mathbf{m}_{i_2}\right) = \frac{S_{m_{r1},m_{r2}}}{n_r + \sum_{i \neq r} I_{m_{i1},m_{r2}}}, \tag{11}$$

where $\mathbf{m}_{i1} = (m_{11}, m_{21}, ..., m_{N_c 1})$, $\mathbf{m}_{i2} = (m_{12}, m_{22}, ..., m_{N_c 2})$, $i \in \mathcal{I}_{-r}$ are the positions of the cochannel interfering mobile users in all interfering cells and, n_r is the background noise power.

Under these conditions the corresponding link capacity will be denoted as $c_r(m_{r_1}, m_{r_2}, \mathbf{m}_{i_1}, \mathbf{m}_{i_2})$ and can be calculated by

$$c_r(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2}) = \log\left(1 + SINR_{m_{r1},m_{r2}}(\mathbf{m}_{i1}, \mathbf{m}_{i2})\right); \; i \in \mathcal{I}_{-r}. \tag{12}$$

We start by considering that there is no power control and later on, in Section 2.7 we will remove this assumption. So, in the case when the transmission power is the same for the referent user and interference users $P_{m_{r_1},m_{r_2}} = P_{m_{i_1},m_{i_2}}$, we have

$$SINR_{m_{r1},m_{r2}}\left(\mathbf{m}_{i_1},\mathbf{m}_{i_2}\right) = \left(N_r^{-1} + \sum_{i \neq r} G_{m_{i1},m_{r2}} / G_{m_{i1},m_{i2}}\right)^{-1}, \tag{13}$$

where $N_r = P / n_r = SNR$.

In general, links between the mobile and the *AP* may require less power if space time coding is used. The modifications to include this fact are straightforward.

We define now the multihop (*H* hops) route as a series of relaying transmissions

$$\Re_r(m_{r1}, m_{r2}, ..., m_{rH-1}, m_{rH}, \mathbf{m}_{i1}, \mathbf{m}_{i2}, ..., \mathbf{m}_{iH-1}, \mathbf{m}_{iH}). \tag{14}$$

The above equations correspond to the conventional relaying scheme (CONR), in the first phase of the communication (*i.e.*, first hop), an intermediate user from m_{rh} receives the transmission of user m_{rh-1} that is destined to the base station AP_r. In the next phase (hop), the intermediate user m_{rh} simply forwards the information to the following intermediate user m_{rh+1}. The users continue relaying to the next user until the signal that they have received during the previous phases reaches the base station AP_r. There can be many concurrent transmissions in each phase (within the same cell and in the adjacent cells). The choice of the most appropriate relays to satisfy certain network requirements will be address in next section. The extension to cooperative relaying will be introduced in Section 2.7.

The capacity of the route is then defined as

$$c_{\Re_r} = \min_h c_{\Re_r,r}(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2}) \tag{15}$$

$$c_{\Re_r,r}(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2}) = \log\left(1 + SINR_{m_{r1},m_{r2}}(\mathbf{m}_{i1}, \mathbf{m}_{i2})\right), \tag{16}$$

which is equal to the minimum link capacity on the route. The optimum set of relaying routes is the one that maximizes the capacity of the network and is defined as

$$\{\Re_r\} = \max_{H,\Re_r} C_{\Re}; \quad \text{where } C_{\Re} = \sum_r c_{\Re_r,r}; \ r \in \mathcal{I}. \tag{17}$$

Due to the fact that a node can not receive and transmit the signal simultaneously on the same channel, only a subset of transmissions can be active simultaneously. For that reason a scheduling in different time slots will be introduced in the sequel.

2.3 Two dimensional relaying topology

Full real time knowledge of all cross channel coefficients between all users would provide complete insight into the mutual interference and basis for optimal scheduling. However the channel measurements, messaging overhead and transmission coordination (scheduling) for such a scheme would be very complex. For this reason we suggest an alternative (simplified) scheme that requires only positioning. In this section, we start with a simpler model that requires only the distance from the base station/access point, which is an information relatively easy to acquire and already available in existing systems. High resolution model that requires full information (r,φ) on the user position will be presented in Chapter 4. In this section, we have approximated the continuum of possible

positions of the terminals with a discreet set of values, represented by rings in the cells. For such a scheme we include the impact of the fading into the model.

The cochannel interference can be further reduced by scheduling different transmissions in different subchannels (time slots, frequency bins). All necessary transmissions between all users and their respective access points should be completed in B slots (*scheduling cycle*).

Motivating example: Limited interference relaying topology with same data rate/source.

As an illustration, for the two cells scenario and notation shown in Fig. 5, a possible (feasible) topology is shown in Fig. 5b. For a systematic presentation of the problem, the cell area is divided into concentric rings (e.g. three rings for each cell in Fig. 5), and we assume that the users are located on those rings. It is assumed that there is one cochannel user in each ring and each user has the same amount of information to transmit.

The topology consists of four partial topologies representing transmissions in four consecutive time slots ($B = 4$). In the first time slot (the first partial topology) there are two simultaneous transmissions: packet originating from ring 3 in reference cell r is transmitted from ring 3 to ring 2 and at the same time packet originating from ring 1 of interfering cell i is transmitted from ring 1 to access point AP_i. In the second time slot (the second partial topology), packet originating from ring 3 in cell r is transmitted from ring 2 to AP_r and at the same time packet originating from ring 2 of cell i is transmitted from ring 2 to access point AP_i. Similarly the same notation is used for transmission in time slot 3 and 4.

Limited interference in this context means that the distance between the active receiver (m_{r2}) and the closest point reached by the omnidirectional interfering transmission (m_{i1}), referred to as interfering distance d_i, equals to $d_i d_0$. The minimum interference distance in the previous topologies is $d_i = 2$.

So, in the above example, there will be limited interference transmission for 3 users per each cell in four channels (4 time slots in Fig. 5b), giving the “intercell throughput” 3/4 (3 users/4 time slots), as opposed to the $3/6 = 1/2$ in a conventional TDMA system where each cell uses a half of available channels (slots), and each ring transmits in separate channel (slot).

These four partial topologies together are referred to as a *possible* or *feasible two dimensional* (*time and space*) *topology* and will be represented in the sequel by a given topology index t. For this concept (17) becomes

$$\left\{\Re_r^{(2)}\right\} = \max_{B,H,\Re_r^{(2)}} C_{\Re^{(2)}}; \text{where } C_{\Re^{(2)}} = \sum\nolimits_r c_{\Re_r^{(2)}}; r \in \mathcal{I} \quad (18)$$

and $\Re^{(2)}$ is two dimensional relaying topology to be elaborated in more detail in the next section. For each slot $b = 1,...,B$, parameter $c_{\Re_r^{(2)}}$ is given by corresponding (15)-(16).

TSL program is developed to find the optimum relaying topology for the problem described in the previous example. It performs a search through all feasible topologies, gives users different priorities to transmit, all possible options for relaying and coordinates the transmission between the different cells, to find the best topology in accordance with a given network utility function.

To generalize this modeling and further explain the meaning of the scheduling interval *B*, we use analogy with standard contention graph modeling of MAC layer operation by representing users in the relaying rings from the previous examples as nodes in the network. Designing a MAC protocol can be modeled as a bandwidth allocation problem at the link layer. When considering link layer flows, contention relations between the links can be represented by a *link conflict graph*. In such a graph, vertices represent link flows and edges between vertices denote contention between links, which is the situation where there is interference between either the sender or the receiver of one link with either the sender or the receiver of the other link. A fully connected subgraph in a conflict graph is referred to as a *clique* and *maximal clique*, is a clique not contained in any larger clique. Therefore, a maximal clique represents a “channel resource”, which has a given fixed capacity. The basic requirement for feasibility of a schedule or bandwidth assignment is that the total flow rate in each clique does not exceed the clique’s capacity, subject to the conflict constraints. In addition, the bandwidth allocation should satisfy some performance requirement such as fairness.

Assuming all nodes use omnidirectional antennas to transmit packets in the same shared wireless channel, a link conflict graph can be used to describe the contention relations between link flows, and each maximal clique is treated as a “channel resource” with a given fixed capacity. The capacity of a clique depends on the topology of the network, and the fairness principle under consideration. In the previous example, transmissions within the same clique are shared on the time scale by scheduling transmissions in different time slots. After *B* slots (scheduling cycle) the transmissions can be repeated in the same order.

2.4 Cooperative relaying scheme

In this section, we assume that the users are equipped with cooperative diversity receivers, so they can exploit the broadcast nature of the wireless transmission to cooperate and improve the SNR at the intended receiver. Cooperative multihop networks can improve the performanve via cooperative diversity, reducing the transmission power and increasing the capacity via multiple communication links to a single destination [14], [46].

The Cooperative Relaying Scheme (COOR) proposed in this section works as follows. In the first phase, in the reference cell r the set of intermediate users located in $\mathcal{M}_{r1} = \{m_{r2}, ..., m_{rH}\}$ and AP_r listen to the transmission coming from m_{r1}. In the next phase, the set of intermediate users in $\mathcal{M}_{r2}$ and AP_r listen to the transmission of intermediate user located on m_{r2}. The same scheme continues until the last intermediate user from m_{rH} transmits to AP_r. This transmission model is presented in Fig. 6. Meanwhile in the same cell and in adjacent cells, many concurrent transmissions may take place. We assume that the set of intermediate users in $\mathcal{M}_{rj}$ and AP_r combine the received signals in a diversity receiver [176], so that the equivalent $SINR^{coor}$ at AP_r after H hops is

$$SINR^{coor}_{m_{rh},r}(\mathbf{m}_{i_1}, \mathbf{m}_{i_2}) = \sum_{h=1}^{H} \frac{G_{m_{rh},r} P_{m_{rh},r}}{n_r + \sum_{i \in \mathcal{I}} \sum_{p=1}^{H} G_{m_{ip},r} P_{m_{ip},m_{ip+1}}}, \tag{19}$$

where $P_{m_{rh},r}$ and $P_{m_{ip},m_{ip+1}}$ are the transmission powers of user m_{rh} and m_{ip}, respectively, $G_{m_{rh},r}$ is the channel gain between m_{rh} and AP_r and the same applies for $G_{m_{ip},r}$, and n_r is the background noise power.

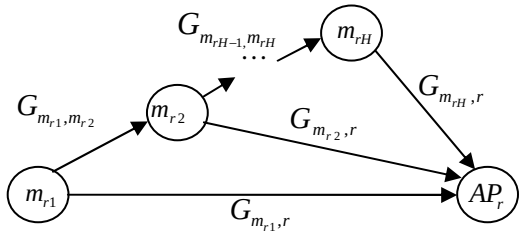

Fig. 6. Cooperative relaying scheme, ([118] [©IEEE 2011]).

The capacity of the route $\Re_r$ is now defined as

$$c_{\Re_r} = \log\left(1 + SINR_{m_{rh},r}(\mathbf{m}_{i1}, \mathbf{m}_{i2})\right). \quad (20)$$

The optimum set of relaying routes is given again by (17) with $c_{\Re_r,r} = c_{\Re_r}$.

2.5 Traffic modeling

In a real network, the traffic will vary in time and space. To model the network traffic dynamics we introduce the following definitions. We denote by n_i the number of rings in cell i and $N = \sum_{i=1}^{N_c} n_i$ the total number of rings in the network where N_c is the number of cells ($N_c = |\mathcal{I}|$). The vector $\boldsymbol{\lambda}_i = (\lambda_{1,i}, ..., \lambda_{n_i,i})$, $\boldsymbol{\lambda}_i \in \mathbf{R}^{n_i}$ defines the amount of generated source traffic by the users situated in the different rings in cell i to be transmitted to the access point AP_i on the uplink. For the same traffic vector $\boldsymbol{\lambda}_i$ the base station can schedule the transmission through different channels (time slots) resulting in temporal and spatial MAC protocol. The overall network traffic on the uplink is defined as $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, ..., \boldsymbol{\lambda}_{N_c}) = (\lambda_1, ..., \lambda_N)$.

The base stations jointly assign an access vector $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_{N_c}) = (a_1, ..., a_N)$ to the different rings to give them permission to transmit, where each component $a_n \in (0,1)$. With a_n = 1 the users from ring n are allowed to transmit otherwise not. In the two cell case $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$, the first half of the coefficients represent the permissions to transmit for the rings in referent cell r and the second half for rings in interfering cell i, $i \in \mathcal{I}_{-r}$. The index pattern is presented in Fig. 7.

2.6 Joint optimization of relaying topology, routing and scheduling

In this section, we define the optimization process by using Network Utility Maximization (NUM) formulation, the definition of utility function and necessary constraints which are result of the physical limitations of the system. The utility function and constraints sets have been further elaborated compared to those in (1) to model MCNs with ICI and traffic awareness.

The examples of topologies presented in Section 2.3 are based on intuition and we need a systematic approach to the system optimization. The optimization of the relaying topology in MCNs should answer the questions “who is transmitting to whom” and “when”, to ensure the best system performance. In order to define the optimization problem, we introduce the following definitions. For $i \in \mathcal{I}$,

We define the topology matrix $\mathbf{T}_i = \left[T(m_{i2}, m_{i1}) \right]$ with $T(m_{i2}, m_{i1}) = 1,$ if node m_{i1} is transmitting to m_{i2} and 0 otherwise, with indexes $m_{i1}, m_{i2} = 0,1,2,...,n_i$, and $m_{i2} < m_{i1}$. Each (m_{i2}, m_{i1}) pair is represented by a specific link index *l* as shown in Fig. 7 for the case of two cells. With this notation the column vector of equivalent (source + relayed) rates in cell *i* becomes

$$\mathbf{x}_i = \left[x_{m_{i2}} \right] \text{ where } x_{m_{i2}} = x_{sm_{i2}} + \sum\nolimits_{m_{i1}} T(m_{i2}, m_{i1}) x_{m_{i1}}, \tag{21}$$

where $x_{sm_{i2}}$ is the source rate of user m_{i2}, $x_{m_{i2}}$ and $x_{m_{i1}}$ are the overall rate of the traffic sent by user m_{i2} and m_{i1} respectively. The overall topology matrix will be formally defined as $\mathbf{T} = diag\left[\mathbf{T}_i \right]$, and $\mathbf{x} = \left[\mathbf{x}_i \right]$ is the concatenated column vector of the overall aggregate rates.

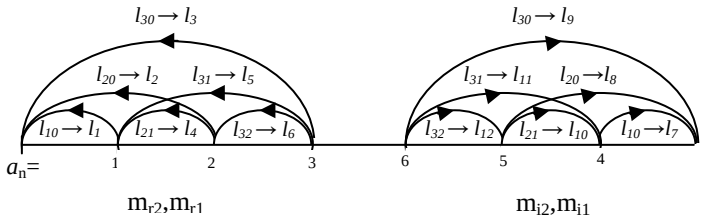

Fig. 7. Link notation.

The routing matrix $\mathbf{R} = [r_{ln}]$ has entries $r_{ln} = 1$ if source *n* (*n = 1,2,..,N*) is using link *l* (*l = 1,2,..,L*) and 0 otherwise. Recall that, parameter *N* is the number of overall rings in the network. Parameters $r_{ln} = r_{lm_{i1}}$ of the routing matrix **R** are calculated as $r_{lm_{i1}} = \bigcup_{m_{i2}} r_{lm_{i2}} T(m_{i2}, m_{i1})$, and the number of links is given by $L = \sum_{m_{i1}} T(m_{i2}, m_{i1})$.

The scheduling set **Π** will be combined with the routing matrix **R** resulting into two dimensional routing protocol characterized by *extended routing matrix* $\mathbf{R}^{(2)} \in \Re^{(2)}$. Depending on the type of scheduler this will result into space time or space frequency routing. By assuming that the scheduling cycle within the maximum clique has *B* steps, the remaining NUM will include:

a) Utility function that characterizes the design objectives

$$U = (1/B) \sum\nolimits_n a_n U_n(x_n) / P_n, \tag{22}$$

where $U_n(x_n) = \log(x_n)$, x_n and P_n are the aggregate rate (source + relays) and aggregate power respectively needed for transmission of information from the source n to the access point, a_n is the access parameter (a_n = 1 indicates that source n is active and a_n = 0 otherwise). We assume that the buffers of the users are infinite, and we maximize the sum of the transmission rates and leave to the internal node fairness policy as a parameter to control the share of the channel between the local and relayed traffic.

b) Constraint $\mathbf{R}^{(2)}\mathbf{x}^{(2)} \le \mathbf{c}^{(2)}(\mathbf{R}^{(2)})$ with the following definitions of extended system parameters

$$\begin{aligned} &\mathbf{x}^{(2)T} = \left(\mathbf{x}^T(1), \mathbf{x}^T(2), ..., \mathbf{x}^T(B)\right); \mathbf{c}^{(2)T} = \left(\mathbf{c}^T(1), \mathbf{c}^T(2), ..., \mathbf{c}^T(B)\right) \\ &\mathbf{R}^{(2)} = diag\left\|\mathbf{R}(b)\right\|,\ b = 1, 2, .., B \\ &\mathbf{R} \in \Re,\ \mathbf{x} \in \mathbf{\Pi} \Rightarrow \mathbf{R}^{(2)} \in \Re^{(2)}, \end{aligned} \tag{23a}$$

where **c** are the logical link capacities calculated as discussed in Section 2.2 for conventional relaying (CONR) and Section 2.4 for cooperative relaying (COOR). They capture the functional dependency of transmission power and interference level in the network and, $\mathbf{R}(b)$ is the partial routing matrix (subset of active users on matrix **R** in slot b).

c) Each component of the set of feasible routes in $\Re^{(2)}$ should provide directional connection for each terminal to the corresponding access point. This means that the sequence of links generated in a clique cycle must provide connection for all terminals to the corresponding access point. To define this constraint explicitly we introduce the link hopping distance h_l and the vector $\mathbf{h} = (h_1, ..., h_L)$. Parameter h_l represents the number of rings that link l is hopping over, from its transmitter/receiver to the corresponding receiver/transmitter. Similarly, the source hopping distance vector is denoted as $\mathbf{d} = (d_1, ..., d_N)$, where d_n is the hopping distance between the source and the access point. The sum of link hopping distances on the route from source n to the access point should be equal to the source hopping distance

$$\sum\nolimits_b \mathbf{R}^T(b)\mathbf{h}(b) = \mathbf{d}. \tag{23b}$$

d) The overall transmission rate is defined as

$$(\mathbf{I} - \mathbf{T})\mathbf{x} = \mathbf{x}_s \leftarrow\ x_{m_{i2}} = x_{sm_{i2}} + \sum\nolimits_{m_{i1}} T(m_{i2}, m_{i1}) x_{m_{i1}}. \tag{23c}$$

Formulation of the problem defined by equations (22), (23a)-(23c) can be summarized as:

$$\begin{aligned} \underset{\mathbf{x},\mathbf{T}}{\text{maximize}} \;& U \\ \text{subject to} \;& \mathbf{R}^{(2)}\mathbf{x}^{(2)} \le \mathbf{c}^{(2)}(\mathbf{R}^{(2)}) \\ & \textstyle\sum_b \mathbf{R}^T(b)\mathbf{h}(b) = \mathbf{d} \\ & (\mathbf{I}-\mathbf{T})\mathbf{x} = \mathbf{x}_s \\ & \mathbf{R} \in \Re,\ \mathbf{x} \in \mathbf{\Pi} \ \Rightarrow \mathbf{R}^{(2)} \in \Re^{(2)}. \end{aligned} \tag{24}$$

The results presented in Section 2.8 show that the optimal relaying topology depends on a given access vector **a.** For this reason for a given spatial traffic distribution the access vector should be varied to provide

$$E(\mathbf{a}(t)) = \boldsymbol{\lambda}(t). \tag{25}$$

The set of feasible topologies obtained by TSL for a given **a** will be denoted as $\Im^{(2)}$. These topologies are represented formally as a collection of partial topologies (links) generated in $B = 1,...,b$ slots $\Im^{(2)} = \bigcup_b PT^{(b)} = \bigcup_b \mathcal{L}^{(b)}$, where $\mathcal{L}^{(b)}$ indicates the set of active links in slot *b*.

The algorithm works as follows:

1. *Calculate the access vector **a** for a given traffic distribution **λ** by using (25).*
2. *Use TSL to generate the set of feasible candidate topologies $\Im^{(2)}$ for **a.***
3. *For each topology in the set $\Im^{(2)}$:*
4. *Use CVX [126] within TSL to optimize the source rates.*
5. *Calculate the aggregate powers needed for users to deliver the information.*
6. *Calculate the utility using (22).*
7. *Go to 3) until the best utility is obtained.*

2.7 Throughput-power trade-off with ICI awareness

In interference limited systems the throughput and power optimization becomes non trivial due to network-wide coupling of the power allocation. Different approximations are presented in [141-143] to convexify the problem. Recently, Papandriopoulus *et al.* [144] proved how the problem can be rendered convex, and that the global optimum can be found through standard Lagrange duality.

In this section, we extend our previous optimization problem (24) to include the optimization of the power allocation. We define a new utility function that strikes a balance between maximizing the sum rate utility U_1, and minimizing the cost (total power consumption) U_2 to model the trade-off between throughout and power allocation in MCNs with ICI awareness. The constraints (23a)-(23c) defined for problem (24) are now extended to include the dependence on the power. This framework will be evaluated in Section 2.8 by using the conventional relaying scheme (CONR) described in Section 2.2 and cooperative diversity relaying scheme (COOR) described in Section 2.4. The optimization of the power allocation further reduces the effects of the interference and prolongs the battery life of the relaying users. The resulting intercell interference management (I^2M) protocols by COOR and CONR schemes will be referred as I^2M-COOR and I^2M-CONR, respectively.

The throughput-power trade-off is defined by assigning weights w_1, w_2 to the utility U_1 and cost U_2

$$\begin{aligned} U &= w_1 U_1 - w_2 U_2 \\ U_1 &= \frac{1}{B}\sum\nolimits_n a_n U_n(x_n); \;\; U_2 = \sum\nolimits_n P_n, \end{aligned} \tag{26}$$

where U_1 is the sum rate utility, and U_2 is the overall power consumption. The resulting optimization problem is defined as utility maximization with pricing

$$\begin{aligned} \underset{\mathbf{x},\mathbf{T},\mathbf{P}}{\text{maximize}} \quad & w_1 \frac{1}{B}\sum\nolimits_n a_n U_n(x_n) - w_2 \sum\nolimits_n P_n \\ \text{subject to} \quad & \mathbf{R}^{(2)}\mathbf{x}^{(2)} \le \mathbf{c}^{(2)}\left(\mathbf{R}^{(2)}, \mathbf{P}^{(2)}\right) \\ & \sum\nolimits_b \mathbf{R}^T(b)\mathbf{h}(b) = \mathbf{d} \\ & (\mathbf{I} - \mathbf{T})\mathbf{x} = \mathbf{x}_s \\ & \mathbf{R} \in \Re,\ \mathbf{x} \in \mathbf{\Pi} \ \Rightarrow \mathbf{R}^{(2)} \in \Re^{(2)} \\ & \mathbf{P}^{\min} \le \mathbf{P} \le \mathbf{P}^{\max}, \end{aligned} \tag{27}$$

where w_1, w_2 are the weights controlling the trade-off between the throughput and the power allocation, a_n is the component of the access vector assigned by the BS, U_n is defined as $U_n(x_n) = \log(x_n)$, P_n is the overall power consumed by user from ring n. The first constraint has been modified with respect to (24) as $\mathbf{c}^{(2)}\left(\mathbf{R}^{(2)}, \mathbf{P}^{(2)}\right)$ to include the dependence with the power. $\mathbf{R}$ is the routing matrix, b is the index of time slot, $\mathbf{x}$ is the vector of source rates, $\mathbf{T}$ is the topology matrix,

$\Re$ is the overall set of feasible routes, $\mathbf{P}^{\min}$ and $\mathbf{P}^{\max}$ are the minimum and maximum values of the trasmission power, respectively.

The optimization is solved in three steps:

1. The topologies are generated for a given access vector a by TSL.
2. To render the problem formulation (27) convex, we apply an exponential variable transformation as in [144] $x_n \leftarrow e^{\tilde{x}_n}$ and $P_n \leftarrow e^{\tilde{P}_n}$ along with a log-transformation of the rate constraints to arrive at

$$\begin{aligned} &\text{maximize} \quad w_1 \sum_n a_n U_n(e^{\tilde{x}_n}) - w_2 \sum_n e^{\tilde{P}_n} \\ &\text{subject to} \quad \log(\sum_n e^{\tilde{x}_n} \sum_l r_{ln}) \le \log(\sum_l c_l(e^{\tilde{P}_l})), \end{aligned} \tag{27a}$$

   where

$$P_n = \sum_{l \in \bigcup_b \mathcal{L}^{(b)}(n)} P_l$$

   and $\mathcal{L}^{(b)}(n)$ is the set of links used by node n in slot b. After the *log*-transformation the constraint set is now convex. As the utility function used is (log, x)-concave, the transformed problem (27a) is convex, see [144, theorem 2]. Under these conditions problem (27a) can be solved to optimality via standard Lagrange duality.
3. By Dual decomposition the optimization problem is decomposed in the following subproblems:

$$\begin{aligned} D_1(\tilde{\boldsymbol{\mu}}) &= \max_{\tilde{\mathbf{x}} \ge 0} \left\{ L_{\tilde{x}}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}}) \triangleq \gamma_1 \frac{1}{B} \sum_n a_n U_n(e^{\tilde{x}_n}) - \sum_n e^{\tilde{x}_n} \sum_l r_{ln} \tilde{\mu}_l \right\} \\ &= \sum_n \max_{\tilde{\mathbf{x}} \ge 0} \left\{ L_{\tilde{x}}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}}) \triangleq \gamma_1 \frac{1}{B} a_n U_n(e^{\tilde{x}_n}) - e^{\tilde{x}_n} \sum_l r_{ln} \tilde{\mu}_l \right\} \end{aligned} \tag{28}$$

$$D_2(\tilde{\boldsymbol{\mu}}) = \max_{\tilde{\mathbf{P}}} \left\{ L_{\tilde{P}}(\tilde{\mathbf{P}}, \tilde{\boldsymbol{\mu}}) \triangleq \sum_l c_l(e^{\tilde{P}_l}) \tilde{\mu}_l - \gamma_2 \sum_n e^{\tilde{P}_n} \right\}, \tag{29}$$

where D_1 is the routing subproblem in the network layer, D_2 is the power allocation subproblem, $\tilde{\mathbf{x}}$ is the vector of source rates, $\tilde{\mathbf{P}}$ is the vector of allocated powers and $\tilde{\boldsymbol{\mu}}$ is the vector of Lagrange multipliers.

Due to the strict concavity of the partial Lagragian functions $L_{\tilde{x}}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\mu}})$ and $L_{\tilde{p}}(\tilde{\mathbf{P}}, \tilde{\boldsymbol{\mu}})$, by [145, Prop. 6.1.1] the dual functions $D_1(\tilde{\boldsymbol{\mu}})$ and $D_2(\tilde{\boldsymbol{\mu}})$ are differenciable everywhere. The dual function can be written as $D(\tilde{\boldsymbol{\mu}}) = D_1(\tilde{\boldsymbol{\mu}}) + D_2(\tilde{\boldsymbol{\mu}})$. We can solve the dual problem

$$\begin{aligned} &\text{minimize} && D(\tilde{\boldsymbol{\mu}}) \\ &\text{subject to} && \tilde{\boldsymbol{\mu}} \geq 0 \end{aligned}$$

using the gradient-descendent method where the gradient of $D(\tilde{\boldsymbol{\mu}})$ can be shown to be

$$\phi_l(\tilde{\mu}_l) = \sum_l c_l(e^{\tilde{P}_l^*})\tilde{\mu}_l - \sum_n e^{\tilde{x}_n^*} \sum_l r_{ln}\tilde{\mu}_l \tag{30}$$

with $\tilde{x}_n^*$ and $\tilde{P}_l^*$ the optimal solutions in (28) and (29). Given the locally optimal solutions $\tilde{x}_n^*(t)$ and $\tilde{P}_l^*(t)$, the algorithm updates $\tilde{\mu}_l$ as follows

$$\tilde{\mu}_l(t+1) = \left[\tilde{\mu}_l(t) - \beta(t)\phi_l(\tilde{\mu}_l)\right]^+, \tag{31}$$

where $\beta(t)$ is the appropriate step-size in iteration t. The algorithm for the optimization problem in (27) iteratively updates the dual variable by (31), and solves the subproblems (28) and (29) until the globally optimal solutions are achieved.

The dual decomposition method leads to a distributed algorithm, where user keeps track on its own transmission rate and power.

The Intercell Interference Management (I^2M) algorithm works as follows:

1. *Generate the feasible set of topologies $\Im^{(2)}$ for a given access vector* **a** *by using TSL program.*
2. *For each topology in the set $\Im^{(2)}$:*
3. *Solve subproblem (28) and (29) by using CVX [126] with the capacity defined by (15) for I^2M-CONR and (20) for I^2M-COOR.*
4. *Update the dual variable* $\tilde{\boldsymbol{\mu}}$ *by (31). Go to 3) until the subproblems converge to the optimum solutions* $\tilde{\mathbf{x}}^*$ *and* $\tilde{\mathbf{P}}^*$.
5. *Evaluate the objective function in (26) for the optimal values* $\tilde{\mathbf{x}}^*$ *and obtained in 3), by using certain weights w_1 and w_2. Go to 2) until we find the optimal values for all feasible topologies.*
6. *As result, the optimum topology* $\mathbf{T}^*$, *the optimal* $\tilde{\mathbf{x}}^*$ *and* $\tilde{\mathbf{P}}^*$ *are obtained for a given access vector* **a**.

The optimum throughput is defined as

$$Thr^* = \frac{1}{B}\sum_n a_n x_n^*, \tag{32}$$

where $\mathbf{x}^*$ is the optimum vector of source rates.

2.8 Performance evaluation

In this section, we provide several numerical examples to illustrate the performance of the proposed algorithms. We consider the scenario presented in Fig. 5 and assume that the users are situated in the line that connects both BS, which represents the worst case in terms of interference.

2.8.1 Joint optimization of relaying topology, routing and scheduling

We start by considering the conventional relaying scheme (CONR) and calculate the link capacities $c_r(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2})$ as specified in Section 2.2. For this calculus we need the channel gains $G_{m_{i1},m_{r2}}$ and $G_{m_{i1},m_{i2}}$ that depend on distance between the transmitter and receiver, and fading. While the analysis is general, for simplicity in this section we model only propagation losses as $G_{m_{i1},m_{r2}} \sim 1/d^{\alpha}_{m_{i1},m_{r2}}$ where $d_{m_{i1},m_{r2}}$ is the distance between the interfering transmitter in ring m_{i1} and reference receiver in ring m_{r2}, and α is the propagation constant. In the simulations we use $\alpha = 4$. The calculation of $d_{m_{i1},m_{r2}}$ is straightforward from the geometry presented in Fig. 5a.

In Figs. 8, 9 and 10 we present the utility for different access vectors **a** versus the topology index (*t*) for the scenario presented in Fig. 5a. The topology index (*t*) represents each feasible topology in the set $\Im^{(2)}$ obtained by TSL for a given **a**, and defines a certain combination of the active links. The ring and link notation used for the access vectors $\mathbf{a}(a_1,...,a_n)$, $a \in (0,1)$ is shown in Fig. 7 where *n* corresponds to the index of the ring.

In Fig. 8, we present the utility versus *t* for **a** = [001001]. In this case, users located at the border of the cell are allowed to transmit. Different values of the utlity are obtained for different feasible topologies *t*. The maximum utility is obtained for $t = 100$ ($u_{100} = 0.4366$) defined by the set of links $\mathcal{L} = \{\{l_{12}\},\{l_3, l_8\}\}$ where the first subset corresponds to the partial topology in the first slot and the second subset to the second slot.

As an illustration, we present in Fig. 8a the transmission pattern that corresponds to this optimum topology. We can also see that the number of feasible topologies obtained by TSL for this **a** is 106.

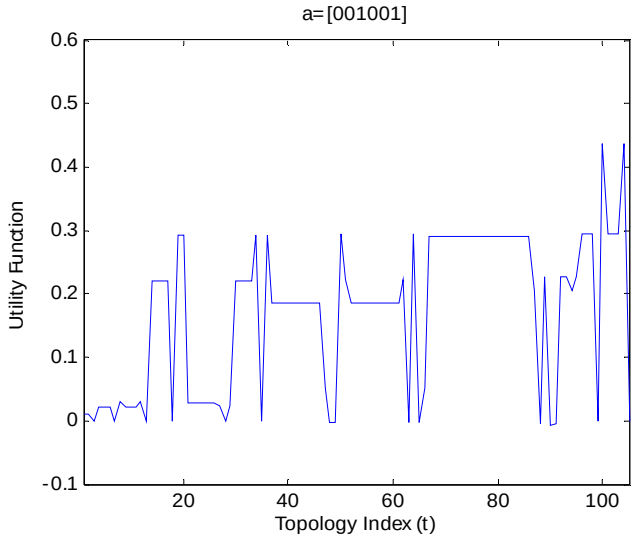

Fig. 8. Utility function versus topology index for access vector a = [001001]. The maximum utility corresponds to topology index *t* = 100 (U_{100} = 0.4366) defined by the set of links $\mathcal{L} = \{\{l_{12}\},\{l_3, l_8\}\}$, ([116] [©IEEE 2009]).

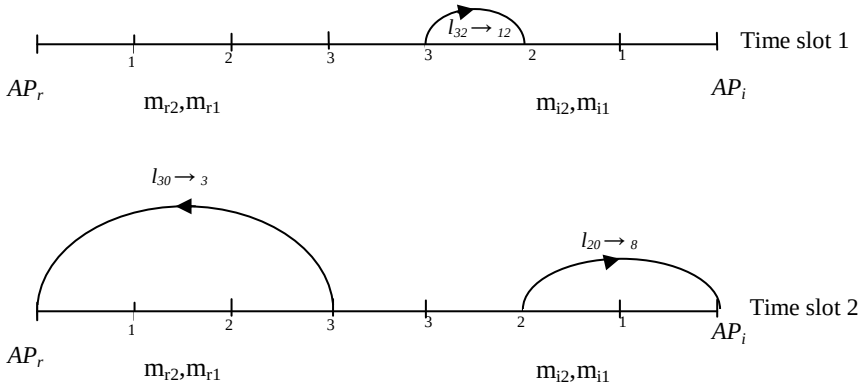

Fig. 8a. Representation of the transmission pattern defined by the topology index *t* = 100.

In Fig. 9, access to 3 rings is allowed (**a** = [011010]) and we can see that the maximum utility is obtained for t = 400 ($u_{400} = 0.5888$). The optimum topology is now given by the set of links $\mathcal{L} = \{\{l_6\},\{l_4\},\{l_1,l_8\}\}$. As we increase the interference level by allowing more concurrent transmissions (activating more rings in the access vector) we can see that the improvement in the system performance is higher. We should also notice that the number of feasible topologies (t) generated changes for different access vectors **a**. This demonstrates that for different traffic distribution vectors, the algorithm outputs different optimum topologies. In this example, we have obtained near 700 feasible topologies, and the value of the utility significantly changes depending on the topology. As the utility includes the logarithm of the rates, negative value of the utility is obtained when the rates are low ($0 < \mathbf{x} \leq 1$). More details about the different topologies for this **a** are provided in Table 1 where $\mathcal{L}^{(b)}$ indicates the active links (active = 1, non active = 0) in slot b.

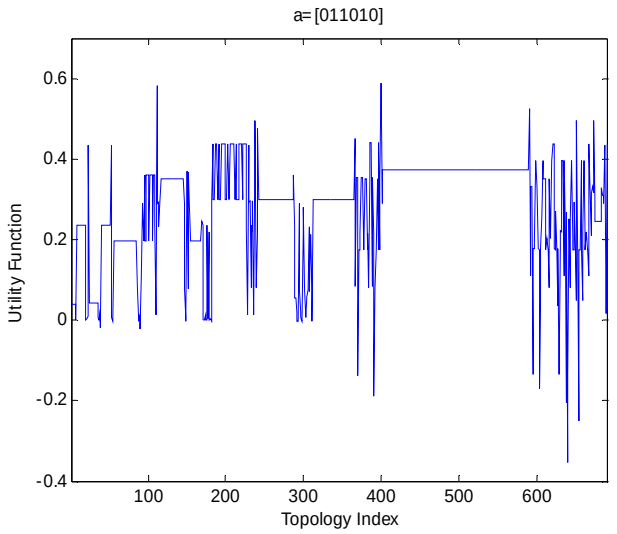

Fig. 9. Utility function versus topology index for access vector a= [011010]. The best utility corresponds to topology index t = 400 (u_{400} = 0.5888), defined by the set of links $\mathcal{L} = \{\{l_6\},\{l_4\},\{l_1,l_8\}\}$, ([116] [©IEEE 2009]).

Table 1. Representation of different topologies for the access vector a = [011010], ([116] [©IEEE 2009])

(*t*, *b*)	$\mathcal{L}^{(b)}$	Utility Function
(1,1)	{0,1,0,0,0,0,0,0,0,0,0,0}	
(1,2)	{0,0,1,0,0,0,0,0,0,0,0,0}	
(1,3)	{0,0,0,0,0,0,0,1,0,0,0,0}	0.04
(7,1)	{0,1,1,0,0,0,0,1,0,0,0,0}	0.0009
(62,1)	{0,0,0,0,1,0,0,0,0,0,0,0}	
(62,2)	{0,1,0,0,0,0,0,0,0,0,0,0}	
(62,3)	{1,0,0,0,0,0,0,0,0,0,0,0}	
(62,4)	{0,0,0,0,0,0,0,0,0,1,0,0}	
(62,5)	{0,0,0,0,0,0,1,0,0,0,0,0}	0.1966
(400,1)	{0,0,0,0,0,1,0,0,0,0,0,0}	
(400,2)	{0,0,0,1,0,0,0,0,0,0,0,0}	
(400,3)	{1,0,0,0,0,0,0,1,0,0,0,0}	0.5888
(402,1)	{0,0,0,1,0,0,0,0,0,0,0,0}	
(402,2)	{1,0,0,0,0,0,0,0,0,0,0,0}	
(402,3)	{0,0,0,0,0,1,0,0,0,0,0,0}	
(402,4)	{0,0,0,1,0,0,0,0,0,0,0,0}	
(402,5)	{1,0,0,0,0,0,0,0,0,0,0,0}	
(402,6)	{0,0,0,0,0,0,0,0,0,1,0,0}	
(402,7)	{0,0,0,0,0,0,1,0,0,0,0,0}	0.3745
(640,1)	{1,0,0,0,0,1,0,0,0,1,0,0}	
(640,2)	{0,0,0,1,0,0,1,0,0,0,0,0}	
(640,3)	{1,0,0,0,0,0,0,0,0,0,0,0}	-0.3529

- For topology index $t = 1$, the topology is given by the set of links $\mathcal{L} = \{\{l_2\},\{l_3\},\{l_8\}\}$ split in three time slots (*b*=1, 2, 3). These are the links that connect the users activated by **a** directly to their respective access points. This topology corresponds to conventional system in TDMA mode, and the utility is $u_1 = 0.04$, which is far from the best utility for this **a** ($u_{400} = 0.5888$). The explanation to this behaviour is that these links require that the users transmit with the highest power.
- For topology index $t = 7$, the topology is given by the set of links $\mathcal{L} = \{l_2, l_3, l_8\}$ where all users transmit simultaneous (one slot) using the links that connect them directly to their access point. This topology corresponds to conventional system in CDMA mode. The utility for this topology is $u_7 = 0.009$ which is low because the interference level is high, resulting in low capacity of the links.

- For the topology index $t = 62$, the following set of links $\mathcal{L} = \{\{l_5\},\{l_2\},\{l_1\},\{l_{10}\},\{l_7\}\}$ is activated. These links required less transmission power and generate no mutual interference. The utility now is higher than in the two previous cases ($u_{62} = 0.1966$), but still is not the optimum because we are using 5 times slots.
- In the topology with index $t = 400$, we get the optimum topology with utility $u_{400} = 0.5888$, defined by the set of links $\mathcal{L} = \{\{l_6\},\{l_4\},\{l_1, l_8\}\}$. The users transmit with low power and use 3 time slots. The optimum topology gives an utility significantly higher than for conventional systems in TDMA ($u_1 = 0.04$) and CDMA ($u_7 = 0.009$).
- In the topology with index $t = 402$, the topology is defined by $\mathcal{L} = \{\{l_4\},\{l_1\},\{l_6\},\{l_4\},\{l_1\},\{l_{10}\},\{l_7\}\}$. We need 7 time slots for the transmission. All users transmit doing relaying and they transmit their own information and the relaying information in different time slots. The utility is $u_{402} = 0.3745$. We have the same utility for indexes $t = 402$-590 that corresponds to different combinations of the same set of links as we can see in Fig. 9.
- In the topology with index $t = 640$, we have the smallest utility $u_{640} = -0.3529$. The topology is defined by the set of links $\mathcal{L} = \{\{l_1, l_6, l_{10}\},\{l_4, l_7\},\{l_1\}\}$. Although the transmission power is low, we have three simultaneous transmissions in the first time slot and two in the second slot which results in high interference level.

In Fig. 10, we present the utility when 4 rings have permission to transmit (**a** = [101110]) and an additional improvement is obtained. The best utility is obtained for topology index $t = 602$ ($u_{602} = 0.6991$), defined by the set of links $\mathcal{L} = \{\{l_6\},\{l_4\},\{l_1\},\{l_{10}\},\{l_7\}\}$. It is worth noticing that as traffic increases, the topology that contains isolated and short range transmisions is favored. In Fig. 11 we present the overall capacity, defined as

$$\sum_{l}\sum_{b} c_l(b) = \sum_{l,b} c^{(2)},$$

for the same **a** as before**.** We can see that the overall capacity of the system improves by a factor 6 compared with the worst case.

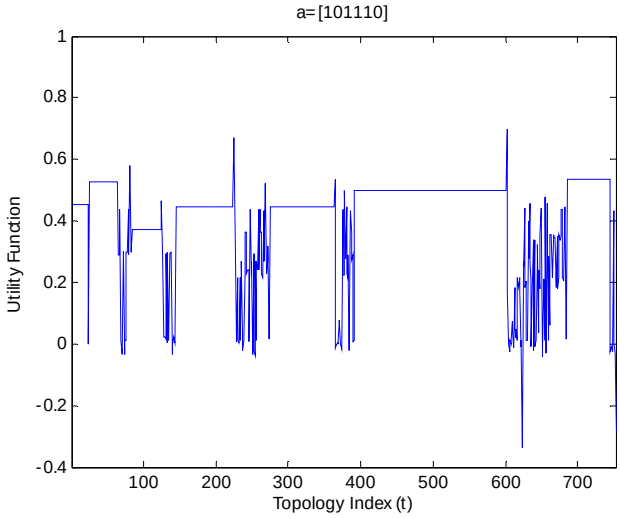

Fig. 10. Utility function versus topology index for access vector a = [101110]. The best utility corresponds to topology index *t* = 602 (u_{602} = 0.6991), defined by the set of links $\mathcal{L} = \{\{l_6\},\{l_4\},\{l_1\},\{l_{10}\},\{l_7\}\}$, ([116] [©IEEE 2009]).

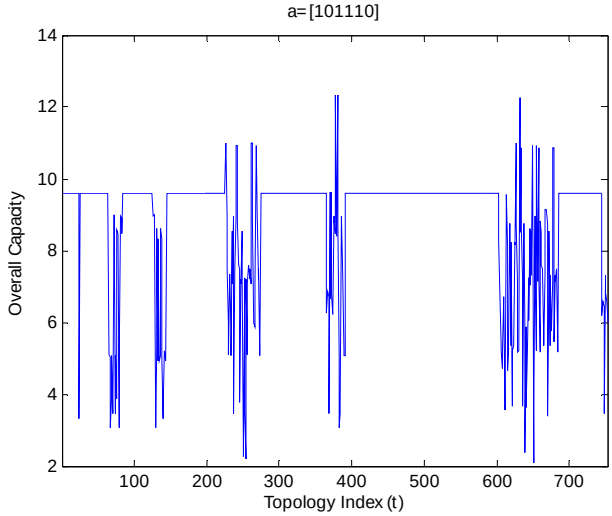

Fig. 11. The overall capacity versus topology index for access vector a= [101110], ([116] [©IEEE 2009]).

2.8.2 Throughput-power trade-off with ICI awareness

In the results presented in this subsection, we have considered the same scenario as in the previous subsection and extended those results to include cooperative relaying scheme (COOR). The link capacities $c_r(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2})$ are calculated as specified in Section 2.2 and Section 2.4 for CONR and COOR schemes, respectively.

In Figs. 12, 13 and 14 we present the utility defined by (26), throughput and power consumption respectively versus the topology index t for I^{2}M-CONR and I^{2}M-COOR schemes. The traffic distribution in the network is given by access vector **a** = [001001] and the weights for optimization are $w_1 = 1$ and $w_2 = 0.2$. The optimum utility is obtained for topology index $t = 18$ given by the following set of links $\mathcal{L} = \{\{l_3, l_{12}\}, \{l_{10}\}, \{l_7\}\}$ where the first subset corresponds to the partial topology in the first slot and the same applies for the second and third subset.

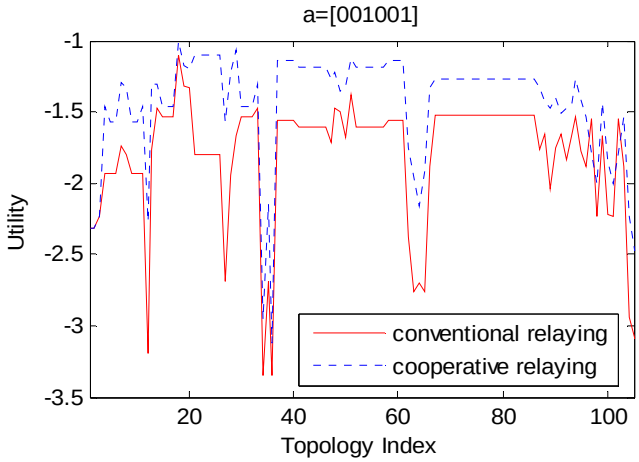

Fig. 12. Utility versus topology index for a = [001001], ([118] [©IEEE 2011]).

We can also see that the optimum topology has the larger utility because the power consumption is quite low (about 4), and the throughput is about 0.5. The topology index that obtains the highest rates is topology index $t = 12$, but the power consumption for this topology is very high which results in low value of the utility. We can see an improvement of at least 4 times in terms of network throughput and the power consumption is reduced 1.5 times by using I^{2}M-COOR

with respect to I^2M-CONR scheme. The variation in the power consumption for the optimum topology is 3 times inferior with respect to non-optimum options, and the variation in the throughput is about 6 times with respect to the maximum one.

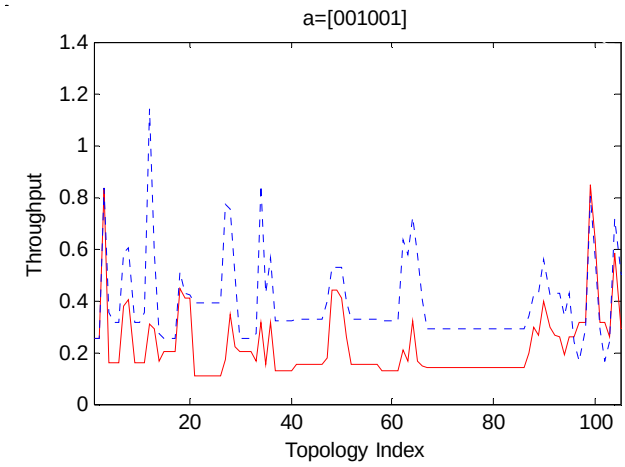

Fig. 13. Throughput versus topology index for a = [001001], ([118] [©IEEE 2011]).

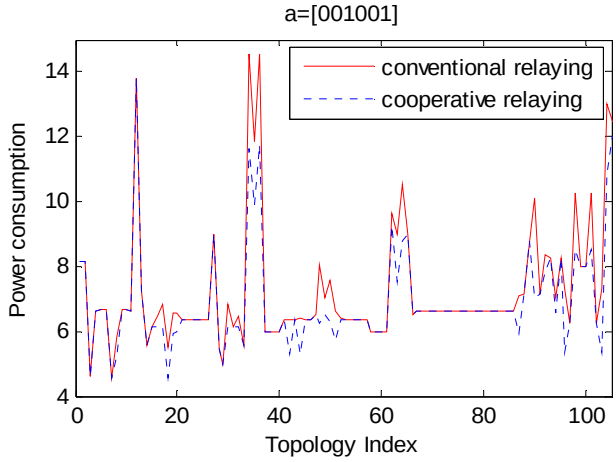

Fig. 14. Power consumtpion versus topology index for a = [001001], ([118] [©IEEE 2011]).

In Fig. 15, we present the trade-off between U_1 and power for the optimum topology (topology index 18), versus w_2 with $w_1 = 1$. We can see that for $w_2 = 0.13$ the power consumption by cooperative relaying is 4.83 compared to 6.62 by conventional relaying. We can see also from the graphic that by reducing the power for a factor less than 2, U_1 will be reduced by factor one (throughput approximately by factor 10 due to the logarithmic scale). By increasing w_2 for a fixed value of w_1, the power is reduced because we are increasing the weight of its minimization, until the power gets the minimum value. At the same time, by highly decreasing the power we also decrease U_1. Due to the level of interference in the network, by slightly decreasing the power (from 7 to 5) we are still keeping good value of U_1 because this power reduction is reducing the effects of the intercell interference. If the power is lower than certain threshold (P = 5) then U_1 gets to its minimum values too.

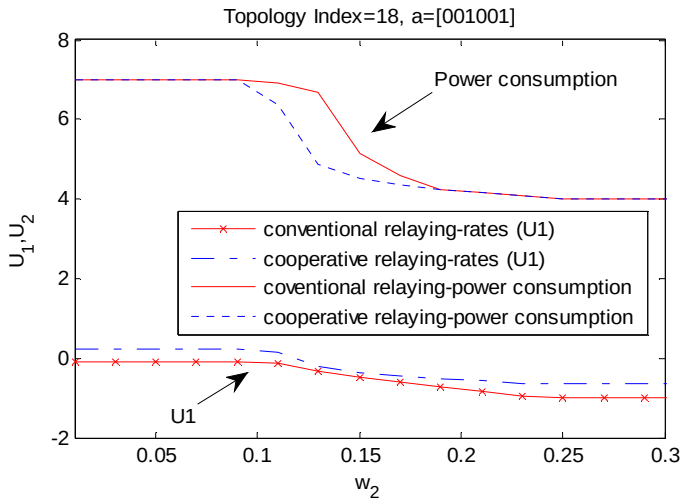

Fig. 15. U_1 and power consumption U_2 versus w_2 for topology index t = 18, ([118] [©IEEE 2011]).

In Fig. 16 and 17 we present the utility and throughput respectively versus the topology index when the traffic distribution in the network is changed and is given by access vector **a** = [001110], for $w_1 = 1$ and $w_2 = 0.2$. We can see again significant improvement by using I^{2}M-COOR with respect to I^{2}M-CONR.

In this case, there are more options for the feasible topologies and the optimum topology is given by t = 120 ($\mathcal{L} = \{\{l_7\},\{l_{10}\},\{l_7\},\{l_6\},\{l_4\},\{l_1\}\}$).

In Fig.18 we show the improvement in the power consumption obtained for this optimum topology by using I^{2}M-COOR, we can see from the graphic that the reduction in power consumption for w_2 from $w_2 = 0.01$ to 0.28 is about 3 times with respect to I^{2}M-CONR.

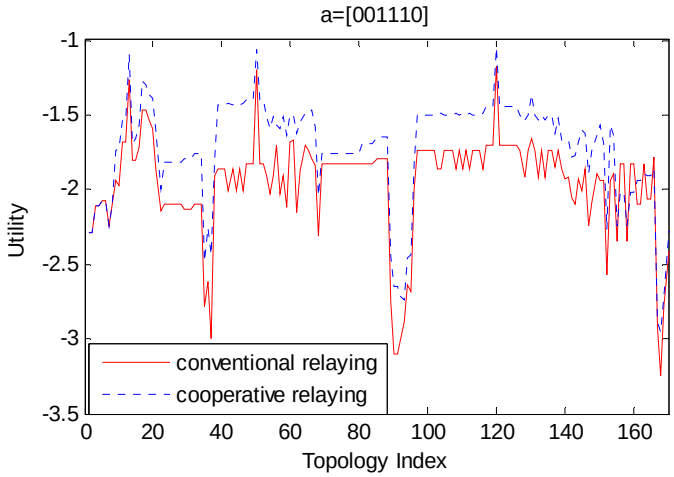

Fig. 16. Utility versus topology index for access vector a = [001110], ([118] [©IEEE 2011]).

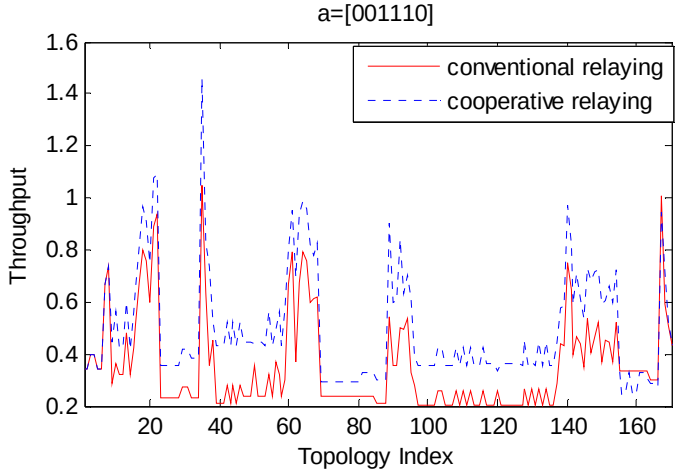

Fig. 17. Throughput versus topology index for a = [001110], ([118] [©IEEE 2011]).

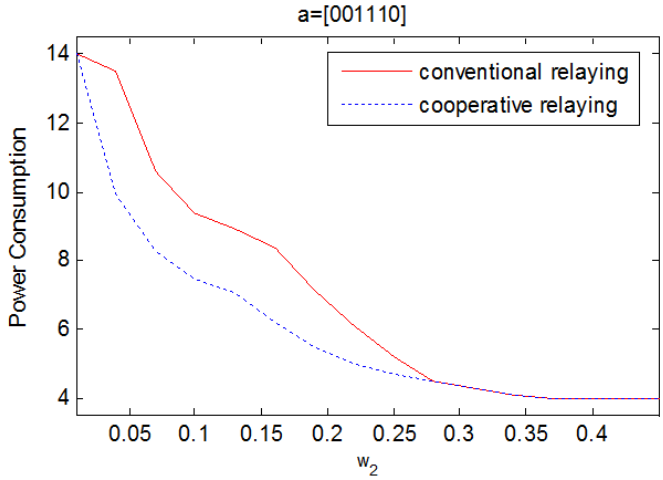

Fig. 18. Power consumption versus w_2 for topology index t = 120, ([118] [©IEEE 2011]).

2.9 Chapter summary

In this chapter, we have presented an optimization framework for relaying topology control which is aware of the ICI requiring coordinated action between the cells and resulting in multicell jointly optimal relaying topology. The algorithm jointly chooses the relaying topology and scheduling in the adjacent cells to minimize the system performance degradation due to ICI. The utility function under consideration includes data rate, power consumption and delay. The cooperation between the cells does not require additional exchange of messages except to synchronize the slots and follow the scheduling as specified by the protocol.

Numerical results demonstrate that a reconfigurable relaying topology provides the network utility improvements and presents the framework for quantifying these improvements for spatially and temporally varying traffic. As the traffic increases in the network, higher utility is obtained for topologies that favour isolated and short range transmissions. The optimum topology gives a utility 10 times higher than conventional TDMA and 60 times higher than CDMA systems for certain traffic distributions. The improvement obtained in the overall capacity can reach up to 6 times with respect to nonoptimum solutions. The overall optimization problem is defined and solved by using the combination of a

new topology search TSL (Oulu) program and convex optimization CVX program [126].

In practice the overall optimization program may be implemented in different ways depending on the specification of the available information about the network and acceptable amount of overhead traffic allowed for necessary messaging for optimization purposes. An interesting option is also to run TSL program in centralized way because it requires the exchange of information between the access points and CVX in distributed way using the Lagrange dual problem. The extension of this approach would be to precompute optimum topologies for all possible access vectors **a** and for a given **a** use a look up table to find the optimum topology. For a given topology, rate allocation optimization may be run by using distributed approach.

As an extension of the previous framework, we have included the power allocation into the optimization problem. By adding weights in the utility function we analyze the trade-off between throughput and power consumption. We consider conventional relaying (CONR) and cooperative relaying (COOR) schemes and presented two intercell interference management protocols based on those schemes referred to as I^2M-CONR, and I^2M-COOR, respectively. Numerical results show an improvement of at least 4 times in terms of throughput and a reduction on power consumption up to 3 times by using I^2M-COOR compared to I^2M-CONR for different traffic distributions.

3 Sequential genetic algorithm for dynamic topology reconfiguration in MCNs

In a real network, where the traffic distribution changes (in time and space) due to the changes in available link capacities and load demands, the network topology should be reconfigured dynamically to track the variations in the network and guarantee good network performance. When considering a multicell scenario with nonuniform traffic distribution in MCNs, the search for the optimum topology becomes an NP-hard problem. For such problems, exact algorithms based on exhaustive search are only useful for small models, so heuristic algorithms such as genetic algorithms (GAs) must be used in practice.

Topology control in this scenario involves the computation of new topologies, rerouting and rescheduling to dynamically optimize the network performance. GAs have shown to be efficient in solving problems where the space of all potential solutions is too large to be searched exhaustively in any reasonable amount of time. GAs provide optimal or good suboptimal results in a short period of time [146].

For this purpose, we present a novel Sequential Genetic Algorithm (SGA) for joint relaying topology, routing and scheduling optimization in multihop cellular network aware of the intercell interference and the spatial traffic distribution dynamics. SGA dynamically adjusts the relaying topology to the traffic variations in the network, reducing considerably the number of operations required by exhaustive search, and improving the network scalability.

The topologies are encoded as a set of chromosomes and, special crossover and mutation operations are developed to search for the optimum topology. The performance is measured by a fitness function that includes throughput, power consumption and delay. Improvement in the fitness function is sequentially controlled as newer generations evolve and whenever the improvement is sufficiently increased the current topology is updated by the new one having higher fitness. As a result, a specific encoding and fitness control in a SGA is developped for relaying topology update.

A comprehensive scenario is considered including uplink/downlink transmission and network coding [147] to reduce the number of slots needed for the users to complete their transmission. The procedure, that reconfigures the relaying topology based on observation of the temporal traffic in the network, will be referred to as Traffic Cognitive Topology Control $(TC)^2$ and network based on this approach as Traffic Cognitive Network (TCN).

Numerical results show that SGA provides both high performance improvements in the system and fast convergence (at least one order of magnitude faster than exhaustive search) in a dynamic network environment. Different options are discussed for the initialization of the algorithm to show the robustness of SGA with respect to the initial state of the network i.e., the algorithm can be initiated with any feasible topology. Different options for implementation, discussion on system dynamics and comparison with other heuritics are also provided.

The rest of the chapter is organized as follows. An overview and background on genetic algorithms is given in Section 3.1. In Section 3.2, the system model and assumptions are presented. This includes the network model, the physical layer model and traffic dynamics. Section 3.2 presents the bidirectional relaying topology with network coding. Section 3.4 defines the system optimization while Section 3.5 presents the new topology search program based on genetic algorithm SGA-TSL. Traffic Cognitive Topology Control algorithm $(TC)^2$ is developed in Section 3.6. Details on algorithm implementation are also provided in this section. Section 3.7 presents numerical examples to demonstrate the efficiency of the algorithms, discussion on system dynamics and other heuristics. Finally, Section 3.8 concludes the chapter.

3.1 Overview and background

A comprehensive survey of the previous work on interference avoidance topology control is presented in [Section 1.2.3, Chapter 1]. The potential for network performance improvement by topology reconfiguration in time-varying environments is emphasized and the limited research work done so far in topology control for MCNs is outlined.

Recently, there have been increasing interests in applying biologically inspired approaches to topology control in MSNs and MANETs [60]-[62]. However, the problems associated with these types of networks are different from those of cellular networks as already mentioned in [Section 1.2.3, Chapter 1]. For this reason, in this chapter we present a sequential genetic algorithm for dynamic topology control in MCNs. The efficiency of our algorithm is achieved by considering dynamic joint optimization of relaying topology, routing (power) and inter relay scheduling in multicell MCNs.

In the remaining of this subsection, we provide an overview on GAs and their efficiency in solving dynamic problems which justifies its selection to solve our problem.

Basic concepts of Genetic Algorithms

Genetic Algorithms (GAs) are adaptive methods based on the mechanics of natural selection [148]. The basic principles of GAs are described in many texts [148]-[150].

The first step in GA is to encode the problem as a chromosome or a set of chromosomes that consist of several genes. The solution in its original form is referred to as *phenotype*, whereas its binary encoded version is called *genotype* or simply *chromosome*.

Next, a pool of feasible solutions to the problem, called initial population, is created. Each chromosome in the population is associated with a fitness value that is calculated using a fitness function that indicates how good the chromosome is.

Genetic operators' *selection*, *crossover*, and *mutation* operate on the population to generate a new generation of population, *i.e.*, a new set of feasible solutions, from the old ones. Good feasible solutions are selected with higher probability to the next generation, in line with the idea of *survival of the fittest*. The standard *crossover* operation recombines arbitrarily selected chromosomes pairwise, by interchanging portions of them, producing new chromosomes to explore the search space. An occasional *mutation* operation is performed on a chromosome to facilitate jumping of solutions to new unexplored regions of the search space. As the algorithm continues and newer generations evolve, the quality of solutions improves. The success of GA is explained by the schema theorem and building-block hypothesis in [148].

Genetic algorithms for dynamic environments

Genetic Algorithms have been shown to be a useful alternative to traditional search and optimization methods, especially for problems where the space of all potential solutions is too high to be searched exhaustively in any reasonable amount of time. They are very efficient in directing the search towards relatively prospective regions of the search space. Empirical studies have shown that GAs do converge on global optima for a large class of NP-hard problems [146].

A key element in a genetic algorithm is that it maintains a population of candidate solutions that evolves over time. The population allows the genetic algorithm to continue to explore a number of regions of the search space that appear to be associated with high performance solutions. The distributed nature of the genetic search provides a natural source of power for searching in changing environments. As long as the population remains distributed over the search space, there is good reason to expect the genetic algorithm to adapt to changes in the utility function by reallocating future search effort towards the region of the search space that is currently favored by the utility function.

A number of previous studies have addressed the use of genetic algorithms in changing environments such as dynamic shortest path routing in MANETs [62], dynamic coverage and connectivity problem in WSN [151], dynamic resource allocation problem in cellular networks [152] and dynamic network coding problems [153].

In the sequel, a novel sequential genetic algorithm is presented for dynamic topology reconfiguration in MCNs. A special encoding scheme, crossover and mutation operations are proposed to search for the optimum topology when the traffic in the network changes. Improvement in the fitness function is sequentially controlled as newer generations evolve and whenever the improvement is sufficiently increased the current topology is updated by the new one having higher fitness. Numerical results show that SGA provides high performance improvements in a dynamic network environment.

3.2 System model and assumptions

In this section, we extend the model considered in [Section 2.2, Chapter 2] to include duplex transmission (uplink/downlink) in MCNs. We consider a cellular network with a set $\mathcal{I} = \{i\}$ of base stations. It is assumed that the area of the cell is divided into concentric rings as before with index m_r for the reference cell $i = r$ and m_i for the interfering cell i, $i \in \mathcal{I}_{-r} = \{i \neq r\}$. We assume that one cochannel user from each ring has bidirectional connection with the corresponding access point. Conventional relaying scheme (CONR) is used. Let us consider that a reference *mobile* user m_{r1} (located in ring m_{r1}) is transmitting (relaying) to another *mobile* user m_{r2} (located in ring m_{r2}) in the reference cell r and, at the same time, a cochannel *interfering mobile* user m_{i1} is transmitting to another *mobile* user m_{i2} in cell i.To keep the notation general, in the uplink last-hop transmission m_{r2} denotes

the reference base station AP_r, and on the downlink AP_r refers to m_{r1}. The same applies for the interfering base station AP_i.

By this generalization, the physical layer model is described again by (8)-(17). As an illustration, for the two cells scenario, the extension to downlink topology is shown in Fig. 19b.

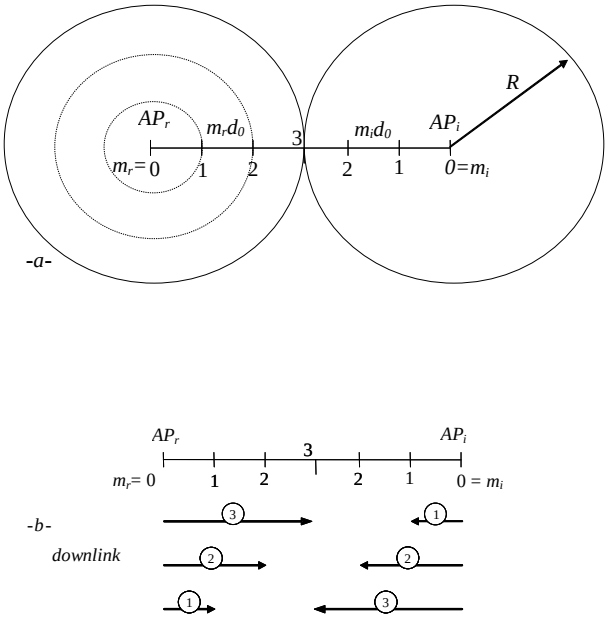

Fig. 19. a) Modeling interfering users positions for 2-cells; b) Possible transmission schedule for downlink.

In the example shown in Fig. 5b for uplink, 4 slots were needed to complete the transmission. In this case, for the scenario presented in Fig. 19b for downlink, as the hopping distance is larger, the transmission is completed in 3 slots. So, only 7 time slots are required for all 6 users to transmit on the up and down link. Classical TDMA scheme would require 6 + 6 = 12 slots. The optimum set of relaying routes is defined as in (18).

To model the network traffic dynamics, we extend the notation presented in [Section 2.5, Chapter 2] for uplink and we denote by $\boldsymbol{\delta}_i = (\delta_{i,1},...,\delta_{i,n_i})$, $\boldsymbol{\delta}_i \in \mathbf{R}^{n_i}$ the traffic that the access point AP_i is transmitting to the users on the downlink where

n_i is the number of rings in cell *i*. The overall network traffic on the uplink and downlink is defined as $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, ..., \boldsymbol{\lambda}_{N_c}) = (\lambda_1, ..., \lambda_N)$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, ..., \boldsymbol{\delta}_{N_c}) = (\delta_1, ..., \delta_N)$, respectively where N_c is the total number of cells ($N_c = |\mathcal{I}|$) and N is the total number of rings in the network obtained as $N = \sum_{i=1}^{N_c} n_i$. For the same traffic vectors $\boldsymbol{\lambda}_i$, $\boldsymbol{\delta}_i$, the base station schedule the transmission through different channels (time slots) which results in temporal and spatial MAC protocol.

The base stations jointly assign an access vector $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_{N_c}) = (a_1, ..., a_N)$ to the different rings to give them permission to transmit, where each component $a_n = (a_{\lambda_n}, a_{\delta_n})$ with $a_{\lambda_n}, a_{\delta_n} \in (0,1)$. With $a_{\lambda_n} = 1$ the users from ring *n* are allowed to transmit uplink otherwise not, and with $a_{\delta_n} = 1$ the users from ring *n* are allowed to transmit downlink otherwise not. In the two cell case $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$, the first half of the coefficients represent the permissions to transmit for the rings in referent cell *r* and the second half for rings in interfering cell *i*, $i \in \mathcal{I}_{-r}$.

Symmetric bidirectional transmission is considered in the sense that the access point will only transmit to the users situated in the rings activated by **a** where both components of $a_n = (a_{\lambda_n}, a_{\delta_n})$ are active, or not active, simultaneously. Recall that the index pattern for **a** is presented in Fig. 7. The link notation for the bidirectional case will be address in the following sections.

3.3 Bidirectional relaying topology with physical layer network coding

In this section, network coding [147] is additionally introduced and combined with the previous results on optimum relaying to reduce the number of slots needed for the users to complete their transmissions and achieve further improvements of the system performance.

Let us assume that the hops are indexed in increasing order for uplink as $h^{(up)}$ and for downlink as $h^{(down)}$. By combining the uplink and downlink traffic from the previous hop at hop *h* as $y_h^{(down,up)} = y_{h-1}^{(down)} \oplus y_{h-1}^{(up)}$ the number of overall time slots needed for transmission in *B* cycles can be reduced.

To elaborate this concept in more detail, an example of possible topology that includes network coding is shown in Fig. 20 for two cells. The traffic between users and access point is bidirectional. So given a schedule that alternates the transmissions between the different rings, after certain number of time slots all intermediate users (m_i, $i \in \mathcal{I}$) have information frames buffered for transmission in both directions. Whenever an opportunity arises, the intermediate users combine two information frames, one for each direction, with a simple *XOR*

operation and send it to its neighbors in a single omnidirectional transmission. Both receiving nodes already know one of the frames combined (have it stored from the previous transmission), while the other frame is new. Thus, one transmission allows two users to decode a new packet, effectively doubling the capacity of the path, reducing the power consumption of the transmitter node and reducing the number of time slots required to complete the transmission.

The transmission schedule presented in Fig. 20 defines a possible topology for two cell scenario and access vector **a** = **1**. In this case all rings have duplex connection and the topology consists of eight partial topologies representing transmissions in eight consecutive time slots. In the first time slot (the first partial topology) there are two simultaneous transmissions; packet originating from the access point AP_r (addressed to user in ring 3) is transmitted to ring 2 in cell *r* and at the same time packet originating from ring 2 (addressed to access point AP_i) of cell *i* is transmitted from ring 2 to ring 1. In the second time slot (the second partial topology), packet originating from access point AP_r (addressed to user in ring 1) is transmitted to ring 1, at the same time packet originating from ring 3 (addressed to access point AP_r) is transmitted from ring 3 to ring 2 and, packet originating at AP_i (addressed to user in ring 2) is transmitted to ring 1 in the adjacent cell. Similarly the same notation is then used for transmissions in time slots from 3 to 8. As already discussed earlier these eight partial topologies together are referred to as a *possible two dimensional* (*time and space*) *topology* and will be represented in the sequel by a given topology index (*t*).

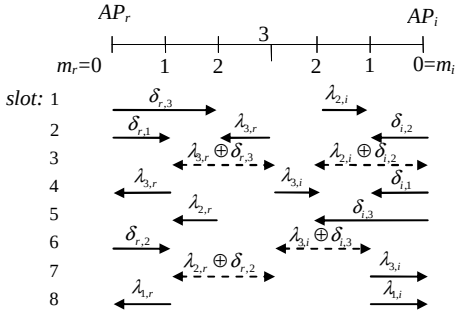

Fig. 20. Possible schedule by using network coding.

So there will be limited interference transmission for 3 users per cell in 8 channels (8 slots in Fig. 20) giving the intercell throughput 6/8 = 3/4, as opposed to the 6/12 = 1/2 in a conventional TDMA system where each cell uses a half of available channels (slots). Although scheduling in Fig. 5b and Fig. 19b requires 7 time slots in total (for uplink and downlink) it also assumes transmissions over three rings which requires higher power.

The optimization process defined by (18) now becomes

$$\begin{aligned} &\left\{\Re_r^{(2)}\right\} = \max_{B,H,\oplus,\Re_r^{(2)}} C_{\Re^{(2)}};\ \text{where}\ C_{\Re^{(2)}} = \sum_r c_{\Re_r^{(2)}} \\ &\Re = \Re^{(up)} \cup \Re^{(down)}, \end{aligned} \tag{33}$$

where $\Re^{(2)}$ is a two dimensional relaying topology to be elaborated in more detail in the next section, B is the number of slots needed for the users to complete their transmission uplink/downlink (scheduling length), H is the number of hops and $\oplus$ stands for the network coding operation. For each slot $b = 1,...,B$, parameter $c_{\Re_r^{(2)}}$ is given by corresponding (15)-(16) and $C_{\Re^{(2)}}$ is the network capacity.

3.4 System optimization

The extension of the optimization problem presented in [Section 2.6, Chapter 2] by (24) to bidirectional traffic is straighforward. An independent set of equations (23) should be written for both directions and (24) should be modified to include the overall utility function

$$U = U^{(up)} + U^{(down)} \tag{34}$$

with separate set of constrains for both directions. For each direction of the traffic, U is given by (22) and includes data rate, power consumption and delay.

TSL algorithm has been modified to search for the optimum topology for the scenario presented in Fig. 20. In a real network, where the traffic is dynamic, the network topology should be reconfigured to track the variations in the network and guarantee good network performance. In multicell MCNs with nonuniform traffic distribution, the search for the optimum topology becomes an NP-hard problem. TSL algorithm results in high complexity and for this reason in the next section we define an evolutionary sequential genetic algorithm that can be used for readjustment of the topology due to traffic variation.

3.5 SGA-TSL algorithm

In this section, a Sequential Genetic Algorithm (SGA) is presented to dynamically adjust the optimum topology to the traffic variations in the network. As result, routing and scheduling will be implicitly adjusted with the relaying topology. SGA represents a suboptimum solution which provides performance close the optimum with less complexity or with shorter computational time.

3.5.1 Encoding scheme

For simplicity of presentation we start this section by considering only uplink transmission which will be further extended to the bidirectional case. The topologies are encoded as a set of chromosomes, where each chromosome defines a partial topology.

A chromosome consists of a number of gene-instances $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_{N_c}) = (\gamma_1, ..., \gamma_N)$ where N_c is the total number of cells and N is the total number of rings in the network. Each vector $\boldsymbol{\gamma}_i$ consists of n_i components, where n_i is the number of rings in cell i. These gene-instances in our design correspond to mobile users that are transmitting from specific rings. We use binary coding scheme and the value of the gene will be 1 if the corresponding user is transmitting in that time slot or 0 otherwise.

On the other hand, the phenotype information for a genotype instance is represented by the set of active links that the corresponding users are activating in each time slot. For the two cell example, using the notation of the links shown in Fig. 7 for uplink transmission, the phenotype of gene γ_1 is 1 (link l_1 is used), for gene γ_2 can be 2 or 4 (link l_2 or l_4 can be used), and so on.

With this scheme the topology is given by a set of chromosomes that define the partial topologies generated in B time slots and are denoted by $PT_{\gamma}^{b}(t)$, where t is the index of the topology, b is the index of the time slot $b = 1,\ldots,B$ and γ the index of the gene.

To illustrate the encoding scheme a simple example of a possible topology (t = 1) is considered, for two cell case and **a** = **1**, that consists of the set of links $T(1) = \{PT_{\gamma}^{b}(1)\} = \{\{l_1\},\{l_2\},\{l_3\},\{l_7\},\{l_8\},\{l_9\}\}$ as shown in Fig. 21.

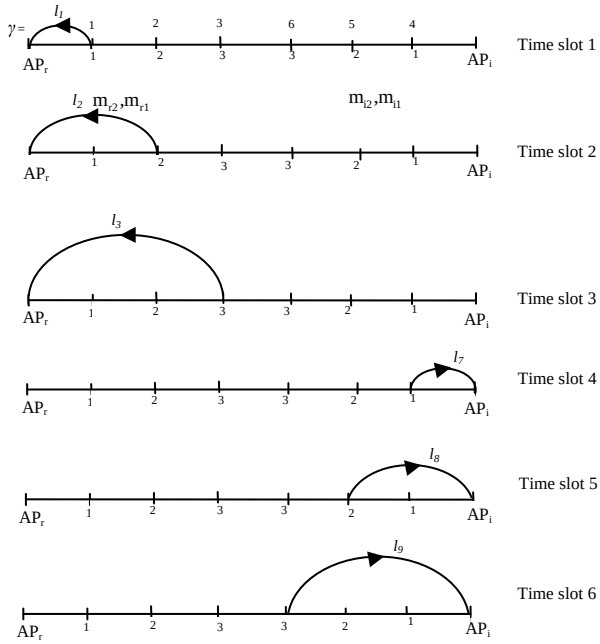

Fig. 21. The transmission pattern for topology index t = 1.

In this case every user transmits in a separate time slot directly to its access point. The partial topologies are:

$$PT_{\gamma}^{b=1}(1) = \{l_1\} \rightarrow genotype = (100000) \rightarrow phenotype = 1\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0$$

$$PT_{\gamma}^{b=2}(1) = \{l_2\} \rightarrow genotype = (010000) \rightarrow phenotype = 0\text{-}2\text{-}0\text{-}0\text{-}0\text{-}0$$

$$PT_{\gamma}^{b=3}(1) = \{l_3\} \rightarrow genotype = (001000) \rightarrow phenotype = 0\text{-}0\text{-}3\text{-}0\text{-}0\text{-}0$$

$$PT_{\gamma}^{b=4}(1) = \{l_7\} \rightarrow genotype = (000100) \rightarrow phenotype = 0\text{-}0\text{-}0\text{-}7\text{-}0\text{-}0$$

$$PT_{\gamma}^{b=5}(1) = \{l_8\} \rightarrow genotype = (000010) \rightarrow phenotype = 0\text{-}0\text{-}0\text{-}0\text{-}8\text{-}0$$

$$PT_{\gamma}^{b=6}(1) = \{l_9\} \rightarrow genotype = (000001) \rightarrow phenotype = 0\text{-}0\text{-}0\text{-}0\text{-}0\text{-}9$$

As we can see, the previous topology consists of a set of six chromosomes (partial topologies) that give the information of which user is transmitting in each time slot and which link is being used.

To extend the previous notation to bidirectional links, the genes are duplicated. So, every partial topology is defined by two genotypes (uplink/downlink) and two phenotypes. The vector of gene-instances now should include also the access point $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_{N_c}) = (\gamma_0, \gamma_1, ..., \gamma_{N+N_c-1})$.

Each vector $\boldsymbol{\gamma}_i$ consists now of n_i + 1 components, where n_i is the number of rings in cell *i* and the first entry of $\boldsymbol{\gamma}_i$ represents transmissions from/to the *i*th access point. To illustrate the encoding scheme with bidirectional links, we present an example where the transmission pattern presented in Fig. 25a is defined by the following partial topologies:

$$PT_{\gamma}^{b=1}(t) = \left\{ l_1^{(down)} \right\} \rightarrow \begin{cases} genotype^{(up)} = \mathbf{0}; phenotype^{(up)} = \mathbf{0} \\ genotype^{(down)} = (10000000); phenotype^{(down)} = 1\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0 \end{cases}$$

$$PT_{\gamma}^{b=2}(t) = \left\{ l_{10}^{(up)} \right\} \rightarrow \begin{cases} genotype^{(up)} = (00000010); phenotype^{(up)} = 0\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0\text{-}10\text{-}0 \\ genotype^{(down)} = \mathbf{0}; phenotype^{(down)} = \mathbf{0} \end{cases}$$

$$PT_{\gamma}^{b=3}(t) = \left\{ l_4^{(up)} \right\} \rightarrow \begin{cases} genotype^{(up)} = (00100000); phenotype^{(up)} = 0\text{-}0\text{-}4\text{-}0\text{-}0\text{-}0\text{-}0\text{-}0 \\ genotype^{(down)} = \mathbf{0}; phenotype^{(down)} = \mathbf{0} \end{cases}$$

$$PT_{\gamma}^{b=4}(t) = \left\{ l_7^{(down)} \right\} \rightarrow \begin{cases} genotype^{(up)} = \mathbf{0}; phenotype^{(up)} = \mathbf{0} \\ genotype^{(down)} = (00001000); phenotype^{(down)} = 0\text{-}0\text{-}0\text{-}0\text{-}7\text{-}0\text{-}0\text{-}0 \end{cases}$$

$$PT_{\gamma}^{b=5}(t) = \left\{ l_1^{(up)} \oplus l_4^{(down)}, l_7^{(up)} \oplus l_{10}^{(down)} \right\}$$

$$\rightarrow \begin{cases} genotype^{(up)} = (01000100); phenotype^{(up)} = 0\text{-}1\text{-}0\text{-}0\text{-}0\text{-}7\text{-}0\text{-}0 \\ genotype^{(down)} = (01000100); phenotype^{(down)} = 0\text{-}4\text{-}0\text{-}0\text{-}0\text{-}10\text{-}0\text{-}0 \end{cases}$$

where upper index (*up*) corresponds to uplink transmission and (*down*) to downlink.

3.5.2 Population: structure and initialization

The initial population consists of topologies (sets of chromosomes). Combinations of the chromosomes from the initial population enable definition of all possible routes to/from the access point. In this way the reproduction operators will produce as result new chromosomes that define all possible feasible topologies through the pass of the generations. For example, for the two cell case in Fig. 7 and **a** = 1, the initial population is formed by:

- $T(1) = \{PT_\gamma^b(1)\} = \{\{\boldsymbol{l}_1\},\{\boldsymbol{l}_2\},\{\boldsymbol{l}_3\},\{\boldsymbol{l}_7\},\{\boldsymbol{l}_8\},\{\boldsymbol{l}_9\}\}$ as defined in the previous section, with the genotypes and phenotypes for uplink/downlink transmission and $\boldsymbol{l}_j = \{\{l_j^{(up)}\},\{l_j^{(down)}\}\}$.
- $T(2) = \{\{\boldsymbol{l}_6\},\{\boldsymbol{l}_4\},\{\boldsymbol{l}_1\},\{\boldsymbol{l}_{12}\},\{\boldsymbol{l}_{10}\},\{\boldsymbol{l}_7\}\}$ where the genotypes and phenotypes are defined in the same way as before.
- $T(3) = \{\{\boldsymbol{l}_5\},\{\boldsymbol{l}_1\},\{\boldsymbol{l}_2\},\{\boldsymbol{l}_{11}\},\{\boldsymbol{l}_7\},\{\boldsymbol{l}_8\}\}$.

In the general case for N_c cells, and n_i rings per cell, the initial population is formed by $\bigcup_{h=1}^{H}\{T(h) = \{\boldsymbol{l}_\varsigma\}\}$ with link index ς given by

$$\varsigma = \begin{cases} m_{r2}, (N_c - 1)\cdot L_i + m_{i2} & \text{if } H = 1 \\ m_{rh} + h, m_{rh-1} + h - 1, ..., m_{r1}, (N_c - 1)\cdot L_i + m_{ih} + h, \\ (N_c - 1)\cdot L_i + m_{ih-1} + h - 1, ..., (N_c - 1)\cdot L_i + m_{i1} & \text{if } H > 1 \end{cases}$$

where $m_{r2} = 1, 2, ..., n_r$, $m_{i2} = 1, 2, ..., n_i$, and $L_i = \sum_{m_i=1}^{n_i} m_i$ is the number of links per cell.

A new generation of population is formed by applying reproduction operators' crossover and mutation that combine the partial topologies PT_γ^b (chromosomes) or modify them respectively, to form new feasible topologies.

3.5.3 Fitness function and SGA

The quality of the solution is judged by the fitness function (*f*) defined as in (34) and is included in the global optimization defined by (24) with *f* = *U* for a "*given* ***T***". The topologies that correspond to the best values of the fitness are kept in the pool for possible future reconfiguration of the network, and the optimum topology which corresponds to the highest fitness will be assigned to the network.

In order to compare convergence of GA and exhaustive search method, we use as "*given* ***T***" an optimum topology found by exhaustive search. In Section 3.7 we discuss other options to calculate ***T***. In a practical implementation of GA this variable is not known so we run GA as long as we find a topology that provides $\Delta f \geq threshold$ where Δf is the difference in utility function between the previously accepted solution and the new one. This algorithm will be referred to as SGA since the usable topology T_{new} is produced as a sequence of solutions satisfying $\Delta f \geq threshold$.

3.5.4 Crossover operator

In our algorithm, we adopt *arithmetic crossover* [149] that defines new chromosomes as a result of the following arithmetic operation:

$$\{T_{new}\} = \left\{PT_{\gamma}^{b}(\mathbf{t})\right\} = \left\{P\mathbf{T}_{\gamma_k}^{b_i}(t)\right\} \; OR \; \left\{P\mathbf{T}_{\gamma_k}^{b_j}(t)\right\}, \tag{35}$$

where $b = 1,...,B-1$; $b_i, b_j = 1,...,B$ with $b_i \neq b_j$ and $k = 0,...,N + N_c - 1$.

From each initial topology, we generate a new set of topologies $\{T_{new}\}$ that consists of $\binom{B}{2}$ topologies as a result of *OR* operation between all pairs of chromosomes PT_{γ}^{b} that define certain topology $T(t)$. If the new topology is not feasible it will be excluded from the population.

To guarantee that all feasible routes are generated through the pass of the generations, we allow certain level of elitism and we let the chromosomes of the initial population survive to the next generation.

As an illustration, we consider the transmission pattern for topology $T(1) = \{PT_{\gamma}^{b}(1)\} = \left\{\{\boldsymbol{l}_1\},\{\boldsymbol{l}_2\},\{\boldsymbol{l}_3\},\{\boldsymbol{l}_7\},\{\boldsymbol{l}_8\},\{\boldsymbol{l}_9\}\right\}$ as shown in Fig. 21. The crossover operation defined as $PT_{\gamma}^{b_i=1}(1)$ *OR* $PT_{\gamma}^{b_j=2}(1) = \{\boldsymbol{l}_1, \boldsymbol{l}_2\}$ results in a partial topology that is not feasible, because the *AP* cannot receive two signals at the same time. The same situation occurs for the crossover defined as $PT_{\gamma}^{b_i=1}(1)$ *OR* $PT_{\gamma}^{b_j=3}(1) = \{\boldsymbol{l}_1, \boldsymbol{l}_3\}$. So, in this case the first feasible topology resulting from the crossover operation is $PT_{new,\gamma}^{b=1}(\mathbf{t}) = PT_{\gamma}^{b_i=1}(1)$ *OR* $PT_{\gamma}^{b_j=4}(1) = \{\boldsymbol{l}_1, \boldsymbol{l}_7\}$, and the overall topology is shown in Fig. 22.

By analogy other topologies can be obtained as result of the crossover defined as $PT_{\gamma}^{b_i=1}(1)$ *OR* $PT_{\gamma}^{b_j=5}(1)$, $PT_{\gamma}^{b_i=1}(1)$ *OR* $PT_{\gamma}^{b_j=6}(1)$, $PT_{\gamma}^{b_i=2}(1)$ *OR* $PT_{\gamma}^{b_j=4}(1)$ and so on.

3.5.5 Mutation operation

Mutation probability is normally set to be small. This is because even though mutation can enable the algorithm to search a new search space, it actually destroys the current pattern. If the optimum topology is not obtained after successive crossover operations, we mutate the chromosome by choosing for a given gene γ a different phenotype. This means that for a certain user (gene) we select another link from the set of possible links represented in Fig. 7 for the two cell scenario. The maximum number of mutations that we can perform in this scenario is 3 mutations per cell, *i.e.* from $l_3 \rightarrow l_5$, from $l_5 \rightarrow l_6$ and from $l_2 \rightarrow l_4$ as

shown in Fig. 23. The same applies for the adjacent cell. In general, the maximum number of mutations is $\sum_i \sum_{m_i=1}^{n_i} (m_i - 1)$, $i \in \mathcal{I}$. After mutating one chromosome, if the resulting topology is not fit enough ($\Delta f \geq threshold$), then we perform the crossover operation with this new mutated chromosome.

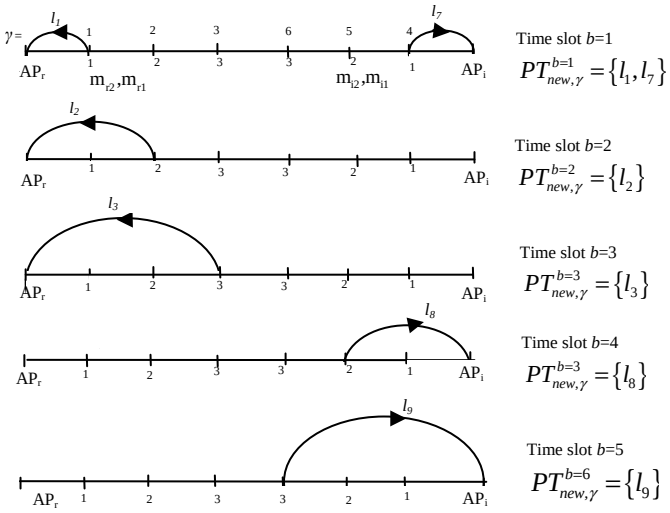

Fig. 22. Topology obtained after the crossover operation.

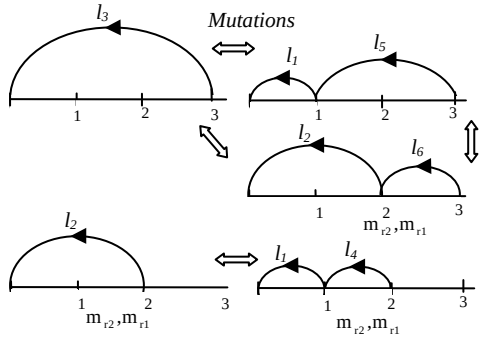

Fig. 23. Possible mutations.

3.6 Traffic cognitive topology control

After modeling our dynamic topology search problem by genetic algorithm GA-TSL, in this section the functioning of the network based on this approach is described. The network will be referred to as Traffic Cognitive Network (TCN). More specifically we discuss the operation of the Traffic Cognitive Topology Control $(TC)^2$ algorithm to adjust the topology to the traffic variations in the network by sequential genetic algorithm. The term cognitive refers to the awareness of the traffic variation.

To model the traffic in a TCN, we use the traffic vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ described in Section 3.2. For a given spatial traffic distribution the access vector should be varied in time to provide

$$E(\mathbf{a}_{\lambda}(t)) = \boldsymbol{\lambda}(t) \text{ and } E(\mathbf{a}_{\delta}(t)) = \boldsymbol{\delta}(t). \tag{36}$$

The variation of the traffic in the network is defined by the vector $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, ..., \boldsymbol{\Delta}_{N_c}) = (\Delta_1, ..., \Delta_N)$ where N_c is the total number of cells and N is the total number of rings in the network**.** If the traffic in the network changes due to users that became inactive then the corresponding component of Δ_n where the change occurs is negative. On the other hand if a new source appears in the network this component is positive.

The differential access vector corresponding to the traffic variation is given by $\mathbf{a}' = \mathbf{a}_I \oplus \mathbf{a}_F$, where $\mathbf{a}_\mathrm{I}$ is the access vector corresponding to the initial traffic and $\mathbf{a}_\mathrm{F}$ is the access vector after the traffic has changed (final). We assume that the traffic distribution is observed (cognition) in time intervals short enough to detect each change in the traffic so that traffic change only in one ring is assumed in a given observation instant.

$(TC)^2$ algorithm described below uses an exhaustive search (TSL) algorithm to find initial optimal topology and SGA-TSL for tracking traffic variations. Different options to initialize SGA-TSL are discussed in the following section.

The $(TC)^2$ algorithm works as follows:

1) Calculate the access vector $\mathbf{a}_\mathrm{I}$ *for a given traffic distribution* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\delta}$*, by using (36).*

2) Use TSL to generate the set of feasible candidate topologies $\Im^{(2)}$ *for* $\mathbf{a}_\mathrm{I}$**.**

3) For each topology in the set $\Im^{(2)}$*:*

4) Calculate the aggregate powers needed for users to deliver the information.

5) Calculate the utility using (23).

6) Use CVX [126] to optimize the source rates in (34).

7) Go to 3) until the optimum topology $\mathbf{T}^*$ *is obtained or if the traffic distribution in the network has changed (* $\mathbf{a}' \neq \mathbf{0}$ *), use SGA-TSL to find the new optimum topology* $\mathbf{T}^*$ *and go to 3).*

The operation of *SGA-TSL* program can be summarized as:

1. *B_1: number of time slots of the initial optimum topology* $\mathbf{T}^{B_1}$ *for* $\mathbf{a}_I$ (initial traffic)
2. *f_1: value of the fitness function (utility) associated with the initial topology* $\mathbf{T}^{B_1}$
3. *B': number of time slots of the optimum topology* $\mathbf{T}^{B'}$ *for* $\mathbf{a}'$ (differential access vector)
4. *f_0: value of the fitness function (utility) associated with the initialization of the new topology* $\mathbf{T}^{B_0}$ (after traffic variation)
5. *N_A: number of active rings in the network after the traffic variation* $N_A = \sum_n \mathbf{a}_{Fn}$
6. **procedure** *Check_traffic_variation*
7. $E(\mathbf{a}_{F\lambda}(t)) = \boldsymbol{\lambda}(t);\ E(\mathbf{a}_{F\delta}(t)) = \boldsymbol{\delta}(t)$ *{Assign access vector* $\mathbf{a}_F$ *based on the existing traffic}*
8. $\mathbf{a}' = \mathbf{a}_I \oplus \mathbf{a}_F$
9. *If* $\mathbf{a}' \neq \mathbf{0}$
10. *Apply TSL(**a'**) to obtain the optimum topology for **a'**.*
11. *Check the value of* $\boldsymbol{\Delta}$ *to initialize the new topology:*
12. *- If* $\Delta_n > 0$*, initialize the new topology as the set of chromosomes* $\mathbf{T}^{B_0} = \{P\mathbf{T}^{B_1}, P\mathbf{T}^{B'}\}$.
13. *- If* $\Delta_n < 0$*, the new topology is initialized as the set of the different*
14. *chromosomes in both sets* $\mathbf{T}^{B_0} = \{P\mathbf{T}^{B_1} - P\mathbf{T}^{B'}\}$
15. *Calculate the fitness function (f_0) by using (23) and (34) with $f_0 \rightarrow U$*
16. *Initialize $f = f_0$; $n\_m = 0$;* $\mathbf{a}_I = \mathbf{a}_F$
17. *end*
18. **end**
19. **procedure** *Calculate_fitness*
20. *Calculate f by using (24) and (34) with $f \rightarrow U$*
21. *If ($f - f_0$) > threshold*
22. *Reconfigure the system with* $\boldsymbol{T}^* = \boldsymbol{T}_{new}$ *that corresponds to the fitness* $f_{opt} = f$
23. $f_0 = f$;

24. end
25. **end**
- - - - - - - - - - - - - SGA-TSL algorithm - - - - - - - - - -
26. While (1)
27. Check_traffic_variation
28. for b = 1 to B_0
29. $\boldsymbol{T}_{new} = crossover(\, P\mathbf{T}^{b}, P\mathbf{T}^{(b+1) \bmod B_0} \,);$
30. Calculate_fitness
31. end
32. If (n_m < N_A*) {number of mutations < number of rings actived}*
33. $\boldsymbol{T}_{new} = mutation(\, \mathbf{T}^{B_0}, \gamma_{n\_m});$
34. n_m++;
35. Calculate_fitness
36. end
37. end

Lines 1 to 5 define the variables used in the program. From line 6 to 18, the procedure *Check_traffic_variation* is defined. In line 7, the access vector $\mathbf{a}_F$ is assigned depending on the traffic variation. Line 8 calculates the differential access vector $\mathbf{a}'$. Line 9 examines if the traffic has changed. In line 10, TSL program is used to obtain the optimum topology associated with the traffic variation in the network, $\mathbf{T}^{B'}$. This topology is needed to initialize the algorithm. This topology has only one active source and TSL program should complete the search in few iterations. From line 11 to 14 the topology depending on the traffic variation is initialized to start the program. If a new source has appeared, the new topology $\mathbf{T}^{B_0}$ is initialized as the set (union) of partial topologies (chromosomes) of the initial topology $\mathbf{T}^{B_1}$ and the set of partial topologies corresponding to $\mathbf{T}^{B'}$. On the other hand, if a source has become inactive then the new topology $\mathbf{T}^{B_0}$ is initialized as the difference between the two sets of partial topologies. In line 15, the fitness function f_0 for the new topology is calculated. In line 16 the instant fitness value f, the number of mutations n_m and access vector $\mathbf{a}_I$ are initialized.

From lines 19 to 25 the procedure *Calculate_fitness* is defined. In line 20, the fitness value f is calculated as result of the optimization described by (23). Line 21 checks if the new fitness value f is higher than the previous one f_0 plus certain threshold. The threshold can be zero or a positive value depending on how often the traffic changes in the network. If the previous condition holds, line 22 reconfigures the system with the new topology $\boldsymbol{T}_{new}$.

From lines 26 to 37, SGA-TSL algorithm is described. Line 27 checks if the traffic in the network has changed. In line 28, the time slots of the new topology $\mathbf{T}^{B_0}$ are assigned to index *b*. The crossover of the partial topologies corresponding to $\mathbf{T}^{B_0}$ is performed in line 29 to obtain the new topology $\boldsymbol{T}_{new}$. In line 30, procedure Calculate_fitness is called.

Lines 32 to 36 perform the mutation operation if *f* did not follow the previous requirements for the topology to be updated. As the mutation is performed gene by gene, line 32 checks if the number of mutations is less than the number of active rings in the network. Line 33 realizes the mutation operation over gene γ_{n_m}. Line 34 updates the number of mutations. In line 35 procedure *Calculate_fitness* is called again to check if the new fitness value *f* has improved compared to the previous one f_0 to reconfigure the network with the topology associated to *f*. Finally, the algorithm goes back to 26 to check again the traffic or to continue with the crossover and mutation operations to find a better fitness.

An extensive set of examples is given in Section 3.7 to illustrate the performance of SGA-TSL algorithm for different traffic variations.

Implementation

The algorithm may be implemented in one of the base stations, where cooperating base stations must exchange information about the traffic distribution ($\boldsymbol{\lambda}_i$, $\boldsymbol{\delta}_i$) in their respective cells *i*. The base station or an equivalent coordinating unit should pass the information about the resulting access vector $\mathbf{a}_F$ back to the cooperating base stations.

The system can be implemented in different ways opening still the possibilities for different proprietary rights. For the functioning of the protocol the access point needs to know the position of active nodes which is information already available in these systems. This information should be exchanged between the adjacent access points. The transmissions should also be slot synchronous. The optimum schedule is communicated to the terminals so that they know the slot index that they can use. The simplest way to implement messaging is to use separate control channel that covers the whole cell (already exist in cellular networks) and use relaying only for data channel. This can be afforded since the data rate in the control channel is significantly lower than in the data channel so that the power consumption needed for full cell coverage is not excessive. The topology where the control plane (C) and data plane (D) are

implemented on different concepts will be referred to as Inter System Networking or *InSyNet(C,D)*.

This combination of cellular network concept (in C plane) and elements of ad hoc network (D plane) is the main advantage over any other solution that would be based exclusively on one of these concepts. Cellular networks need relaying in order to reduce the terminal power especially for high data rates. Ad hoc networks provide relaying type transmissions but there is no awareness of the overall network state and mutual impact between different transmissions in different parts of the network. The awareness of dynamic changes in the traffic distribution and possible network topology due to the node appearance/disappearance [138, Chapter 13] propagates slowly throughout the network. For all these reasons ad hoc networks routing algorithms [155]-[157] are not efficient. The same applies for bio inspired routing protocols [158] like swarm intelligence based routing in ant colony system models.

3.7 Performance evaluation

3.7.1 Numerical examples

In this section, we provide some examples to evaluate the performace of the proposed SGA-TSL algorithm. The link capacities $c_r(m_{r1}, m_{r2}, \mathbf{m}_{i1}, \mathbf{m}_{i2})$ are calculated as specified in Section 3.2. The channel gains $G_{m_{i1},m_{r2}}$ and $G_{m_{i1},m_{i2}}$ depend on the distance between the transmitter and receiver, and fading. For simplicity, we adopt the same model as in Section 2.8 where only propagation losses are consirered. The channel gains are defined as $G_{m_{i1},m_{r2}} \sim 1/d^{\alpha}_{m_{i1},m_{r2}}$ where $d_{m_{i1},m_{r2}}$ is the distance between the interfering transmitter in ring m_{i1} and reference receiver in ring m_{r2}, and α is the propagation constant ($\alpha = 4$). The calculation of $d_{m_{i1},m_{r2}}$ is straightforward from the geometry presented in Fig. 19.

In Fig. 24, the utility function versus the topology index *t* for access vector **a** = [010010] is shown. With this access vector user in ring 2, in both, cell *r* and *i* have permission to transmit. As the number of possible topologies obtained for this access vector is very high, we plot the segment of topologies close to the optimum topology. With no coding the maximum utility is $u = 0.5826$ while with network coding maximum utility is increased up to $u = 0.6991$. We can see that different topology indexes can provide the same value of utility. This is due to the

fact that those topologies consist of different combination of the same active links in different slots.

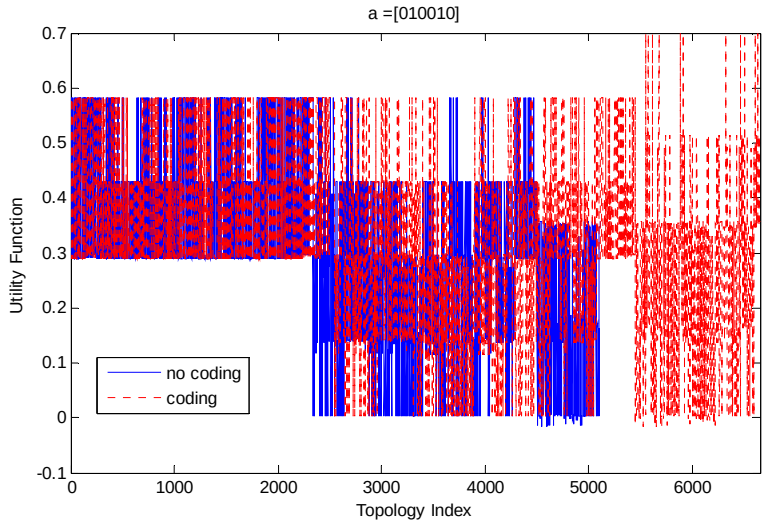

Fig. 24. Utility function for access vector a = [010010].

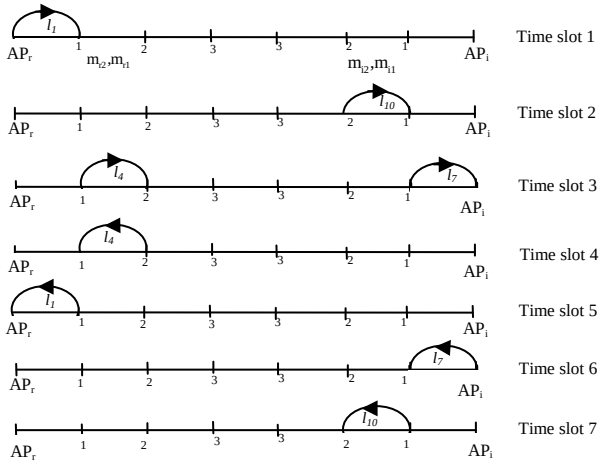

Fig. 24a. Representation of the transmission pattern defined by the topology index t = 7.

In Fig. 24a the transmission pattern is shown for one of the optimum topologies (topology index t = 7) in the case with no coding, for the previous access vector, defined by the set of links $T_7 = \left\{\{l_1^{(down)}\},\{l_{10}^{(up)}\},\{l_4^{(down)},l_7^{(up)}\},\{l_4^{(up)}\},\{l_1^{(up)}\},\{l_7^{(down)}\},\{l_{10}^{(down)}\}\right\}$. We can see that isolated short range transmissions are favored which can simultaneously reduce the intercell interference and power consumption.

In Figs. 25a and 25b the transmission patterns for two topology indices that correspond to the maximum utility with coding (t = 5571 and t = 5621) for the previous access vector are presented. The optimum topologies for the two cases are given by $T_{5571} = \left\{\{l_1^{(down)}\},\{l_{10}^{(up)}\},\{l_4^{(up)}\},\{l_7^{(down)}\},\{l_1^{(up)} \oplus l_4^{(down)},l_7^{(up)} \oplus l_{10}^{(down)}\}\right\}$ and $T_{5621} = \left\{\{l_1^{(down)}\},\{l_4^{(up)}\},\{l_7^{(down)}\},\{l_1^{(up)} \oplus l_4^{(down)},l_{10}^{(up)}\},\{l_7^{(up)} \oplus l_{10}^{(down)}\}\right\}$.

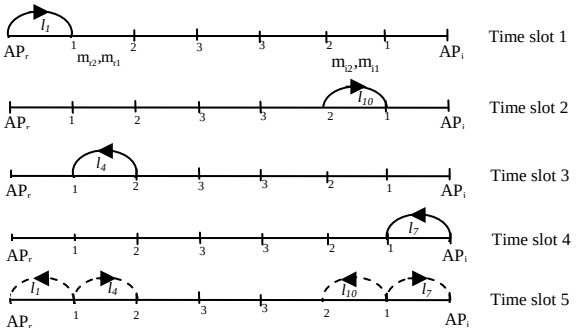

Fig. 25a. Representation of the transmission pattern defined by the topology index t = 5571.

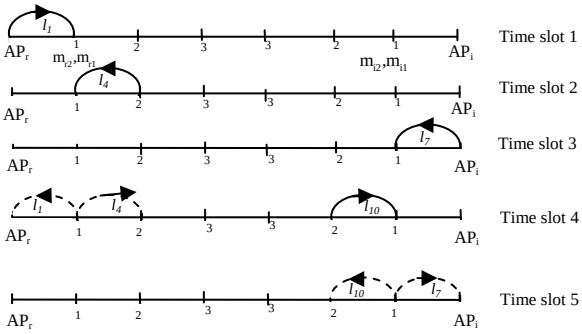

Fig. 25b. Representation of the transmission pattern defined by the topology index t = 5621.

We can see an improvement in the number of slots needed when network coding is used (5 slots) compared to 7 slots in the case with no coding. So, for the same type of isolated and short range transmissions the utility function is improved with network coding by reducing the number of slots.

In Fig. 26 the overall capacity for the previous access vector **a** is presented. We can see that the overall capacity of the system obtained for the optimum topologies is improved by a factor 1.5 when network coding is used compared with the case with no coding. The overall capacity obtained for optimum topologies is 4 times higher than for non optimum solutions.

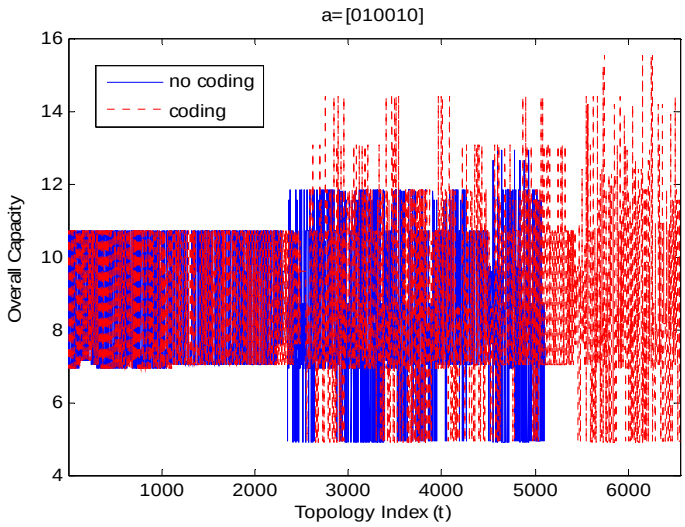

Fig. 26. Overall capacity for a = [010010].

In Fig. 27, the utility function is shown for **a** = [010100]. With this access vector user from ring 2 in cell *r* and user from ring 1 in cell *i* have permission to transmit. The maximum utility is obtained for topology index $t = 478$ ($u_{478} = 0.8739$) by using network coding. We can see a significant improvement compared with the maximum utility with no coding, obtained for topology index $t = 215$ ($u_{215} = 0.6640$). Both utilities are higher than in the previous case due to lower interference level.

The optimum topologies for both cases are given by $T_{478} = \left\{ \{l_1^{(down)}\}, \{l_7^{(down)}\}, \{l_4^{(up)}\}, \{l_1^{(up)} \oplus l_4^{(down)}, l_7^{(up)}\} \right\}$ and $T_{215} = \left\{ \{l_1^{(down)}\}, \{l_4^{(up)}\}, \{l_4^{(down)}, l_7^{(up)}\}, \{l_7^{(down)}\}, \{l_1^{(up)}\} \right\}$.

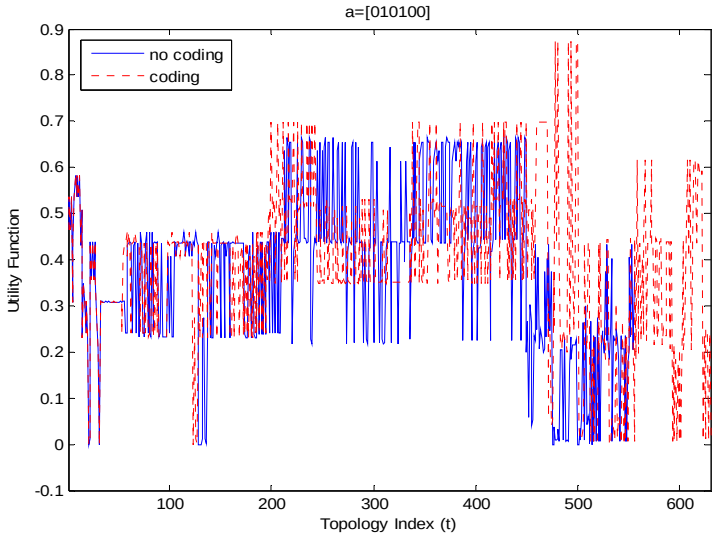

Fig. 27. Utility function for access vector a = [010100].

In Table 2 the reconfiguration results obtained with SGA-TSL algorithm are presented for uplink transmission and localized variations in the traffic Δ_n. The initial state in the network is defined by the optimum topology T^{B_1} associated with the access vector $\mathbf{a}_I$. As mentioned in Section 3.6, we consider that the traffic distribution is observed in time intervals short enough to detect each change in the traffic so that traffic change only in one ring is assumed in a given observation instant. The traffic change is given by access vector $\mathbf{a}'$.

The new optimum topology T_{new}, as a result of the reconfiguration of the initial optimum topology, is associated with the access vector $\mathbf{a}_F$.

As the first step, we show the number of generated topologies needed to find the optimum one by exhaustive search N_{es} and by GA N_{ga}, and the fitness value f_{opt} associated to the new optimum topology.

Table 2. Different topologies after the traffic variation Δ.

Δ_n	$\mathbf{a}_I$	t	$\mathbf{a}'$	$\mathbf{a}_F$
> 0		1	[010000] $T^{B'} = \{l_4, l_1\}$	[110001] $T_{\text{new}} = \{\{l_{12}\}, \{l_4\}, \{l_1\}, \{l_1, l_8\}\}$ $N_{es} = 209, N_{ga} = 8, f_{opt} = 0.6554$
> 0	[100001] $T^{B_1} = \{\{l_{12}\}, \{l_1\}, \{l_8\}\}$	2	[001000] $T^{B'} = \{l_6, l_2\}$	[101001] $T_{\text{new}} = \{\{l_{12}\}, \{l_6\}, \{l_1\}, \{l_2, l_8\}\}$ $N_{es} = 412, N_{ga} = 11, f_{opt} = 0.6554$
> 0		3	[000100] $T^{B'} = \{l_7\}$	[100101] $T_{\text{new}} = \{\{l_{12}\}, \{l_1\}, \{l_8\}, \{l_7\}\}$ $N_{es} = 59, N_{ga} = 1, f_{opt} = 0.6554$
> 0		4	[000010] $T^{B'} = \{l_{10}, l_7\}$	[100011] $T_{\text{new}} = \{\{l_{12}\}, \{l_1\}, \{l_{10}\}, \{l_8\}, \{l_7\}\}$ $N_{es} = 163, N_{ga} = 9, f_{opt} = 0.5243$
> 0	[010100] $T^{B_1} = \left\{\left\{\overset{\downarrow}{l}_4\right\}, \{l_1, l_7\}\right\}$	5	[000010] $T^{B'} = \{l_{10}, l_7\}$	[010110] $T_{\text{new}} = \left\{\{l_{10}\}, \{l_7\}, \left\{\overset{\downarrow}{l}_2, l_7\right\}\right\}$ $N_{es} = 70, N_{ga} = 8, f_{opt} = 0.5888$
> 0	[010001] $T^{B_1} = \left\{\{l_4\}, \left\{l_1, \overset{\downarrow}{l}_9\right\}\right\}$	6	[000100] $T^{B'} = \{l_7\}$	[010101] $T_{\text{new}} = \left\{\left\{\overset{\downarrow}{l}_{12}\right\}, \{l_1, l_8\}, \{l_7\}, \{l_4\}\right\}$ $N_{es} = 205, N_{ga} = 13, f_{opt} = 0.5243$
> 0	[001001] $T^{B_1} = \left\{\{l_6\}, \{l_2\}, \left\{\overset{\downarrow}{l}_{12}\right\}, \{l_8\}\right\}$	7	[010000] $T^{B'} = \{l_4, l_1\}$	[011001] $T_{new} = \left\{\{l_6\}, \{l_4\}, \left\{\overset{\downarrow}{l}_{11}, l_1\right\}, \{l_7, l_2\}\right\}$ $N_{es} = 1579, N_{ga} = 27, f_{opt} = 0.4406$
> 0	[010010] $T^{B_1} = \left\{\{l_4\}, \left\{l_1, \overset{\downarrow}{l}_8\right\}\right\}$	8	[001000] $T^{B'} = \{l_6, l_2\}$	[011010] $T_{\text{new}} = \left\{\{l_6\}, \{l_4\}, \left\{l_1, \overset{\downarrow}{l}_{10}\right\}, \{l_2, l_7\}\right\}$ $N_{es} = 560, N_{ga} = 4, f_{opt} = 0.4957$
> 0	[001100] $T^{B_1} = \{\{l_6\}, \{l_2\}, \{l_7\}\}$	9	[010000] $T^{B'} = \{l_4, l_1\}$	[011100] $T_{\text{new}} = \{\{l_6\}, \{l_4\}, \{l_2\}, \{l_1\}, \{l_7\}\}$ $N_{es} = 154, N_{ga} = 9, f_{opt} = 0.5243$
> 0	[110010] $T^{B_1} = \{\{l_{10}\}, \{l_2, l_7\}, \{l_1\}\}$	10	[000100] $T^{B'} = \{l_7\}$	[110110] $T_{new} = \{\{l_7\}, \{l_{10}\}, \{l_2, l_7\}, \{l_1\}\}$ $N_{es} = 257, N_{ga} = 1, f_{opt} = 0.6600$

Δ_n	$\mathbf{a}_I$	t	$\mathbf{a}'$	$\mathbf{a}_F$
< 0		11	[000100] $T^{B'} = \{l_7\}$	[100011] $T_{\text{new}} = \{\{l_{12}\},\{l_{10}\},\{l_1\},\{l_7\},\{l_8\}\}$ N_{es} = 163, N_{ga} = 2,f_{opt} = 0.5243
< 0	[100111] $T^{B_1} = \{\{l_{12}\},\{l_{10}\},\{l_8\},\{l_7\},\{l_1,l_7\}\}$	12	[000010] $T^{B'} = \{\{l_{10}\},\{l_7\}\}$	[100101] $T_{\text{new}} = \{\{l_1\},\{l_7\},\{l_{12}\},\{l_8\}\}$ N_{es} = 59, N_{ga} = 5,f_{opt} = 0.6554
< 0		13	[000001] $T^{B'} = \{\{l_{12}\},\{l_8\}\}$	[100110] $T_{\text{new}} = \{\{l_7\},\{l_{10}\},\{l_1,l_7\}\}$ N_{es} = 22, N_{ga} = 1,f_{opt} = 0.6815
< 0	[100011] $T^{B_1} = \{\{l_{12}\},\{l_{10}\},\{l_1\},\{l_7\},\{l_8\}\}$	14	[000001] $T^{B'} = \{\{l_{12}\},\{l_8\}\}$	[100010] $T_{\text{new}} = \{\{l_{10}\},\{l_1,l_7\}\}$ N_{es} = 8, N_{ga} = 3,f_{opt} = 0.5823
< 0	[001110] $T^{B_1} = \left\{\{l_7\},\{l_{10}\},\left\{\overset{\downarrow}{l}_3,l_7\right\}\right\}$	15	[000010] $T^{B'} = \{\{l_{10}\},\{l_7\}\}$	[001100] $T_{\text{new}} = \left\{\left\{\overset{\downarrow}{l}_6\right\},\{l_2\},\{l_7\}\right\}$ N_{es} = 20, N_{ga} = 2,f_{opt} = 0.5826
< 0	[101110] $T^{B_1} = \left\{\{l_1\},\{l_6\},\{l_2\},\{l_7\},\left\{\overset{\downarrow}{l}_8\right\}\right\}$	16	[001000] $T^{B'} = \{l_6,l_2\}$	[100110] $T_{\text{new}} = \left\{\{l_7\},\left\{\overset{\downarrow}{l}_{10}\right\},\{l_1,l_7\}\right\}$ N_{es} = 23, N_{ga} = 11,f_{opt} = 0.6815
< 0	[110110] $T^{B_1} = \{\{l_7\},\{l_{10}\},\{l_2,l_7\},\{l_1\}\}$	17	[100000] $T^{B'} = \{l_1\}$	[010110] $T_{\text{new}} = \{\{l_7\},\{l_{10}\},\{l_2,l_7\}\}$ N_{es} = 71, N_{ga} = 1,f_{opt} = 0.5888
< 0	[011001] $T^{B_1} = \left\{\left\{\overset{\downarrow}{l}_6\right\},\{l_4\},\left\{l_1,\overset{\downarrow}{l}_{11}\right\},\{l_2,l_7\}\right\}$	18	[010000] $T^{B'} = \{l_4,l_1\}$	[001001] $T_{new} = \left\{\{l_{12}\},\left\{\overset{\downarrow}{l}_3,\overset{\downarrow}{l}_8\right\}\right\}$ N_{es} = 105, N_{ga} = 4,f_{opt} = 0.4369
< 0	[111010] $T^{B_1} = \left\{\{l_6\},\{l_4\},\{l_1\},\left\{\overset{\downarrow}{l}_{10},l_1\right\},\{l_2,l_7\}\right\}$	19	[010000] $T^{B'} = \left\{\overset{\downarrow}{l}_4,l_1\right\}$	[101010] $T_{\text{new}} = \left\{\{l_6\},\left\{\overset{\downarrow}{l}_2,\overset{\downarrow}{l}_8\right\},\{l_1\}\right\}$ N_{es} = 166, N_{ga} = 9,f_{opt} = 0.5816
< 0	[011010] $T^{B_1} = \left\{\{l_6\},\{l_4\},\left\{\overset{\downarrow}{l}_{10},l_1\right\},\{l_2,l_7\}\right\}$	20	[001000] $T^{B'} = \{l_6,l_2\}$	[010010] $T_{\text{new}} = \left\{\{l_4\},\left\{l_1,\overset{\downarrow}{l}_8\right\}\right\}$ N_{es} = 22, N_{ga} = 4, f_{opt} = 0.4462

– Entries from $t = 1$ to $t = 4$ in Table 2 represent the reconfiguration results for $\mathbf{a}_I = [100001]$ when a new source appears in ring 2 ($\mathbf{a}' = [010000]$), 3 ($\mathbf{a}' = [001000]$), 4 ($\mathbf{a}' = [000100]$), or 5 ($\mathbf{a}' = [000010]$), respectively. In the first case, topology index $t = 1$, $N_{es} = 209$ topologies are generated to find the optimum one by exhaustive search when a new source appears in ring 2 compared to $N_{ga} = 8$ generated by GA. For $t = 2$, as a new source appears in ring 3, we have more routes than in the previous case so $N_{es} = 412$, and $N_{ga} = 11$. For $t = 3$, as the new source just introduces one more route we obtain $N_{es} = 59$ and $N_{ga} = 1$. For $t = 4$, we have $N_{es} = 163$ and $N_{ga} = 9$. The optimum topologies for $\mathbf{a}_I$, $\mathbf{a}'$ and $\mathbf{a}_F$ are presented in Table 2. From $t = 5$ to 10, we show more examples for different traffic patterns and traffic variations for $\Delta_n > 0$.
– From $t = 11$ to 13, the reconfiguration results for $\mathbf{a}_I = [100111]$ are shown when a source becomes inactive in ring 4 ($\mathbf{a}' = [000100]$), 5 ($\mathbf{a}' = [000010]$), or 6 ($\mathbf{a}' = [000001]$), respectively. For $t = 11$, one route from the initial topology is eliminated to obtained the optimum one after the traffic has changed and we have $N_{es} = 163$ compared to $N_{ga} = 1$. For $t = 12$, more routes are eliminated, so we need to generate more topologies than in the previous case to obtain the optimum topology resulting in $N_{es} = 59$ and $N_{ga} = 5$. For $t = 13$, the source in the border of the cell became inactive so the number of routes is significantly reduced and we obtained $N_{es} = 22$ compared to $N_{ga} = 1$. From $t = 14$ to 20, other examples are shown for different traffic patterns and traffic variations for $\Delta_n < 0$.

We can see that the number of generated topologies by using GA in the search for the optimum one is significantly reduced. In some of the examples the mutation operation is needed to obtain the optimum topology, and is indicated by arrow "↓" over the mutated link l_k. Due to the symmetry of the scenario presented in Fig. 19, $\mathbf{a}_{F_1} = (\mathbf{a}_1, \mathbf{a}_2)$ and $\mathbf{a}_{F_2} = (\mathbf{a}_2, \mathbf{a}_1)$ will produce symmetrical topologies.

The previous examples represent some illustrative cases to show the performance of the algorithm for different traffic variations. In order to calculate the total number of possible combinations, we denote by $|\mathbf{a}_1|$ and $|\mathbf{a}_2|$ the number of rings active in cell 1 and 2, respectively. For $\Delta_n > 0$**,** taking into account the symmetry of the scenario considered, the number of combinations where the changes can occur is

$$\sum_{|\mathbf{a}_1|=0}^{n_1}\sum_{|\mathbf{a}_2|=|\mathbf{a}_1|}^{n_2}\binom{n_1}{|\mathbf{a}_1|}\binom{n_2}{|\mathbf{a}_2|}\binom{n_1+n_2-|\mathbf{a}_1|-|\mathbf{a}_2|}{1}.$$

For the previous examples, $n_1 = n_2 = 3$ (3 rings per cell). For $\Delta_n < 0$**,** the number of combinations is

$$\sum_{|\mathbf{a}_1|=1}^{n_1}\sum_{|\mathbf{a}_2|=|\mathbf{a}_1|}^{n_2}\binom{n_1}{|\mathbf{a}_1|}\binom{n_2}{|\mathbf{a}_2|}\binom{|\mathbf{a}_1|+|\mathbf{a}_2|}{1}.$$

The performance of the algorithm is evaluated using two performance parameters: the success ratio $SR = 1/N_t$ where N_t is the number of generated topologies in the search for the optimum solution and, the improvement factor F given by $F = N_{es}/N_{ga}$. These performance parameters for the previous examples, described in Table 2, are shown in Table 3. The computational time *Tc* needed for SGA-TSL algorithm to find the optimum topology is also presented in Table 3 and it will be analyzed in the following subsection.

- Entry $t = 1$, in Table 3, shows $SR_{ga} = 1/8$ compared to $SR_{es} = 1/209$ to obtain the optimum topology for $\mathbf{a}_F = [110001]$ which gives an improvement factor $F = 209/8$.
- Entry $t = 2$, in Table 3, shows $SR_{ga} = 1/11$ compared to $SR_{es} = 1/412$ to obtain the optimum topology for $\mathbf{a}_F = [101001]$. In this case the number new of routes introduced by **a**' is higher than in the previous case, so N_t is increased and it gives an improvement factor $F = 412/11$.
- Entry $t = 3$, F is significantly increased ($F = 59/1$) because the new optimum topology is obtained by concatenating the active links from the initial topology and $T^{B'}$, which is the first operation of SGA-TSL algorithm for $\Delta_n > 0$. Equivalently, for $\Delta_n < 0$ the first operation to obtain the new optimum topology is eliminating the active links of $T^{B'}$ from the initial topology. This is the case of $t = 13$ ($F = 22/1$), and $t = 17$ ($F = 71/1$).

In all cases, independently of the location of the traffic variation in the network, N_{ga} is at least one order of magnitude less than N_{es}.

Table 3. Performance evaluation of GA-TSL.

t	$\mathbf{a}_F$	SR_{ga}	$(Tc)_{ga}$ (CPU seconds)	SR_{es}	F
1	[110001]	1/8	9.6253	1/209	209/8
2	[101001]	1/11	13.4005	1/412	412/11
3	[100101]	1/1	1.4196	1/59	59/1
4	[100011]	1/9	10.8265	1/163	163/9
5	[010110]	1/8	9.4693	1/70	70/8
6	[010101]	1/13	16.4737	1/205	205/13
7	[011001]	1/27	35.8802	1/1579	1579/27
8	[011010]	1/4	4.3212	1/560	560/4
9	[011100]	1/9	10.9825	1/154	154/9
10	[110110]	1/1	1.7784	1/257	257/1
11	[100011]	1/2	1.4040	1/163	163/2
12	[100101]	1/5	4.7892	1/59	59/5
13	[100110]	1/1	1.4352	1/22	22/1
14	[100010]	1/3	1.7472	1/8	8/3
15	[100010]	1/2	0.8268	1/20	20/2
16	[100110]	1/11	13.2133	1/23	23/11
17	[010110]	1/1	1.4508	1/71	71/1
18	[001001]	1/4	1.7628	1/105	105/4
19	[101010]	1/9	9.3133	1/166	166/9
20	[010010]	1/4	2.6988	1/22	22/4

So far, we have initialized SGA-TSL algorithm by the optimum topology calculated by exhaustive search. In more complicated scenarios exhaustive search is not practical. In the sequel, we calculate the optimum topology for the cases t = 1,2,3,4 from Table 2 and 3, when the initial topology is any of the feasible topologies for $\mathbf{a}_I$ = [100001]. In Fig. 28 we present the utility versus the topology index for access vector $\mathbf{a}_I$. We can see that there are 20 feasible topologies for that access vector, and the optimum topologies correspond to indexes 16 to 18 in the figure. To show the robustness of our algorithm we initialized GA with any feasible topology (SGA-AFT), to calculate the optimum topology after the traffic changes as in the cases t = 1,2,3,4 from Table 2 and 3.

In Fig. 29 we show that to obtain the optimum topology for $\mathbf{a}_F$ = [1100001] starting for any feasible topology, in the worst case we need to generate $N_{ga\text{-}AFT}$ = 31, compared to N_{ga} = 11 needed for SGA-TSL. For $\mathbf{a}_F$ = [101001] in the worst case we need $N_{ga\text{-}AFT}$ = 15 to obtain the optimum topology, compared to N_{ga} = 8 needed by SGA-TSL. For $\mathbf{a}_F$ = [100101] we obtain $N_{ga\text{-}AFT}$ = 13, compared to N_{ga} = 1. Finally, for $\mathbf{a}_F$ = [100011], SGA-AFT in the worst case needs $N_{ga\text{-}AFT}$ = 17,

compared to N_{ga} = 14. We can see that our proposed algorithm outperforms exhaustive search even when the initial topology is not the optimum one.

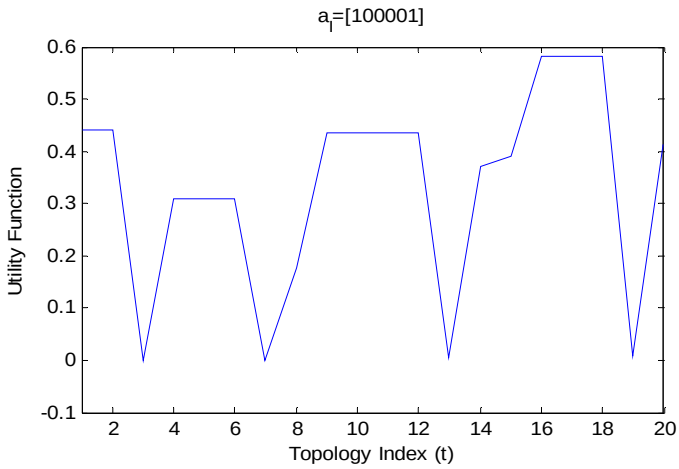

Fig. 28. Utility function for access vector a_I = [100001].

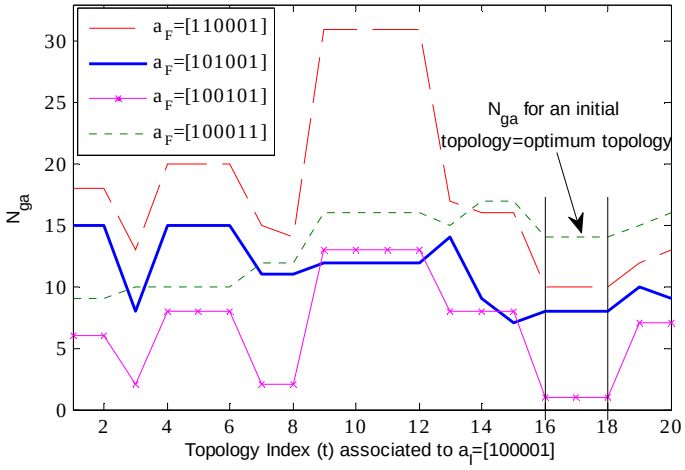

Fig. 29. N_{ga} to obtain the optimum topology for access vectors a_F = [110001], [101001], [100101], [100011] when a_I = [100001] and *GA* is initialized by any feasible topology t.

3.7.2 Comparisons

We compare the results obtained by SGA-TSL algorithm in Table 2, with the conventional routing protocols, used in ad hoc networks, which are based on collecting the information from the nearest neighbor. In this case, the topology is reconfigured in such a way that the users relay to their nearest neighbor. The results of this comparison are shown in Fig. 30. The fitness function obtained by nearest neighbor heuristic is significantly lower than the optimum value obtained by SGA-TSL. In some cases the value obtained by nearest neighbor heuristic can reach up to 50%.

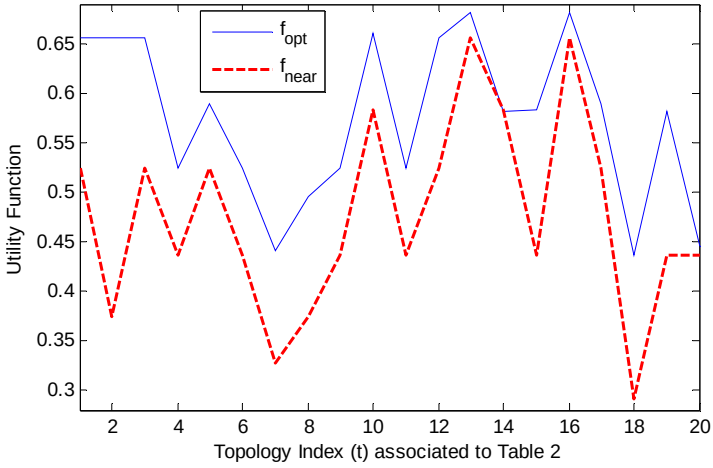

Fig. 30. Utility function for topologies *t* associated to Table 2 by SGA-TSL and nearest neighbor heuristics.

3.7.3 System dynamics

The main objective of any optimization process is to find the best possible solution and performance limits of a system that would be used as a bench mark for any suboptimal solution that may be simpler and more affordable to implement. Even if not immediately implementable optimal solutions set up the target for the further development of the technology that would enable implementation of such solutions. In the sequel we argue that our optimum

solutions are feasible for practical implementation even at the current state of the art of the technology.

In order to get an insight into the system dynamics the computation time *Tc* for SGA-TSL algorithm on Intel Core i7 620M 2,66 Ghz processor with 4GB RAM memory is shown in Table 3. As one can see the convergence of the algorithm depends on the changes in the traffic vector and for typical examples from Table 3, *Tc* varies in the range 1-30 seconds with more than half of these values being below 10 sec.

- *Traffic dynamics:* For the typical call arrival rate $\lambda_m = 0.01$calls/s[159], $Tc < 1/\lambda_m$ and the algorithm will be able to compensate the changes in traffic vector, resulting in proper reconfiguration of the network topology.
- *Mobility:* The equivalent traffic vector will be also changed due to mobility (terminal moves from one ring to the next neighboring ring). As long as $vT_c < \Delta d$, where Δd is the distance between two adjacent rings, and v the terminal velocity, the proper reconfiguration of the network topology is possible.
- *Channel Defading:* The program run time for rate optimization for a given topology is of the order of 0.01 sec. Furthermore, for the purpose of this analysis we can model the channel as a Rician fading which steady state component is increasing as the hopping distance reduces. So, for the hopping distance short enough the random component will vanish, which is referred to as the channel defading, and the algorithm will be able to converge. More details on channel defading will be provided in Chapter 4.

Impact of mobility and fading on the system efficiency can be further reduced by splitting the traffic into two groups, static/low dynamics and high dynamic and applying the above algorithm only on static/low dynamic users. Another possibility is to use more parallel processing and reduce *Tc*, or faster processor.

3.7.4 Discussion on other heuristics

Over the years, several approaches have been developed for evolutionary algorithms to address dynamic environments, such as maintaining diversity during the run via random immigrants [150], using memory schemes to reuse stored useful information [160], applying multi-population and speciation schemes to search in different regions of the search space, and adapting (the

parameters of) operators to quickly respond to a new environment [154]. All these approaches result into different heuristics to solve dynamic problems [154].

Changing environments usually require GAs to keep a certain population diversity level to maintain their adaptability. To address this problem, the random immigrants approach has been proposed in [150]. It maintains the diversity level of the population through replacing some individuals of the current population with random individuals, called *random immigrants*, in every generation.

However, with the traffic model that we assume in this chapter where the traffic distribution is observed in time intervals short enough to detect each change in the traffic, the introduction of random immigrants may divert the searching force of the GA because individuals in the previous environment may still be quite fit in the new environment. Our Sequential GA allows certain level of elitism and lets the chromosomes of the initial population survive to the next generation to maintain the population diversity, adapt to the changing environment, and thus guarantee that all feasible routes are generated through the pass of the generations.

Other heuristics include memory schemes *i.e.*, MEGA [160] that store useful information from the current environment to reuse it later in a new environment. Storing the optimum topology for each possible access vector in a real network would be very inefficient and would require huge amount of memory.

We have shown the robustness of our algorithm to the initial state of the network and, even when SGA is restarted from scratch after a change is detected, it outperforms considerably exhaustive search.

3.8 Chapter summary

In this chapter, we have presented a dynamic joint optimization of relaying topology, routing (power) and inter relay scheduling in MCNs. As a result we have developed a specific encoding and fitness control in a sequential genetic algorithm for relaying topology update. Depending on the traffic load, there may be situations where searching for the new optimum topology will be NP-hard.

Through numerical simulations we have shown that by using SGA the number of operations required to reconfigure the optimum topology is significantly reduced independently of the initial topology of the network. The utility function used in the optimization process drives the solution towards the topology favoring simultaneously isolated and short range transmissions. As

expected, within these solutions further improvements are obtained by using network coding to reduce the number of slots needed for transmission.

In addition to optimum performance in terms of network utility, numerical results demonstrate also significant improvements in the convergence rate of the new algorithm. The number of generated topologies in the search for the optimum one by using SGA-TSL is at least one order of magnitude less than by exhaustive search. The same order of improvement is obtained independently of the initialization of SGA-TSL. We also have compared the performance of SGA-TSL with nearest neighbor heuristic and the value of the fitness obtained is about 50% lower than with SGA-TSL.

SGA may be implemented in one of the base stations. Cooperating base stations must exchange information about the traffic distribution, and the coordinating base station should pass the information about the resulting access vector $\mathbf{a}_F$ back to the cooperating base stations. This level of coordination between the base stations seems to be already considered in practice *i.e.*, coordinated multipoint transmission, where a cluster of base stations jointly perform beamforming in order to reduce intercell interference.

4 Context aware nano scale modeling of MCNs for high resolution optimization

In this chapter, the model from Chapter 3 is further extended to provide higher resolution optimization of MCNs. Instead of using rings, we apply a hexagonal tessellation for inner partitioning of the cell into smaller subcells of radius *r*. Subcells may be several orders of magnitude smaller than e.g. micro cells resulting in, what we refer to as, a Nano Scale Network Model (NSNM). A special Nano Scale Channel Model (NSCM), enabling relay hop distance optimization, is developed for this application. By adjusting the radius of the subcell *r* we obtain different hopping ranges which directly affect the throughput, power consumption and interference.

We first apply this scheme to downlink multicast transmission in MCNs. By decreasing the radius of the subcells *r*, the number of hops needed for the BS to deliver the information to the final users is increased, which increases the delay and consequently decreases the throughput. On the other hand, as the hopping range is reduced, less transmission power is needed in each time slot which produces less interlink interference and enables larger number of concurrent transmissions.

By developing relations between the geometry of the cell (tessellation factor *r*) and the physical layer model, routing and topology, the network can be optimized by using one single control parameter *r*. With *r* as the optimization parameter, we jointly optimize scheduling, routing and power control, aware of the intercell and intersession interference to obtain the optimum trade-off between throughput, delay and power consumption in multicast MCNs. New solutions to the problem of reducing the impact of intercell or intersession interference are presented including a spatial interleaving SI MAC protocol to coordinate the multiple concurrent transmissions due to multihop relaying.

A set of numerical results demonstrates that the nano scale network model enables high resolution optimization of the system and an effective use of the context awareness. In general these solutions depend on the location of sinks and potential relays willing to cooperate. The sink location matrix **D** is used in the system utility function to relate optimal tessellation to the location of the set of the multicast receivers in the network. The relay availability and willingness to cooperate is controlled by the relay availability matrix **A** which is included in the constraints of the optimization program. Incorporation of these two additional context aware matrices, along with a new definition of directed flooding routing

protocol (DFRP) and inter flooding network coding (IFNC) for such a system, further improve the optimization process and system performance. An intercell flooding coordination (ICFC) protocol is developped to reduce the intercell interference.

A detailed discussion on network implementation is included addressing control data collection/distribution, mobility, signaling overhead, cell design, comparison with other protocols and node availability.

The system performance in MCNs is affected by various factors such as mobility, battery power, coverage, interference and density of nodes. Due to these factors, users may be temporally unavailable to relay the message which degrades the network performance and, establishing routes in such network efficiently is an important and challenging research issue. For this purpose, the NSNM is used to present a new concept for route discovery protocols which is aware of the mutual impact of all routes in the cell. The efficiency of the Nano Route Discovery Protocol (NRDP) is measured in terms of the utility function that includes: throughput, power consumption, terminal time to live (depletion) and delay. The numerical results show that the proposed algorithm is superior when compared to other existing route discovery protocols adapted to this scenario.

Finally, the NSNM is used to analyze the trade-off between cooperative diversity and spatial reuse in multihop cellular networks in order to maximize the throughput in the network. The increased number of concurrent transmissions, enabled by spatial cell partitioning, increases the system throughput but also increases the level of interference that reduces the capacity of simultaneously used links in the network. The radius of the subcells r determines the relaying hop range and the amount of interlink interference. All transmissions are recorded by the neighboring receivers and combined in a cooperative diversity transmission. The increased number of hops increases the diversity order ξ but at the same time reduces the throughput per user since the network capacity has to be shared between the increased number of users. By introducing a utility function as a ratio of the network throughput and the overall power consumption we can simultaneously optimize these parameters and the packet delivery delay, as a function of relaying range. The optimum relaying range defines the optimum tessellation factor r and the spatial reuse in the network.

The rest of the chapter is organized as follows. An overview and background in this area is given in Section 4.1. In Section 4.2, the context aware nano scale optimization of multicast MCNs is presented. In Section 4.2.1, we introduce the system model and notation, the model assumptions and its justifications. Section

4.2.2 presents the physical layer model with the nano scale channel model for this application. The network layer model is developed in Section 4.2.3 which includes the definition of the routing schemes, topology and intercell flooding coordination (ICFC) protocol. In Section 4.2.4, joint optimization of tessellation, scheduling, routing and power control for multicast MCNs is shown. Context aware route discovery protocol is developed and analyzed in Section 4.3. The trade-off between cooperative diversity and spatial reuse is analyzed in Section 4.4. Numerical results that validate our theoretical claims in certain representative scenarios are provided in Section 4.5. Implementation issues are discussed in Section 4.6. Finally, some concluding remarks are offered in Section 4.7.

4.1 Overview and background

This chapter provides a systematic approach to network optimization to study the gains and trade-offs associated with MCNs. To be able to deal with this level of details the network model from Chapter 3 is extended by first applying a hexagonal tessellation for inner partitioning of the cell into smaller subcells of radius r as shown in Fig. 31. This partitioning can be considered as a special form of surface tessellation technique used in conventional network information theory [44], [161] so that a number of results from that area can be adapted for the evaluation of the proposed system. By adjusting the radius of the subcell r (tessellation factor r) we obtain different hopping ranges which directly affect the throughput, power consumption and interference. By developing relations between the geometry of the cell (tessellation factor r) and the physical layer model, routing and topology we can optimize the network by using one single control parameter r.

We first apply this tessellation scheme to jointly optimize scheduling, routing and power control in multicast cellular networks. Multicast in MCNs has received more and more research attention [44], [163]-[165]. Inspired by the research for unicast traffic, a number of hybrid architectures have been proposed for multicast applications [163], [166]–[167]. In [163] near optimal multicast relay strategies are proposed to improve multicast throughput using ad hoc relays in a single cell environment. [166] develops a routing algorithm to find *ad hoc* paths from proxies to cellular multicast receivers. A multicast group selection algorithm is presented in [167] to guarantee certain QoS when multiple groups coexist in the cell. Significant amount of study on multicast has been focused on higher layer issues such as efficient design of routing protocols [168]-[171]. Several works has

been done on power aware routing since the nodes are characterized by their limited battery-power [169], [171]. In [169], the minimum-energy multicast problem is studied by exploiting the “wireless multicast advantage”. [171] finds a path for each node-pair connecting the source node and the destination set such that any node in each path does not run out of its power during the transmission of packets.

Multicast protocols have been shown to considerably reduce the bandwidth and power consumption compared with unicast counterparts [164]. For further enhancement of overall throughput, lower layers are also required to efficiently combat the adverse fading effect on multicast transmission. In [164], [165] opportunistic multicast schemes have been proposed to balance the trade-off between multiuser diversity and the multicast gain. With the best user approach base station transmits with the maximum rate to the user with the best channel. In this approach there is a risk of excessive delays for the worst user. In the worst user approach the base station transmits with the minimum rate so that all users can simultaneously receive the signal. Besides low rate this approach faces the problem of intercell interference.

Our proposed scheme provides new solutions to the problem of reducing the impact of intercell or intersession interference including the problem of coordinating multiple concurrent transmissions due to multihop relaying.

The nano scale network model is further extended to include a conventional resource reuse scheme used for cellular networks to design a new concept of route discovery protocols aware of the mutual impact of all routes in the cell. The same model is also used to provide an analysis of optimum hopping range, and analyze the trade-off between spatial reusability and cooperative diversity.

4.2 Context aware nano scale optimization of multicast MCNs

4.2.1 System model and assumptions

The downlink transmission is considered in a multicast cellular network with hexagonal cells of radius R and uniform distribution of the mobile users across the cell. In addition, hexagonal tessellation is applied in order to divide each cell into inner hexagonal subcells of radius r as shown in Fig. 31. For simplicity we assume the same tessellation factor r for all cells. We consider that the location of the mobile users is fixed during the scheduling cycle, and a potential, ready to

cooperate, transmitter/receiver is in average situated in the centre of those inner cells. The actual presence or absence of such user in a subcell will be characterized by entry 1 or 0 respectively in spatial user distribution matrix denoted as $\boldsymbol{\beta} = \|\beta(h,\theta)\|$ where the hopping index h and angle θ define the spatial position of the user in polar coordinates with respect to the base station $(h = 0,\ \theta = 0^\circ)$.

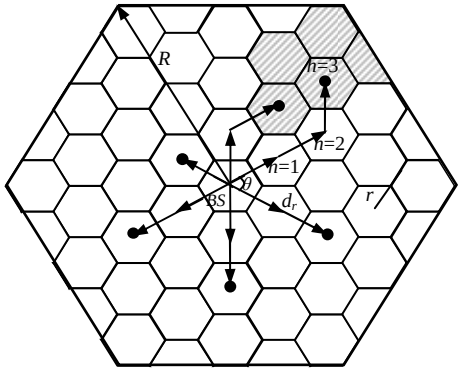

Fig. 31. Multihop transmission scheme.

In each cell, the network topology is represented by a directed graph $G = (\mathcal{N}, \mathcal{L})$ with a set of potential nodes (subcells) $\mathcal{N}$ and a set of links $\mathcal{L}$. The actual realization of the network will depend on $\boldsymbol{\beta} = \|\beta(h,\theta)\|$. Each link is associated with a nonnegative real number c_l representing its transmission capacity in bits per unit time. Since the capacity depends on the level of interference, this parameter will be also controlled by the same parameter r.

As we can see from Fig. 31, the BS is surrounded by $n_{h=1} = 6$ adjacent subcells, situated one hop $h = 1$ away from the BS, denoted as first ring of subcells. We assume that the rings are circular and concentric to the BS. In each 60° angular segment in the first hop, there is an additional subcell in the second hop (shaded), two new subcells in the third hop and so on. So, the number of subcells per ring is $n_h = 6 + (h-1)6 = 6 \cdot h$ and the number of subcells per cell

$$N = \sum_{h=1}^{H} n_h = \sum_{h=1}^{H} 6 \cdot h = 3H(H+1), \tag{37}$$

where H is the total number of rings in the cell. Different from Chapter 2 and 3, with this tessellation scheme the users hopping distance is the same in each hop so that number of rings in all cells is H. We should notice that with this

definition of N we are including the subcells situated in the border of the cell that are part of two different cells. The transmission between different subcells is controlled by introducing an intercell flooding coordination scheme (ICFC) that will be explained later in Section 4.2.3.C.

The angular separation between two adjacent subcells in ring h is $\hat{\theta}_h = 360^\circ / n_h = 60^\circ / h$. The transmission/relaying or hopping distance between adjacent rings is denoted by d_r as shown in Fig. 31 and is related to the subcell radius as $d_r = \sqrt{3} \cdot r$. The total number of rings in the cell is given by

$$H = \lfloor \mathcal{H} / d_r \rfloor = \lfloor \mathcal{H} / (\sqrt{3} \cdot r) \rfloor = \lfloor R / (2 \cdot r) \rfloor, \tag{38}$$

where $\mathcal{H}$ is the apothem of the hexagonal cell. The subcell radius r determines the number of rings H and the amount of interference. These two aspects will be elaborated in detail in the next section.

In polar coordinates (h, θ) we denote the location of a mobile user situated in ring h at angle θ as

$$m(h, \theta) = h \cdot d_r \cdot e^{j\theta} \tag{39}$$

$$\theta \in \Theta^{(h)} = \left\{ \theta_n^{(h)} \right\} = \left\{ \theta_1^{(h)}, ..., \theta_{n_h}^{(h)} \right\}, \tag{40}$$

where $\Theta^{(h)}$ is the set of angles θ for the users located in hop h, and the number of elements in this set $\left| \Theta^{(h)} \right| = n_h$. The detailed computation of the elements of set $\Theta^{(h)}$ is elaborated in Appendix A.

With this notation, the spatial user distribution matrix $\boldsymbol{\beta} = \left\| \beta(h, \theta) \right\|$ has dimensions $H \times \overline{\theta}_H$, where $\overline{\theta}_H$ is the overall number of different user angles (see Appendix A, *e.g.*, (A3)). For the example shown in Fig. 31 for single cell case with $H = 4$, $\overline{\theta}_H$ is 36.

In general, the user at location $m(h, \theta)$ is denoted as $m_{h,\theta}$ and any user in ring h (for any angle θ) as m_h. The BS is referred to as m_0 and its location is $m(0,0)$.

In multihop transmission, user at location $m(h, \theta)$ relays the information to any of its adjacent users $m(h', \theta')$. As we are considering the transmission in one direction (downlink) we assume that the user will relay to any of its 5 adjacent users on the way to the destination user $m(h_d, \theta_d)$. If we denote the location of user $m(h, \theta)$ in vector representation as $\vec{m}(h, \theta) = h \cdot d_r \cdot e^{j\theta}$, the adjacent relay user at location $m(h', \theta')$ is calculated in vector form as

$$\vec{m}(h', \theta') = h' \cdot d_r \cdot e^{j\theta'} = \vec{m}(h, \theta) + d_r \cdot e^{j\left(\frac{\pi}{2}\operatorname{sgn}(\pi - \theta) + \alpha_{h,\theta}\frac{\pi}{3}\right)} \tag{41}$$

$$\theta' = arg(\vec{m}(h',\theta')) = \arg\left\{\vec{m}(h,\theta) + d_r e^{j(\frac{\pi}{2}\mathrm{sgn}(\pi-\theta)+\alpha_{h,\theta}\frac{\pi}{3})}\right\};\ \alpha_{h,\theta} = 0,\pm1,\pm2, \qquad (42)$$

where parameter $\alpha_{h,\theta}$ is the routing parameter at subcell (h,θ), and indicates the direction to which of the 5 adjacent users we are relaying. The routing across the cell will be characterized by the routing matrix $\boldsymbol{\alpha} = \|\boldsymbol{\alpha}(-2)\ \boldsymbol{\alpha}(-1)\ \boldsymbol{\alpha}(0)\ \boldsymbol{\alpha}(1)\ \boldsymbol{\alpha}(2)\| = \|\boldsymbol{\alpha}(\alpha_{h,\theta})\|$ which is (1x5 block matrix) with dimensions $H \times \left(\bar{\theta}_H\right)^5$. As an example, we write $\boldsymbol{\alpha}$ for the scenario presented in Fig. 31.

$$\boldsymbol{\alpha}(-2) = \boldsymbol{\alpha}(2) = \mathbf{0}$$

$$\boldsymbol{\alpha}(-1) = \begin{array}{c} h\backslash\theta \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0^\circ & 10^\circ & 15^\circ & \ldots & 90^\circ & \ldots & 345^\circ & 350^\circ \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \ldots & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 \\ \vdots & & & & & & & \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 \end{bmatrix} \end{array}; \quad \boldsymbol{\alpha}(1) = \begin{array}{c} h\backslash\theta \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 0^\circ & 10^\circ & \ldots & 30^\circ & \ldots & 345^\circ & \ldots\ 350^\circ \end{array} \\ \begin{bmatrix} 0 & 0 & \ldots 0 & \ldots & 0 & \ldots & 0 \\ 0 & 0 & \ldots 1 & \ldots & 0 & \ldots & 0 \\ \vdots & & & & & & \\ 0 & 0 & \ldots 0 & \ldots & 0 & \ldots & 0 \end{bmatrix} \end{array}$$

$$\boldsymbol{\alpha}(0) = \begin{array}{c} h\backslash\theta \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccccccc} 0^\circ & 10^\circ & \ldots & 30^\circ & \ldots & 210^\circ & \ldots 270^\circ & \ldots & 330^\circ & \ldots & 350^\circ \end{array} \\ \begin{bmatrix} 0 & 0 & \ldots & 1 & \ldots & 1 & \ldots 1 & \ldots & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & \ldots & 0 & \ldots 0 & \ldots & 0 & \ldots & 0 \\ \vdots & & & & & & & & & & \\ 0 & 0 & \ldots & 0 & \ldots & 0 & \ldots 0 & \ldots & 0 & \ldots & 0 \end{bmatrix} \end{array}.$$

In this example $\boldsymbol{\alpha}(-2) = \boldsymbol{\alpha}(2) = \mathbf{0}$. If the traffic from user $m_{h,\theta}$ is relayed in the direction with $\alpha_{h,\theta} = -1$, the entry of the matrix $\boldsymbol{\alpha}(-1)$ with indices h,θ is $\alpha_{h,\theta}(-1) = 1$ and zero otherwise. We can see in Fig. 31 that the user in (h = 1, θ = 30º) relays to user in (h' = 2, θ = 30º) so $\alpha_{1,30^\circ}(0) = 1$. The same applies for users in h = 1, θ = 210º, θ = 270º and θ = 330º. The user in (h = 1, θ = 90º), relays with $\alpha_{1,90^\circ}(-1) = 1$ to user in (h' = 2, θ' = 60º), with θ' calculated as (42), and so on.

If users at locations $m(h,\theta)$ and $m(h',\theta')$ are the origin and destination respectively of link $l \in \mathcal{L}$, then the link is represented as $l\left(m_{h,\theta} \to m_{h',\theta'}\right)$ and its vector representation as

$$\vec{l}\left(m_{h,\theta} \to m_{h',\theta'}\right) = \vec{m}(h',\theta') - \vec{m}(h,\theta) = d_r \cdot e^{j(\frac{\pi}{2}\mathrm{sgn}(\pi-\theta)+\alpha_{h,\theta}\frac{\pi}{3})},\ \alpha_{h,\theta} = 0,\pm1,\pm2. \qquad (43)$$

The values of $\alpha_{h,\theta}$ will depend also on the relaying scheme used as it will be explained in Section 4.2.2.

In addition we define the destination matrix $\mathbf{D} = \|D_{h,\theta}\|$ with dimensions $H \times \overline{\theta}_H$, where $D_{h,\theta} = 1$ if the user at location (h,θ) is a destination user (sink), or 0 otherwise and $\overline{\theta}_H$ is the number of different user angles defined in Appendix A by (A3) for $h = H$. Destination users can also act as relays to transmit the data to another destination. We denote by $\Re_{h_d,\theta_d}$ the set of links that form the relaying route from user in location $m(0,0)$ to the destination (final) user at location $m(h_d,\theta_d)$. One should notice that for $h' = h + 1$ every route is effectively defined by a particular choice of matrix $\boldsymbol{\alpha} = \|\alpha_{h,\theta}\|$. The set of all routes to all destination users $\mathcal{D}$ is denoted by

$$\Re^{\mathcal{D}} = \bigcup_{h_d,\theta_d \in \mathcal{D}} \Re_{h_d,\theta_d}. \tag{44}$$

The relay availability and willingness to cooperate is characterized by the relay availability matrix $\mathbf{A} = \|A_{h,\theta}\|$ where $A_{h,\theta} = 1$ if user at location (h,θ) is available and willing to cooperate as relay, or 0 otherwise. The issue of stimulating the user cooperation is studied extensively in the literature [172, Chapter 16] and is in the area of pricing. Game theory is used as a tool for the analysis and billing with awards for relaying as a method to implement it. The interest to cooperate may be also significantly influenced by its battery state. All together these issues are out of the scope of this thesis.

A network topology graph $G = (\mathcal{N},\mathcal{L})$, a set of link capacities $C = \{c_l, l \in \mathcal{L}\}$, a set of sinks $\mathcal{D}$ at positions $\mathbf{D} = \|D_{h,\theta}\|$ and a set of multicast connection requirements (capacity, throughput) $\mathcal{C}$ specify a multicast connection problem.

As the location of the mobile users is assumed to be fixed during the duration of the scheduling cycle, the sink location **D** and the user availability **A** are also assumed to be fixed during that period.

4.2.2 Physical layer model

After presenting the notation and the cell tesellation scheme, in this section we focus on the channel model proposed for this application and the relaying schemes used.

A) Nano scale channel model

The channel model introduced in this section, which is referred to as Nano Scale Channel Model (NSCM), consists of a Rician channel with progressive decrease of the steady state component with increase of the transmission range. In other words, when the number of hops used to reach a given destination is small (longer transmission range) the channel has stronger random component due to reflections, but on the other hand when the number of hops increases (the transmission range decreases) the random component decreases and the steady (line of sight) component becomes predominant. Let us denote by H_0 the smallest number of hops needed to make the transmission range, in a cell with radius *R*, such that the random component of the channel is eliminated (details on implementation are given is Section 4.6). Such tessellation is referred to as channel defading (CD) tessellation. For any other tessellation, the channel model can be represented as

$$g_{h,h'} = g^0_{h,h'} + g^{\chi}_{h,h'} = \sqrt{\frac{1}{(d_r)^{\alpha}}}\left(\frac{H-1}{H_0-1} + \left(1-\frac{H-1}{H_0-1}\right)\chi\right);\ H \le H_0 \tag{45a}$$

$$\chi \sim N\left(0, \sigma^2_{\chi}\right)$$

$$G_{h,h'} = G^0_{h,h'} + G^{\chi}_{h,h'} = \frac{1}{(d_r)^{\alpha}}\left(\frac{H-1}{H_0}\right)^2 + \frac{1}{(d_r)^{\alpha}}\left(1-\frac{H-1}{H_0}\right)^2 \sigma^2_{\chi}, \tag{45b}$$

where (45a) defines the overall channel gain, and $g^0_{h,h'}$, $g^{\chi}_{h,h'}$ are the steady state and the random component of the channel gain respectively. In (45b), $G^0_{h,h'}$, $G^{\chi}_{h,h'}$ are the power gain for the steady state and the random component respectively, and χ is circular complex Gaussian process. For $H = 1$ the channel is pure Rayleigh channel, for $1 < H < H_0$ Rician channel and for $H \ge H_0$ we have Gaussian channel (no fading).

Determining H_0 for different types of channel (urban, suburban, …) belongs to the channel modeling and is not within the scope of this thesis.

In a CD tessellation the interference power produced by a cochannel interfering user m_h, transmitting at the same time from the position (h,φ), $\varphi \neq \theta$ is $I_{h',\theta'} = G_{h,\varphi\neq\theta;h',\theta'}P_h$ where $G_{h,\varphi\neq\theta;h',\theta'}$ is the channel gain between referent receiver $m_{h',\theta'}$ and the cochannel interfering user $m_{h,\varphi\neq\theta}$ approximated by $G_{h,\varphi\neq\theta;h',\theta'} \simeq 1/(d_{h,\varphi\neq\theta;h',\theta'})^{\alpha}$. Parameter α is the propagation loss and $d_{h,\varphi\neq\theta;h',\theta'}$ is the mutual interfering distance given by

$$
\begin{aligned}
d_{h,\varphi\neq\theta;h',\theta'} &= \left|\vec{m}(h',\theta') - \vec{m}(h,\varphi\neq\theta)\right| \\
&= d_r\sqrt{(h')^2 + (h)^2 - 2\cdot(h')\cdot h\cdot\cos(\theta',\varphi)} \\
&= d_r\cdot Z_{h,\varphi\neq\theta;h',\theta'}.
\end{aligned}
\tag{46}
$$

The modification for the general case of non CD tessellation is straightforward and will be elaborated in next section.

The previous generations of mobile communications have been also using different approaches for channel defading. CDMA based technology [172] converts selective fading channel into a Gaussian channel by first extracting multipath component and after that combining them in a Rake receiver. OFDMA based technology [172] splits a given bandwidth into a number of subchannels by choosing the number of the subcarriers large enough so that the channel on any subcarrier is nonselective. In this chapter, in the nano model of cellular system, we split the size of the cell into a number of subcells large enough for channel defading. While the previous technologies achieve the channel defading in the physical layer, the channel defading tessellation achieves this effect on the network layer by proper combination of topology and routing.

B) Conventional Relaying with SI MAC (Protocols A & B)

Directed flooding routing protocols (DFRP) presented in this section are based on the conventional relaying scheme where the signal propagates through the cell from ring h to $h' = h+1$ as illustrated in Fig. 32 and Fig. 33 (protocol A and B, respectively). In these figures, transmissions in a given hop h take place in a given time slot b, so that $\mathbf{\Psi}^b$ (b = 1,2,…,B) represent the positions of active transmitters in slot b (delay) and small empty circles represent the positions of the receivers where the signal is received with no collision in that particular hop.

We introduce the protocol transmission matrix $\mathbf{\Psi}^b = \left\|\mathbf{\Psi}^b(h,\theta)\right\|$ with entries $\mathbf{\Psi}^b(h,\theta) = 1$ if a given protocol assumes a transmitter at location (h,θ) in slot b and zero otherwise. Matrix $\mathbf{\Psi}$ defined as $\mathbf{\Psi} = \bigcup_b \mathbf{\Psi}^b$ includes transmitters in all slots. The dot matrix product $(\cdot *)$ defined as $\mathbf{\Psi}' = \mathbf{\Psi}\cdot *\mathbf{A} = \left|\Psi'_{h,\theta}\right| = \left|\Psi_{h,\theta}A_{h,\theta}\right|$ has value $\Psi'_{h,\theta} = 1$ if the required transmitter is available and 0 otherwise.

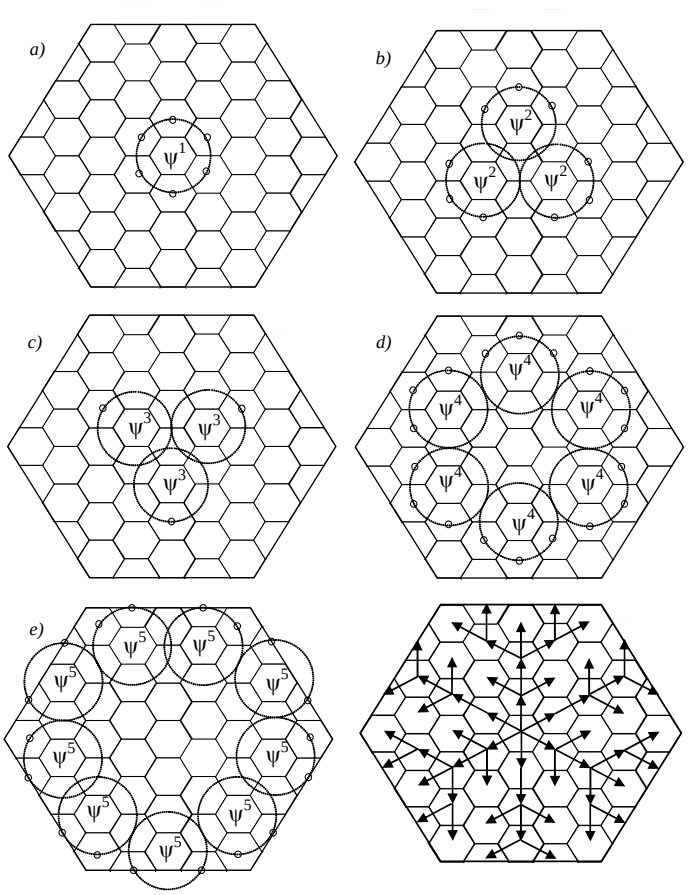

Fig. 32. a)-e) Network topology schedule for broadcast case (protocol A); f) All routes available using protocol A, ([122] [©IEEE 2012]).

It should be noted that protocols A and B, shown in Fig. 32 and Fig. 33 respectively, have different scheduling patterns. The mutual interference between the links is controlled by proper spatial selection of the transmission sites, which is referred to as spatial interleaving (*SI*). In order to avoid the collision between adjacent subcells, in each ring *h* the simultaneously active transmitters, situated in the center of the subcell, are separated (interleaved) in angular domain

$$\hat{\phi}_h = \frac{360^\circ}{(n_h / 2)} = 120^\circ / h$$

with $n_h / 2$ simultaneously active transmitters per ring *h*.

The protocol is aware of the location of the mobile users in the cell and assigns the transmission turn in such a way that the collision between different transmitters is avoided. More details about the implementation are given in Section 4.6.

By inspection of Fig. 32 we can see that the minimum interfering distance in Fig. 32b (slot $b = 2$) denoted as $d_i(\Psi^{b,p})$ where *b* is the slot and *p* denotes the protocol, is $d_i(\Psi^{2,A}) = 2d_r$, and the minimum interfering distance for subsequent transmissions is $d_i(\Psi^{3,A}) = \sqrt{(2d_r)^2 + (3r)^2}$, $d_i(\Psi^{4,A}) = 3r$ and $d_i(\Psi^{5,A}) = 3r$. Similar relations can be derived for protocol B shown in Fig. 33. We can observe from Fig. 32 and 33, that although in Fig. 33 one more slot is needed to complete the transmission, the interference produced in $\Psi^{4,B}$ is smaller than in $\Psi^{4,A}$ due to larger distance between the transmitters of interfering users $d_i(\Psi^{4,B}) = 2d_r$, $d_i(\Psi^{5,B}) = \sqrt{(d_r / 2)^2 + (9r / 2)^2}$.

The signal to interference plus noise ratio SINR at user $m_{h'}$ is defined as

$$\begin{aligned} SINR^{con}_{h',\theta'}(\mathbf{P}, r) &= G_{h,h'} P_h / \left(\sum_{\varphi \neq \theta} G_{h,\varphi \neq \theta;h',\theta'} P_h + N_{h'} \right) \\ &= \frac{P_h / (d_r)^\alpha}{\sum_{\varphi \neq \theta} P_h / (d_r \cdot Z_{h,\varphi \neq \theta;h',\theta'})^\alpha + N_{h'}} \\ &= \frac{P_h}{\sum_{\varphi \neq \theta} P_h / Z^\alpha{}_{h,\varphi \neq \theta;h',\theta'} + N_{h'} \cdot (\sqrt{3} \cdot r)^\alpha} \end{aligned} \tag{47}$$

and depends on the transmission power of user m_h, given by $P_h, \forall h, \theta$, and on the tessellation factor *r*. In a non CD tessellation $P_h = S_{h'} / G_{h,h'} \rightarrow P_h = S_{h'} / G^0_{h,h'}, \forall \theta$ and $N_{h'} \rightarrow N_{h'} + G^\chi_{h,h'}$, where $G^0_{h,h'}$ and $G^\chi_{h,h'}$ are given by (45b) and $N_{h'}$ is the Gaussian noise power. The constant $Z_{h,\varphi \neq \theta;h',\theta'}$ is defined as in (46). Based on this definition of the SINR, the capacity of link $l\left(m_{h,\theta} \rightarrow m_{h',\theta'}\right) \in \Re_{h_d,\theta_d}$ is

$$c_l = \log(1 + SINR^{con}_l(\mathbf{P}, r)) = \log(1 + SINR^{con}_{h',\theta'}(\mathbf{P}, r)), \ \forall l \in \Re_{h_d,\theta_d} \tag{48}$$

where *h'* = *h* + *1* and *θ'* is given by (42).

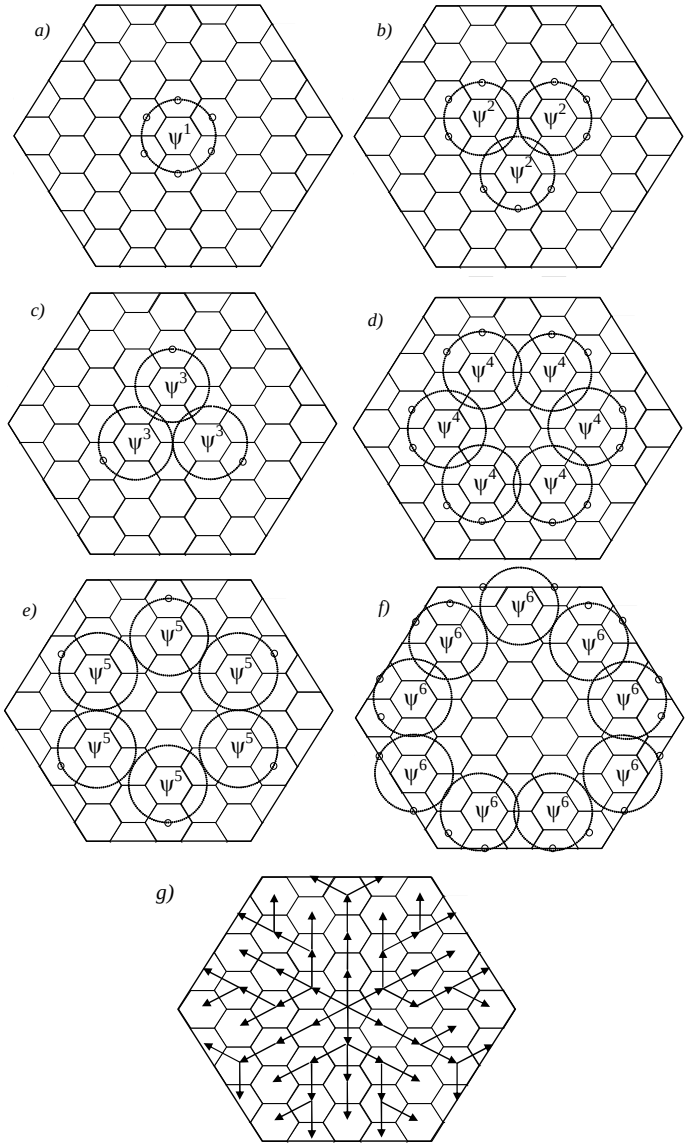

Fig. 33. a)-f) Network topology schedule for broadcast case (protocol B); g) All routes available using protocol B, ([122] [©IEEE 2012]).

The Shannon channel capacity is approximated in practice by a specific discrete value depending on the overall receiver capabilities (amount of coding, modulation…). The capacity of the route $\Re_{h_d,\theta_d}$ is the minimum of the link capacities on that route

$$c_{\Re_{h_d,\theta_d}} = \min_l c_l,\ \forall l \in \Re_{h_d,\theta_d}. \tag{49}$$

The multicast (broadcast) capacity of the overall network is the minimum of the link capacities to all destination users $m(h_d,\theta_d) \in \mathcal{D}$

$$C = c_{\Re^D} = \min_{h_d,\theta_d} c_{\Re_{h_d,\theta_d}},\ \forall \Re_{h_d,\theta_d} \in \Re^{\mathcal{D}} \tag{50}$$

and the throughput $Thr = C/B$ where B is the number of slots needed for the protocol to reach all destination users $m(h_d,\theta_d) \in \mathcal{D}$.

A) Cooperative Relaying (Protocol C)

The multicast transmission by cooperative diversity relaying scheme is represented in Fig. 34 and referred to as protocol C. User $m(h',\theta')$ receives the signal simultaneously transmitted by $m(h,\theta)$ and from different transmitters from the positions $m(h,\varphi),\ \varphi \neq \theta$ transmitting in the same ring. It combines all received signals by using diversity receiver. With this scheme the transmission delay to the border of the cell is reduced to 4 instead of being 5 or 6 as it was in Fig. 32 and Fig. 33 respectively. The signal to noise ratio SNR at user $m_{h'}$ is also improved and now becomes

$$SNR^{coo}_{h',\theta'}(\mathbf{P},r) = \left(G_{h,h'}P_h + \sum_{\varphi \neq \theta} G_{h,\varphi\neq\theta;h',\theta'}P_h + \sum_{\eta<h,\ \phi} G_{\eta,\phi;h',\theta'}P_\eta \right) / N_{h'} \tag{51}$$

with $P_\eta = P_h$ and CD tessellation. The third term is due to accumulated signals received in the previous transmissions. The capacity is given again by (48) and (49). We should notice that here in some transmissions $h' = h$ as there is no collision due to the diversity receivers and θ' is given by (42). In this scheme the spatial interleaving between simultaneously active transmitters is reduced with respect to the previous schemes due to the diversity receivers.

The fundamental concept at the physical layer is the capacity region $\mathcal{C}$ which characterizes the trade-off between power allocation and link capacity. In the case of multihop transmission we also need to include parameter H that is related to the tessellation factor r by (38). We consider slotted time, such that each packet

transmission takes one time slot (one hop per slot), and multihop packet propagation in a store and forward manner. The set $(\mathbf{c},\mathbf{P},r)$ defines the system with link capacities **c**, links power **P** provided to the transmitting users for a given tessellation factor *r*.

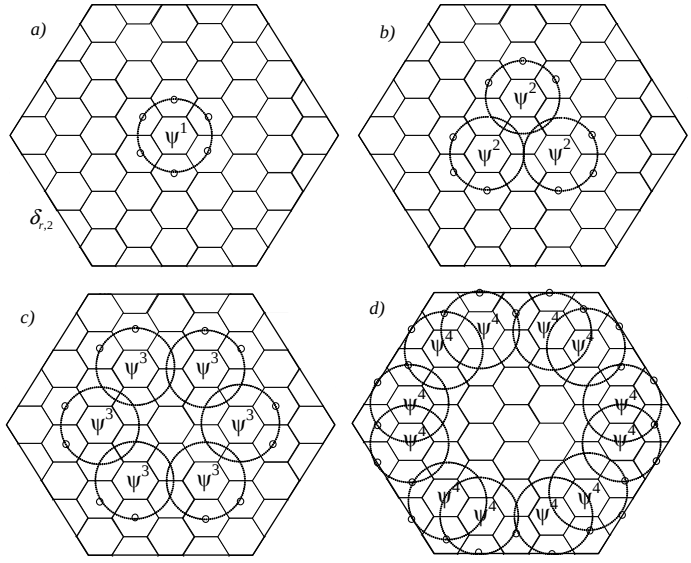

Fig. 34. Network topology schedule for broadcast case and cooperative diversity receivers (Protocol C), ([122] [©IEEE 2012]).

4.2.3 Network layer model

In this section we elaborate the route discovery process and routing schemes for the nano scale model. The available routes for protocol A and B are presented in Fig. 32f and Fig. 33g, respectively. The representation of the available routes for protocols C and D (introduced later) are omitted but can be easily calculated by following the same logic.

By defining the topology matrix for each slot we can describe the routes that the BS will use to reach different users as final destinations **D**. We define the network topology matrix in time slot *b* as $\mathbf{T}^b = \left\| \mathbf{T}_{tx}^b(h,\theta)\ \ \mathbf{T}_{rx}^b(h',\theta') \right\|$ with

dimensions $(H+1)\times(\bar{\theta}_H+\bar{\theta}_H)$, where $\bar{\theta}_H$ is the number of angular positions of the transmitters and receivers and is defined in Appendix A by (A3) for $h = H$ and $h'=h+1$. The topology matrix corresponds to a 1x2 block matrix, where the first block corresponds to the transmitters and the second one to the receivers. In this case the number of rows is $(H+1)$ because we are including the BS $m(0,0)$ situated at hop $h = 0$. So, the transmitters are situated in hop $h=0,1,...,H-1$, and the receivers in hop $h'=1,2,...,H$, thus $\mathbf{T}^b_{tx}(H,\theta)=\mathbf{0}$, $\mathbf{T}^b_{rx}(0,\theta)=\mathbf{0}$.

The detailed calculation of the topology matrix for protocols A, B and C presented in Fig. 32, 33 and 34 is elaborated in Appendix B.

An alternative representation of $\mathbf{T}$ can be obtained by $\mathbf{T}^b=\left\|\mathbf{T}_{h,h'}(\theta,\theta')\right\|$, with $\sum_b \mathbf{T}^b=\boldsymbol{\beta}$ and θ' calculated as in (42). The topology matrix $\mathbf{T}$ should provide that all destination nodes can be reached. In the broadcast case, $\mathbf{D}=\boldsymbol{\beta}$.

A) Route discovery process

A source node m_0 multicasts information simultaneously to all receivers at rate *x*. The unicast relaying route $\Re_{h_d,\theta_d}$ from user m_0 to destination user $m(h_d,\theta_d)\in\mathcal{D}$, is defined by a sequence of routing coefficients $\Re_{h_d,\theta_d}=\left\{\alpha_{1,\theta^{(1)}},\alpha_{2,\theta^{(2)}},...,\alpha_{h_d-1,\theta^{(h_d-1)}}\right\}$ with $h=1,2,..,h_d-1$. The relaying route is obtained from the topology matrix and matrices $\boldsymbol{\beta},\mathbf{A},\boldsymbol{\alpha}$ and $\mathbf{D}$ in the following way:

Route discovery protocol

1) For each destination user $m(h_d,\theta_d)\in\mathcal{D}$, depending on the protocol used, we check the topology matrix $\mathbf{T}^b=\left\|\mathbf{T}^b_{tx}(h,\theta)\ \mathbf{T}^b_{rx}(h_d,\theta_d)\right\|$ to obtain the candidate transmitter $m(h,\theta)$ = $m(h=h_{d-1},\theta=\theta_{d-1})$ for that user $m(h_d,\theta_d)\in\mathcal{D}$.

2) If the transmitter is available $A(h,\theta)=1$, then we calculate $\alpha_{h,\theta}$ as solution to (42), and we get the first component of the relaying route $\Re_{h_d,\theta_d}=\left\{\alpha_{h,\theta}\right\}$[3].

3) If user $m(h,\theta)\neq m(0,0)$ then go to 1) with $h_d\leftarrow h_{d-1}$, $\theta_d\leftarrow\theta_{d-1}$, and the process is repeated until the route is complete.

[3] Modifications of the algorithm in the case when a node is not available are further elaborated in Section 4.6 since the choices depend on acceptable performance degradations. Also, a nano route discovery protocol is presented in Section 4.3 to deal with this situation.

The route $\Re_{h_d,\theta_d}$ from user m_0 to destination user $m(h_d,\theta_d)\in\mathcal{D}$ is feasible (feasibility $F_{h,\theta}$) if all required transmitters on the route are available so that $F_{h,\theta}=\prod_{k=1}^{h}\Psi'_{k,\theta}=1$.

The multicast route $\Re^{\mathcal{D}}$ from m_0 to the set of sinks $\mathcal{D}$ is defined as in (44). In the case of multisource multicast we have M information sources at m_0, M groups of sinks $\mathcal{D}=\{\mathcal{D}^1,\mathcal{D}^2,...,\mathcal{D}^k,..,\mathcal{D}^M\}$ and M independent data flows $f\in F=\{f^1,f^2,...,f^k,..,f^M\}$. The flow from user m_0 to each destination user $m(h^k_{di},\theta^k_{di})\in\mathcal{D}^k$, $i=1,2\ldots$, $|\mathcal{D}^k|$ is $f^k=\{f^k_{0,0;h^k_{di},\theta^k_{di}}\}=\{f^k_{\Re^k_i}\}$ with $m(0,0)=m(0)$ and $\Re^k=\{\Re^k_i\}$ the set of relaying routes (multicast tree) for destination users in $\mathcal{D}^k$. Each flow component on that set will be denoted as f^k_l (flow on link l in session k).

Let $\mathbf{x}=\{x^k\}=\{f^k/B^{k,p}\}$ be the vector of multicast throughputs for each session k and $B^{k,p}$ is the maximum number of slots needed in session k to reach the border of its broadcast area which also depend on the protocol (p) used. The routing region $\Re$ defines the set $(\mathbf{x},\mathbf{f})$ such that flow rates $\mathbf{f}$ can support multicast throughput $\mathbf{x}$. The characterization of the routing region $\Re$ is a fundamental concept at the network layer, and depends on the specific choice of the protocol $p\in\wp(\text{A,B,C,D})$.

In the sequel, we include some additional constraints on the flows (*flow conservation equations*). We denote by $\mathcal{R}$ the set of relaying nodes (a destination node can act as relay too).

$$f^k_{0,0;h^k_{di},\theta^k_{di}}=f^k_{0,0;h^k_{dj},\theta^k_{dj}};\forall i,j,\ m(h^k_{di},\theta^k_{di}),\ m(h^k_{dj},\theta^k_{dj})\in\mathcal{D}^k \tag{52}$$

$$\sum_k f^k_l\le c_l \tag{53}$$

$$f_{h,\theta;0,0}=0;\ \forall h,\theta \tag{54}$$

$$\sum_{h,\omega}f_{h-1,\omega;h,\theta}-\sum_{h,\theta}f_{h,\theta;h',\theta'}=0;\forall\omega;\ m(h,\theta)\in\mathcal{R}^k, \tag{55}$$

where (52) indicates that there is one flow per session, (53) requires that the overall flow on each link should be less than the link capacity, (54) constraints all flows directed into the source node to be equal to 0, and (55) indicates that the net-in flow for the relays nodes must be equal to the net-out flow.

B) Nano Scale Routing with Interflood Network Coding (Protocol D)

In general, it is theoretically possible to apply network coding on multiple incoming streams of different sessions [173] which is referred to as intersession

network coding. However, we argue against this possibility, and use *coding by superposition* [174], *i.e.*, network coding applied only to incoming streams of the same session. This argument is mainly supported by the computational intractability of the optimal throughput problem if inter-session coding is allowed and it is not practical to code data streams from different applications either.

In this section, we introduce Interflooding Network Coding protocol (IFNC) which is explained by using the example shown in Fig. 35 (Protocol D). In the first time slot $b = 1$, the base station transmits two bits (y_a and y_b) to different sectors as shown in Fig. 35a. This transmission is designated as $\Psi^1_{y_a/y_b}$. For the second slot $b = 2$, transmission (flooding) protocol A from Fig. 32 is used to transmit bit y_a as shown in Fig. 35b. For $b = 3$, protocol B from Fig. 33 is used for flooding bit y_b as in Fig. 35c. For $b = 4$, *interflooding network coding* is used by transmitting $y_c = y_a + y_b$ from positions designated as $\Psi^4_{y_c}$ as represented in Fig. 35d. By using the same notation the three following floodings are designated as $\Psi^5_{y_a}$, $\Psi^6_{y_c}$ and $\Psi^7_{y_a}$ as indicated in Fig. 35e, Fig. 35f and Fig. 35g respectively. One can see from the figure that in this way all subcells will be able to decode both, bit y_a and bit y_b in seven slots.

We define the topology matrix with network coding as

$$\begin{aligned} \mathbf{T}^b &= \left\| \mathfrak{T}^b_{tx}(h,\theta) \;\; \mathfrak{T}^b_{rx}(h',\theta') \right\| \\ \mathfrak{T}^b_{tx} &= \left\| \mathbf{T}^b_{tx,y_a} \;\; \mathbf{T}^b_{tx,y_b} \;\; \mathbf{T}^b_{tx,y_c} \right\|, \;\; \mathfrak{T}^b_{rx} = \left\| \mathbf{T}^b_{rx,y_a} \;\; \mathbf{T}^b_{rx,y_b} \;\; \mathbf{T}^b_{rx,y_c} \right\|. \end{aligned} \tag{56}$$

The topology matrix $\mathbf{T}^b$ includes submatrix: $\mathfrak{T}^b_{tx}$ that indicate the location of the transmitters in slot b, as with network coding we transmit bits y_a, y_b and $y_c = y_a + y_b$. The location of this transmitters is indicated by submatrix $\mathbf{T}^b_{tx,y_a}$, $\mathbf{T}^b_{tx,y_b}$ and $\mathbf{T}^b_{tx,y_c}$ respectively and the same applies for the location of the receivers $\mathfrak{T}^b_{rx}$ with submatrices $\mathbf{T}^b_{rx,y_a}$, $\mathbf{T}^b_{rx,y_b}$ and $\mathbf{T}^b_{rx,y_c}$. The detailed content of these submatrices can be easily calculated based on the topologies from protocols A, B and C as presented in Appendix C.

The network capacity by using network coding is

$$C = 2c_{\Re^D} = 2\min_{h_d,\theta_d} c_{\Re_{h_d,\theta_d}}, \; \forall \Re_{h_d,\theta_d} \in \Re^{\mathcal{D}}, \tag{57}$$

where $c_{\Re_{h_d,\theta_d}}$ is the minimum of the links capacities on that route $\Re_{h_d,\theta_d}$ calculated as in (49). Factor 2 is due to the fact that the protocol will deliver two bits y_a, y_b to the cell boundary after B^D slots.

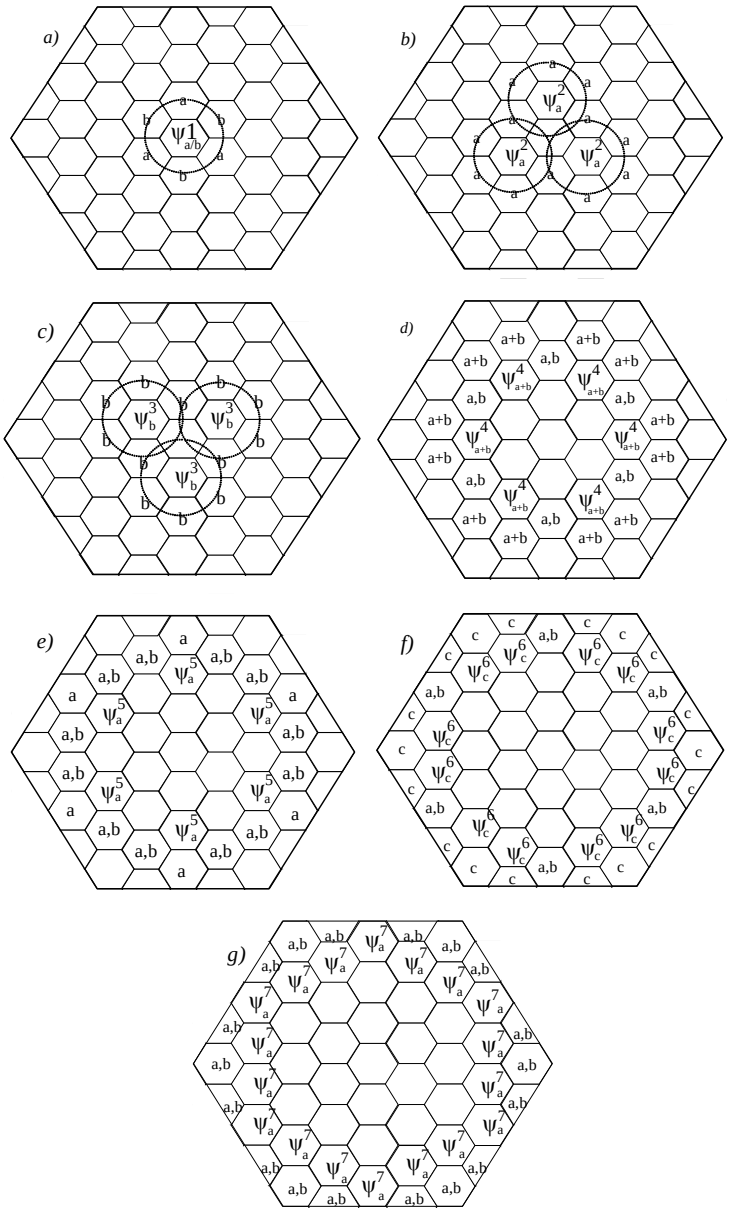

Fig. 35. Network topology schedule for broadcast case and network coding with $y_c = y_a + y_b$ (Protocol D), ([122] [©IEEE 2012]).

C) Intercell Flooding Coordination (ICFC)

In a multicell scenario, in order to avoid intercell interference the initiation of the flooding in the cell should be coordinated within the cluster of seven cells as illustrated in Fig. 36 for H = 1, 2 and 3 hops. The hops are represented as rings, simultaneous transmissions are shown with the same color and the numbers represent the time slots when the corresponding segments of the cells are active. We can see that in these figures there are no simultaneous transmissions at the borders of the cells.

Formally, the intercell reuse factor *irf* is defined as $irf = \tau^{(n)} / \tau^{(1)}$ where $\tau^{(n)}$ is the time needed for all cells in the multicell network to schedule their transmission in the last ring and $\tau^{(1)}$ is the time for an isolated single cell to complete transmission from the BS to the border of the cell. Let us first assume in general that the protocols spend equal time τ (amounts of slots) in each hop (ring of the cell).

In general, we require that for multihop transmission no two adjacent cells are allowed to transmit in the last hop $h = H$ at the same time. One can see from Fig. 36 that for one hop multicast (H = 1) the intercell reuse factor is *irf* = 3 since $\tau^{(1)} = 1$ and $\tau^{(n)} = 3$. So, 3 slots are needed to complete the transmission in all cells.

In this three examples the transmission within the cell is completed within 3 slots for the whole network giving *irf* = 3/2 for H = 2 and *irf* = 3/3 = 1 for H = 3. In the case when the protocol spends uneven times in different rings this discussion requires more details. For this discussion we introduce the following parameters: B^p - number of slots needed for protocol p to complete the transmission from the BS to the border of the cell and $B^{H,p}$ - number of slots protocol p spends in the last ring. From Fig. 36 we can see that there are three constellations of the multiple cells transmitting simultaneously: outer ring marked in 1) gray, 2) dashed or 3) white. For a given B^p network needs $3B^{H,p}$ slots to schedule the transmissions in the last ring. Based on this we have $irf = max(3B^{H,p}, B^p) / B^p$ and $irf = 1$ as long as $3B^{H,p} \le B^p$. Therefore in a multicell network the throughput defined in the previous section should be further modified by factor 1/*irf*.

The above examples are demonstrated in Fig. 36 for the cluster of 7 cells for simplicity. By using the same definition for *irf* and condition for *irf* = 1 as before, these parameters can be calculated for any network.

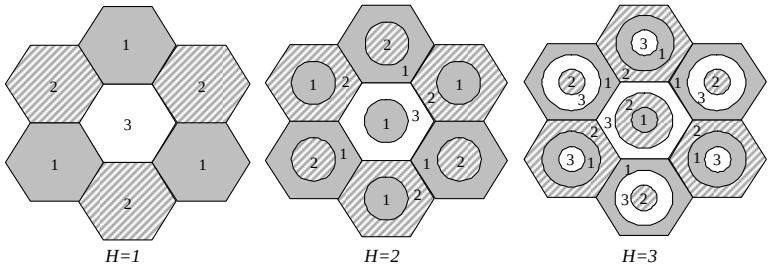

Fig. 36. Intercell flooding scheduling, ([122] [©IEEE 2012]).

4.2.4 Joint optimization of tessellation, scheduling, routing and power control

First, we assign a utility function U which measures the degree of satisfaction based on the achieved flow rates $\mathbf{f} = \left\{ f^k \right\}$, the total power consumption of all nodes $\sum_h P_h$ and finally the delay $B^{k,p}$. This delay depends on the number of hops *H*, which is a function of the radius of the subcells *r*. It also depends on the protocol used *p*. So, the utility can be written as

$$U\left(\mathbf{f},\mathbf{P},r,\wp\right) = \frac{1}{irf}\sum_{k=1}^{M}\frac{\left|\mathcal{D}^k\right| Y \log\left(f^{k,p}\right)}{B^{k,p}\sum_h P_h},\ p \in \wp \tag{58}$$

where *irf* is the intercell reuse factor, *Y* is the number of bits (*Y* = 1 for protocols A, B and C, and *Y* = 2 for protocol D) and, *M* refers to the number of sessions. The optimization problem can be formulated as

$$\text{maximize} \quad U\left(\mathbf{f},\mathbf{P},r,\wp\right) \tag{59}$$

$$\text{subject to} \quad \sum_{k=1}^{M} f_l^k \le c_l,\ \forall l \in \Re^{\mathcal{D}^i} \tag{60}$$

$$c_l = \log(1 + SINR_l(\mathbf{P},r)),\ \forall l \in \Re^{\mathcal{D}^i} \tag{61}$$

$$x^k = f^k / B^{k,p} \tag{62}$$

$$(\mathbf{x},\mathbf{f}) \in \wp\ ;\ (\mathbf{c},\mathbf{P},r) \in \mathcal{C} \tag{63}$$

$$\mathbf{P}_{\min} \le \mathbf{P} \le \mathbf{P}_{\max};\ 0 \le r \le R, \tag{64}$$

where the constraint (60) means that the physical layer can support the network traffic if and only if the aggregated flow on each link is less than the capacity.

Equation (61) defines the capacity of link *l* and (62) is the definition of the multicast throughput in session *k*. Equation (63) models the relation between the achievable multicast throughput **x**, the data flow routing scheme **f**, and the relation between the link capacity **c**, the link power consumption **P** and the tessellation factor *r*. The constraints (64) define the allowed power levels, and *r* is the tessellation factor. We should notice that the topology, scheduling and routing are implicitly included in the protocol type $p \in \wp$ and the user availability is included on the set of routes $\Re^{\mathcal{D}^k}$.

For certain tessellation factor *r*, we obtain the optimum value for the transmission power that maximizes the utility. As we are assuming that the tessellation factor *r* is the same for all cells, the transmission power obtained is the same for all users. We also obtain which is the optimum scheduling (protocol *p*) to achieve the maximum multicast throughput **x** ($x^k = f^k / B^{k,p}$). The BS assigns the transmission turn following scheduling *p* and the transmission power *P* that the users need to achieve that optimum throughput. The implementation will be elaborated in more detail in Section 4.6.

A) Multicast Gain

For fair comparisons of the different systems' implementations we assume a constant number of users $\mathcal{U}$ per cell. The user density is then defined as $\rho_u = \mathcal{U} / S$ where *S* is the cell area. The multicast gain is defined as the number of users reached by one transmission $g_m = \mathcal{U} / \sum_b \left|\boldsymbol{\Psi}^b\right|$, where $\sum_b \left|\boldsymbol{\Psi}^b\right|$ is the number of transmissions per a complete flood of the cell and $\left|\boldsymbol{\Psi}^b\right|$ designates the number of ones in matrix $\boldsymbol{\Psi}^b$ *i.e.* $\left|\boldsymbol{\Psi}^b\right| = \sum_{h,\theta} \Psi^b(h,\theta)$.

In accordance with the definition of the multicast gain we can define the multicast gain per session as $g_m^k = \left|\mathcal{D}^k\right| / \sum_b \left|\boldsymbol{\Psi}^b\right|$ where $\left|\mathcal{D}^k\right|$ is the number of destinations in session *k*. In this case for $P_h = P$ (58) becomes

$$U(\mathbf{f}, \mathbf{P}, r, \wp) = \frac{1}{irf} \sum_{k=1}^{M} \frac{g_m^k Y \log\left(f^{k,p}\right)}{B^{k,p} P}, \quad p \in \wp. \tag{65}$$

B) Broadcast Directivity

If all receivers are located within the cluster in angular domain φ the number of activated transmitters will be proportional to $\varphi / 2\pi = 1 / \Omega_d$ where Ω_d is the broadcast directivity and (65) becomes

$$U(\mathbf{f},\mathbf{P},r,\wp) = \frac{1}{irf}\sum_{k=1}^{M}\frac{2\pi g_m^k Y \log\left(f^{k,p}\right)}{\varphi B^{k,p} P} = \frac{1}{irf}\sum_{k=1}^{M}\frac{\Omega_d g_m^k Y \log\left(f^{k,p}\right)}{B^{k,p} P}, \quad p \in \wp. \tag{66}$$

As it can be seen from (66), larger directivity in the spatial distribution of the users increases (improves) the utility function since $\Omega_d > 1$. This results from the fact that all users can be reached only by activating the transmitters in the direction of the users. This is equivalent to some extend to the gain that can be achieved in the classical broadcast system if instead of omnidirectional antenna a beam forming is used with the spatially limited lobe only in the direction of the users. The segment of the cell not covered by this broadcast can be reused for other sessions.

4.3 Context aware route discovery protocol

4.3.1 System model and assumptions

The route discovery protocol presented in Section 4.2.3 assumed that all required transmitters on the route to the destination users are available. In multihop cellular networks (MCNs), users may be temporally unavailable to relay the messages due to mobility, battery power, coverage, interference and density of nodes. The nodes might also belong to different operators which are not willing to cooperate and relay each other's traffic. Establishing routes in such network efficiently is an important and challenging research issue.

To systematically address this problem, we apply a conventional resource reuse scheme used for cellular networks to our nano scale network model as shown in Fig. 37 and Fig. 38 for resource reuse factor $\Gamma = 4$ and 7, respectively. The clustering factor Γ, equivalent to the frequency reuse factor in a cellular network, partitions the network into clusters of Γ different types of users and enables $\Gamma - 1$ possible options for rerouting/rescheduling. The type of user m^Γ is determined by its position within the cluster (in Fig. 37 $m^\Gamma = 1,2,...,\Gamma$).

The location of a user m^Γ is approximated (polar coordinates) as

$$m^\Gamma(h,\theta) \approx h \cdot d_r \cdot e^{j\theta_h^\Gamma}, \tag{67}$$

where the user location depends also on Γ.

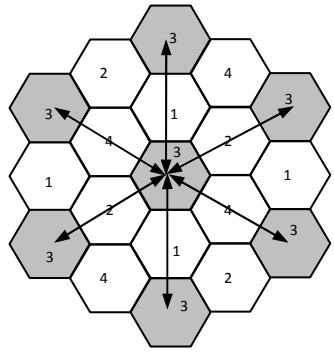

Fig. 37. N-cell reuse pattern for $\Gamma = 4$.

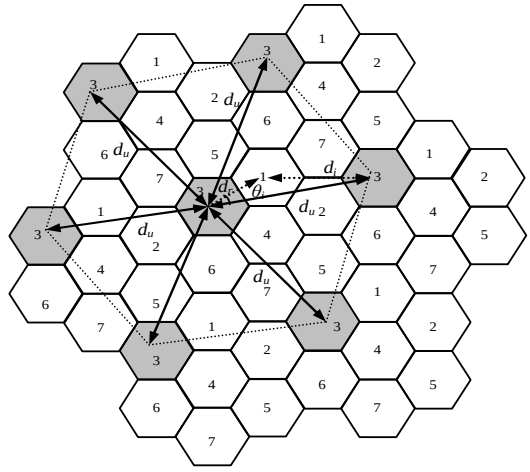

Fig. 38. N-cell reuse pattern for $\Gamma = 7$.

For the motivating example presented in Fig. 39, let us assume uplink transmission and a potential, ready to cooperate, transmitter/receiver is in average situated in the centre of those subcells. The adjacent users (adjacent subcells), which are on the way to the destination (BS), relay the transmission towards the receiver.

The transmissions are scheduled in time slots of index *b* and the scheduling cycle B_i represents the number of slots needed for a packet sent by user *i* (route *i*) to reach the base station BS. It is assumed that the location of the users is fixed during the scheduling cycle. As before, the actual availability of such user in a subcell is characterized by entry 1 or 0 respectively in the availability matrix $\mathbf{A} = \|A(h,\theta)\|$.

Based on the previous definitions, we denote the location of any user m^{Γ} adjacent to the base station (first cluster of users) for the scenario presented in Fig. 39, as

$$m^{\Gamma}(1,\theta_1^{\Gamma}) = m^{\Gamma}(0,0) + d_r \cdot e^{j(\pi/6+\pi/3(m^{\Gamma}-2))},\ m^{\Gamma} = 2,\ \ldots, 7 \qquad (68)$$

$m^{\Gamma}(0,0)$ is the location of BS that corresponds in this example to user type $m^{\Gamma} = 1$.

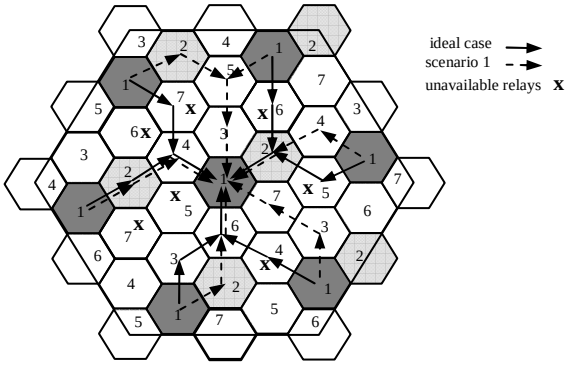

Fig. 39. N-cell reuse pattern for $\Gamma = 7$ and transmission schedule for $m^{\Gamma} = 1$.

We define the minimum distance between users using the same resource (subchannel) as $d_u = d(m^{\Gamma}(h,\theta), m^{\Gamma}(\eta,\varphi))$ where $m^{\Gamma}(h,\theta)$ is the reference transmitter and $m^{\Gamma}(\eta,\varphi)$ a cochanel interfering user. The distance between adjacent relay users is denoted as $d_r = d(m^{\Gamma}(h,\theta), m^{\Gamma'}(h',\theta'))$. The resource reuse factor Γ can be used to express the relation between d_u and d_r as follows [175]:

$$d_u = \sqrt{\Gamma} \cdot d_r = \sqrt{3\Gamma} \cdot r. \qquad (69)$$

The interference distance between the reference receiver $m^{\Gamma'}(h',\theta')$ and the cochanel interfering user $m^{\Gamma}(\eta,\varphi)$ is denoted by $d_i = d(m^{\Gamma'}(h',\theta'), m^{\Gamma}(\eta,\varphi))$. If we apply the second cosine law to the calculation of d_i we obtain

$$d_i = \sqrt{d_r^2 + d_u^2 - 2d_r \cdot d_u \cdot \cos\theta_i}, \tag{70}$$

where $\theta_i = \measuredangle(d_r, d_u)$ and θ_1 is calculated from the geometry as presented in Fig. 38 for $\Gamma = 7$. For the other 5 interferers, θ_i is given by

$$\theta_i = \theta_{i-1} + 60^{\circ}. \tag{71}$$

We disregard the other interferences except the interference from 6 nearest interfering nodes (1-tier). So, the interference power at the position of the relay user $m^{\Gamma'}(h',\theta')$ due to the interfering cochannel signal transmitted by user $m^{\Gamma}(\eta,\varphi)$, can be presented as

$$I(m^{\Gamma}(\eta,\varphi), m^{\Gamma'}(h',\theta')) = I_i = \sum_{i=1}^{6} \frac{1}{d_i^{\alpha}} P_i, \tag{72}$$

where we consider the effects of the propagation looses, but not the effects of fading due to the proximity (*CD network*) between the users. The signal to interference noise ratio at any relaying user by using the conventional relaying scheme presented in the previous section is given by

$$\begin{aligned} SINR_{m^{\Gamma'}}^{con} &= \frac{P / d_r^{\alpha}}{N_r + I_i} = \\ &= \left(\frac{N_r d_r^{\alpha}}{P} + \sum_{i=1}^{6} \left(\sqrt{\frac{d_r^2}{d_r^2 + d_u^2 - 2d_r d_u \cos\theta_i}} \right)^{\alpha} \right)^{-1} \end{aligned} \tag{73}$$

By substituting (69) into (73), we have

$$SINR_{m^{\Gamma'}}^{con} = \left(\frac{N_r d_r^{\alpha}}{P} + \sum_{i=1}^{6} \left(\sqrt{\frac{1}{1 + \Gamma - 2\sqrt{\Gamma}\cos\theta_i}} \right)^{\alpha} \right)^{-1}, \tag{74}$$

where N_r is the background noise power, and $d_r = \sqrt{3} \cdot r$.

4.3.2 Nano route discovery protocol

In order to illustrate the behaviour of the Nano Route Discovery Protocol (NRDP), we use the example presented in Fig. 39. It is assumed that in certain time instant, users type $m^{\Gamma} = 1$ transmit to the BS by relaying to their adjacent users on the way to the BS. In the *ideal case* where all users are available for relaying, the transmission follows the route indicated as →. The transmission is scheduled in such a way that in the first slot, all users type $m^{\Gamma} = 1$ transmit simultaneously, and in the next slots the relaying transmissions should be scheduled in such a way that any two transmitters can transmit simultaneously if $d_i \geq d_r$. In this case d_i refers to the interference distance between the reference receiver $m^{\Gamma'}(h',\theta')$ and any cochannel interfering user $m^{\kappa}(\eta,\varphi)$. In a real scenario, where the users marked as **x** are unavailable, an efficient route discovery protocol is needed to guarantee a good network performance.

The route discovery protocol proposed in this section is based on the fact that the highest interference distance is obtained when the relaying is made between users of the same type Γ'. So, when the adjacent users that are on the minumum distance d_{min} from the BS are unavailable to relay, the transmission should be rerouted/rescheduled in such a way that all transmitters relay to the same type of adjacent relay available. In the case that the available relay at the minimum available distance is located at hop $h' > h$, the user will not reschedule its transmission to avoid the loop in the route.

This tessellation enables $\Gamma - 1$ possible options for rerouting/rescheduling, and it can be easily extended to any tessellation Γ. After the relays are found, the schedule is performed in the same fashion as in the ideal case where the only constraint is $d_i \geq d_r$ as explained above. We will show in Section 4.5 that although the routes found by NRDP may not be the shortest ones, the gain obtained in the increased capacity due to the controlled interference compensates the impact of the increased delay on the throughput.

In order to define the protocol, we denote by $\mathcal{T}^b = \{m^{\Gamma}(h,\theta)\}$ the set of candidate users to transmit in slot *b*, and $\Re^b_{m^{\Gamma}(h,\theta)}$ denotes the set of routes to adjacent users that receive the transmission from user $m^{\Gamma}(h,\theta)$ in slot *b*. Then, the route discovery protocol can be summarized as follows:

1. $b = 1$
2. ***while*** $h > 0$
3. *all users* $m^{\Gamma} \in \mathcal{T}^b$ *initiate the transmission to adjacent users* $\Re^b_{m^{\Gamma}(h,\theta)}$
4. ***if*** *there is any user* $m^{\Gamma}(h,\theta)$ *such that there is no* $m^{\Gamma'}(h',\theta') \in \Re^b_{m^{\Gamma}(h,\theta)}$ *that satisfies* $d(m^{\Gamma'}(h',\theta'), m^1(0,0)) = d_{\min}$ *then*
5. ***while*** $j < 7$
6. $m^{\kappa} = (m^{\Gamma} + 1) \bmod 7$

   find which is the common user type m^{κ} *to all users* $m^{\Gamma} \in \mathcal{T}^b$ *to reschedule*

   $m^{\kappa}(h',\theta') = m^{\Gamma'}(h',\theta') + d_r e^{j(\pi/6+\pi/3(m^{\Gamma'}-2))}$

   $j = j + 1;$
7. ***end***
8. $m^{\Gamma} = m^{\kappa}$
9. ***end***
10. $b = b + 1;$ $h \leftarrow h'$
11. *Define* $\mathcal{T}^b$ *such that* $m^{\Gamma}(h,\theta), m^{\kappa}(\eta,\varphi) \in \mathcal{T}^b$ *if* $d(m^{\Gamma'}(h',\theta'), m^{\kappa}(\eta,\varphi)) \geq d_r$
12. ***end***

Details on the protocol implementation are given in Section 4.6.

4.3.3 Performance analysis of nano route discovery protocol

The calculation of the path cost is a critical component in the route discovery and directly affects the resource consumption in the network. We define the cost function of route *i* to reflect the impact of the traffic load on the power depletion and also the overall power consumption as

$$cost_i = w_1 \cdot load_{i\max} + w_2 \cdot power\_cost_i = w_1 \cdot load_{i\max} + w_2 \cdot P(\sum_{i=1}^{h} load_i). \qquad (75)$$

Here, coefficients w_1 and w_2 are the weighting factors. The first term takes into account the maximum number of transmissions from one node on route *i* given by $load_{i\max}$. In order to reduce the energy depletion of the terminal this should be kept low. The second term gives the power cost on route *i* ($power\_cost_i$) as the

sum of the overall power used for necessary transmissions on route i, where P is the users' transmission power and $load_i$ represents the traffic load at node i. By varying weighting factors w_1 and w_2, we can change the relevancy of these cost metrics during route discovery.

The capacity of the network C is defined as the sum of the minimum link capacities of the different routes

$$C = \sum_i C_i = \sum_{i=1} \min_i c_l, \tag{76}$$

where c_l is the link capacity defined as $c_l = \log(1 + SINR_l(P))$. We evaluate the performance of the algorithm in terms of the network throughput $Thr = C/B$ where B is the delay, and the utility that includes the path cost as

$$U = \sum_i U_i \quad \text{where} \quad U_i = \frac{C_i}{B_i \cdot cost_i}. \tag{77}$$

Parameter U_i is the utility of user i. The aim is to choose the route that favors the least-cost path, where the path cost metric reflects the optimization criteria.

4.4 Joint optimization of cooperative diversity and spatial reuse factor in MCNs

In this section, the model presented in the previous section is extended to address the trade-off between cooperative diversity and spatial reuse in MCNs. We consider uplink transmission in a MCN as shown in Figs. 37 and 38.

To determine the number of concurrent transmissions in the cell N_s, we first define the number of subcells per cell N as a ratio of the area of the cell and the area of the subcell, so

$$N = (R/r)^2. \tag{78}$$

This definition of N is equivalent to the one given by (37). The number of users transmitting packets simultaneously N_s, is determined by the inner cell cluster size Γ, defined as

$$\Gamma = N / N_s. \tag{79}$$

We assume cooperative relaying scheme as explained in [Chapter 2, Section 2.4]. This scheme exploits spatial reuse and cooperative diversity. The diversity order is denoted by ξ which is defined as the number of previous transmissions heard by

a certain intermediate user and it depends on spatial reuse factor. The transmissions produced by other users situated at distances larger than the 1-tier are discarded as shown in Figs. 37 and 38.

By applying the definitions presented in the previous section (69)-(72), the equivalent SINR with the cooperative relaying scheme $SINR^{co}$ at a diversity receiver is given by [176]

$$\begin{aligned} SINR_{BS}^{co} &= SINR_{m_\xi,BS}^{co} + \ldots + SINR_{m_1,BS}^{co} = \sum_{h'=\xi}^{1} \frac{P/(h'd_r)^\alpha}{N+I_i} \\ &= \sum_{h'=\xi}^{1} \left(\frac{N(h'd_r)^\alpha}{P} + \sum_{i=1}^{6} \left(\sqrt{\frac{(h')^2}{1+\Gamma-2\sqrt{\Gamma}\cos\theta_i}} \right)^\alpha \right)^{-1}, \end{aligned} \tag{80}$$

where ξ is the diversity order that depends on the routing and in general is higher for larger Γ.

As the BS is surrounded by 6 subcells (we are considering hexagonal cells), if we split the cell into 6 regions as shown in Fig. 37 and 38, the number of subcells in each of those triangular regions is

$$N_t = \frac{1}{6}N = \frac{1}{6}(R/r)^2. \tag{81}$$

The arrival rate of user *m* is denoted by λ_m, and the rate received at any user situated in the adjacent subcells to the *BS* after relaying the information is $N_t \cdot \lambda_m$. As the 6 subcells, around the base station, are capacity bottleneck, then the average cell capacity is determined by the relaying capability of 6 subcells as

$$N_t \cdot \lambda_m = C/6, \tag{82}$$

where the Shannon capacity is

$$C^{con} = \log\left(1 + SINR^{con}(P)\right) \tag{83}$$

$$C^{co} = \log\left(1 + SINR^{co}(P)\right) \tag{84}$$

for conventional and cooperative relaying respectively.

The throughput per user *region* is defined as

$$Thr_m = N_t \cdot \lambda_m. \tag{85}$$

By taking into account the previous analysis, we define the utility as

$$U = \frac{Thr_m}{\sum P} = \frac{C}{N_t \cdot P \cdot 6}. \tag{86}$$

By using this utility function we can simultaneously optimize the system throughput, power consumption and packet delivery delay, as a function of relaying range d_r which is a function of the radius r of the subcell.

In order to define: a) the optimum relaying range d_r, which implicitly determines the optimum maximum number of hops $h_{max} = H$ and, b) the spatial reuse (which determines the maximum diversity order for cooperative transmission) in the network, in the sequel we present the joint optimization of cooperative diversity and spatial reuse, as

$$\begin{aligned} \underset{P,r}{\text{maximize}} \quad & U = \frac{\log\left(1 + SINR_r(P)\right)}{N_t \cdot P \cdot 6} \\ \text{subject to} \quad & P_{\min} \le P \le P_{\max}, \\ & 1 \le r \le R \end{aligned} \tag{87}$$

where $P_{\min}$ and $P_{\max}$ are the minimum and maximum power respectively, and R is the radius of the cell.

The optimum subcell radius r defines the optimum subcell partitioning and the spatial reuse in the network. The number of hops ξ that can be used to improve cooperative transmission (diversity order) depends on routing through the inner cells cluster and in general is higher for higher reuse factor Γ. In fact, the maximum and minimum number of hops $\xi_{\min}$ and $\xi_{\max}$ respectively, can be calculated as

$$\xi_{\min} = \left\lfloor \sqrt{\Gamma} \right\rfloor \text{ and } \xi_{\max} = \left\lceil \sqrt{\Gamma} \right\rceil. \tag{88}$$

Other forms of tessellation with regular polygons are possible, *i.e.*, square, triangle tessellation. Hexagonal tessellation has been used because the hexagon is the regular polygon that minimizes the covering problem between the cells.

4.5 Performance evaluation

A set of numerical results is presented in this section to validate our models.

4.5.1 Context aware nano scale optimization of multicast MCNs

A) Network tessellation

We consider the scenario presented in Figs. 32-35 for $R = 1000m$ and $\alpha = 2$. In Figs. 40a)-d) the throughput in the network is presented with intercell flooding coordination (ICFC) versus the number of hops H with $\mathbf{D} = \boldsymbol{\beta}$ (broadcast) for protocols A, B, C and D and different transmission powers. We can see that for protocols A and B, with results shown in Figs. 40a) and 40b) respectively, the maximum network throughput depends on both, the transmission power P and number of hops H. The optimum number of hops is $H = 3$ for as long as $P \leq 0.5$. As we further increase the number of hops, the throughput is reduced due to increased delay and increased number of mutually interfering transmitters. Similarly, if the power is further increased ($P > 0.5$), one hop becomes better option since the level of interference due to excessive power in multihop network is higher. On the other hand, for protocols C (cooperative relaying) and D (network coding) with results shown in Fig. 40c) and 40d), respectively the optimum number of hops is $H = 4$ for all power levels. This is because there is no interference coming from adjacent transmissions, and as we increase the number of hops the number of transmitters is increased, so there are more useful signals to combine in the diversity receivers. If the number of hops is further increased the performance start to degrade because the throughput is inversely proportional to the number of hops and improvements due to diversity are lower than the degradations due to increased delay.

Although for cooperative relaying and network coding the system performance is better for $H = 4$, the largest increment of network throughput $\Delta Thr / \Delta H$ is obtained again for $H = 3$, suggesting these values as the best choice when it comes to trade-off between performance and complexity. We can also see that the highest throughput is obtained by protocol D. From Figs. 40a)-d) one should notice that protocols C and D provide more than 3 times higher maximum throughput (for $H = 4$) than protocols A and B (with maximum value in $H = 1$ or $H = 3$) for the same transmission power. Or equivalently, protocols C and D (for

$H = 4$) can provide same throughput than protocols A and B (for $H = 1$ or $H = 3$) with ten times less power.

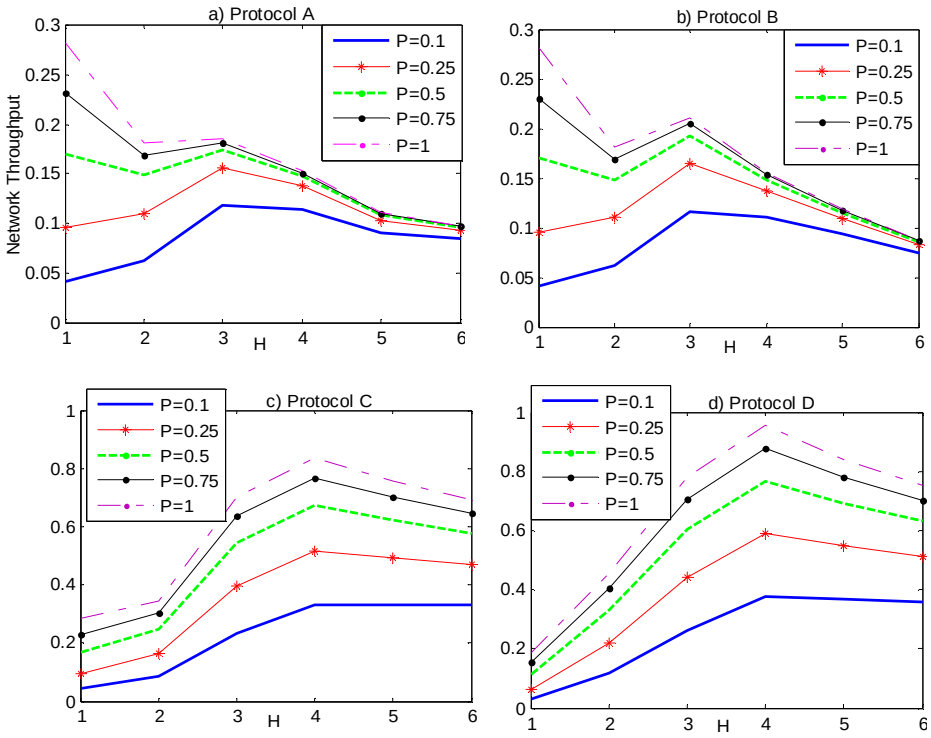

Fig. 40 a)-d) Network throughput versus the number of hops *H* with $\mathbf{D} = \boldsymbol{\beta}$ for protocols A, B, C and D, respectively and different power values *P*, ([122] [©IEEE 2012]).

B) Channel defading

In Fig. 41 we compare optimum tessellation for ideal channel (no fading) and a fading channel defined by (45b) using protocol A. In the case of ideal channel, optimal tessellation is with $H = 3$ while in fading channel optimal tessellation is obtained for fully defaded channel. In this example we have two types of channels, one that is defaded for $H > H_0 = 3$ and another one for $H > H_0 = 6$. For $H_0 = 6$, the potential loss in throughput due to increased number of hops is overcompensated

by the improved link capacities. For more details on network cell design and channel defading see Section 4.6.

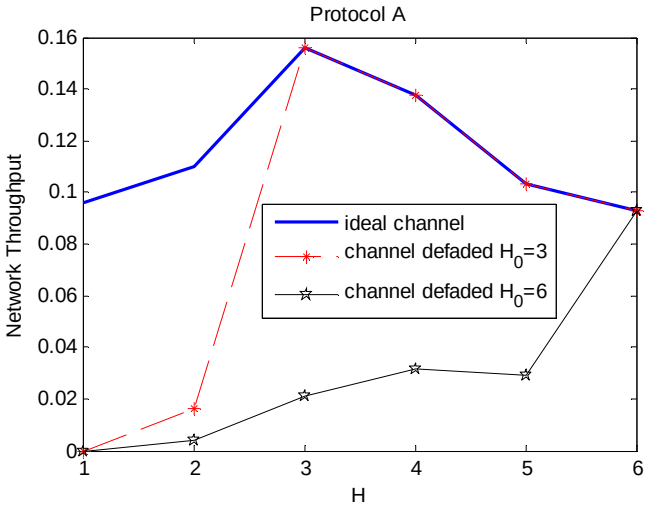

Fig. 41. Comparison of the throughput for both channel schemes for P = 0.25 and protocol A, ([122] [©IEEE 2012]).

C) Multicast Directivity

In Fig. 42 we present 4 multicast scenarios. In each scenario, there are 6 destination users situated in different locations within the cell which results in different broadcast directivity Ω_d. The location of the destination users in each scenario defines the destination matrix **D**. The routes to the destinations, as indicated in Fig. 42, follow protocol A. These routes can be easily obtained for protocols B, C and D from the schemes shown in Figs. 33 to 35. In these scenarios we have assumed that all required transmitters on the route to the destination are available.

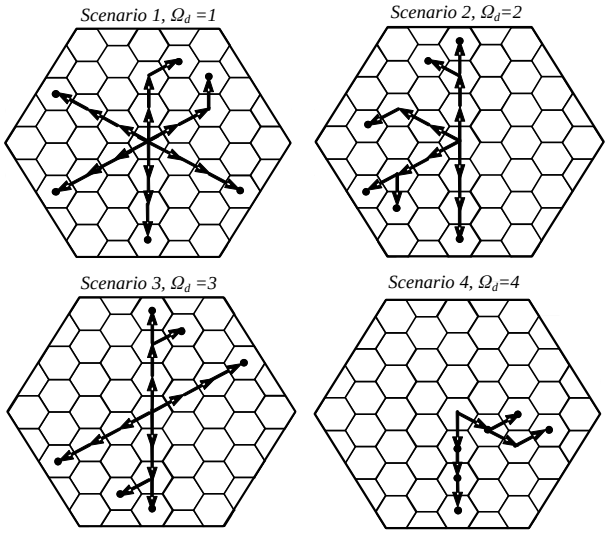

Fig. 42. Multicast scenarios 1 to 4. The routes to the destination follow protocol A, ([122] [©IEEE 2012]).

In Figs. 43-46, we present the network utility obtained by (66) versus the power *P* for scenarios 1-4 as shown in Fig. 42, respectively and protocols A, B, C and D.

In Fig. 43, the network utility is shown for scenario 1 (Fig. 42) where the destination users are located within $\varphi \approx 360^{\circ}$ which results in $\Omega_d = 1$. For this scenario, the highest utility is obtained by protocol C and D, while protocols A and B give lower results. Similar results are obtained in Fig. 44 for scenario 2. In this case, the destinations are located within $\varphi \approx 180^{\circ}$ so that $\Omega_d = 2$ and the utility is higher than in the previous case. In Fig. 45 we show the results corresponding to scenario 3, with $\Omega_d = 3$ (destinations located within $\varphi \approx 120^{\circ}$) which gives higher utility. In this case, protocol B needs more transmitters than protocol A to reach the destinations and also one more slot than protocol A, so that the lowest utility is obtained in this case by protocol B.

Finally, in Fig. 46 the network utility is shown for scenario 4 where the users are located in the cluster $\varphi \approx 90^{\circ}$. The highest utility is obtained again for protocol C. In this case the performance obtained by protocol D is considerably degraded due to the additional number of transmitters needed only for the network coding

function. We should notice that in all these scenarios the maximum network utility is obtained for $P \in [0.15, 0.25]$.

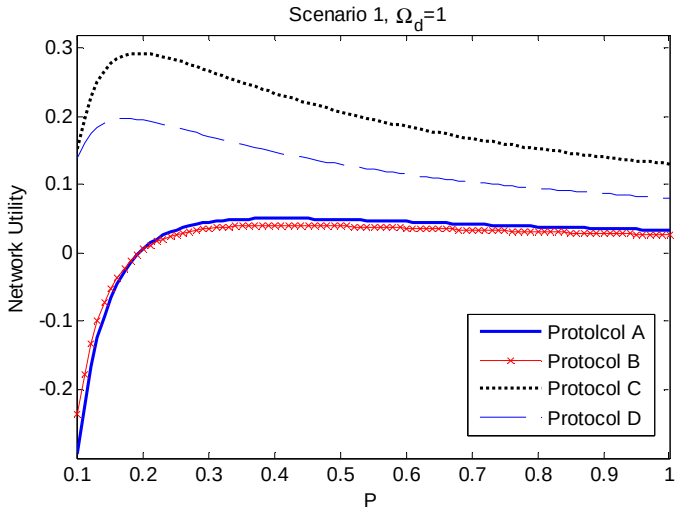

Fig. 43. Network utility obtained by (66) versus *P* for the scenario 1 presented in Fig. 42, and Protocols A,B,C and D, ([122] [©IEEE 2012]).

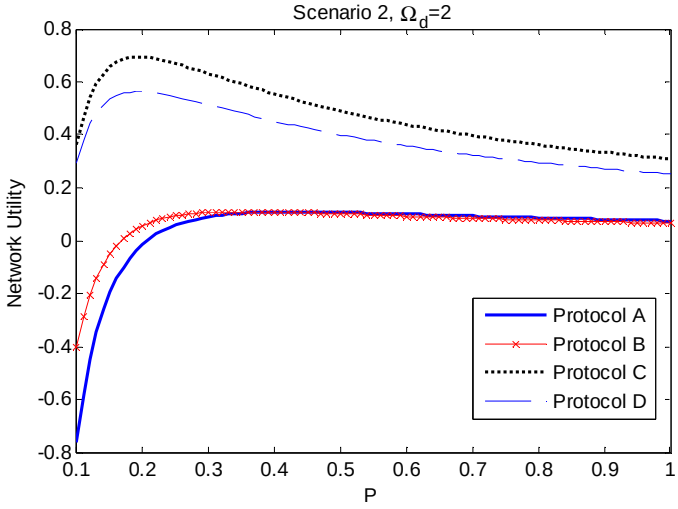

Fig. 44. Network utility obtained by (66) versus *P* for the scenario 2 presented in Fig. 42, and Protocols A, B, C and D, ([122] [©IEEE 2012]).

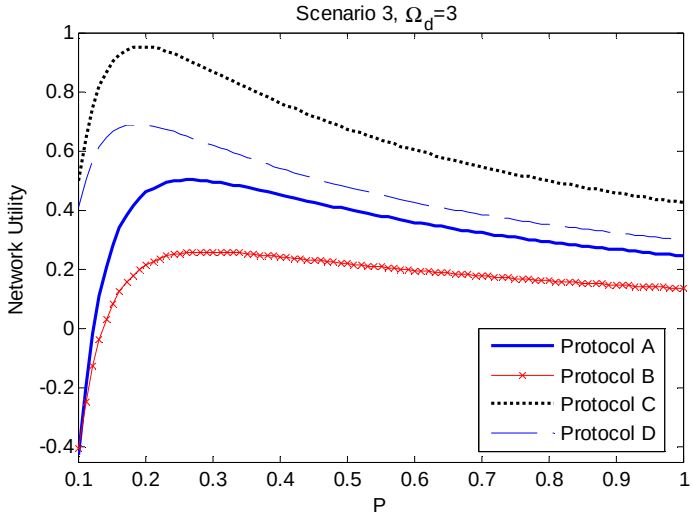

Fig. 45. Network utility obtained by (66) versus *P* for the scenario 3 presented in Fig. 42, and Protocols A, B, C and D, ([122] [©IEEE 2012]).

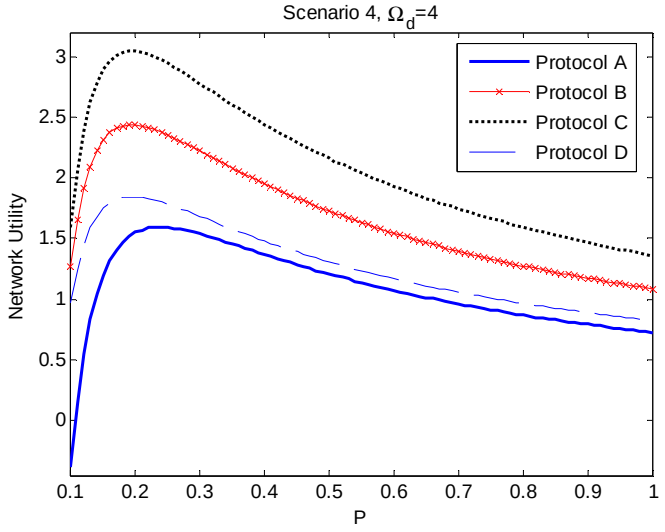

Fig. 46. Network utility obtained by (66) versus *P* for the scenario 4 presented in Fig. 42, and Protocols A, B, C and D, ([122] [©IEEE 2012]).

In Fig. 47 we present again the network utility obtained by (66) for a multisession scenario (scenario 5) consisting of four simultaneous sessions where each session corresponds to one of the four scenarios presented in Fig. 42. The largest increment in the performance is obtained for low values of power and for $P > 0.3$ there is no further significant improvement. The density of destination users in this case is quite high (24 destinations) and we can see that the best result is obtained by protocol D. The largest difference between the protocols is for low power. One should be aware of the log scale for flow rates which means that for e.g. $P = 0.1$, protocol D offers two orders of magnitude larger flow rates than protocol A.

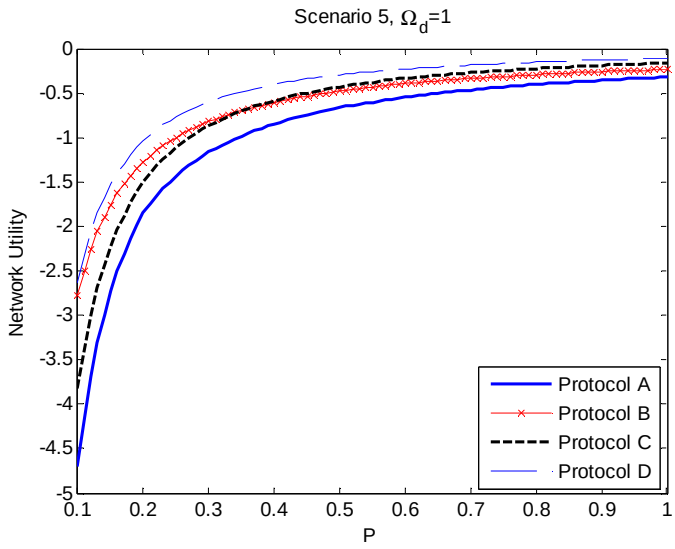

Fig. 47. Network utility obtained by (66) versus *P* for the multisession scenario (scenario 5) consisting of 4 sessions where each session corresponds to one of the four scenarios presented in Fig. 42, ([122] [©IEEE 2012]).

4.5.2 Nano route discovery protocol

Different scenarios, with different sets of *unavailable users*, as presented in Table 4, are simulated to show the performance of the NRDP. The tessellation presented in Fig. 39 is used with $R = 100m$, $r = 15m$ and propagation loss coefficient $\alpha = 3$. We assume that users type $m^{\Gamma} = 1$ are transmitting and the rest of the available

terminals act as relays. This can be easily modified for any other user type m^{Γ} or their combination. In each of these scenarios the rescheduling has been done through different users type m^{Γ} = 2, 3, …, 7 to cover all possible options.

Scenario 1 corresponds to Fig. 39 where the unavailable users are marked with **x**. We compare these results for NRDP with another two discovery protocols referred to as Shortest Available Path Routing (SAPR) and Load Aware Routing (LAR). In SAPR, the protocol finds the shortest available route for rescheduling irrespective of the interference level that this choice produces in the network. In LAR protocol, traffic load and power depletion are taken into account in the route discovery, so the protocol finds the route in such a way that the traffic is uniformly distributed through the whole network.

Table 4.

scenario	unavailable users	rescheduling
1 (Fig. 39)	$x=\begin{cases} m^5(2,0^\circ), m^6(2,30^\circ), m^7(2,120^\circ), m^6(2,150^\circ) \\ m^5(1,210^\circ), m^7(2,120^\circ), m^4(2,300^\circ) \end{cases}$	$m^{\Gamma} \to 2$
2	$o=\begin{cases} m^5(2,0^\circ), m^2(1,30^\circ), m^6(2,60^\circ), m^2(3,110^\circ) \\ m^7(2,120^\circ), m^7(2,210^\circ), m^2(2,270^\circ) \end{cases}$	$m^{\Gamma} \to 3$
3	$p=\begin{cases} m^5(2,0^\circ), m^6(3,60^\circ), m^5(2,90^\circ), m^2(3,110^\circ) \\ m^2(2,0^\circ), m^3(2,240^\circ), m^6(1,270^\circ), m^3(2,330^\circ) \end{cases}$	$m^{\Gamma} \to 7$
4	All users type m^{Γ} = 2 and 3	$m^{\Gamma} \to 5$
5	$n=\begin{cases} m^4(2,30^\circ), m^7(2,120^\circ), m^2(3,110^\circ), m^2(2,180^\circ) \\ m^7(2,210^\circ), m^2(2,270^\circ), m^7(1,330^\circ), m^3(2,330^\circ) \end{cases}$	$m^{\Gamma} \to 6$
6	$z=\begin{cases} m^5(2,0^\circ), m^6(2,60^\circ), m^7(3,50^\circ), m^7(2,120^\circ), \\ m^2(3,110^\circ), m^7(2,210^\circ), m^7(3,270^\circ), \\ m^2(2,270^\circ), m^3(2,330^\circ), m^7(1,330^\circ) \end{cases}$	$m^{\Gamma} \to 4$

In Fig. 48 and 49 the capacity and throughput are presented respectively for the ideal case where all users are available for relay. Those results are compared with more realistic scenarios where different set of users are not available for relaying as shown in Table 4. The performance obtained by NRDP is compared to SAPR and LAR protocols. In nonideal case, the highest capacity and throughput are obtained by NRDP. By SAPR, the users experience the shortest delay per

route but on the other hand there is no control of the traffic distribution through the network, so there is more interference between adjacent links and consequently the capacity is lower. The capacity obtained by LAR is larger than with SAPR and although more slots are needed to complete the transmission with LAR, the gain obtained in distributing the traffic in some scenarios compensates the delay as we can see in Fig. 49.

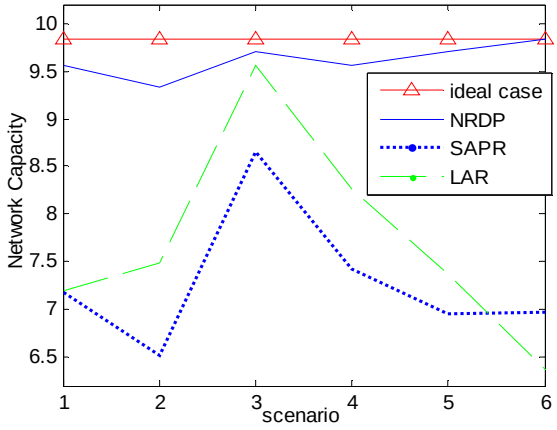

Fig. 48. Capacity versus the scenario as described in Table 4.

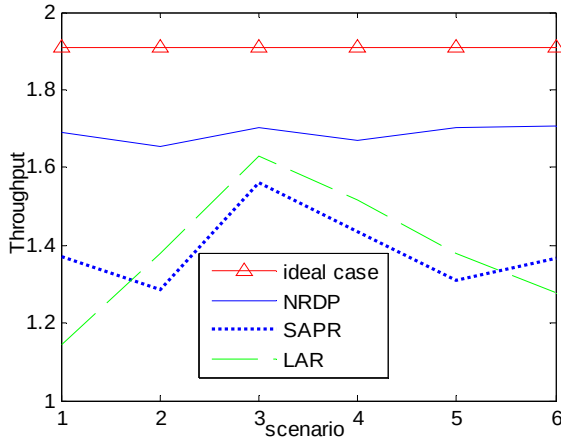

Fig. 49.Throughput versus the scenario as described in Table 4.

In Fig. 50 we present the utility with $w_1 = 1$ and $w_2 = 0$ to measure the performance of our algorithm in terms of the traffic load. We can see that with NRDP we obtained better results in most of the scenarios. LAR outperforms SAPR for as long as there are users available to split the traffic in an efficient way. In scenario 6, the routes found were not able to distribute the traffic uniformly due to the lack of users available in certain areas of the cell.

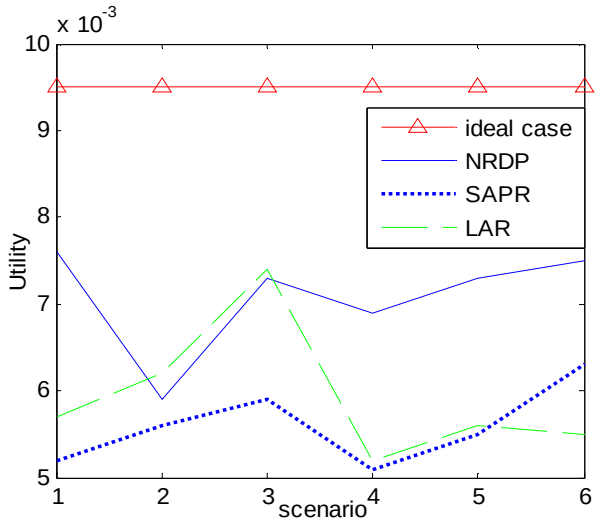

Fig. 50. Utility with $w_1 = 1$, $w_2 = 0$ versus the scenario as described in Table 4.

In Fig. 51 the utility is presented for the case $w_1 = 0$ and $w_2 = 1$, to measure the overall power cost of the proposed protocol. Again we can see that the best results are obtained by NRDP. In most of the cases SAPR performs better than LAR because the routes found are shorter, so less users were needed for relaying and thus, there is less overall power consumption.

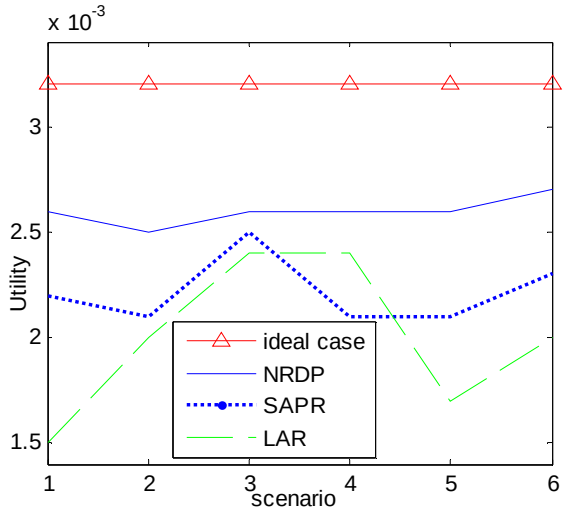

Fig. 51. Utility with $w_1 = 0$, $w_2 = 1$ versus the scenario as described in Table 4.

4.5.3 Joint optimization of cooperative diversity and spatial reuse factor in MCNs

In this subsection, simulation results are presented to show the performance of this optimization problem for $R = 100m$, $\alpha = 3$ and different values of the cell reuse factor Γ. In Fig. 52 we show the utility defined as in (86) versus the radius of the subcell r, for different transmission powers and $\Gamma = 4$. The reuse scheme corresponding to factor $\Gamma = 4$ is represented in Fig. 37. One can see that there is a clear optimum value for the format of cell partitioning in terms of parameter r. As expected, if for the given power the radius of the subcell is too low, there will be a high level of interference and utility will be low due to low link capacity. On the other hand if r is too high, the propagation losses for the useful signal will be excessive and although the interference level is lower, the received signal level is lower and the capacity will be reduced resulting in lower utility.

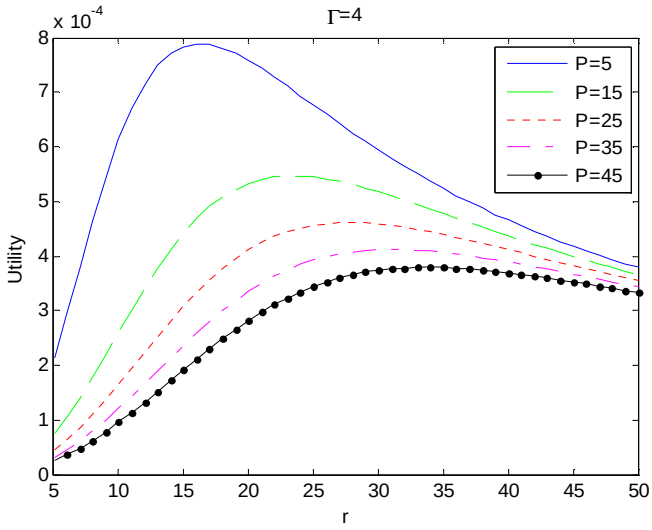

Fig. 52. Utility versus *r* for different power values, $\Gamma = 4$ and conventional relaying scheme, ([124] [©IEEE 2010]).

In general, for lower power the utility peak is higher since the utility function is inversely proportional to power. For a given cell radius *R*, parameter *r* is related to the maximum number of hops *H* in the cell. Figs. 53a and 53b present the results for utility function versus *H* for the case of conventional and cooperative relaying respectively. One can see that in the case of cooperative relaying the utility is higher which means that with less power higher capacity can be achieved. This is a direct consequence of signal accumulation in cooperative diversity. Fig. 54a and 54b show that the capacity is higher for cooperative relaying and is increased as the number of hops is increased.

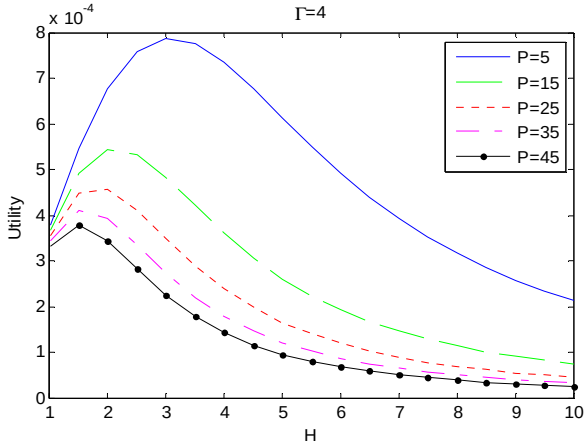

Fig. 53a. Utility versus *H* for different power values and $\Gamma = 4$ with conventional relaying, ([124] [©IEEE 2010]).

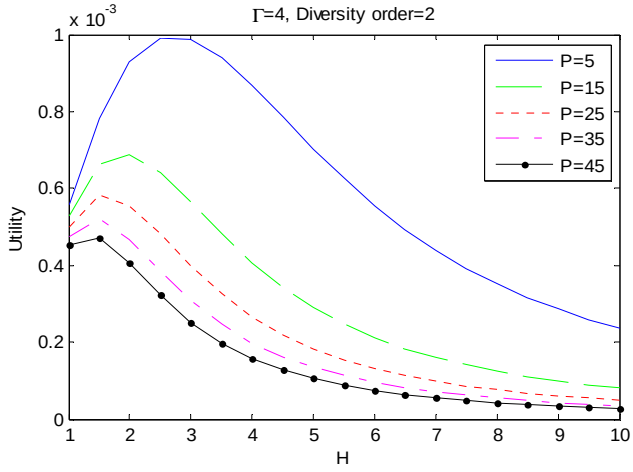

Fig. 53b. Utility versus *H* for different power values, $\Gamma = 4$ and with cooperative relaying, ([124] [©IEEE 2010]).

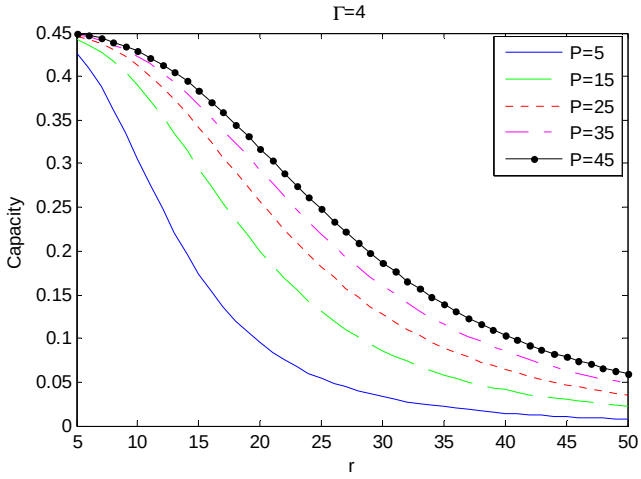

Fig. 54a. Capacity versus *r* for different power values, $\Gamma = 4$ and conventional relaying scheme, ([124] [©IEEE 2010]).

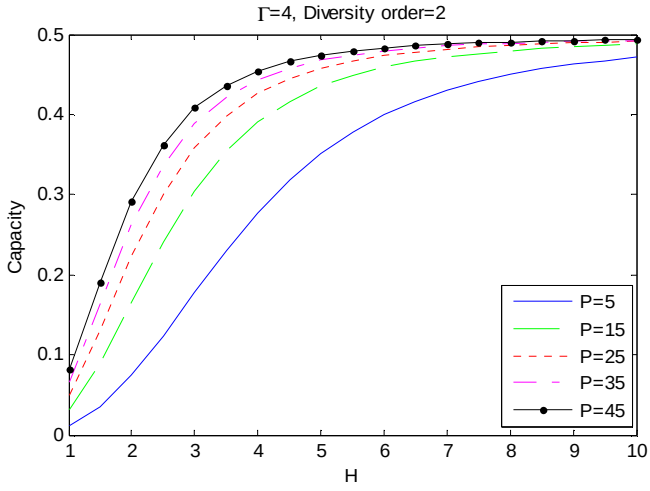

Fig. 54b. Capacity versus *H* for different power values, $\Gamma = 4$ and cooperative relaying scheme, ([124] [©IEEE 2010]).

In Figs. 55, 56a, 56b, 57a and 57b, we present the same set of results for $\Gamma = 7$. The scenario corresponding to this reuse factor is presented in Fig. 38.

In general the shape of the curves remains the same but the values of the utility and capacity are now higher than in the case when $\Gamma = 4$. One should notice that in this case the maximum diversity order is $\xi = 3$ which contributes to better results.

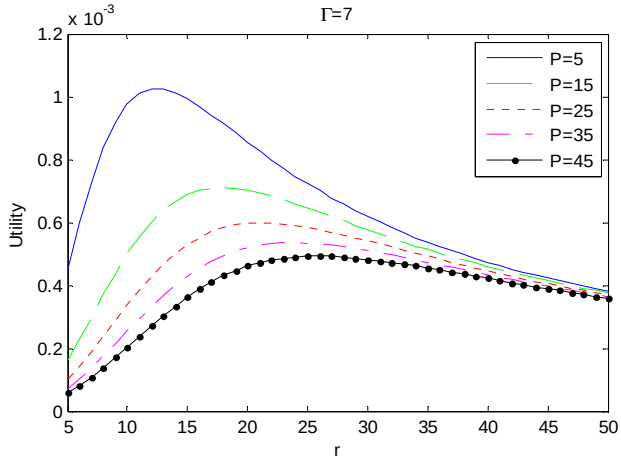

Fig. 55. Utility versus *r* for different power values, $\Gamma = 7$ and conventional relaying scheme, ([124] [©IEEE 2010]).

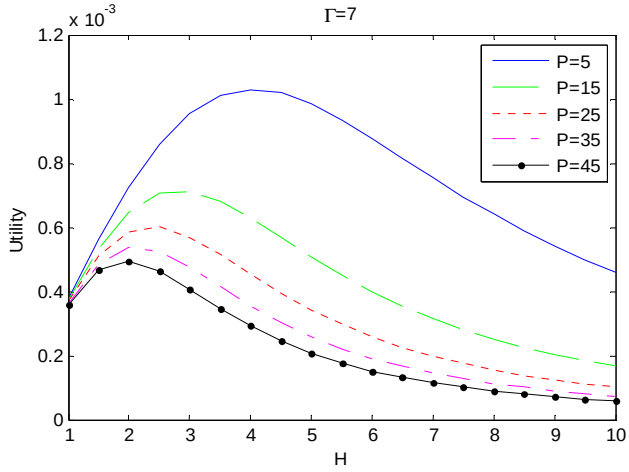

Fig. 56a. Utility versus *H* for different power values and $\Gamma = 7$ with conventional relaying, ([124] [©IEEE 2010]).

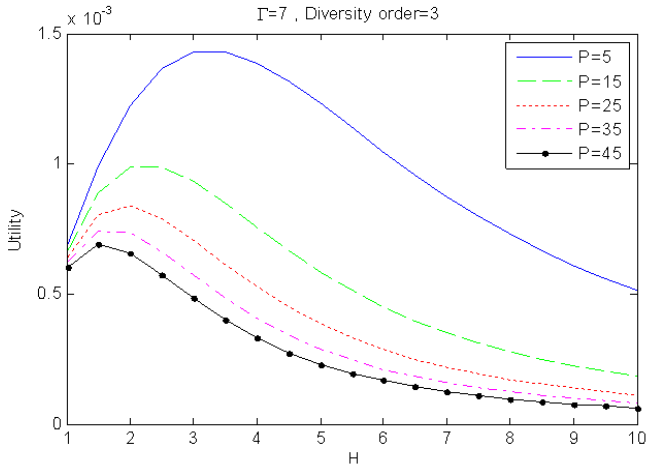

Fig. 56b. Utility versus *H* for different power values and $\Gamma = 7$ with cooperative relaying, ([124] [©IEEE 2010]).

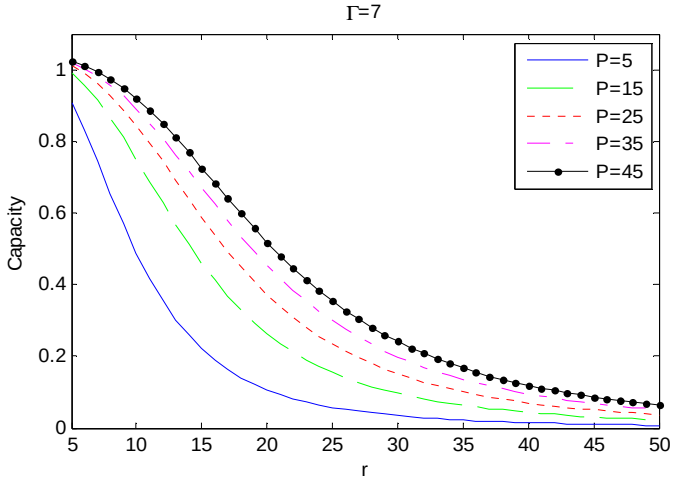

Fig. 57a. Capacity versus *r* for different power values, $\Gamma = 7$ and conventional relaying scheme, ([124] [©IEEE 2010]).

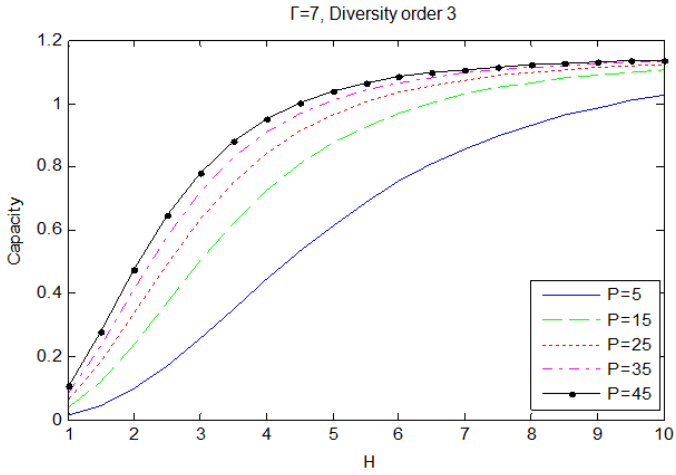

Fig. 57b. Capacity versus *H* for different power values, $\Gamma = 7$ and cooperative relaying scheme, ([124] [©IEEE 2010]).

Figs. 58a and 58b show the utility versus *H* for conventional and cooperative relaying, respectively for Γ = 12. In Fig. 59 the capacity is shown for the same Γ. Figs. 60a, 60b and Fig. 61 present the utility and capacity for Γ = 19. The schemes of the N-cell reuse patterns for these reuse factors can be found in [175].

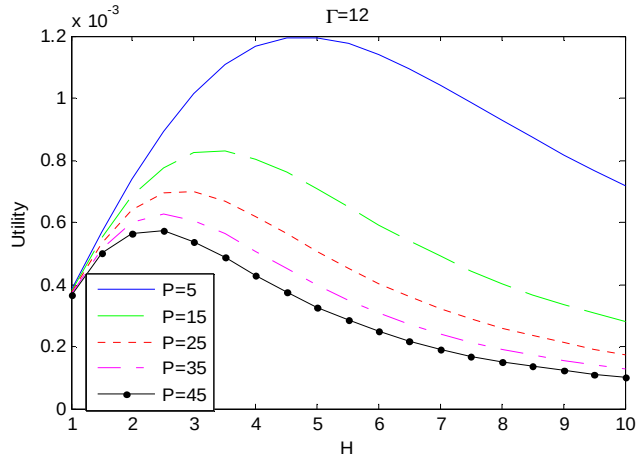

Fig. 58a. Utility versus *H* for different power values and $\Gamma = 12$ with conventional relaying, ([124] [©IEEE 2010]).

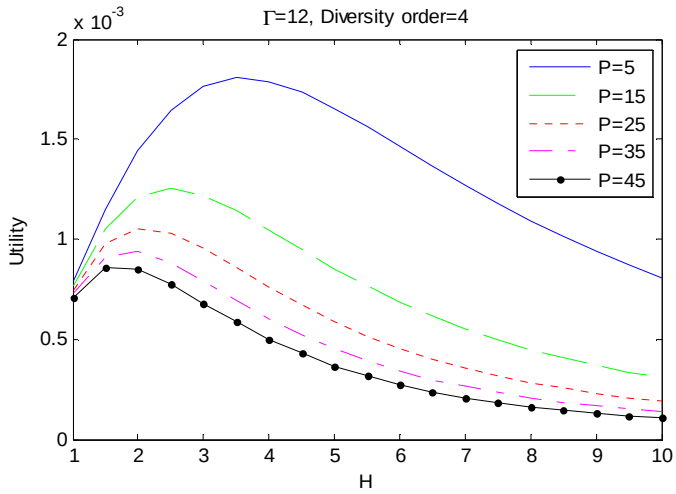

Fig. 58b. Utility versus *H* for different power values and $\Gamma = 12$ with cooperative relaying, ([124] [©IEEE 2010]).

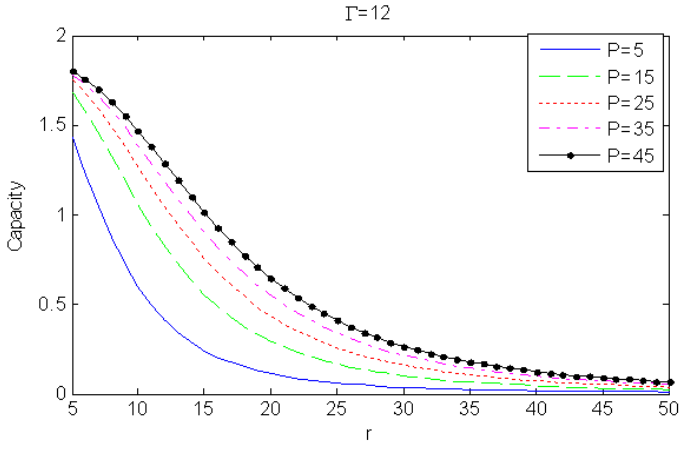

Fig. 59. Capacity versus *r* for different power values, $\Gamma = 12$ and conventional relaying scheme, ([124] [©IEEE 2010]).

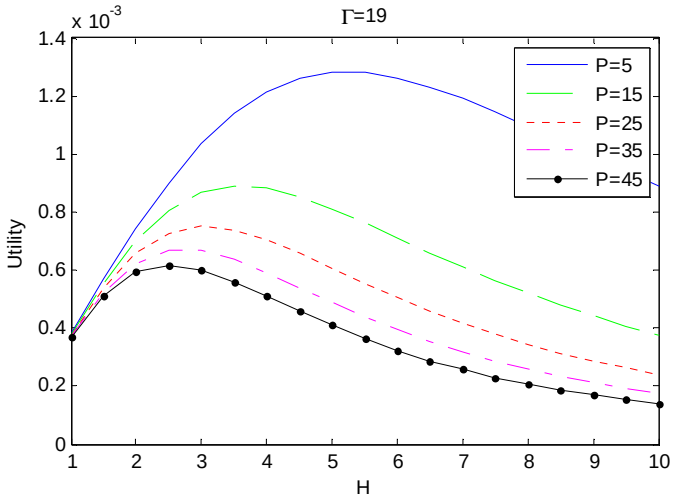

Fig. 60a. Utility versus *H* for different power values, $\Gamma = 19$ and conventional relaying, ([124] [©IEEE 2010]).

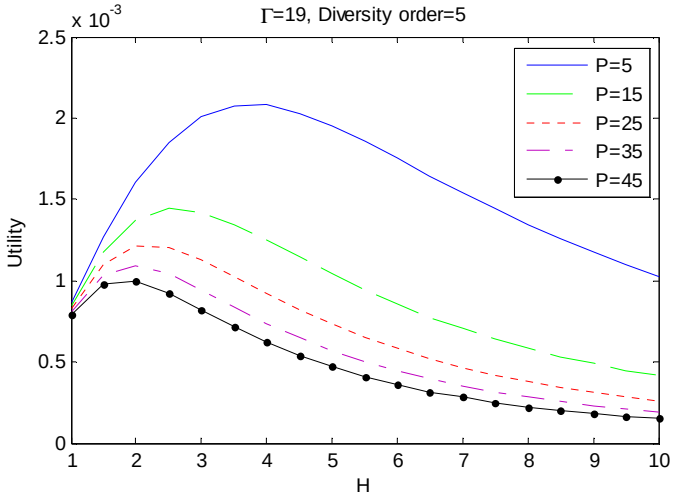

Fig. 60b. Utility versus *H* for different power values and $\Gamma = 19$ by cooperative relaying, ([124] [©IEEE 2010]).

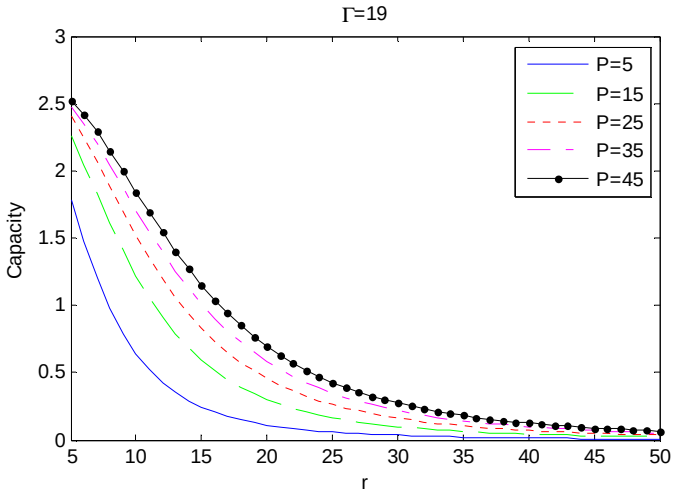

Fig. 61. Capacity versus r for different power values, $\Gamma = 19$ and conventional relaying scheme, ([124] [©IEEE 2010]).

In general, one can see from Fig. 52 to Fig. 60, that utility function for cooperative relaying is higher than for conventional relaying. Also, it should be noted that the peak of the utility for conventional relaying is uniformly increased in the range 8×10^{-4}, 10^{-3}, 1.2×10^{-3}, 1.3×10^{-3} as Γ is increased in the range 4, 7, 12, 19 respectively. For the same range of Γ, the peak of the utility function for cooperative relaying is uniformly increased in the range 10^{-3}, 1.4×10^{-3}, 1.8×10^{-3}, 2.1×10^{-3} which is approximately 50% higher than in the case of conventional relaying. The optimum number of hops is higher for higher Γ.

Finally, Fig. 62 shows the number of concurrent transmissions (level of spatial reuse in the network) for Γ = 4, 7, 12 and 19 versus *r*. We can see that as we increase the radius of the inner cells *r*, the number of inner partitions *N* and, consequently the number of concurrent transmissions N_s are decreased.

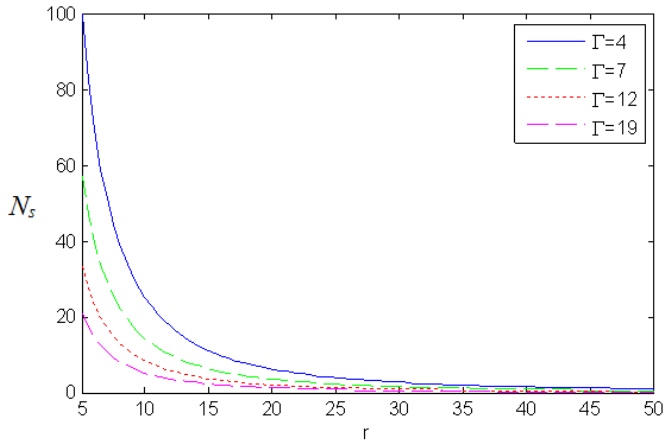

Fig. 62. Number of concurrent N_s transmissions versus r for $\Gamma = 4,7,12$ and 19, ([124] [©IEEE 2010]).

4.6 Implementation

4.6.1 Context aware nano scale optimization of multicast MCNs

The objective of this optimization is to present optimum nano scale network model of MCNs and suggest corresponding transmission protocols for multicast channel. The suggested solutions still leave a variety of practical implementation options for competing for proprietary rights for potential industry manufacturers. Even so, as illustrations in this subsection we indicate some possible implementation directions and their limitations.

A) Control data collection/distribution

The user position and willingness to cooperate is communicated on the conventional uplink signaling (control) channel. Based on this information base station determines the slot index (Ψ^1, Ψ^2, Ψ^3,…) for each user. The index of the current valid protocol (A, B, C or D) and slot index for transmission are communicated to the user on the downlink conventional signaling (control) channel. The potential transmitter/receiver in the nano cell is chosen to be the most static and the most centric (closest to the center of the nano cell) user. Nano

scale network with a specific uplink/downlink signaling channel, implemented separately from the data channel in the form of single hop transmission, is referred to as *InSyNet (Inter System Networking).* The name indicates the fact that the overall nano network is implemented by integrating two different systems: single hop network for control channel and multihop network for data channel as already explained in [Section 3.6, Chapter 3].

B) Updating rate/mobility

The control data updating rate is equal to the broadcast cycle $B \cdot T_r$ where B is the maximum number of slots needed to reach the cell border and T_r is the relaying delay between two nodes.

The length of the message, transmitted in one hop, should be designed so that the updating rate is high enough to compensate mobility. In other words the change in the position due to mobility, characterized by the terminal velocity *v,* should be less than the nano cell radius *r, i.e.* $v \cdot B \cdot T_r < r$.

C) Signaling overhead

In SI MAC on the downlink, 3 bits are needed to transmit its slot index to the user (for the optimum H obtained, $H = 3$, we need 4, 5 or 6 slots depending if the protocol used is A, B or C respectively) and 2 bits to transmit protocol index (4 different protocols). So less than a byte ($5 < 8$) of signaling information per update is needed.

On the uplink the position is defined by the hop range h and the angle of the cell. Since the BS has no direct contact with all terminals, positioning can be calculated in the terminal by using separate, beacon like signals transmitted from the BS. The coverage of these transmissions depends on the type of the positioning algorithms. For conventional triangulation the terminal should be able to see at least three beacons. This will require higher power but the updating rate is rather high since the beacon is available all the time and mobility is no limiting factor. Another option is to use a beam former for the beacon and to transmit the beacon with $\overline{\theta}_H$ different angles. Each transmission will carry inherently the beacon (angle) index and strongest beacon will indicate the angular position of the terminal. The distance (parameter h) can be measured in the terminal if the beacon is modulated by the pseudorandom sequence. This approach would require lower transmission power from the beacon but the beam rotation time will

be restricting factor for the location updating rate and so the acceptable mobility rate. Regarding the overhead, for $H = 3$, we need 2 bits to transmit value of $h < H$ and 5 bits to transmit $\overline{\theta}_H = 24 < 2^5$ values of the angle which sums up to $2 + 5 = 7 < 8 = 1$byte of information. One additional bit can be used to indicate readiness/not readiness to cooperate (1/0) so that $7 + 1 = 8$bits = 1byte of overhead is sufficient to transmit all necessary information. So the bandwidth of the control channel on the uplink would be $1/B \cdot T_r$ bytes = $8/B \cdot T_r$ bits/s.

If *InSyNet* architecture is used then these functions can be implemented on the control signals directly since the control plan covers the entire cell in a single hop.

An alternative solution for positioning would be GPS system. In this case there would be no need for beacons but (x,y) coordinates obtained in the terminal should be passed to the BS with more precision. This would require more overhead bits. This increased overhead may be now reduced by more sophisticated messaging where the complete (x,y) information would be transmitted at the beginning (x_0,y_0) of messaging and in the sequel only the changes in the position $(\Delta x_0, \Delta y_0)$ would be sent to the base station. Dimensioning of the message in this case should be straightforward engineering work. Using these two type of messaging, resulting into reduced overhead, would handle higher mobility in the system.

D) Channel defading and cell design

Fig. 41 illustrates that the best system performance is obtained for the complete channel defading. The channel defading distance d_{CD} is the given channel parameter and cannot be changed. It depends on propagation conditions and represents the longest distance that the signal can be transmitted to on which the random component of the channel transfer function can be neglected with respect to its steady component. For the cell geometry used so far we have $d_{CD} = d_r = \sqrt{3}r$ and H_0 can be obtained by (38). For a given channel, characterized by d_{CD}, and given technology that includes conventional receivers for which optimum $H_0 = 3$ or diversity receivers for which $H_0 = 4$, we can find optimum cell radius R. If the cell is given in advance, characterized by R and given the channel with d_{CD}, we can find H_0.

E) Other protocols

The conventional multicast/broadcast protocols in cellular networks use one hop transmissions ($H = 1$) and can be compared directly with our approach for optimum H on Figs. 40a)-d). The superiority of optimum multihop schemes is evident.

Recently there has been works on opportunistic broadcast and multicast schemes [177]. In this family two specific protocols stand out, *the worst case user* and *the best case user* protocol.

The worst case user protocol adjusts the broadcast rate to the receiving capabilities of the user with the worst channel which can be approximated with our protocols for $H = 1$ which is inferior compared with the optimum H protocol.

The best user protocol adjusts the broadcast rate to the channel receiver capability of the user with the best channel. This rate is still lower of the rate achievable in multihop network with defaded channel. In addition the opportunistic protocol assumes that all users due to mobility will have a chance to become the best user in acceptably short period of time. This cannot be controlled by the network and depends solely on the user mobility.

F) Node availability/Route discovery protocol modifications

The results presented in this section are based on the assumption that all required transmitters on the route to the destination users are available. There are a number of options how the network can deal with the situation where this assumption does not hold and modify the route discovery protocol:

a) First hopping distance can be increased (lower H) for all users that would accordingly degrade the system performance as shown in previous figures.
b) The hopping distance can be changed only for a portion of users that will degrade performance less and result into an equivalent system throughput/utility corresponding to some reduced noninteger $H = p_1H_1 + p_2H_2$. The control of such a network would be also significantly more complicated.
c) Another possibility is to do rerouting of the broadcast, away from the directions with no terminal available without changing the hopping distance but increasing the number of hops H. This will increase the time needed for the broadcast to reach the boarder of the cell.

d) In the case of protocol C using transmit diversity nonavailability of a node will reduce signal to noise ratio (capacity) of a link due to reduced number of combined useful signals but the route can be maintained without presence of the node.

e) In the case of protocol D we can use either rerouting if the node is not available or accept the reduced capacity and keep the same route to simplified control of the network. In this case the capacity will be reduced due to reduced signal to noise ratio in the receiver diversity combiner, but also due to the fact that network decoding might not be possible which will for itself reduce the capacity by factor ½.

   Depending on the compromise between the performance degradation and the system complexity the route discovery protocol may be extended (modified accordingly).

f) A nano route discovery protocol is presented in Section 4.3 to deal with this situation. The implementations details are given below.

4.6.2 Nano route discovery protocol

To avoid the excessive signaling and preserve the advantages of the multihop network, the signaling plan is designed as conventional single hop cellular network and data plan as multihop cellular network as explained in the previous section. The route discovery protocol is operated by the BS based on the terminal location information. The availability for relaying and the user location is calculated in the same fashion as for the multicast case. In this case, the user location also depends on Γ and the location of any user type m^{Γ} can be easily obtained from (67) and (68). Based on the users' availability for relaying, the BS will send to the candidate transmiters which is the next relaying user available as defined by NRDP. Once all transmiters are one hop away from BS (first cluster of users), the cell partitioning enables the use of a specific Round Robin MAC protocol within the cluster of subcells making the system feasible for implementation.

4.6.3 Joint optimization of cooperative diversity and spatial reuse factor in MCNs

In practical implementation, for a given cell geometry (radius R) and type of terminals (power P), the optimum value of parameter r (or H) can be obtained from the optimization process (see numerical results in Section 4.5). For the network operation, users should be frame (6 slots) synchronized and aware of their turn (slot) in the Round Robin scheduling. This can be controlled either by the base station or in a distributive way by the mobile users. In both cases the controller should be aware of the transmitting user position in the cell.

4.7 Chapter summary

We have presented a nano scale model for high resolution optimization of MCNs. Four different protocols are developed and analyzed for the multicast scenario by using different relaying schemes and network coding. A spatial interleaving SI MAC protocol is introduced for context aware interlink interference management. The directed flooding routing protocol (DFRP) and inter flooding network coding (IFNC) are proposed for such network model. By adjusting the radius of the subcell r we obtain different hopping ranges which directly affects the throughput, power consumption and interference. With r as the optimization parameter, we jointly optimize scheduling, routing and power control to obtain the optimum trade-off between throughput, delay and power consumption in multicast cellular networks. The additional context awareness is characterized by two parameters: the sink location matrix $\mathbf{D}$ and the relay availability matrix $\mathbf{A}$. The major results of the chapter can be summarized as follows:

a) In the broadcast case ($\mathbf{D} = \boldsymbol{\beta}$): For conventional relaying in nonfading channel the optimum number of hops to reach the border of the cell is $H = 3$; For cooperative relaying and network coding the system performance is significantly better. The maximum throughput for these protocols is obtained for $H = 4$. However the largest increment of network throughput $\partial Thr/\partial H$ is obtained again for $H = 3$, suggesting this value as the best choice when it comes to trade-off between performance and complexity.
b) In the multicast case, the best performance is obtained by protocol C (cooperative relaying) or protocol D (network coding) depending on the

density of destination users in the network. The locations of the destination users have significant impact on the performance.

c) The nano scale channel model (NSCM) gives an insight into how the transformation of the fading process, resulting from increasing the number of hops in the cell, impacts the system performance. It was shown that the best system performance is obtained with the number of hops needed to completely transform fading channel into a non fading channel. This particular choice of r is referred to as channel defading tessellation. Network design procedure for channel defading is also elaborated.

d) The nano scale network model is further extended to include a conventional resource reuse scheme used for cellular networks to design a new concept of route discovery protocols aware of the mutual impact of all routes in the cell. The novel nano route discovery protocol (NRDP) performs rerouting/rescheduling when there are users temporally unavailable for relaying. NRDP provides results close to the ideal case (all users available for relaying) in terms of network capacity and throughput. Also, NRDP is the most efficient in terms of traffic load, power consumption and delay.

e) We provide an analysis of optimum hopping range in MCNs and we have optimized the trade-off between cooperative diversity and spatial reuse to maximize the throughput in the network. The increased number of concurrent transmissions, enabled by spatial cell partitioning, increases the system throughput but also increases the level of interference that reduces the capacity of simultaneously used links in the network. The radius of subcells r determines the relaying hop range and the amount of interlink interference. All transmissions are recorded by the neighboring receivers and combined in a cooperative diversity transmission. The increased number of hops increases the diversity order ξ but at the same time reduces the throughput per user since the network capacity has to be shared between the increased number of users. By introducing a utility function as a ratio of the network throughput and overall power consumption we can simultaneously optimize these parameters and the packet delivery delay, as a function of relaying range. The optimum relaying range defines the optimum subcell partitioning and the spatial reuse in the network. The main results of the analysis show the following:

   – There is a clear optimum value for the format of cell partitioning in terms of inner cell radius r.

- In general, for lower power the utility peak is higher since the utility function is inversely proportional to power consumption.
- In the case of cooperative relaying, the utility is higher which means that with less power higher capacity can be achieved.
- The capacity is higher for cooperative relaying and is increased as the number of hops is increased. Higher capacity is obtained for higher power.
- The peak of the utility is uniformly increased as the cluster factor Γ is increased. The optimum number of hops is higher for higher Γ.

5 Enhancing multicast performance in MCNs to support multimedia applications

In this chapter, we continue with our study of multicast in MCNs and extend it to support multimedia applications (transmission of very large files). As the network load in this type of applications requires special attention, “store-carry-forward” paradigm has been suggested to provide load balancing in MCNs at the expense of higher message delivery delay [104]. Since mobile users are equipped with different interfaces, cellular service providers may offload big multimedia files to an auxiliary network to alleviate bottlenecks and reduce tranmission costs. “Store-carry-forward” routing was originally conceived as a way to provide communication in *Delay Tolerant Networks* (DTN). We assume that a DTN is used as auxiliary network and new algorithms are presented to improve the performance of DTN multicast in terms of average delivery delay, and energy efficiency.

The most common “store-carry-forward” routing protocol is epidemic routing. Analogous to disease spreading, a user receiving a packet buffers and carries that packet as it moves, *infecting* new users that it encounters. The user possessing the packet is referred to as *infected* user. Once the infected node meets the destination, the network initiates the so called “infection recovery process” in order to remove the delivered packet from the rest of the nodes. The packet is deleted for efficient buffer and bandwidth utilization. On the other hand, a node retains “packet delivered” information in the form of an *anti-packet* that prevents it from accepting another copy of the same packet. Haas and Small [178] suggest the following recovery schemes for unicast applications:

- *immune:* An anti-packet is created at a node only after it meets the destination.
- *immune_TX:* A node carrying an anti-packet transmits it to another node that is carrying the associated obsolete packet to let that node know of packet delivery.
- *vaccine:* A node carrying an anti-packet forwards it to all other nodes including uninfected nodes.

The conventional infection recovery process starts as soon as the packet reaches the first destination which in the case of multicast session may reduce the chances that the rest of the destination nodes receive the message. So, in a multicast application there is a need to delay the initialization of this recovery process in

order to allow more efficient delivery of the information to all intended destinations.

The main focus of the work presented in this chapter is to analytically study the effect of different recovery schemes on the performance of multicast DTN. In addition, new adaptive recovery schemes are developed where the infection recovery process is adjusted to the multicast traffic. In general for a multicast with more destinations the initialization of the recovery process will be postponed longer. The performance of these new algorithms is compared to a number of unicast recovery schemes modified for multicast DTN, which also represents a contribution of this chapter.

Although epidemic routing achieves the highest delivery probability with the least delays, it wastes much energy in excessive duplications. Variations of epidemic routing have been proposed [179] to exploit this trade-off between delivery delay and resource consumption. In [180] a form of network coding and epidemic routing for unicast transmission in DTNs was suggested. In general it is known that in conventional multicast/broadcast networks, network coding improves the performance by taking advantage of the redundancy of packets in the network and combining them by XOR operation reduces the number of new transmissions [147]. The efficiency of network coding in multicast DTN was also showed by [181] through simulations for Spray and Wait forwarding [179]. So, we include in our model network coding too combined with epidemic routing and extend the study to the network behavior in multicast scenario. The resulting routing protocol will be referred to as Polymorphic Epidemic Routing (PER).

The analytical framework considered is based on Ordinary Differential Equations (ODEs) as a fluid limit of Markovian models [182]. Our analytical study is able to provide insights for future designs of recovery process for any routing protocol in multicast DTN.

The performance measures considered include the delivery delay to the destinations, recovery delay from the infection process and energy efficiency in terms of the number of packet copies made until the time of delivery and recovery. Numerical analysis shows the outstanding performance of our new adaptive recovery schemes when cooperative and non-cooperative destinations are used. By adaptive *immune*, *immune_TX* and *vaccine* schemes the delivery delay is reduced up to 3 times compared to the conventional schemes. By adaptive timeout recovery scheme, the reduction in the delivery delay can reach up to 5 times at the expense of larger recovery delay. The results are closely related to the number of destinations.

The rest of this chapter is organized as follows. An overview and background of multicast in DTN is given in Section 5.1. Section 5.2 describes the system model, traffic model and Polymorphic Epidemic Routing (PER). In Section 5.3, we present our adaptive infection recovery schemes along with performance analysis in Section 5.4. Numerical results are shown in Section 5.5. Finally, we conclude this chapter in Section 5.6. A special iterative algorithm, developed to solve the differential nonlinear system of equations (DiNSE-algorithm) for numerical solution of the problem, is presented in the Appendix D.

5.1 Overview and background

Delay Tolerant Networks (DTN) [179], in their basic form, provide communication service in highly challenging scenarios where only intermittent connectivity exists, and it is difficult to maintain paths between any communication source and destination pair. Examples of such networks include sparse sensor networks for wildlife tracking and habitat monitoring [178],[183], vehicular ad hoc networks for road safety and commercial applications [184], mobile social [185], military [186], and deep-space interplanetary networks [187]. In those scenarios, there has been a growing interest in multicast DTN protocols that enable the distribution of data to multiple receivers [187]-[189], *i.e.*, providing update information such as news, weather reports, road, traffic congestion, stock prices to a group of users and in disaster recovery scenes where it is particularly essential to distribute critical information to rescue teams. Traditional ad hoc routing protocols, which rely on the end-to-end paths [190], may fail to work for such networks. In future wireless communication systems low exposure networks will be more and more attractive and conventional cellular and ad hoc networks may integrate concepts with intermittent terminal connectivity for delay tolerant but still delay controlled traffic. In such networks short distance (low power) transmissions will be used whenever possible avoiding the long distance transmissions requiring high power and high exposure of the user to the radiation. As a result, researchers propose a new routing mechanism called store–carry–forward routing [191] to provide communication.

For a detailed survey on the applications of store-carry-forward paradigm for load balancing and multicast support in cellular multicast applications, the reader is refered to [Section 1.2.5, Chapter 1].

Multicast in DTN is a fundamentally different and hard problem compared to multicast in Ad Hoc networks due to the frequent disconnections. Zhao *et al.*

[192] developed several multicast algorithms with different strategies depending on the availability of knowledge about network topology and group membership. They have shown that by using broadcast epidemic routing (BER) they achieve the same delivery ratio under different amounts of available knowledge. We believe that including a learning process in discovering the topology drastically increases the overhead as the number of hops increases. Lee *et al.* [193] study the scalability property of DTN multicast routing and propose *RelayCast* scheme based on 2-hop forwarding. This scheme does not completely exploit the characteristic of opportunistic forwarding, since there is a single relay node for a given packet. An improved scheme, RelayCast with Multicast Receiver Relay (RelayCast-MRR) allows that each relay node can use all nodes for relaying. However, they have shown that RelayCast-MRR cannot improve the delay except in the broadcast case. Gao *et al.* [194] study the multicast problem from the social network perspective. They formulate the relay selections for multicast as a unified knapsack problem and demonstrate the efficiency of the proposed schemes by simulation results. The main drawback of this scheme is the low rate of node contacts in DTN which results into very high delivery delays.

Epidemic Multicast Routing (EMR) applies epidemic algorithm [195] to the multicast communication of DTN. Due to the flooding mechanism, the efficiency of the algorithm will be poor unless some improvements can be done to solve the resource problem. In [181] MIDTONE protocol is proposed based on Spray and Wait fowarding with network coding. The efficiency of network coding was shown through simulations. They also proposed some recovery schemes where the users keep a list of destinations that have received the packets. Our recovery schemes guarantee the delivery without any knowledge of neither the contact information nor packet delivery. As users in DTN may have limited memory and computational capabilities, our recovery schemes reduce the overhead just to the exchange of anti-packets.

The analytical models and feasibility of the system implementation for our proposed schemes presented in the sequel represent significant contribution to the concept of multicast DTN networks paradigm.

5.2 System model and assumptions

In this section, we introduce the traffic model, review the concept of epidemic routing and, define Polymorphic Epidemic Routing (PER) for multicast DTN.

5.2.1 Traffic model

We study a network consisting of $N + 1$ wireless mobile nodes where there is one source and a set of relaying nodes $\mathcal{N}$ ($|\mathcal{N}| = N$) moving within a constrained area according to a random mobility model. We consider multicast communication from the source node to a set of destinations $\mathcal{D} \subseteq \mathcal{N}$ where the destinations can also forward the packet to each other which is referred to as Destination Cooperative Multicast (DCM). Comments on the case of Destination Non-Cooperative Multicast (DNCM) will be provided in Section 5.3. Since the density of nodes is sparse in DTN environment, two nodes can communicate only when they come within the transmission range of each other, which means a communication opportunity to forward packets to each other. As the node density is low, we ignore the interference among nodes.

To facilitate the analysis without loss of generality, we assume that when two nodes meet, the transmission opportunity is only sufficient to completely transmit one data packet per flow. This assumption is justified by choosing the proper packet length (maximum packet length allowed by the rendezvous time) and allowing *only one packet transmission per flow per node* during the nodes' rendezvous. It is straightforward to extend this to the general case where an arbitrary number of packets can be delivered when the opportunity arises. We start by considering that the nodes buffer can accommodate all packets that they receive.

We assume that the time between two consecutive transmissions opportunities (when nodes meet) follows an exponential distribution with a rate λ. This model has been widely adopted in the recent literature, e.g., in [196], [197] and verified by both theoretical analysis [198] and in many practical systems [199]. It also enables the theoretical analysis by using continuous Markov model [182].

5.2.2 Polymorphic Epidemic Routing

In general it is known that in conventional multicast/broadcast networks, network coding improves the performance of the network [147]. A form of network coding and epidemic routing *for unicast* transmission in DTN networks was suggested in [180]. So, we include in our model network coding too and extend the study to the network behavior in *multicast* scenario. In order to be able to have a tractable model for the analysis of the infection recovery schemes we introduce a number

of modifications into the concept of epidemic routing which represent a part of the contributions of this chapter. We assume that a set of destinations $\mathcal{D}$ requests a common packet f from the multicast source. To keep the model simple, we consider the case where the multicast source infects the nodes with two packets a and b ($f = a,b$) and their combination $c = a \oplus b$ where $\oplus$ stands for *XOR* operation on binary data stream. By analogy between epidemic routing and disease spreading, infection with two different packets (agents) is referred to as *polymorphic infection* and *DNA* combination of these agents c is referred to as the mutation. The incentive behind this approach is that now every rendezvous between the two users increases the probability that a useful transmission will take place. User c (infected with agent c) will transmit a useful packet if it meets either user b (since $c = a \oplus b$) or user a (since $c = a \oplus b$). The infection process is illustrated in details in Fig. 63, where the new packet received by the each node is underlined.

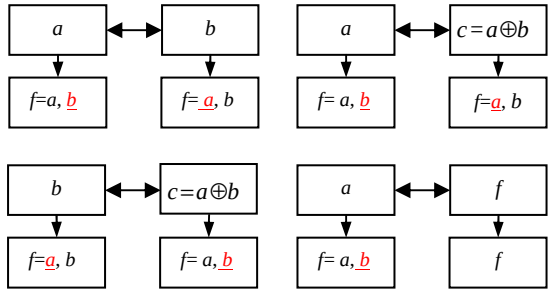

Fig. 63. Polymorphic infection process.

To model *Polymorphic Epidemic Routing* we will use the following notation. We denote by $A(t)$, $B(t)$, $C(t)$, and $F(t)$ the number of users infected by agents a, b, $c = a \oplus b$ and $f = a,b$ respectively in time t. We denote by $I(t) = A(t) + B(t) + C(t) + F(t)$ the overall number of infected users in the network.

We model the infection rate for users a, b, c and f by using ODEs as a fluid limit of the Markovian model [182]. Hence, we have

$$X'(t) = \lambda(X(t) + \frac{1}{3}F(t))(N - I(t)) - \lambda X(t)(I(t) - X(t)) \qquad (89)$$

for $X \in \{A, B, C\}$, and

$$F'(t) = \lambda A(t)(B(t)+C(t)) + \lambda B(t)(A(t)+C(t)) + \lambda C(t)(A(t)+B(t)) + \lambda F(t)\big(X(t)-F(t)\big). \quad (90)$$

In (89) an increment in $X(t)$, denoted as $X'(t)$ for $X \in \{A,B,C\}$, is proportional to two terms. The first term represents the rate at which a node infected by packet x meets a non infected node plus the rate that node infected by f meets a non infected node when it randomly, with probability 1/3, chooses one of the three infection options: packet a, b or c. It cannot infect by f since this would require transmission of two packets. The second term represents the event that we lose packet x (negative increment) which happens when it becomes f if it meets any other packet except x.

An increment in $F(t)$ in (90), denoted as $F'(t)$, is obtained if a meets b, when they exchange the packets and two new f 's will be created. This occurs with rate proportional to $A(t)B(t) + B(t)A(t)$ (first part of the first and the second term in (90)) and it will further propagate randomly one of the three options specified above. On the other hand if f meets a, b or c only one extra f will be created, which is included in the last term of (90). Similarly other terms in (90) can be interpreted. An algorithm for iterative solution of (89) and (90) is given in Appendix D for the initial conditions $A(0) = B(0) = C(0) = 1$ and $F(0) = 0$.

After introducing the scenario and the routing algorithm for multicast DTN, in the sequel we model the infection of the destination users and present the new infection recovery schemes.

5.3 Recovery schemes for multicast DTN

In this section, we first extend the most common recovery schemes applied so far to unicast to multicast and Polymorphic Epidemic Routing (PER) by modifying (89) and (90) to include the recovery process for different schemes. Later on, we present the new adaptive recovery schemes.

5.3.1 Conventional recovery schemes applied to multicast DTN

Once a node delivers a packet to the destination, it should delete the copy from its buffer both to save storage space, and to prevent the node from infecting other nodes. But if the node does not store any information to keep itself from receiving the packet again (*i.e.*, becomes susceptible to the packet), a packet would generally be copied, and the infection would never die out. In order to prevent a

node from being infected by a packet multiple times, an anti-packet can be stored in the node when the node delivers a packet to the destination. Reference [200] refers to this scheme as *immune* scheme. With *immune* scheme, a node stores a packet copy in the buffer until it meets the destination, often long after the first copy of the packet is delivered. A more aggressive approach to delete obsolete copies is to propagate the anti-packets among the nodes. The anti-packet can be propagated (transmitted_TX) only to those infected nodes (*immune_TX* scheme), or also to susceptible nodes (*vaccine* scheme).

Similar to our earlier analysis in Section 5.2, we can derive ODEs to model the infection and recovery process as the limit of Markov models [182]. We denote by $D = |\mathcal{D}|$ the number of destinations. In the sequel we derive the expressions for the number of infected $I(t)$ and recovered nodes $R(t)$ for all three schemes extended to our multicast system and *Polymorphic Epidemic Routing* (PER). These expressions are obtained based on two different assumptions that are referred to as

- *Option 1*. A node can deliver up to two packets to the destination: This assumption is based on the fact that there is no transmission from the destination to the intermediate node so, with the same rendezvous time the destination can received both packets *f = a, b*.
- *Option 2*. The node can deliver only one packet when it reaches the destination: the destination should be already infected by *a* or *b* or *c* in order to become *f*.

The infection rate and recovery rate for users *a*, *b*, *c* and *f*, are obtained for each recovery scheme, as follows:

A) Immune

By using the same logic as before in generating the fluid equations under this condition, we have that the infection rate for users *a, b, c* and *f* for the assumption made in *Option1* are

$$X'(t) = \lambda(X(t) + \frac{1}{3}F(t))(N - I(i) - R(t)) - \lambda X(t)(I(t) - X(t)) - \lambda X(t)D \qquad (91)$$

for $X \in \{A, B, C\}$, and

$$F'(t) = \lambda A(t)(B(t)+C(t)) + \lambda B(t)(A(t)+C(t)) + \lambda C(t)(A(t)+B(t)) + \lambda F(t)(I(t)-F(t)) - \lambda F(t)D. \quad (92)$$

The difference between (89)-(90) and the previous equations (91)-(92), is in the last term in (91) and (92). This term means that the infection rate X(t) or F(t) decreases when a packet $x \in \{a,b,c\}$ or f, respectively meets the destinations D. As users can transmit up to two packets when they meet the destination, users infected by f get recovered when they deliver the packet to the destination independently if the destination was already infected by other packet a, b or c. Thus, the recovery rate for users a, b, c and f is obtained as

$$\begin{aligned} R^x'(t) &= \lambda X(t)D \\ R^f'(t) &= \lambda F(t)D \\ R(t) &= R^a(t)+R^b(t)+R^c(t)+R^f(t). \end{aligned} \quad (93)$$

The infection rate of the destinations infected by $x \in \{a,b,c\}$ when they cooperate and forward the packet to other destinations (DCM) is modeled as:

$$D^x'(t) = \lambda(X(t)+D^x(t))(D-D^i(t)) - \lambda D^x(t)(I(t)-X(t)+D^i(t)-D^x(t)), \quad (94)$$

where D^x is the number of destination users infected by packet x and $D^i(t) = D^a(t)+D^b(t)+D^c(t)+D^f(t)$. The infection rate $D^x'(t)$ in time t is increased when packet x or a destination infected just by x meets a destination that has not been infected $(D-D^i(t))$. On the other hand, $D^x'(t)$ decreases when a destination infected just by x meets other user or destination infected by other type of packet (last term in (94)). The same reasoning applies for the infection rate $D^f'(t)$ which can be written in compact form as

$$D^f'(t) = \lambda(F(t)+D^f(t))(D-D^f(t)) + \sum_{x\in\{a,b,c\}} \lambda(X(t)+D^x(t)) \sum_{y\in\bar{x}} D^y(t). \quad (95)$$

For the case of non-cooperative destinations, (94)-(95) are modified as follows

$$D^x'(t) = \lambda X(t)(D-D^i(t)) - \lambda D^x(t)(I(t)-X(t)) \quad (94a)$$

$$D^f'(t) = \lambda F(t)(D-D^f(t)) + \sum_{x\in\{a,b,c\}} \lambda X(t) \sum_{y\in\bar{x}} D^y(t). \quad (95a)$$

In this case, the infection rate $D^x'(t)$ in time t is increased when packet x meets a destination that has not been infected $(D-D^i(t))$ and it decreases when a

destination infected just by *x* meets a user infected by different packet $y \in \overline{x}$. The same applies for $D^f{}'(t)$.

As destinations are not recovered from the infection, (94)-(95) or (94a)-(95a) are the same independently of the recovery schemes used.

The difference between *Option 1* and *Option 2* for immune comes from the fact that in *Option 2* as the users can transmit just one packet to the destination, the destination should be already infected by *a* or *b* or *c* in order to become *f*. This modifies (93) as

$$R^f{}'(t) = \lambda F(t)(D^i(t)) \tag{96}$$

and consequently the last term of (92) should be modified as in (96). The infection rate of the destination $D^f{}'(t)$ in the case of DCM is also modified as

$$\begin{aligned} D^x{}'(t) &= \lambda(X(t) + \frac{1}{3}F(t) + D^x(t) + \frac{1}{3}D^f(t))(D - D^i(t)) \\ &\quad - \lambda D^x(t)(I(t) - X(t) + D^i(t) - D^x(t)) \\ D^f{}'(t) &= \lambda(F(t) + D^f(t)) \sum_{x \in \{a,b,c\}} D^x(t) \\ &\quad + \sum_{x \in \{a,b,c\}} \lambda(X(t) + D^x(t)) \sum_{y \in \overline{x}} D^y(t), \end{aligned} \tag{97}$$

where $F(t)$ and $D^f(t)$ can transmit with probability 1/3 a packet *x* when they meet an uninfected destination and increase in that way $D^x{}'(t)$. The logic behind $D^f{}'(t)$ can be easily deduced from the previous explanations. Equation (97) can be easily modified for (DNCM).

The rest of the equations for *Option 2* remain the same as in (91) and (94).

Similarly, differential equation models for *immune*_TX and *vaccine* scheme can be derived from Markov model. To simplify the presentation, we provide the infection and recovery rates for each scheme by considering *Option 1*. The expressions for *Option 2* can be easily obtained in the same way as explained for immune scheme.

B) Immune_TX

In this scheme, the anti-packet can be transmitted to those infected nodes, so we obtain a new recovered node when an infected node meets a node that has been recovered or the destination. This modifies (89) and (90) as

$$X'(t) = \lambda(X(t) + \frac{1}{3}F(t))(N - I(t) - R(t)) \\ - \lambda X(t)(I(t) - X(t)) - \lambda X(t)(D + R^x(t) + R^f(t)) \quad (98)$$

for $X \in \{A, B, C\}$, and

$$F'(t) = \lambda A(t)(B(t) + C(t)) + \lambda B(t)(A(t) + C(t)) + \lambda C(t)(A(t) + B(t) \\ + \lambda F(t)(I(t) - F(t)) - \lambda F(t)(D + R^f(t)), \quad (99)$$

where the last term in (98) and (99) indicates the reduction in the number of infected nodes due to recovery. In this scheme we assume that a node that has been recovered from *f* can recover a node infected by $x \in \{a, b, c\}$. This is justified by the fact that if packet *f* has been received, there is no need to transmit more packets *a*, *b* or *c*. Thus, the recovery rates are given by

$$R^x'(t) = \lambda X(t)(D + R^x(t) + R^f(t)) \\ R^f'(t) = \lambda F(t)(D + R^f(t)) \quad (100) \\ R(t) = R^a(t) + R^b(t) + R^c(t) + R^f(t).$$

The infection of the destinations is modeled again as in (94)-(95) for DCM or (94a)-(95a) for DNCM. For *Option 2*, the equations can be generated by analogy to *immune* scheme.

C) Vaccine

In this scheme, in addition to the previous schemes, we also vaccinate the uninfected users that are susceptible to receiving the packet. We now additionally recover (vaccinate) the users that have neither been infected nor recovered $(N - I(t) - R(t))$ when they meet a recovered node or a destination that has been infected by that packet. The infection rates for users *a*, *b*, *c* and *f* are defined as in *immune_TX*, given by (98) and (99). But the recovery rates are now obtained as

$$R^x'(t) = \lambda X(t)(D + R^x(t) + R^f(t)) + \lambda(N - I(t) - R(t))(D^x(t) + R^x(t)) \\ R^f'(t) = \lambda F(t)(D + R^f(t)) + \lambda(N - I(t) - R(t))(D^f(t) + R^f(t)), \quad (101)$$

where the last term in (101) indicates the fact that the recovery rates $R^x'(t)$ and $R^f'(t)$ are increased if a uninfected node $(N - I(t) - R(t))$ meets the destination infected by *x* and *f*, or a recovered node from *x* and *f*, respectively. The infection

rate of destination users is modeled as in (94)-(95) for DCM or (94a)-(95a) for DNCM.

The previous recovery schemes start deleting the packets when the first destination is reached. This slows down the infection process in multicast DTN as the packets are recovered before all destinations have received them. In the sequel, we present different options for improvement depending on the level of signaling available in the network. The issue of signaling will be additionally discussed in the sequel.

5.3.2 Adaptive recovery schemes

We denote by $p_r(t)$ the recovery probability in time *t*. In other words, when a node meets a destination, the destination will send the anti-packet to the node with probability $p_r(t)$. In the existing recovery schemes, $p_r(t) = p_r = 1$. The aim of the adaptive recovery schemes is to modify $p_r(t)$ based on the number of destinations *D*, so that the recovery is performed in such a way that the packets are removed slower while the infection process is still being performed or, the recovery is delayed until all (most) destinations have received the packets.

We introduce a time dependent probability of packet recovery

$$p_{r_e}(t) = 1 - e^{-\frac{\lambda N}{D}t}, \tag{102}$$

where the decay parameter is proportional to the meeting rate λ and *N*, and inversely proportional to the number of destinations *D*. This approach requires low level of signaling as all parameters λ, *N* and *D* are known in the network.

As an alternative, we also propose to delay the recovery for certain time T_D^f where T_D^f is estimated as the time needed to deliver the packets to the destination (delivery delay). The calculations for estimating T_D^f will be elaborated in the next section. In this case, we assume that certain level of signaling is available in the network (provided by the cellular network) so when the last destination receives the packet *f* can signal the source and then, the recovery process will start with probability

$$p_{r_T}(t) = \begin{cases} 1, & t \geq T_D^f \\ 0, & t < T_D^f \end{cases}. \tag{103}$$

The integration of the cellular network and DTN results into a *Inter System Networking paradigm* referred to as *InSyNet*(C,D) where the control plane (C) is

designed to have full coverage through the cell and the data plane (D) follows the conventional DTN. This concept has been already introduced in the previous chapters. The signaling plane handles much lower data rates. By using *InSyNet* the packet delivery to the last destination could be signalled instantaneously to the source and to all users in the network.

Equations presented in the previous section should be modified by replacing the meeting rate between the recovered node and other nodes by $\lambda \to \lambda(t) \to \lambda p_r(t)$ to model *immune*, *immune_TX* and *vaccine* under different $p_r(t)$.

5.3.3 Timeout recovery scheme

This scheme was introduced in [178] and referred to as just-TTL recovery scheme. In this section, we extend this scheme to PER for multicast DTN. The scheme behaves as follows: when a node receives a packet, it starts a timer with duration drawn from an exponential distribution with rate υ, after the time expires the packet will be removed from the buffer and the node stores an anti-packet to avoid future infections by the same packet. So, the node recovers from the infection *x* after the timer associated to *x* expires, and there is no need for explicit transmission of anti-packets. This is modeled by the following ODEs:

$$\begin{aligned} X'(t) &= \lambda(X(t) + \frac{1}{3}F(t))(N - I(i) - R(t)) - \lambda X(t)(I(t) - X(t)) \\ &\quad - \lambda X(t)D - \upsilon(X(t) - 1) \\ F'(t) &= \lambda A(t)(B(t) + C(t)) + \lambda B(t)(A(t) + C(t)) + \lambda C(t)(A(t) + B(t)) \\ &\quad + \lambda F(t)(I(t) - F(t)) - \lambda F(t)D - \upsilon(F(t) - 1) \end{aligned} \tag{104}$$

for $X \in \{A, B, C\}$, where the last term in (104) indicates the number of packets recovered, and it is obtained as

$$\begin{aligned} R^x(t) &= \upsilon(X(t) - 1) \\ R^f(t) &= \upsilon(F(t) - 1) \\ R(t) &= R^a(t) + R^b(t) + R^c(t) + R^f(t). \end{aligned} \tag{105}$$

The infection rate of the destinations is defined again as in (94)-(95) for DCM or (94a)-(95a) for DNCM.

The recovery schemes are evaluated by different performance metrics, *e.g.*, delivery delay, energy consumption, time efficiency of the recovery schemes. All these metrics are explained in more details in next section.

5.4 Performance analysis

5.4.1 Delivery delay

We define the packet delivery delay T_D^f as the time from the moment when packet *a*, *b* and *c* are generated at the source to the time when $f = a,b$ is received by all destinations *D*, and its Cumulative Distribution Function (CDF) as $P_D^f(t) = Prob(T_D^f < t)$.

We start by considering the delay when there is one destination $m_d \in \mathcal{D}$ (unicast case), and later on we extend it for multicast.

Let us denote by $P_N(t)$ the CDF of T_{m_d} when the number of nodes in the system is *N + 1*. Then, we can derive the following expression

$$\begin{aligned} P_N(t+dt) - P_N(t) &= Prob\{t \le T_{m_d} < t+dt\} \\ &= \text{Prob\{destination receives the packet f in } [t,t+dt] \,|\, T_{m_d} > t \,\} \\ &= \text{Prob\{destination receives the packet f in } [t,t+dt] \,\}\,(1-P_N(t)) \\ &= \text{E\{Prob\{ destination receives the packet f in } [t,t+dt] \,|\, D^f(t) \,\}\} \\ &\approx E\{\lambda D^f(t)dt\}(1-P_N(t)) \;= \lambda E\{D^f(t)\}(1-P_N(t))dt = \lambda E\left\{D^f(t)\right\}(1-P_N(t))dt \end{aligned}$$

where $D^f(t)$ is given by (95) or (97), depending on the delivery option considered.

Hence the following equation holds for $P_N(t)$:

$$\frac{dP_N}{dt} = \lambda E\left\{D^f(t)\right\}(1-P_N(t)). \tag{106}$$

As *N* increases, $P_N(t)$ converges to the solution of the following equation:

$$P^f{}'(t) = \lambda D^f(t)(1-P^f(t)), \tag{107}$$

where $P_{m_d}^f(t) = P^f(t)$ is the cumulative probability of the time needed for the packets $f = a,b$ to reach the destination $m_d \in \mathcal{D}$. This can be solved by using the iterative procedure presented in the Appendix D. Solving (95) or (97) for *immune*,

immune_TX, *vaccine* and *timeout* schemes together with (107), gives $P(t)$ with initial condition $P(0) = 0$.

From $P_{m_d}^f(t)$, the average delivery delay can be explicitly found in closed form as:

$$E[T_{m_d}^f] = \int_0^\infty (1 - P_{m_d}^f(t))dt. \tag{108}$$

In the multicast case, with the set of destination nodes $\mathcal{D}$ of size $D = |\mathcal{D}|$, (107) for each destination node m_d gives $P_{m_d}^f(t)$. The multicast delay is defined as the time needed for all destinations to receive $f = a,b$. Formally, it can be defined as $T_D^f = \max_{m_d} T_{m_d}^f$.

The CDF of the time needed for the double packet f to reach all destinations can be expressed as

$$P_D^f(t) = \left(P_{m_d}^f(t)\right)^D. \tag{109}$$

Finally, the average delay for multicast

$$E(T_D^f) = \int_0^\infty \left(1 - P_D^f(t)\right)dt. \tag{110}$$

Another metric that quantifies how efficient are the recovery schemes is the *average lifetime*. We define the average *lifetime* L^f, of a packet f as the time from when packet a, b and c are generated at the source node to the time when all copies of the packets are removed (*i.e*, there are no more infected nodes by packets a, b, c and f in the network). So, the lifetime of packet f is numerically calculated as

$$\begin{aligned} L^f &= \max_{a,b,c,f}\left\{ t_{\Delta R_t = 0} \right\} \approx \max_{a,b,c,f}\left\{ t_{\Delta R_t \le \varepsilon} \right\} \\ \Delta R_t &= R_t - R_{t-1} \\ R &= R^a, R^b, R^c, R^f, \end{aligned} \tag{111}$$

where ΔR_t are obtained for *immune*, *immune_TX*, *vaccine* and *timeout* schemes by solving (93), (100), (101) and (105) respectively, as indicated in the Appendix D.

The ratio

$$\varepsilon_t = T_D^f / L^f \tag{112}$$

will be referred to as *system time efficiency*. We also define the recovery delay as $T_R^f = L^f - T_D^f$. We will show in the numerical results that for those schemes with slow recovery $L^f > T_D^f$, and $T_R^f > 0$. On the other hand, when the recovery is faster than the infection $L^f < T_D^f$, and $T_R^f < 0$. Our adaptive schemes guarantee the delivery to all destinations D by adjusting the recovery probability to the available network parameters.

5.4.2 Energy consumption

Two metrics related to the energy consumption are considered: the number of times a packet is copied in its entire lifetime G_{L^f} and, the number of times a packet is copied at the time of delivery $G_{T_D^f}$. These are random variables taking value between $[0,\infty]$. The energy consumption grows linearly with the number of transmissions.

The energy efficiency of the system will be defined as

$$\varepsilon_e = G_{T_D^f} / G_{L^f}. \tag{113}$$

We obtain G_{L^f} for each recovery scheme as

$$\begin{aligned} G_{L^f} &= \sum_{t=0}^{L^f} \Delta I_t + \Delta R_t \\ \Delta I_t &= I_t - I_{t-1} \\ \Delta R_t &= R_t - R_{t-1}, \end{aligned} \tag{114}$$

where ΔI_t, and ΔR_t are calculated for *immune*, *immune* _TX, *vaccine* and *timeout* schemes. In other words in each time slot transmissions will increase the number of deliveries to the destinations and the number new infections including those that has been recovered.

Similarly, the number of times that a packet is copied in the network until the time that the packet is received by all destinations D

$$G_{T_D^f} = \sum_{t=0}^{T_D^f} \Delta I_t + \Delta R_t, \tag{115}$$

where T_D^f is the delivery delay given by (110). Equations (114)-(115) counts in each time slot all transmissions. Part of these transmissions are visible as an increase in the number of infected packets but part of these infections are erased by recovery process so both terms should be included in (114)-(115).

5.5 Performance evaluation

In order to compare the system performance for different recovery schemes, in this section we present numerical results, generated by solving the fluid equations for different models using the algorithm presented in the Appendix D. We set the meeting rate $\lambda = 0.004$, $N = 100$ and the number of destinations from $D = 1,\ldots, N$.

In Fig. 64 the average delivery delay T_D^f is shown versus the number of destinations D obtained by polymorphic epidemic routing. We compare the result to basic epidemic routing without network coding (equations for this scheme are presented in the Appendix E) and we can see that T_D^f without network coding is in average about 10% higher than with our proposed scheme. For $D = 1$ (unicast case), the improvement obtained by network coding is almost insignificant as already noticed in [180]. So, performance improvement by network coding which has been proved in multicast scenarios can be also noticed in multicast DTN. The improvement increases as we increase the number of destinations D. The f infection with network coding propagates faster which enables the packets to reach the destination sooner.

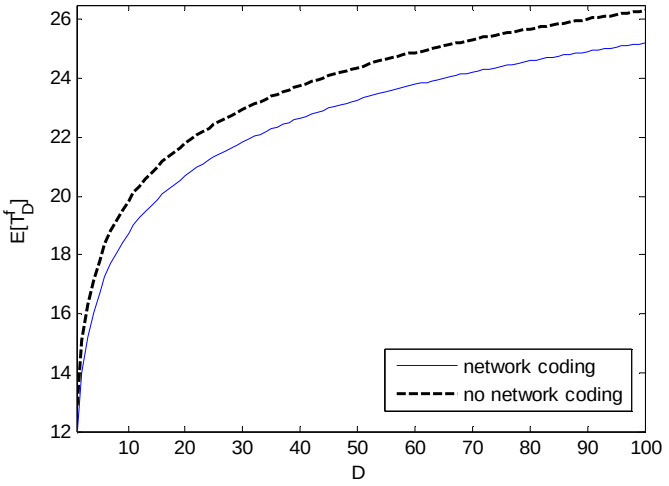

Fig. 64. Average delivery delay versus *D*.

The efficiency of *immune*, *immune_TX* and *vaccine* schemes is shown in Fig. 65 in terms of average delivery delay and lifetime when the recovery probability is fixed to $p_r(t) = 1$. We can observe a number of interesting phenomena.

By *immune* scheme, $E[T_D^f] < L^f$ for any number of destinations *D*. This is because in *immune* scheme the recovery from infection is very slow and all destinations receive the packet before all packets are recovered.

By *immune_TX*, we can see that for *D*>50, $E[T_D^f] > L^f$. The recovery now works faster than in *immune* scheme and the packets are recovered before the infection of set $\mathcal{D}$ is completed.

Finally, *vaccine* is the faster recovery scheme and $E[T_D^f] > L^f$ for any *D*. The average delay $E[T_D^f]$ for *vaccine* is the largest one as the number of infected packets is significantly reduced during the infection process.

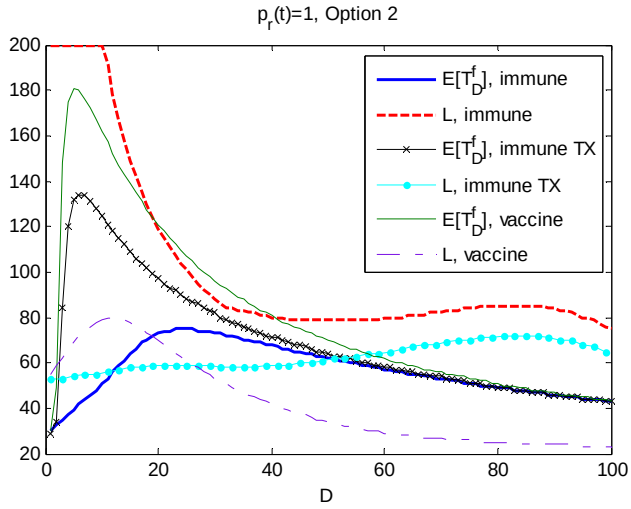

Fig. 65. Average delivery delay $E[T_D^f]$ and lifetime *L* versus *D*.

The average number of times that a packet is copied in its entire lifetime G_L and, at the time of delivery G_{T_D} are shown in Fig. 66 for *immune*, *immune_TX* and *vaccine* for $p_r(t) = 1$. For small number of destinations *D* and *immune* scheme, $E[G_L] > E[G_{T_D}]$. As mentioned before, the recovery with *immune* is very slow and many transmissions are made after the packets are delivered to the destinations ($t>T_D$). The values obtained for $E[G_L]$ and $E[G_{T_D}]$ for *immune_TX* are practically the same. For *vaccine*, as the recovery process finishes before the delivery to all destinations is completed (this is more evident for larger *D*), the destinations continue infecting each other until all destinations have received the packet even

when the rest of the users are recovered from the infection. For this reason, we can see that $E[G_L] < E[G_{T_D}]$ for large D.

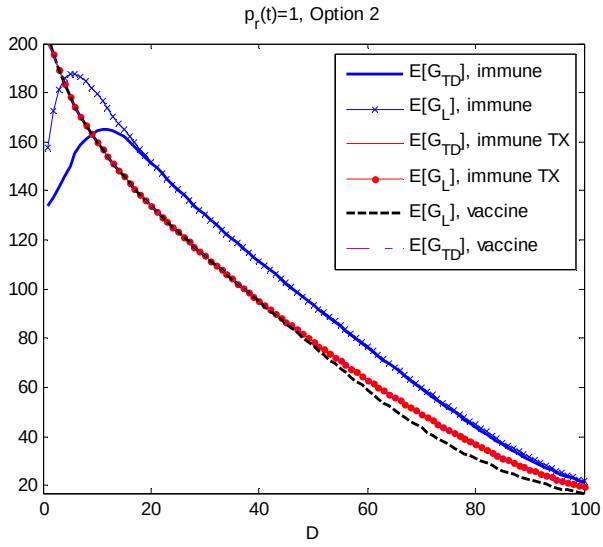

Fig. 66. Average number of copies $E[G_L]$ and $E[G_{T_D}]$ versus *D*.

In Fig. 67, we show the effects of removing the assumption that destinations can transmit to each other (DNCM) on the destination infection rate $D^f(t)$ for *immune TX* scheme. Similar effects were noticed with *immune* and *vaccine* but to simplify the presentation those figures are not presented. We assume that $D = 30$, and we can see that when $p_r(t) = 1$, $D^f(t) = 18$ for $t \to \infty$. As the packets are recovered while the infection to the destinations is taking place, just 18 destinations out of 30 get infected by packet *f*. When adaptive *immune TX* is used with $p_r(t) = p_{r_e}(t)$ or $p_r(t) = p_{r_T}(t)$ we can see that the performance is significantly improved and all destinations received *f*.

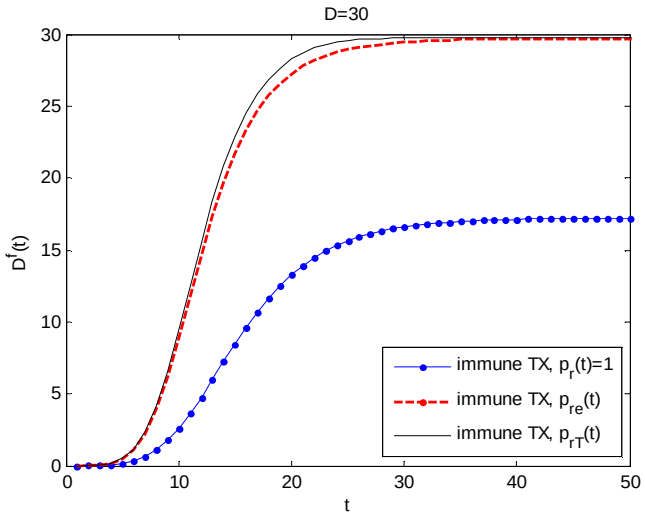

Fig. 67. Infection rate of destinations infected by f, $D^f(t)$, versus t when the destinations cannot infect other destinations.

In Figs. 68 to 70, we show the behavior of *immune*, *immune_TX*, *vaccine* and *timeout recovery* scheme for different recovery probabilities $p_r(t)$. We assume that D = 30 and that destinations can infect each other (DCM).

In Fig. 68, the recovery from infection for packet *a* is presented versus the time *t*. We can see that for *immune*, *immune_TX* and *vaccine*, $R^a(t)$ decreases for $p_r(t) = p_{r_e}(t)$ and $p_r(t) = p_{r_T}(t)$ compared to the case with fixed $p_r(t)$ = 1. This is because with these adaptive recovery schemes, the recovery is slower while the infection of the destination users is still taking place, so the number of users infected by *a*, *b*, or *c* decreases with *t* while a number of new packets *f* are created.

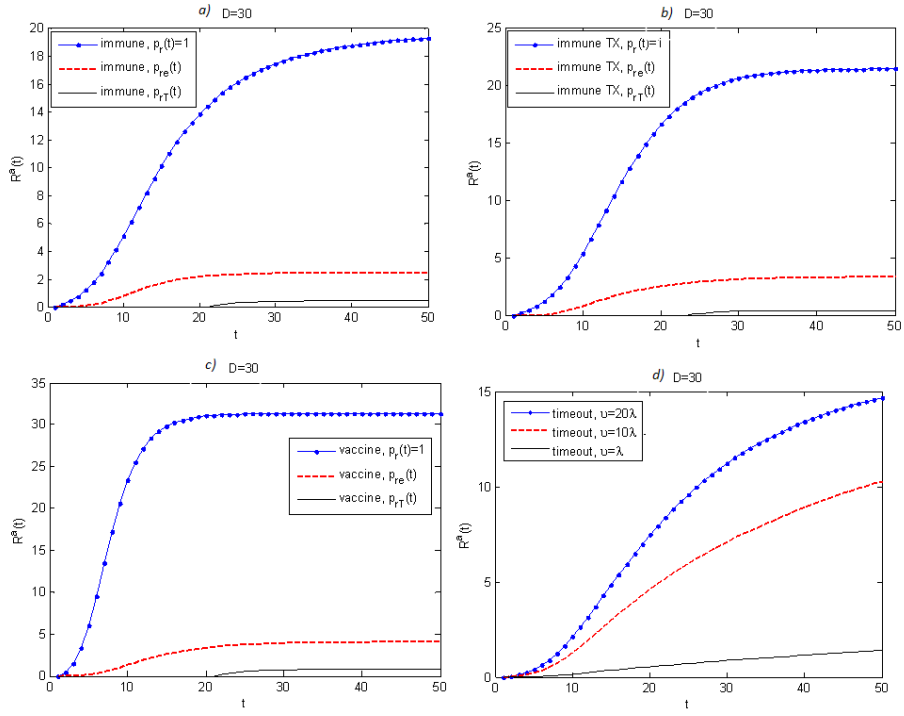

Fig. 68. $R^a(t)$ **versus t for a)** ***immune*****, b)** ***immune TX*****, c)** ***vaccine*** **for different values of** $p_r(t)$ **and, d)** ***timeout*** **recovery scheme.**

We can also see this effect in Fig. 69 where $R^f(t)$ is shown for the same schemes. The highest number of recovered packets is obtained by *vaccine* scheme. It also worth noticing that by $p_r(t) = p_{r_T}(t)$, the recovery is delayed and starts in $t > T_D^f$. For *timeout* recovery scheme, the number of packets recovered depends on the timeout factor μ, and the recovery is much slower than with any other scheme.

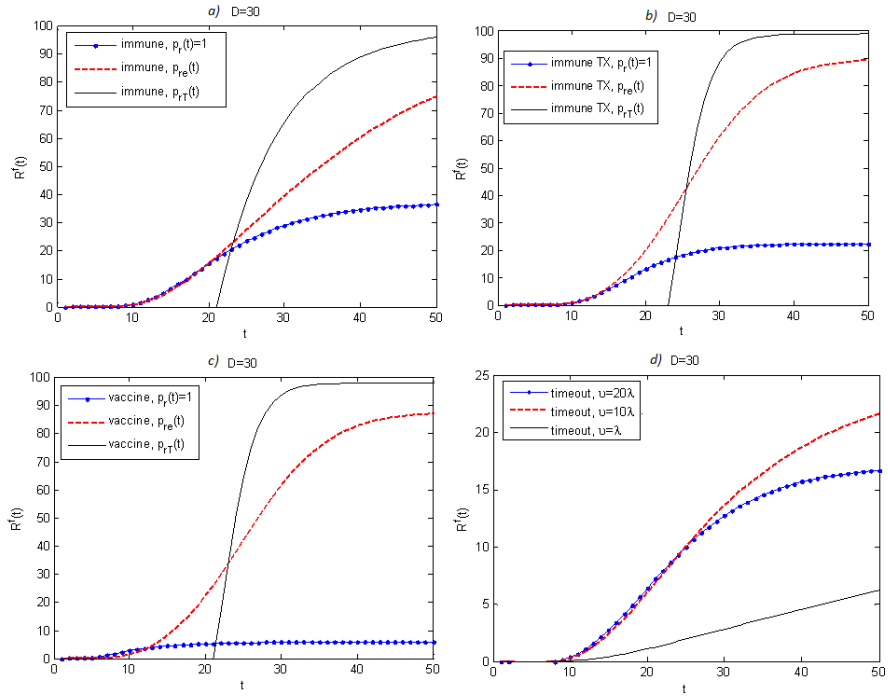

Fig. 69. $R^f(t)$ **versus *t* for a) *immune,* b) *immune TX,* c) *vaccine* for different values of $p_r(t)$ and, d) *timeout* recovery scheme.**

In Fig. 70, the average delivery delay $E[T_D^f]$ is shown for the previous schemes. The highest $E[T_D^f]$ is obtained for *vaccine* and *immune_TX* scheme with fixed $p_r(t)$, while the lowest delay is obtained by *timeout* recovery with $\upsilon = \lambda$ and with adaptive *immune* schemes, at the expense of larger recovery delays. We can see that the improvement obtained by using adaptive schemes compared to those with fixed $p_r(t)$ can reach up to 50% for *immune* scheme, 30% for *immune_TX* and 75% for *vaccine* scheme, and when $D < 30$. For higher D, $E[T_D^f]$ decreases in the same proportion for all schemes as there are more destinations to propagate the infection within themselves. We can also see that the choice of parameter υ also results in different values of $E[T_D^f]$.

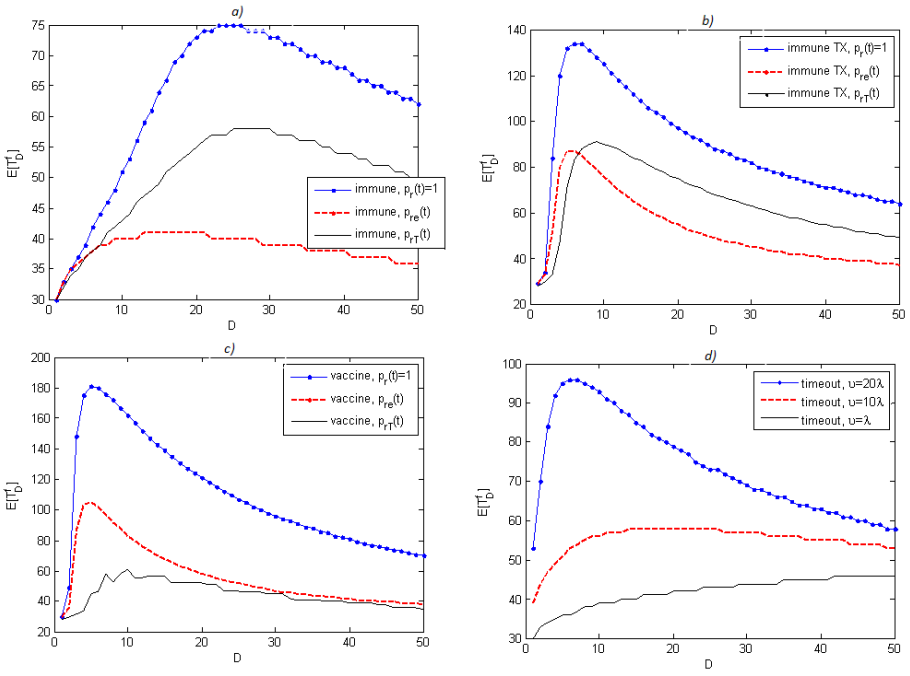

Fig. 70. Average delivery delay $E\left[T_D^f\right]$ versus *t* for a) *immune*, b) *immune TX*, c) *vaccine* for different values of $p_r(t)$ and, d) *timeout* recovery scheme.

5.6 Chapter summary

In this chapter, we have considered a Delay Tolerant Network (DTN) integrated within the cellular network to provide support for multimedia applications (transmission of very large files). The integration between both networks is referred to as *InSyNet*(C,D) where the control (C) plane is designed to have full coverage over the entire cell and the data (D) plane is the conventional concept of DTN.

We have presented an analytical framework to study the performance of different recovery schemes for multicast DTN and developed new adaptive recovery schemes where the infection recovery process is adjusted to the multicast traffic. Different recovery probabilities are used depending on the level of signaling available in the network. The performance of these new algorithms was compared to a number of unicast recovery schemes modified for multicast DTN. Our analytical framework can be easily extended to model the recovery

process for different multicast routing schemes. The network model considered enables us to discuss the above schemes in combination with some additional advanced techniques that have been recently considered in this field like network coding. So, we have included in our model network coding combined with epidemic routing and extended the study to the network behavior in multicast scenario. The resulting routing protocol is referred to as Polymorphic Epidemic Routing (PER).

The performance measures considered include the delivery delay to the destinations, recovery delay from the infection process, and energy efficiency in terms of the number of packet copies made until the time of delivery and recovery. Numerical results showed that the adaptive schemes can reduce the delivery delay by 2 times compared to *immune* scheme, 1.5 times for *immune_TX* and 3 times for *vaccine* scheme with fixed recovery probability. By *timeout* recovery scheme, the reduction in the delivery delay can reach up to 5 times compared to adaptive vaccine at the expense of larger recovery delay.

The presented analytical models based on a system of nonlinear differential equations and the iterative algorithm used to solve the system numerically, represent a contribution to the development of the analytical tools for the theoretical analysis of these systems. We believe that further research can benefit from these tools and result in additional development of specific applications based on the proposed multicast architecture.

6 Conclusions and future work

In this chapter, we summarize the most important contributions and results obtained in this thesis, and we point out some future research directions.

A number of new network paradigms for future MCNs were presented in this thesis. The main contributions include solutions for relaying topology control optimization, dynamic topology reconfiguration, scheduling, multihop routing protocols, intercell interference management, network coding and different proposals for multicast traffic optimization in cellular networks as well as the integration of different types of networks within the MCN. Several optimization frameworks and network models were developed to study the potentials associated with MCNs. Physical layer issues, such as a new channel model for multihop networks including channel defading and new interference management schemes were presented. Power consumption awareness was also considered through the whole thesis.

The first chapter highlights the motivation for the research and provides a comprehensive survey of the main research results and issues addressed in MCNs so far. In Chapter 2, we have introduced an optimization framework for relaying topology control which is aware of the ICI requiring coordinated action between the cells and results in multicell jointly optimal relaying topology. The algorithm jointly chooses the relaying topology and scheduling in the adjacent cells in such a way to minimize the system performance degradation due to intercell interference. The utility function includes throughput, delay and power consumption. The set of constraints in the optimization program depends on relaying specific system parameters and temporal and spatial nonuniform traffic distribution. A new topology search TSL program is developed to find the best topology in accordance with a given objective function. The overall optimization problem is solved by combining TSL and CVX program [126].

Numerical results show that as the traffic increases in the network, higher utility is obtained for topologies that favour isolated and short range transmissions. The optimum topology can provide up to 10 times higher utility than conventional TDMA and 60 times higher than CDMA for certain traffic distributions. The improvement obtained in network capacity can reach up to 6 times when the optimum topology is used. These results show that a reconfigurable relaying topology provides the network utility improvements and presents the framework for quantifying these improvements for spatially and temporally varying traffic.

This framework was further elaborated to include the optimization of the power allocation. The network performance is compared by using cooperative diversity relaying scheme (COOR) and conventional relaying scheme (CONR), resulting in two intercell interference management protocols I^2M-COOR and I^2M-CONR, respectively. By including weights in the utility function we analyze the trade-off between throughput and power allocation. Numerical results show that I^2M-COOR offers an improvement in the network throughput of at least 4 times and a reduction of power consumption of at least 3 times compared to I^2M-CONR.

These results are extended in Chapter 3 to dynamically reconfigure the relaying topology to the traffic variations in the network. As a result we have developed a specific encoding and fitness control in a Sequential Genetic Algorithm (SGA) for relaying topology update. We encoded the topologies as a set of chromosomes and new crossover and mutation operations were developed to search for the optimum topology. Improvement in the utility funtion is sequentially controled as newer generations are created, and once the improvement is sufficiently high the current topology is updated by the new one having higher fitness (utility).

The utility function used in the optimization process drives the solution towards the topology favoring simultaneously isolated and short range transmissions. Numerical results show an improvement in the utility for the optimum topology up to 1.5 times when network coding is used compared to the case without network coding. The utility function is improved with network coding by reducing the number of slots needed to complete the transmission. Similar improvement is also obtained for network capacity. In addition to optimum performance in terms of network utility, numerical results demonstrate also significant improvements in the convergence rate of the new algorithm (at least one order of magnitude faster than exhaustive search) in a dynamic network environment. We also compared the performance of SGA-TSL to the nearest neighbor heuristic, where the topology is reconfigured in such a way that the users relay to their nearest neighbor. In this case the improvement obtained with SGA-TSL was about 50% higher.

In Chapter 4, we have further extended the model from Chapter 3 to provide a new, more detailed, approach to optimization of MCNs. A nano scale network model (NSNM) is developed for high resolution optimization of multicast MCNs. In the first step we partition the macrocell into a number of subcells and adjust the radius of the subcell r to obtain different hop range which directly affects the

throughput, power consumption and interference. With r as the optimization parameter, we jointly optimize scheduling, routing and power control to obtain the optimum trade-off between throughput, delay and power consumption in multicast cellular networks. A spatial interleaving SI MAC protocol is introduced for context aware interlink interference management. The directed flooding routing protocol (DFRP) and inter flooding network coding (IFNC) are proposed for such network model.

Four different protocols are developed and analyzed for different relaying schemes and network coding. In the broadcast case for noncooperative relaying (protocol A and B) in nonfading channel, the optimum number of hops to reach the border of the cell is $H = 3$. For cooperative relaying (protocol C) and network coding (protocol D) the system performance is significantly better (3 times increase in throughput with respect to noncooperative relaying). The maximum throughput for these protocols is obtained for $H = 4$. However the largest increment of network throughput $\partial Thr / \partial H$ is obtained again for $H = 3$, suggesting this value as the best choice when it comes to trade-off between performance and complexity. In the multicast case, the best performance is obtained by protocol C or protocol D depending on the density of destination users in the network. The locations of the destination users have significant impact on the performance. A nano scale channel model (NSCM) is developed for this application which gives an insight into how the transformation of the fading process, resulting from increasing the number of hops in the cell, impacts the system performance. It was shown that the best system performance is obtained with the number of hops needed to completely transform fading channel into a non fading channel. This particular choice of r is referred to as *channel defading tessellation.* Network design procedure for channel defading is also elaborated.

In addition, the nano scale network model (NSNM) is used to develop a new concept for route discovery protocols in MCNs which is aware of the mutual impact of other routes in the network. The routing protocol resulting from this model is referred to as Nano Route Discovery Protocol (NRDP). The efficiency of NRDP is measured in terms of throughput, power consumption, terminal time to live and delay. We compared the performance of NRDP with other two routing protocols referred to as SAPR (Shortest Available Path Routing) and LAR (Load Aware Routing). Numerical results shown that NRDP provides results very close to the ideal case (all users available to relay) in terms of network capacity and throughput. Also NRDP is the most efficient in terms of traffic load, power consumption and delay. By SAPR, the users experience the shortest delay per

route but on the other hand there is no control of the traffic distribution through the network, so there is more interference between adjacent links and consequently the capacity is lower. The capacity obtained by LAR is larger than with SAPR. Although more slots are needed to complete the transmission with LAR, the gain obtained in distributing the traffic in some scenarios compensates the delay. In terms of the traffic load, LAR outperforms SAPR for as long as there are users available to split the traffic in an efficient way. Nevertheless, NDRP obtained better results in most of the scenarios.

Finally, Chapter 5 integrates a Delay Tolerant Network (DTN) into the cellular network to enhance the performance of multicast applications in MCNs and to provide support for multimedia applications. The integration between both networks is referred to as InSyNet(C,D) where the control (C) plane is designed to have full coverage over the entire cell and the data (D) plane is the conventional concept of DTN.

A new routing scheme referred to as Polymorphic Epidemic Routing (PER) was proposed for multicast DTN and new adaptive recovery schemes were developed to remove the delivered packets from the network (recovery from infection). Different recovery probabilities are used depending on the level of signaling available in the network. The performance of these new algorithms was compared to a number of unicast recovery schemes modified for multicast DTN. Numerical analysis showed the outstanding performance of our new adaptive recovery schemes when cooperative and non-cooperative destinations are used. In particular, by adaptive *immune*, *immune_TX* and *vaccine* schemes the delivery delay can be reduced up to 3 times compared to the conventional schemes. By *timeout* recovery scheme, the reduction in the delivery delay can reach up to 5 times at the expense of larger recovery delay. We believe that collection of all these improvements can provide significant progress in the development of 5G cellular systems.

The comprehensive work, analysis and results presented in this thesis open a number of new research directions for future MCNs. A very interesting extension of our models is in the area of hetereogenous networks [200] where different kind of access technologies (e.g. picocells, microcells, femtocells, WLAN, and distributed antennas) are integrated within the cellular network. The resulting architecture is referred to as Heterogenous Cellular Network (HCN).

ICI management schemes for HCN are especially critical due to the rapid changes in user demands and the different spatial densities, transmit powers, cell sizes, and backhaul capabilities of the different access points. Distributed

algorithms should be designed for each network and different level of overhead would be defined depending on the type of network. For these problems, *InSyNet*(C,D) could be extended to include different signaling priorities. Topology control, routing and scheduling algorithms presented in this thesis can be further modified for application in heterogeneous networks and the utility functions used in optimization process can be easily redefined to optimize such networks. The cost of implementing HCN will also influence the design of the new protocols.

References

1. Lin YD & Hsu YC (2000) Multihop cellular: A new architecture for wireless communications. In: Proc. IEEE INFOCOM: 1273–1282.
2. Pabst R, Walke BH, Schultz DC, Herhold P, Yanikomeroglu H, Mukherjee S, Viswanathan H, Lott M, Zirwas W, Dohler M, Aghvami H, Falconer DD, & Fettweis GP (2004) Relay-based deployment concepts for wireless and mobile broadband radio. IEEE Commun Mag 42(9): 80–89.
3. Newton M & Thompson JS (2005) Relay transmit power control in cellular multi-hop networks. In: Proc. *3G and Beyond*: 41–45.
4. Mobile ad-hoc networks (MANET) charter. IETF. URI: http://www.ietf.org/html.charters/manet-charter.html.
5. Cover TM & El Gamal AA (1979) Capacity theorems for the relay channel. IEEE Trans Inform Theory 25(5): 572–584.
6. Gupta P & Kumar P (2003) Towards an information theory of large networks: an achievable rate region. IEEE Trans Inform Theory 49(8): 1849–1877.
7. Zheng K, Fan B, Ma Z, Liu G, Shen X & Wang W (2009) Multihop cellular networks toward LTE-advanced. IEEE Veh Technol Magazine 4(3): 40–47.
8. Kannan G, Merchant SN & Desai UB (2007) Cross Layer Routing for Multihop Cellular Networks. In: Proc IEEE AINAW: 1-6.
9. Jacobson KR, Krzymien WA (2009) System design and throughput analysis for multihop relaying in cellular systems. IEEE Trans Veh Technol 58(8): 4514–4528.
10. Son K, Chong S & Veciana G (2009) Dynamic association for load balancing and interference avoidance in multi-cell networks. IEEE Trans. Wireless Commun 8(7): 3566–3576.
11. Nohara M, Saito K, Sugiyama K, Shinonaga H, Cho J, Son J, Joo P & Le H (2005) Mobile Multi-Hop Relay Networking in IEEE 802.16. Tech Rep IEEE-C802.16-05/013.
12. Nascimbene A (2006) Broadband Radio Access Networks (BRAN), HiperMAN, Physical (PHY) Layer. Tech. Rep. TS-102-177 V1.3.1.
13. Van der Meulen, EC (1971) Three-terminal communication channels. Adv. Appl. Prob. 3: 120-154.
14. Sendonaris A, Erkip E & Aazhang B (2003) User cooperation diversity-part I: System description. IEEE Trans Commun 51(11): 1927-1938.
15. Chen L & Heinzelman WB (2007) A Survey of Routing Protocols that Support QoS in Mobile Ad Hoc Networks. IEEE Network 21(6): 30–38.
16. Zhao Q & Tong L (2005) Energy efficiency of large-scale wireless networks: Proactive versus reactive networking. IEEE J Select Areas Commun 23(5): 1100–1112.
17. Kwon TJ, Gerla M, Varma V.K, Barton M & Hsing TR (2003) Efficient flooding with passive clustering—An overhead-free selective forward mechanism for ad hoc/sensor networks. Proceedings of the IEEE 91(8): 1210–1220.

18. Cheng RG, Cheng SM & Lin P (2006) Power-efficient routing (PER) mechanism for ODMA systems. IEEE Trans Veh Technol 55(4): 1311–1319.
19. Johnson D & Maltz D (1996) Dynamic source routing in ad hoc wireless networks. Mobile Computing. Norwell MA, Kluwer: 153–181.
20. Perkins CE, Belding-Royer EM & Das S (2003) Ad hoc on-demand distance vector (AODV) routing. RFC 3561.
21. Choi HH & Cho DH (2006) Fast and reliable route discovery protocol considering mobility in multihop cellular networks. In: Proc IEEE Veh Technol Conf 2: 886–890.
22. Romero-Jerez JM, Ruiz-Garcia, M & Diaz-Estrella A (2003) Interference statistics of cellular DS/CDMA systems with base station diversity under multipath fading. IEEE Trans Wireless Commun 2(6):1109-1113.
23. IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for BroadbandWireless Access Systems (2009) IEEE Std. 802.16-2009.
24. Chen CT, Tekinay S & Papavassiliou S (2003) Geocasting in cellular ad hoc augmented networks. In: Proc. IEEE Veh Technol Conf 3: 1858–1862.
25. Wu X, Chan SHG & Mukherjee B (2000) MADF: a novel approach to add an ad-hoc overlay on a fixed cellular infrastructure. In: Proc IEEE WCNC 2: 549–554.
26. Li H, Lott M, Weckerle M, Zirwas W & Schulz E (2002) Multihop communications in future mobile radio networks. In: Proc IEEE PIMRC 1: 54–58.
27. Li H, Yu D & Chen H (2003) New approach to multihop-cellular based multihop network. In: Proc IEEE PIMRC 2: 1629–1633.
28. Aggelou GN & Tafazolli R (2001) On the relaying capability of next-generation GSM cellular networks. IEEE Personal Commun 8(1): 40–47.
29. Hsu YC & Lin YD (2002) Base-centric routing protocol for multi-hop cellular networks. In: Proc. IEEE GLOBECOM 1: 158–162.
30. Lin YD & Hsu YC (2000) Multihop cellular: a new architecture for wireless communications. In: Proc. IEEE INFOCOM 1: 1273–1282.
31. Dawy Z, Davidovic S & Oikonomidis I (2003) Coverage and capacity enhancement of CDMA cellular systems via multihop transmission. In: Proc IEEE GLOBECOM 2: 1147–1151.
32. Sreng V, Yanikomeroglu H & Falconer DD (2003) Relayer selection strategies in cellular networks with peer-to-peer relaying. In: Proc IEEE Veh Technol Conf 3: 1949–1953.
33. Kusuma AANA, Andrew LLH (2002) Minimum power routing for multihop cellular networks. In: Proc IEEE GLOBECOM 1: 37–41.
34. Thodime GRV, Vokkarane VM & Jue JP (2003) Dynamic congestion-based load balanced routing in optical burst-switched networks. In: Proc IEEE GLOBECOM 5: 2628 – 2632.
35. Wu TM & Wang SL (2011) Routing selection with overloading cancellation for multihop cellular systems. IEEE Commun Letters 15(1): 61-63.
36. Viswanathan H & Mukherjee S (2005) Performance of cellular networks with relays and centralized scheduling. IEEE Trans. Wireless Commun 4(5): 2318–2328.

37. Liu W, Hu Y, Ci S & Tang H (2011) Adaptive resource allocation and scheduling for cognitive radio MIMO-OFDMA systems. In: Proc IEEE Veh Technol Conf 1: 1–6.
38. Liu Y, Hoshyar R, Yang X & Tafazolli R (2006) Integrated radio resource allocation for multihop cellular networks with fixed relay stations. IEEE J Select Areas Commun 24(11): 2137–2146.
39. Zhao D & Todd TD (2006) Cellular CDMA capacity with out-of-band multihop relaying. IEEE Trans on Mobile Comp 5(2): 170-178.
40. Cheng RG, Cheng SM & Lin P (2006) Power efficient routing mechanism for ODMA systems. IEEE Veh Technol 55(4): 1311–1319.
41. Hanly SV (1999) Congestion measures in DS-CDMA networks. IEEE Trans Commun 47(3): 426–437.
42. Rouse T, McLaughlin S & Band I (2005) Congestion-based routing strategies in multihop TDD-CDMA networks. IEEE J Select Areas Commun 23(3): 668–681.
43. Cho J & Haas ZJ (2004) On the throughput enhancement of the down-stream channel in cellular radio networks through multihop relaying. IEEE J Select Areas Commun 22(7): 1206–1219.
44. R. Bhatia, Li L, Luo H & Ramjee R (2006) ICAM: Integrated Cellular and Ad Hoc Multicast. IEEE Trans. Mobile Comp 5(8): 1004–1015.
45. Karami E, Glisic S (2011) Joint optimization of scheduling and routing in multicast wireless ad-hoc network using soft graph coloring and non-linear cubic games. IEEE Trans Veh Technol 60(7): 3350–3360.
46. Sendonaris A, Erkip E & Aazhang B (2003) User cooperation diversity-part II: Implementation aspects and performance analysis. IEEE Trans Commun 51(11): 1939–1948.
47. Jantarayat I, Uthansakul P, Uthansakul M (2010) Tradeoffs between handover performance and coverage range of relay stations in multihop cellular networks. In: Proc Int Conf Wireless Commun Signal Processing.
48. Okada H (2007) A route establishment scheme for multi-route coding in multihop cellular networks. In: Proc IEEE Veh Technol Conf 1: 6–10.
49. Zhu D, Mutka MW & Cen Z (2005) Using cooperative multiple paths to reduce file download latency in cellular data networks. In: Proc IEEE GLOBECOM 1: 2480–2484.
50. Santi P (2005) Topology control in wireless ad hoc and sensor networks. ACM Comp Survey 3(2): 164–194.
51. Tseng YC, Ni SY, Chen YS & Sheu JP (2002) The broadcast storm problem in a mobile ad hoc network. Wireless Networks 8: 153–167.
52. Burkhart M, Von Rickenbach P, Wattenhofer R & Zollinger A (2004) Does topology control reduce interference? In: Proc IEEE MobiHoc 1: 9–19.
53. Li XY, Moaveni-Nejad K, Song WZ & Wang WZ (2005) Interference-aware topology control for wireless sensor networks. In: Proc IEEE SECON 1: 263–274.
54. Wu KD & Liao W (2006) On constructing low interference topology in multihop wireless networks. In Proc IEEE ICC 8: 3759–3764.

55. Johansson T & Carr-Motyckova (2005) Reducing interference in ad hoc networks through topology control. In: Proc ACM Foundations of mobile computing 1: 17–23.
56. Marina MK, Das SR & Subramanian AP (2010) A topology control approach for utilizing multiple channels in multi-radio wireless mesh networks. Computer Networks 54(2): 241–256.
57. Subramanian AP, Gupta H, Das SR & Cao J (2008) Minimum-interference channel assignment in multi-radio wireless mesh networks. IEEE Trans. Mobile Comp 7(12): 1459–1473.
58. Duran A, Toril M, Ruiz F, Solera M & Navarro R (2010) Analysis of topology control algorithms in multi-hop cellular networks. In: Proc IB2COM 1: 1–6.
59. Guan Q, Yu FR, Jiang S & Wei G (2010) Prediction-based topology control and routing in cognitive radio networks. In: Proc IEEE INFOCOM 1: 4443–4452.
60. Guo WZ, Gao HL, Chen GL & Yu L (2009) Particle swarm optimization for the degree-constrained MST problem in WSN topology control. In: Proc. International Conf. Machine Learning and Cybernetics 3: 1793–1798.
61. Huang Z, Zhang Z, Zhu H & Ryu B (2006) Topology control for wireless ad hoc networks: a genetic algorithm-based approach. In: Proc Int Conf on Commun and Networking ChinaCom 1: 1–5.
62. Yang S, Cheng H & Wang F (2010) Genetic algorithms with immigrants and memory schemes for dynamic shortest path routing problems in mobile ad hoc networks. IEEE Trans on Systems, Man, and Cybernetics, Part C: Applications and Reviews 40(1): 52–63.
63. Halpern SW (1983) Reuse partitioning in cellular systems. In: Proc IEEE Veh Techn Conf. 33: 322–327.
64. Mobile WiMAX—part I: a technical overview and performance evaluation (2006). WiMAX Forum.
65. Li G & Liu H (2006) Downlink radio resource allocation for multi-cell OFDMA system. IEEE Trans. Wireless Commun 5(12): 3451–3459.
66. Bonald T, Borst S & Proutiere A (2005) Inter-cell scheduling in wireless data networks. In: Proc European Wireless 1: 1–6.
67. Gjendemsj A, Gesbert D, Oien GE & Kiani SG (2008) Binary power control for sum rate maximization over multiple interfering links. IEEE Trans Wireless Commun 7(8): 3164–3173.
68. Gjendemsj A, Oien GE & Gesbert D (2007) Binary power control for multi-cell capacity maximization. In: Proc IEEE SPAWC 1: 1–5.
69. Qiu X & Chawla K (1999) On the performance of adaptive modulation in cellular systems. IEEE Trans Commun 47(6): 884–895.
70. Chiang M, Hande P, Lan T & Tan C (2008) Power control in wireless cellular networks. foundations and trends in networking 2(4): 381–533.
71. Huang J, Berry RA & Honig ML (2006) Distributed interference compensation for wireless networks. IEEE J Select Areas Commun 24(5): 1074–1084.
72. Pischella M & Belfiore JC (2008) Distributed weighted sum throughput maximization in multi-cell wireless networks. In: Proc IEEE PIMRC 1: 1–5.

73. Kiani S, Oien G & Gesbert (2007) Maximizing multicell capacity using distributed power allocation and scheduling. In: Proc IEEE WCNC 1: 1690–1694.
74. Stolyar AL & Viswanathan H (2008) Self-organizing dynamic fractional frequency reuse in OFDMA systems. In: Proc IEEE INFOCOM 1: 691–699.
75. Miao G, Himayat N, Li GY, Koc AT & Talwar S (2009) Interference-aware energy-efficient power optimization. In: Proc IEEE ICC 1: 1–5.
76. Leung K & Sung CW (2006) An opportunistic power control for cellular network. IEEE/ACM Trans Networking 14(3): 470–478.
77. Hande P, Rangan A, Chiang M & Wu X (2008) Distributed uplink power control for optimal sir assignment in cellular. IEEE/ACM Trans. Networking 16(6): 1420–1433.
78. Bendlin R, Huang YF, Ivrlac M & Nossek JA (2010) Cost-Constrained transmit processing in wireless cellular networks with universal frequency reuse. In: Proc Conf Infor Sciences and Systems 1: 1–6.
79. Rashid-Farrokhi F, Liu KR & Tassiulas L (1998) Transmit beamforming and power control for cellular wireless systems. IEEE J Select Areas Commun 16(10): 1437–1450.
80. Schubert M & Boche H (2004) Solution of the multiuser downlink beamforming problem with individual SINR constraints. IEEE Trans Veh Technol 53(1): 18–28.
81. Stridh R, Bengtsson M & Ottersten B (2001) System evaluation of optimal downlink beamforming in wireless communication. In: Proc IEEE Veh Technol Conf 1: 343–347.
82. Koutsopoulos I, Ren T & Tassiulas L (2003) The impact of space division multiplexing on resource allocation: a unified approach. In: Proc. IEEE INFOCOM 1: 533–543.
83. Osseiran A, Ericson M, Barta J, Goransson B & Hagerman B (2001) Downlink capacity comparison between different smart antenna concepts in a mixed service WCDMA system. In: Proc IEEE Veh Technol Conf 3: 1528–1532.
84. Pedersen KI & Mogensen PE (2003) Performance of WCDMA HSPA in a beamforming environment under code constraints. In: Proc IEEE Veh Technol Conf 2: 995–999.
85. Pedersen KI & Mogensen PE (2003) Application and performance of downlink beamforming techniques in UMTS. IEEE Commun Mag 41(10): 134–143.
86. Milosevic N, Lorenzo B, Glisic S & Nikolic Z (2009) Intercell cooperative dynamic radio resource allocation. In: Proc Inter Symp Wireless Personal Multimedia Commun 1: 1–5.
87. Milosevic N, Lorenzo B, Nikolic B & Glisic S (2009) Opportunistic scheduling with spatial traffic shaping. In: Proc IEEE PIMRC 1: 1–5.
88. Foschini GJ, Karakayali MK & Valenzuela RA (2006) Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency. IEE Proceedings-Commun. 153(4): 548–555.
89. Shamai S & Zaidel BM (2001) Enhancing the cellular downlink capacity via co-processing at the transmitting end. In: Proc IEEE Veh Techno Conf 3: 1745–1749.

90. Weingarten H, Steinberg Y & Shamai S (2006) The capacity region of the gaussian multiple-input multiple-output broadcast channel. IEEE Trans Inf Theory 52(9): 3936–3964.
91. NGMN TE WP1 Radio Performance Evaluation Phase 2 Report (2008). URI: http://www.NGMN.org.
92. Olsson M & Skillermark P (2008) Multiple antenna techniques for downlink interference mitigation in cellular networks. In: Proc European Wireless Conference 1: 1–5.
93. Boudreau G, Panicker J, Guo N, Chang R, Wang N & Vrzic S (2009) Interference coordination and cancellation for 4g networks. IEEE Commun Mag 47(4): 74–81.
94. Patra SSM, Roy K, Banerjee S, Vidyarthi DP (2006) Improved genetic algorithm for channel allocation with channel borrowing in mobile computing. IEEE Trans Mobile Comp 5(7): 884–892.
95. Son K, Chong S, Veciana G (2009) Dynamic association for load balancing and interference avoidance in multi-cell networks. IEEE Trans. Wireless Commun 8(7): 3566–3576.
96. Sang A, Wang X, Madihian M & Gitlin R (2004) Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems. In: Proc ACM Mobicom 1: 302–314.
97. Bu T, Li L & Ramjee R (2006) Generalized proportional fair scheduling in third generation wireless data networks. In: Proc IEEE INFOCOM 1: 1–12.
98. Kelly F, Maullo A & Tan D (1998) Rate control in communication networks: shadow prices, proportional fairness and stability. J Operational Research Society 49: 237–252.
99. Son K, Chong S & Veciana G (2009) Dynamic association for load balancing and interference avoidance in multi-cell networks. IEEE Trans Wireless Commun 8(7): 3566–3576.
100. Wu X, Mukherjee B, Chan SHG (2000) MACA-An efficient channel allocation scheme in cellular networks. In: Proc GLOBECOM 3: 1385–1389.
101. Wu H, Qiao C, De S & Tonguz O (2001) Integrated Cellular and Ad Hoc Relaying Systems: iCAR. IEEE J Select Areas Commun 19(10): 2105–2115.
102. Zhou J & Yang Y (2002) PARCelS: pervasive ad-hoc relaying for cellular systems. In: Proc Med-Hoc-Net 1: 1–6.
103. TamYH, Safwat A & Hassanein HS (2005) A load balancing and relaying framework for TDD W-CDMA multi-hop cellular networks. In: Proc Inter IFIP-TC6 Networking Conf 1: 1267–1280.
104. Kolios P, Friderikos V & Papadaki K (2010) Load balancing via store-carry and forward relaying in cellular networks. In: Proc IEEE GLOBECOM 1: 1–6.
105. Fall K & Farrell S (2008) DTN: an architectural retrospective. IEEE J Select Areas Commun 26(5): 828–836.
106. Megna M (2009) AT&T Faces 5,000 Percent Surge in Traffic. URI: http://www.internetnews.com/mobility/article.php/3843001/ATT-Faces-5000-Percent-Surge-in-Traffic.htm.

107. T-Mobile's growth focusing on 3G (2009) URI: http://connectedplanetonline.com/wireless/news/t-mobile-3g-growth-0130.
108. Parkvall S, Englund E, Lundevall M & Torsner J (2006) Evolving 3G mobile systems: Broadband and broadcast services in WCDMA. IEEE Commun Mag 44(2): 30–36.
109. Kornfeld M & May G (2007) DVB-H and IP Datacast – broadcast to handheld devices. IEEE Trans Broadcasting 53(1): 161–170.
110. Law L, Pelechrinis K, Krishnamurthy S & Faloutsos M (2010) Downlink capacity of hybrid cellular ad hoc networks. IEEE/ACM Trans Networking 18(1): 243–256.
111. Lao L & Cui J (2006) Reducing multicast traffic load for cellular networks using ad hoc networks. IEEE Trans Veh Technol 55(3): 317–329.
112. Park J & Kasera S (2005) Enhancing cellular multicast performance using ad hoc networks. In: Proc IEEE WCNC 1: 2175–2181.
113. Bhatia R, Li L, Luo H & Ramjee R (2006) ICAM: Integrated cellular and ad hoc multicast. IEEE Trans Mobile Comp 5(8): 1004–1015.
114. Qin M & Zimmermann R (2010) An adaptive strategy for mobile ad hoc media streaming. IEEE Trans Multimedia 12(4): 317–329.
115. Hua S, Guo Y, Liu Y, Liu H & Panwar S (2009) SV-BCMCS: Scalable video multicast in hybrid 3G/ad-hoc networks. In: Proc GLOBECOM 1: 4662–4667.
116. Lorenzo B & Glisic S (2009) Traffic adaptive relaying topology control. IEEE Trans Wireless Commun 8(11): 5612–5620.
117. Lorenzo B & Glisic S (2008) Optimization of relaying topology in cellular multihop wireless networks. In: Proc IEEE MILCOM 1: 1–8.
118. Lorenzo B & Glisic S (2011) Multi-objective optimization for intercell interference management in advanced multihop cellular networks. In: Proc IEEE PIMRC 1: 1–6.
119. Lorenzo B & Glisic S (2012) Genetic algorithm based dynamic relay-topology control in multihop cellular networks. IEEE Trans Mobile Comp, submitted.
120. Lorenzo B & Glisic S (2009) Genetic algorithm for cognitive traffic topology control in cellular multihop wireless networks. In: Proc Inter Symp Wireless Personal Multimedia Commun 1: 1–5.
121. Lorenzo B & Glisic S (2010) Optimization of common air interface in cellular multihop wireless networks. In: Proc IEEE WCNC 1: 1–6.
122. Lorenzo B & Glisic S (2012) Context aware nano scale modeling of multicast multihop cellular network. IEEE/ACM Trans Networking. In press.
123. Lorenzo B, Zhang H & Glisic S (2008) Opportunistic differential scheduling in cellular multicast channels. In: Proc IEEE MILCOM 1: 1–7.
124. Lorenzo B & Glisic S (2010) Joint optimization of cooperative diversity and spatial reuse in multihop hybrid cellular/ad hoc networks. In: Proc MILCOM 1: 1–8.
125. Lorenzo B & Glisic S (2012) Dynamic reconfiguration of active multicast wireless network with DTN clouds. IEEE/ACM Trans Networking, submitted.
126. Grant M & Boyd S (2012) CVX: Matlab software for disciplined convex programming. URI: http://cvxr.com/cvx.
127. Palomar D & Chiang M (2006) A tutorial on decomposition methods for network utility maximization. IEEE J Select Area Commun 24(8): 1439–1450.

128. Wei DX, Jin C, Low SH & Hegde S (2006) FAST TCP: motivation, architecture, algorithms, performance. IEEE/ACM Trans Networking 14(6): 1246–1259.
129. Griffin TG, Shepherd FB & Wilfong G (2002) The stable path problem and interdomain routing. IEEE/ACM Trans Networking 10(2): 232–243.
130. Lee JW, Chiang M & Calderbank RA (2006) Utility-optimal medium access control: reverse and forward engineering. In: Proc IEEE INFOCOM 1: 1–13.
131. Han H, Shakkottai S, Hollot CV, Srikant R & Towsley D (2006) Multi-Path TCP: A Joint Congestion Control and Routing Scheme to Exploit Path Diversity in the Internet. IEEE/ACM Trans Networking 14(6): 1260–1271.
132. He J, Bresler M, Chiang M & Rexford J (2007) Towards multi-layer traffic engineering: Optimization of congestion control and routing. IEEE J Select Areas Comm 25(5): 868–880.
133. Cruz RL & Santhanam A (2003) Optimal routing, link scheduling, and power control in multihop wireless networks. In: Proc IEEE INFOCOM 1: 702–711.
134. Y. Xi and E. Yeh, Node-based distributed optimal control of wireless networks. In: Proc CISS, Mar. 2006.
135. Chen L, Low SH, Chiang M & Doyle JC (2006) Cross-Layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks. In: Proc IEEE INFOCOM 1: 1–13.
136. Eryilmaz A Srikant R (2006) Joint congestion control, routing and MAC for stability and fairness in wireless networks. IEEE J Select Areas Commun 24(8): 1514–1524.
137. Chiang M (2005) Balancing transport and physical layer in wireless multihop networks: Jointly optimal congestion control and power control. IEEE J Select Areas Commun 23(1): 104–116.
138. Chiang M & Bell J (2004) Balancing supply and demand of bandwidth in wireless cellular networks: Utility maximization over powers and rates. In: Proc IEEE INFOCOM 4: 2800–2811.
139. Chiang M, Low S, Calderbank R & Doyle J (2007) Layering as optimisation Decomposition: A Mathematical Theory of Network Architectures. Proceedings of the IEEE 95(1): 255–312.
140. Lorenzo B, Suliman I, Glisic S & Martinez Varela A (2008) On the application of network utility function decomposition techniques in wireless networks with mobility and channel fading. In: Proc Int Symp Wireless Personal Multimedia Commun 1: 1–5.
141. Glisic S & Lorenzo B (2009) Advanced wireless Networks: 4G cognitive opportunistic and cooperative technology. 2 ed. Jonh Wiley and Sons.
142. Johansson M, Xiao L & Boyd S (2003) Simultaneous routing and power allocation in CDMA wireless data networks. In: Proc IEEE ICC 1: 51–55.
143. Boyd SP & Vandenberghe L (2003) Convex Optimization. Cambridge University Press. URI: http://www.stanford.edu/˜boyd.
144. Papandriopoulos J, Dey S, Evans J (2008) Optimal and distributed protocols for crosslayer design of physical and transport layers in MANETs. IEEE/ACM Trans Networking 16(6): 1392–1405.
145. Bertsekas DP (1995) Nonlinear Programming. Athena Scientific, Massachusetts, USA.

146. Davis L (1991) Handbook of genetic algorithms. New York, Van Nostrand Reinhold.
147. Popovski P & Yomo H (2007) Wireless network coding by amplify-and-forward for bi-directional traffic flows. IEEE Communications Letters 11(1): 16New York18.
148. Holland JH (1975) Adaptation in natural and artificial systems. Cambridge MA, MIT Press.
149. Michalewicz Z (1992) Genetic Algorithms + Data Structures = Evolution Programs. London, Springer-Verlag.
150. Grefenstette JJ (1992) Genetic algorithms for changing environments. In: Proc Int Conf Parallel Problem Solving Nature 1:137–144.
151. Martins FVC, Carrano EG, Wanner EF, Takahashi RHC, Mateus GR (2011) A hybrid multiobjective evolutionary approach for improving the performance of wireless sensor networks. IEEE J Sensors 11(3): 545–554.
152. Mohammed L, Khanbary O & Vidyarthi D (2008) A GA-based effective fault-tolerant model for channel allocation in mobile computing. IEEE Trans Veh Technol 57(3): 1823–1833.
153. Hu X, Leeson M & Hines E (2009) Dynamic network coding problem: an evolutionary approach. In: Proc Int Conf. Wireless Commun, Networking and Mobile Comp 1: 1–8.
154. Yang S & Yao X (2008) Population-based incremental learning with associative memory for dynamic environments. IEEE Trans Evolutionary Comp 12(5): 542–561.
155. Blough DM, Leoncini M, Resta G & Santi P (2006) The k-neighbors approach to interference bounded and symmetric topology control in ad hoc networks. IEEE Trans Mobile Comp 5(9): 1267–1282.
156. Wang F, Thai MT, Li Y, Cheng X & Du D-Z (2008) Fault tolerant topology control for all-to-one and one-to-all communication in wireless networks. IEEE Trans Mobile Comp 7(3): 322–331.
157. Wu J & Dai F (2006) Mobility-sensitive topology control in mobile ad hoc networks. IEEE Trans Parallel and Distributed Systems 17(6): 522–535.
158. Bonabeau E, Dorigo M & Theraulaz G (1999) Swarm intelligence: from natural to artificial systems, New York, Oxford University Press.
159. Tam YH, Benkoczi R, Hassanein HS & Akl SG (2010) Channel assignment for multihop cellular networks: minimum delay. IEEE Trans on Mobile Comp 9(7): 1022–1034.
160. Yang S (2008) Genetic algorithms with memory- and elitism-based immigrants in dynamic environments. Evol Comput 16(3): 385–416.
161. Gupta P & Kumar PR (2000) The capacity of wireless networks. IEEE Trans Inform Theory 46(2): 388–404.
162. Xie LL & Kumar PR (2004) A network information theory for wireless communication: scaling laws and optimal operation. IEEE Transa Inform Theory 50(5): 748–767.
163. Chaporkar P & Sarkar S (2005) Wireless multicast: theory and approaches. IEEE Trans Inform Theory 51(6): 1954–1972.

164. Gopala PK & El Gamal H (2005) On the throughput-delay tradeoff in cellular multicast. In: Proc. Int Conf Wireless Commun, Networking and Mobile Comp 2: 1401–1406.
165. Low TP, Pun M, Hong YWP & Kuo CCJ (2010) Optimized opportunistic multicast schedulling (OMS) over wireless cellular networks. IEEE Trans Wireless Commun 9(2): 791–801.
166. Park JC & Kasera SK (2005) Enhancing cellular multicast performance using ad hoc networks. In: Proc IEEE WCNC 4: 2175–2181.
167. Lao L & Cui JH (2006) Reducing multicast traffic load for cellular networks using ad hoc networks. IEEE Trans Veh Technol 55(3): 822–830.
168. Wan PJ, Calinescu G & Yi CW (2004) Minimum-power multicast routing in static ad hoc wireless networks. IEEE/ACM Trans Networking 12(3): 507–514.
169. Xi Y & Yeh EM (2010) Distributed algorithms for minimum cost multicast with network coding. IEEE/ACM Trans Networking 18(2): 379–392.
170. Junhai L, Danxia Y, Liu X & Mingyu F (2009) A survey of multicast routing protocols for mobile Ad-Hoc networks. IEEE Commun Surveys & Tutorials 11(1): 78–91.
171. Kumar SMD & Kumar BPV (2008) Energy-aware multicast routing in MANETs based on genetic algorithms. In: Proc. IEEE Int Conf Networks 1: 1–5.
172. Glisic S (2007) Advanced Wireless Communications, Cognitive and Cooperative Broadband Technology. 2 ed. Chichester, John Wiley & Sons.
173. Khreishah A, Wang CC & Shroff NB (2010) Rate control with pairwise intersession network coding. IEEE/ACM Trans Networking 18(3): 816–829.
174. Ahlswede R, Cai N, Li SR & Yeung RW (2000) Network Information Flow. IEEE Trans Inform Theory 46(4): 1204–1216.
175. Lee WCY (1989) Mobile cellular telecommunications: analog and digital systems. New York, McGraw-Hill.
176. Wu Y, Niu Z (2009) Exploting cooperative diversity and spatial reuse in multihop cellular networks. In: Proc IEEE ICC 1: 1–5.
177. Gopala PK & El Gamal H (2005) On the throughput-delay tradeoff in cellular multicast. In: Proc Int Conf Wireless Commun, Networking and Mobile Comp 1: 1401–1406.
178. Haas ZJ & Small T (2006) A new networking model for biological applications of ad hoc sensor networks. IEEE/ACM Trans Networking 14(1): 27–40.
179. Khabbaz M, Assi C & Fawaz W (2011) Disruption-Tolerant Networking: a comprehensive survey on recent developments and persisting challenges. IEEE Commun Surveys & Tutorials 14(2): 607–640.
180. Lin Y, Li B & Liang B (2008) Stochastic analysis of network coding in epidemic routing. IEEE J Select Areas Commun 26(5): 794–808.
181. Narmawala Z & Srivastava S (2009) MIDTONE: Multicast in delay tolerant networks. In: Proc Int Conf Commun and Networking 1: 1–8.
182. Zhang X, Neglia G, Kurose J & Towsley D (2007) Performance modeling of epidemic routing. Comp Networks 51(10): 2867–2891.

183. Martonosi M(2006) Embedded systems in the wild: Zebranet software, hardware, and deployment experiences. In: Proc ACM SIGPLAN Not 41(7): 1.
184. Zhao J & Cao G (2008) VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. IEEE Trans Veh Technol 57(3): 1910–1922.
185. Gao W, Li Q, Zhao B & Cao G (2012) Social-Aware Multicast in Disruption-Tolerant Networks. IEEE/ACM Trans Networking (99): 1.
186. Krishnan R, et al. (2007) The SPINDLE disruption-tolerant networking system. In: Proc IEEE MILCOM 1: 1–7.
187. Burleigh S, Hooke A, Torgerson L, Fall K, Cerf V, Durst B, Scott K & Weiss H (2003) Delay-tolerant networking: An approach to interplanetary Internet. IEEE Commun Mag 41(6): 128–136.
188. Ioannidis S, Chaintreau A & Massoulie L (2009) Optimal and scalable distribution of content updates over a mobile social network. In: Proc IEEE INFOCOM 1: 1422–1430.
189. Chaintreau A, Le Boudec J & Ristanovic N (2009) The age of gossip: Spatial mean field regime. In: Proc ACM SIGMETRICS 1: 109–120.
190. Hu L, Boudec J & Vojnovic M (2010) Optimal channel choice for collaborative ad-hoc dissemination. In: Proc IEEE INFOCOM 1: 1–9.
191. Fall K & Farrell S (2008) DTN: An architectural retrospective. IEEE J Select Areas Commun 26(5): 828–836.
192. Zhao W, Ammar M & Zegura E (2005) Multicasting in delay tolerant networks: semantic models and routing algorithms. In: Proc ACM SIGCOMM 1: 268–275.
193. Lee U, Oh S, Lee K & Gerla M (2008) Scalable multicast routing in delay tolerant networks. In: Proc IEEE ICNP 1: 218–227.
194. Gao W, Li Q, Zhao B & Cao G (2009) Multicasting in delay tolerant networks: A social network perspective. In: Proc ACM MobiHoc 1: 299–308.
195. Vahdat A, Becker D (2000) Epidemic Routing for Partially Connected Ad Hoc Networks. Duke Technical Report.
196. Lin Y, Li B & Liang B (2008) Stochastic analysis of network coding in epidemic routing. IEEE J Select Areas Commun 26(5): 794–808.
197. Zhu H, Fu L, Xue G, Zhu Y, Li M & Ni LM (2010) Recognizing exponential inter-contact time in VANETs. In: Proc IEEE INFOCOM 1: 101–105.
198. Cai H & Eun DY (2007) Crossing over the bounded domain: from exponential to power-law inter-meeting time in MANETs. In: Proc ACM MOBICOM 1: 159–170.
199. Karagiannis T, Le Boudec JY & Vojnovic M (2007) Power law and exponential decay of inter contact times between mobile devices. In: Proc ACM MOBICOM 1: 183–194.
200. Ren W, Zhao Q, Swami A (2011) Connectivity of heterogeneous wireless networks. IEEE Trans Inf Theory 57(7): 4315–4332.

Appendix A

In this appendix we elaborate the computation of the user location in polar coordinates (h,θ) for the context aware nano scale network model presented in Chapter 4.

For the first hop $h = 1$, the set of angles $\Theta^{(1)} = \{\theta_n^{(1)}\}$ is $\theta_1^{(1)} = 30^\circ$; $\theta_n^{(1)} = \theta_{n-1}^{(1)} + \hat{\theta}_1 = \theta_{n-1}^{(1)} + 60^\circ$, $n = 2,...,n_h$, where $\theta_1^{(1)}$ is the first angle of the set. As we can see from Fig. 31 in the first ring of users $h = 1$, the first user is located in 30º with respect to BS, and the separation between users in the first hop is $\hat{\theta}_1$ = 60º/1 = 60º. The set of angles $\Theta^{(h)}$ from $h = 2$ to H is calculated following the same reasoning as

$$\begin{aligned}
&\theta_1^{(2)} = 0^\circ;\ \ \theta_n^{(2)} = \theta_{n-1}^{(2)} + 30^\circ,\ \ n = 2,...,n_h \\
&\theta_1^{(3)} = 10^\circ;\ \theta_n^{(3)} = \theta_{n-1}^{(3)} + 20^\circ;\ \theta_1^{(4)} = 0^\circ;\ \ \theta_n^{(4)} = \theta_{n-1}^{(4)} + 15^\circ \\
&\theta_1^{(5)} = 6^\circ;\ \ \ \theta_n^{(5)} = \theta_{n-1}^{(5)} + 12^\circ,\ ... \\
&\theta_n^{(h)} = \theta_{n-1}^{(h)} + \hat{\theta}_h = \theta_{n-1}^{(h)} + 60^\circ / h,\quad n = 2,...,n_h \\
&\theta_1^{(h)} = \begin{cases} 30^\circ / h, & if\ \ h = 2p+1,\ p = 0,1,...,\left\lceil \dfrac{H-1}{2} \right\rceil. \\ 0^\circ, & otherwise \end{cases}
\end{aligned} \tag{A1}$$

By inspection of Fig. 31, we can see that users situated in different rings h can have the same angle θ. For example there are users situated in h = 1,2, ..., H and θ = 30º. To properly dimension the spatial distribution matrix **β** and other parameters defined through Chapter 4, we need to know the number of new angles θ in each hop. By observing the geometry of the scenario in Fig. 31 we can obtain the following relations between the set of angles

$$\begin{aligned}
&\Theta^{(1)} \subset \Theta^{(m)},\quad m = 1,2,3,...,H \\
&\Theta^{(2)} \subset \Theta^{(2m)},\ m = 1,2,3,...,\lfloor H/2 \rfloor \\
&\Theta^{(4)} \subset \Theta^{(4m)},\ m = 1,2,3,...,\lfloor H/4 \rfloor \\
&\Theta^{(6)} \subset \Theta^{(6m)},\ m = 1,2,3,...,\lfloor H/6 \rfloor,\ \ ... \\
&\Theta^{(h)} \subset \Theta^{(hm)},\ m = 1,2,3,...,\lfloor H/h \rfloor,\ h = 1,2,4,6,...H.
\end{aligned} \tag{A2}$$

The set $\Theta^{(1)}$ corresponding to $h = 1$ is included in any hop m = 1,…,H, the set $\Theta^{(2)}$ corresponding to $h = 2$ is included in any even hop, and so on.
Consequently, the set of new angles $\tilde{\Theta}^{(h)}$ in each hop h can be calculated as

$$\left|\tilde{\Theta}^{(1)}\right| = n_1 = 6 \cdot 1 = 6; \qquad \left|\tilde{\Theta}^{(5)}\right| = n_5 - \left|\tilde{\Theta}^{(1)}\right| = 30 - 6 = 24;$$

$$\left|\tilde{\Theta}^{(2)}\right| = n_2 - \left|\tilde{\Theta}^{(1)}\right| = 12 - 6 = 6; \qquad \left|\tilde{\Theta}^{(6)}\right| = n_6 - \left|\tilde{\Theta}^{(2)}\right| = 36 - 12 = 24;$$

$$\left|\tilde{\Theta}^{(3)}\right| = n_3 - \left|\tilde{\Theta}^{(1)}\right| = 18 - 6 = 12; \qquad \left|\tilde{\Theta}^{(7)}\right| = n_7 - \left|\tilde{\Theta}^{(1)}\right| = 42 - 6 = 36;$$

$$\left|\tilde{\Theta}^{(4)}\right| = n_4 - \left|\tilde{\Theta}^{(2)}\right| = 24 - 12 = 12; \qquad \left|\tilde{\Theta}^{(8)}\right| = n_8 - \left|\tilde{\Theta}^{(4)}\right| = 48 - 24 = 24, ...$$

And the accumulated number of new angles until *h*

$$\bar{\theta}_h = \sum_{p=1}^{h} \left|\tilde{\Theta}^{(p)}\right|. \qquad \text{(A3)}$$

Appendix B

By using the notation presented in Section 4.2.3, the network topology schedules for protocols A, B and C presented in Fig. 32, 33 and 34 respectively, can be written as:

A) Initialization:

$H = 4,\ \ \bar{\theta}_H = \sum_{h=1}^{4} \left|\tilde{\Theta}^{(h)}\right| = 36$ where $\bar{\theta}_H$ is the number of angular positions of the transmitters and receivers.

We start the transmission in slot $b = 1$ from the BS $m(0,0)$ to the first ring of users situated in $h' = 1$. In this case we have entries 1 for the transmitter position ($h = 0, \Theta^{(0)}$) and receiver positions ($h' = 1, \Theta^{(1)}$). For the rest of the positions the corresponding entries in the topology matrix are zero.

$$\mathbf{T}^{1,A} = \begin{array}{c|c|c|} h \backslash \Theta^{(0)} 0^\circ & \Theta^{(1)} \ 30^\circ\ 90^\circ\ 150^\circ\ 210^\circ\ 270^\circ\ 330^\circ \\ 0 & 1 \ \ 0\ 0\ 0\ 0\ 0\ 0 \\ 1 & 0 \ \ 1\ 1\ 1\ 1\ 1\ 1 \\ 2 & 0 \ \ 0\ 0\ 0\ 0\ 0\ 0 \\ 3 & 0 \ \ 0\ 0\ 0\ 0\ 0\ 0 \\ 4 & 0 \ \ 0\ 0\ 0\ 0\ 0\ 0 \end{array} = \left\| \mathbf{T}^1_{tx}(h, \Theta^{(h)}) \ \middle|\ \mathbf{T}^1_{rx}(h', \Theta^{(h')}) \right\|;$$

$$h = 0, \Theta^{(0)} = \{0^\circ\}; \quad h' = 1, \Theta^{(1)} = \{30^\circ, 90^\circ, 150^\circ, 210^\circ, 270^\circ, 330^\circ\}$$

The topology matrix in slot $b = 1$ is the same for protocols A, B and C, $\mathbf{T}^1_B = \mathbf{T}^1_C = \mathbf{T}^1_A$.

B) Protocol A represented in Fig. 32 for the following slots b.

The topology matrix in $b = 2$ with (h=1, h'=2) is:

$$T^{2,A}_{tx}(1, \Theta^{(1)}) = 1; \quad T^{2,A}_{rx}(2, \Theta^{(2)}) = 1$$

$$h = 1,\ \theta^{(1)}_1 = 90^\circ;\ \theta^{(1)}_n = \theta^{(1)}_{n-1} + \hat{\phi}_1 = \theta^{(h)}_{n-1} + 120^\circ \text{ for } n = 2, ..., n_h / 2;$$

$$\Theta^{(1)} = \{\theta^{(1)}_n\} = \{90^\circ, 210^\circ, 330^\circ\};$$

$$h' = 2,\ \theta^{(2)}_n = eq.(42);\ \alpha_{h,\theta} = 0, \pm 1;$$

$$\Theta^{(2)} = \{\theta^{(2)}_n\} = \{0^\circ, 60^\circ, 90^\circ, 120^\circ, 180^\circ, 210^\circ, 240, 300^\circ, 330^\circ\}$$

$$
\begin{array}{c c c|c}
 & h\backslash\Theta^{(1)} & 90^\circ\ 210^\circ\ 330^\circ & \Theta^{(2)}\ 0^\circ\ 60^\circ\ 90^\circ\ 120^\circ\ 180^\circ\ 210^\circ\ 240\ 300^\circ\ 330^\circ \\
 & 0 & 0\quad 0\quad 0 & 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0 \\
 & 1 & 1\quad 1\quad 1 & 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0 \\
\mathbf{T}^{2,A} = & 2 & 0\quad 0\quad 0 & 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1\quad 1 \\
 & 3 & 0\quad 0\quad 0 & 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0 \\
 & 4 & 0\quad 0\quad 0 & 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0
\end{array}
$$

$$
= \left\| \mathbf{T}_{tx}^{2}(h,\Theta^{(h)}) \;\middle|\; \mathbf{T}_{rx}^{2}(h',\Theta^{(h')}) \right\|
$$

For b = 3,

$$
\mathbf{T}_{tx}^{3,A}(1,:) = \boldsymbol{\beta}(1,:) - \mathbf{T}_{tx}^{2,A}(1,:);\;\; \mathbf{T}_{rx}^{3,A}(2,:) = \boldsymbol{\beta}(2,:) - \mathbf{T}_{rx}^{2,A}(2,:)
$$

where $\boldsymbol{\beta}(1,:)$ represent the location of all users in the first hop h = 1, and $\boldsymbol{\beta}(2,:)$ in h = 2.

In slot b = 4, with the transmitters situated in h = 2, and receivers h' = 3, the topology matrix is calculated in the same fashion as

$$
T_{tx}^{4,A}(2,\Theta^{(2)}) = 1;\quad \mathbf{T}_{rx}^{4,A}(3,\Theta^{(3)}) = \boldsymbol{\beta}(3,:)
$$

$$
\theta_1^{(2)} = 30^\circ;\; \theta_n^{(2)} = \theta_{n-1}^{(2)} + \hat{\phi}_3 = \theta_{n-1}^{(2)} + 120^\circ/3 \;\text{ for }\; n = 2,\ldots,n_h/2;
$$

$$
\Theta^{(2)} = \{\theta_n^{(2)}\};\; \theta_n^{(3)} = eq.(42);\; \alpha_{h,\theta} = 0,\pm 1;\; \Theta^{(3)} = \{\theta_n^{(3)}\}
$$

In δ = 5, h = 3 and h' = 4

$$
T_{tx}^{5,A}(3,\Theta^{(3)}) = 1;\quad \mathbf{T}_{rx}^{5,A}(4,\Theta^{(4)}) = \boldsymbol{\beta}(4,:)
$$

$$
\theta_1^{(3)} = 30^\circ;\; \theta_n^{(3)} = \theta_{n-1}^{(3)} + \hat{\phi}_h = \theta_{n-1}^{(3)} + 120^\circ/4 \;\text{ for }\; n = 2,\ldots,n_4/2;
$$

$$
\Theta^{(3)} = \{\theta_n^{(3)}\};\; \theta_n^{(4)} = eq.(42);\; \alpha_{h,\theta} = 0,\pm 1;\; \Theta^{(4)} = \{\theta_n^{(4)}\};
$$

The location of the receivers during the schedule cycle B^A should be such that $\sum_{b=1}^{B^A} \mathbf{T}_{rx}^{b,A} = \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the spatial user distribution matrix. In the broadcast case $\mathbf{D} = \boldsymbol{\beta}$.

In Fig. 32 we are assuming that in each slot b, all required transmitters are available, $\boldsymbol{\Psi}_{h,\theta} A_{h,\theta} = 1,\ \forall h,\theta$.

C) Protocol B represented in Fig. 33.

The network topologies needed to generate protocol B can be calculated by rotating the corresponding topologies (**T**) from protocol A by an angle ω. This

operation will be denoted as *rotate* [**T**, ω], resulting into $\mathbf{T}^{b,B} = rotate\left[\mathbf{T}^{b_A,A}, \omega\right]$. So, we obtain

$$\mathbf{T}^{1,B} = \mathbf{T}^{1,A},\ \mathbf{T}^{2,B} = rotate[\mathbf{T}^{2,A}, 60^\circ],\ \mathbf{T}^{3,B} = rotate[\mathbf{T}^{3,A}, 60^\circ],$$
$$\mathbf{T}_{tx}^{4,B} = rotate[\mathbf{G}_{tx}^{4,A}, 30^\circ],$$
$$\mathbf{T}_{rx}^{4,B}(3,:) = \boldsymbol{\beta}(3,:) - \boldsymbol{\beta}(3,\{30^\circ, 90^\circ, 150^\circ, 210^\circ, 270^\circ, 330^\circ\})$$
$$\mathbf{T}_{tx}^{5,B}(2,:) = \boldsymbol{\beta}(2,:) - \mathbf{T}_{tx}^{4,B}(2,:),\ \mathbf{T}_{rx}^{5,B}(3,:) = \boldsymbol{\beta}(3,:) - \mathbf{T}_{rx}^{4,B}(3,:),$$
$$\mathbf{T}^{6,B} = rotate[\mathbf{T}^{5,A}, 20^\circ], \quad \sum_{b=1}^{B^B} \mathbf{T}_{rx}^{b,B} = \boldsymbol{\beta}.$$

D) Protocol C: Cooperative Diversity Receiver

By using the topologies already defined in protocol A and B, the topology matrix for protocol C can be easily obtained as

$$\mathbf{T}^{1,C} = \mathbf{T}^{1,A},\ \ \mathbf{T}^{2,C} = \mathbf{T}^{2,A},\ \ \mathbf{T}_{tx}^{3,C} = \mathbf{T}_{tx}^{4,B},$$
$$\mathbf{T}_{rx}^{3,C}(3,:) = \mathbf{T}_{rx}^{4,B}(3,:),\ \mathbf{T}_{rx}^{3,C}(2,:) = \boldsymbol{\beta}(2,:) - \mathbf{T}_{rx}^{2,C}(2,:)$$
$$\mathbf{T}_{tx}^{4,C}(3,:) = \mathbf{T}_{rx}^{3,C}(3,:),\ \ \mathbf{T}_{rx}^{4,C}(4,:) = \boldsymbol{\beta}(4,:),$$
$$\mathbf{T}_{rx}^{4,C}(3,:) = \boldsymbol{\beta}(3,:) - \mathbf{T}_{tx}^{4,C}(3,:), \quad \sum_{b=1}^{B^C} \mathbf{T}_{rx}^{b,C} = \boldsymbol{\beta}.$$

Appendix C

The topology matrix $\mathbf{T}^b$ for protocol D consists on the following submatrices,

$T^1_{tx,y_a}(0,\Theta^{(0)}_{y_a})=1, T^1_{tx,y_b}(0,\Theta^{(0)}_{y_b})=1, \mathbf{T}^1_{tx,y_c}=\mathbf{0}$

with $\Theta^{(0)}_{y_a}=\Theta^{(0)}_{y_b}=\{0^\circ\}$

$T^1_{rx,y_a}(1,\Theta^{(1)}_{y_a})=1, T^1_{rx,y_b}(1,\Theta^{(1)}_{y_b})=1,\ \mathbf{T}^1_{rx,y_c}=\mathbf{0}$

with $\Theta^{(1)}_{y_a}=\{90^\circ,210^\circ,330^\circ\}; \Theta^{(1)}_{y_b}=\{30^\circ,150^\circ,270^\circ\}$

$\mathbf{T}^2_{tx,y_a}=\mathbf{T}^{2,A}_{tx},\ \mathbf{T}^2_{tx,y_b}=\mathbf{0},\ \mathbf{T}^2_{tx,y_c}=\mathbf{0};$

$\mathbf{T}^2_{rx,y_a}(2,:)=\mathbf{T}^{2,A}_{rx}(2,:), \mathbf{T}^2_{rx,y_a}(1,:)=\boldsymbol{\beta}(1,:)-\mathbf{T}^2_{tx,y_a}(1,:), \mathbf{T}^2_{rx,y_b}=\mathbf{0},\ \mathbf{T}^2_{rx,y_c}=\mathbf{0}$

$\mathbf{T}^3_{tx,y_a}=\mathbf{0},\ \mathbf{T}^3_{tx,y_b}=\mathbf{T}^{2,B}_{tx},\ \mathbf{T}^3_{tx,y_c}=\mathbf{0};$

$\mathbf{T}^3_{rx,y_a}=\mathbf{0},\ \mathbf{T}^3_{rx,y_b}(2,:)=\mathbf{T}^{2,B}_{rx}(2,:), \mathbf{T}^3_{rx,y_b}(1,:)=\boldsymbol{\beta}(1,:)-\mathbf{T}^3_{tx,y_b}(1,:), \mathbf{T}^3_{rx,y_c}=\mathbf{0}$

$\mathbf{T}^4_{tx,y_a}=\mathbf{0},\ \mathbf{T}^4_{tx,y_b}=\mathbf{0},\ \mathbf{T}^4_{tx,y_c}=\mathbf{T}^{4,B}_{tx};$

$\mathbf{T}^4_{rx,y_a}(2,:)=\boldsymbol{\beta}(2,:)-\mathbf{T}^2_{rx,y_a}(2,:),\ \mathbf{T}^4_{rx,y_b}(2,:)=\boldsymbol{\beta}(2,:)-\mathbf{T}^3_{rx,y_b}(2,:),\ \mathbf{T}^4_{rx,y_c}=\mathbf{T}^{4,B}_{rx}$

$\mathbf{T}^5_{tx,y_a}=\mathbf{T}^{5,B}_{tx},\ \mathbf{T}^5_{tx,y_b}=\mathbf{0},\ \mathbf{T}^5_{tx,y_c}=\mathbf{0};$

$\mathbf{T}^5_{rx,y_a}(3,:)=\boldsymbol{\beta}(3,:),\ \mathbf{T}^5_{rx,y_b}=\mathbf{T}^{4,B}_{rx},\ \mathbf{T}^5_{rx,y_c}=\mathbf{0}$

$\mathbf{T}^6_{tx,y_a}=\mathbf{0},\ \mathbf{T}^6_{tx,y_b}=\mathbf{0},\ \mathbf{T}^6_{tx,y_c}=\mathbf{T}^{4,B}_{rx};$

$\mathbf{T}^6_{rx,y_a}=\mathbf{0},\ \mathbf{T}^6_{rx,y_b}(3,:)=\boldsymbol{\beta}(3,:)-\mathbf{T}^6_{tx,y_c}(3,:),\ \mathbf{T}^6_{rx,y_c}(4,:)=\boldsymbol{\beta}(4,:)$

$\mathbf{T}^7_{tx,y_a}=\mathbf{T}^{6,B}_{rx},\ \mathbf{T}^7_{tx,y_b}=\mathbf{0},\ \mathbf{T}^7_{tx,y_c}=\mathbf{0};$

$\mathbf{T}^7_{rx,y_a}(4,:)=\boldsymbol{\beta}(4,:),\ \mathbf{T}^7_{rx,y_b}(4,:)=\boldsymbol{\beta}(4,:),\ \mathbf{T}^7_{rx,y_c}=\mathbf{0}.$

Appendix D

For an iterative solution of (89)-(90), we rewrite (approximate) the differential equations $F(\mathbf{y'},\mathbf{y})$, $\mathbf{y} = (A,B,C,F)$, $X \rightarrow A,B \text{ or } C$ in difference form as $F(\Delta\mathbf{y},\mathbf{y})$:

$$\Delta X = \lambda(X + \frac{1}{3}F)(N - I) - \lambda X(I - X)$$
$$\Delta F = \lambda A(B + C) + \lambda B(A + C) + \lambda C(A + B) + \lambda F\left(X - F\right)$$

or

$$F_i(\Delta\mathbf{y}_k, \mathbf{y}_k) = F_i(\mathbf{y}_k - \mathbf{y}_{k-1}, \mathbf{y}_k)$$

$$X_k - X_{k-1} = \lambda(X_k + \frac{1}{3}F_k)(N - I_k) - \lambda X_k(I_k - X_k) \qquad \text{(D1)}$$

$$F_k - F_{k-1} = \lambda A_k(B_k + C_k) + \lambda B_k(A_k + C_k) + \lambda C_k(A_k + B_k) + \lambda F_k\left(X_k - F_k\right) \qquad \text{(D2)}$$

with a given initial value of $\mathbf{y}_0$ which depends on the initialization of the protocol. In this way the system of nonlinear differential equations is turned into an iterative process where for each $\mathbf{y}_{i-1}$, calculated in the previous iteration, a system of nonlinear equations $F_i(\Delta\mathbf{y}_i, \mathbf{y}_i) = F_i(\mathbf{y}_i - \mathbf{y}_{i-1}, \mathbf{y}_i)$ has to be solved in order to find the new vector $\mathbf{y}_i$. This can be formulized as *DiNSE algorithm*

1. *Initialized* $\mathbf{y}_0 = (1,1,1,0)$.
2. *Solve* $F_i(\Delta\mathbf{y}_i, \mathbf{y}_i) = F_i(\mathbf{y}_i - \mathbf{y}_{i-1}, \mathbf{y}_i)$
3. *if* $\frac{|\mathbf{y}_i - \mathbf{y}_{i-1}|}{|\mathbf{y}_i + \mathbf{y}_{i-1}|} \geq \varepsilon$ *then* $\mathbf{y}_{i-1} = \mathbf{y}_i$ *and go to 2.*

Appendix E

The equivalent representation of (89)-(90) without network coding is defined as

$$X' = \lambda(X + \frac{1}{2}F)(N - I) - \lambda X(I - X) \quad \text{(E1)}$$

$$F' = \lambda AB + \lambda BA + \lambda F\left(X - F\right), \quad \text{(E2)}$$

where $X \in \{A, B\}$ and $I = A + B + F$.

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-9854-7 (Paperback)
ISBN 978-951-42-9855-4 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO