"I-MAC": AN ICN BASED RADIO ACCESS NETWORK ARCHITECTURE

by

VISHAKHA RAMANI

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Dipankar Raychaudhuri

And approved by

____

____

____

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE THESIS

"I-MAC": An ICN based Radio Access Network Architecture

By VISHAKHA RAMANI

Thesis Director:
Dipankar Raychaudhuri

The rapidly increasing connectivity demands of both fixed and mobile devices on the internet have motivated a clean-slate redesign of existing core and radio access networks. The advent of content-oriented applications such as social media applications, on-demand video streaming services, interactive gaming applications, etc. has exposed the limitations in the existing host-centric Internet Protocol (IP) based internet architecture. Information-Centric Networking, a clean-slate future internet architecture, has been extensively researched to show its effectiveness in its handling of content-centric applications in the core-network. This thesis aims to propose ICN based radio access network architecture, "I-MAC", which enables the integration of ICN identifiers and semantics with radio access network in order to achieve efficiency gains in capacity-limited wireless networks which are used by an increasing proportion of Internet traffic. Additionally, the use of a Local Name Resolution Table (LNRS), maintained by each base station in LTE, called as Evolved Node B (eNodeB), to seamlessly map the device GUID (globally unique identifier) with the corresponding radio network identifier allocated during each attach process is also suggested. This design is further extended to specifically support push and pull-based multicast transmissions at the last hop by a novel control signaling which incorporates physical control and data channels of existing cellular architecture. The verification of this design is done via a special pull-based multicast use case which takes into account the characteristics of highway-tunnel topography, traffic conditions, and user behavior. Through extensive simulations using NS-3 coupled with SUMO (Simulation for Urban MObility), it is shown that significant multicast gain is achieved at the eNodeB and considerable amount of bandwidth is saved. The simulation results of this design show that the temporal correlation among many delay tolerant user requests induces about 45% aggregation of user requests, costing the user a delay of merely a second. The results also show the suitability of this design for massive IoT applications where the device uptime is reduced by a factor of twenty from the current cellular multicast architecture, SC-PTM, thus providing considerable device power savings. Analytical measurements show spectral efficiency of this work with a control data occupancy in data channel of 0.07% in contrast with that of SC-PTM which has 1.32%, thereby, ensuring that user data gets a significant share of the bandwidth. In addition, parametric evaluations to check the sensitivity of the design to various parameters introduced in the model have also been performed.

Acknowledgements

First and foremost, I would like to thank my advisor Professor Dipankar Raychaudhuri for his constant guidance, support and kind understanding when the going was tough. He constantly looked at my work in multiple ways and pushed me towards unknown territories in my research that further opened up my mind. The door to his office was always open whenever I ran into trouble, both in my professional and personal life. Without his persistent help, this thesis would not have been possible.

I would also like to thank my committee members Professor Roy Yates and Professor Predrag Spasojevic for the valuable insights on the project and in the courses I took under them. They answered even the simplest of my questions with utmost sincerity and also providing me with many "light-bulb" moments in their classes.

I am immensely grateful for my labmates who conistently helped me in both academic and personal life, especially Parishad for her constant mentoring and motivation. I wouldn't have enjoyed my time in WINLAB if not for these people. Finally, I would like to express my eternal gratitude to my family for their unconditional love, care and support.

Dedication

In loving memory of my grandmother...

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

According to the recent Cisco forecast update, overall mobile data traffic is expected to grow to 77 exabytes per month by 2022, a seven-fold increase over 2017 [2]. The rapidly growing consumption of bandwidth by data-intensive services constitutes the majority of the internet traffic today. Exponential proliferation of smart mobile devices, wearable technology and IoT continues to put enormous traffic load on existing cellular systems and are further influencing the design of nascent 5G cellular systems. In addition to this, both Virtual Reality (VR) and Augmented Reality (AR) are poised to be the next set of the biggest trends in mobile technology. The evolution of edge computing and advancements in wireless networking ranging from the imminent roll out of 5G to highly efficient mobile connectivity solutions, coupled with access to smarter mobile and wearable devices, have all contributed to providing a rich environment for the proliferation and growth of VR and AR. High Bandwidth and low-latency requirements will become increasingly imperative for a high quality VR and AR experience and service providers will need to take a note of this new demand. The existing wired and wireless infrastructures have been able to handle such huge volumes of network load, however, much inefficiently.

As more and more information is made available and accessed on demand, content and service providers face increasing challenges concerning the spectrum efficiency and scaling of the network. Moreover, it can be observed that much of the current massive traffic load in the network is redundant, stemming from duplicate transmissions of popular content from servers to end hosts. A widely accepted solution for this problem is intelligent caching in the network. Although, the IP based network does store replicas of the contents at various sites to reduce re-transmitting load on the main server, it incurs significant operating costs as multiple locations manage these replicas in a different manner. In addition to this, mobile users are consuming more an more heterogeneous content over the wireless medium.. Therefore, the cellular architecture needs to be able to handle such requests much more efficiently than before. Information-Centric Networking (ICN) [3–6] is a recent paradigm conceived for future Internet architectures, which uses the content based approach instead of a host-centric one.

When IP was first introduced during 1970s, its goal was to allow communication between two machines, with one machine hosting resources and the other desiring access to these resources. Existing IP networks are host centric and therefore, the communication is based on static network identifiers (network addresses and interfaces) used by named hosts such as web servers, mobile phones, computers and other devices [7] . In contrast, ICN uses the concept of Named data [6] Named Objects (NO) to identify and route a variety of digital content such as web pages, documents, music, video etc [8, 9]. The ICN architecture leverages interaction models that fundamentally decouple information from its sources using location-identity split and allows multi-party communications based on contents stored over wide areas such as in cloud, in-network cache and even in UEs. It enables in-network caching and favors the deployment of multicast, thus facilitating reduced transmission delay and efficient delivery of information in the network. The primary objective of this thesis is on the efficient delivery of popular content by extending existing and future cellular architectures with ICN, specifically MobilityFirst, deployed as the core network.

Traditional unicast delivery mechanisms require capacity to increase linearly with number of users. Multicasting content to the end devices, on the other hand, is a spectrally efficient technique which can reduce load on the server significantly. Cellular providers have faced major challenges in scenarios of high bandwidth requirement where a large number of users access the cellular system simultaneously or problem of increased server load when providers have to push the content to a large number of end devices. Multicasting the content can provide significant gains in number of vertical sectors such as automotive, Internet of Things (IoT), Video on Demand (VoD), public safety warning systems and more whenever multiple end devices request for the same content

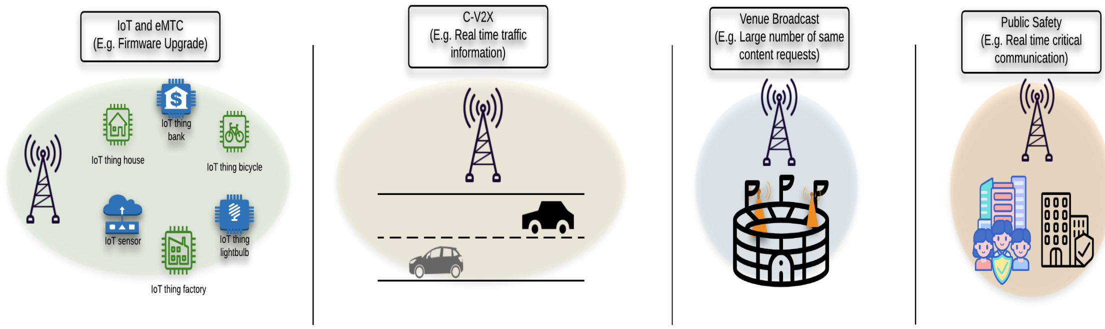


Figure 1.1: Some Use Cases for Multicast/Broadcast

as illustrated in Fig 1.1.

Name based architectures natively support multicast; by using name recursion to store the multicast tree topology, an efficient multicast protocol can be implemented in name-based architectures [10]. We contend that this functionality can be extended to the radio access networks of cellular systems as well. The inspiration of this work is derived from a desire to improve cellular multicast performance relative to the current state of the art called MBMS (Multimedia Broadcast/Multicast Service), which provides some gains but can be improved further with the application of ICN concepts.

1.1 Problem Statement

While a lot of work has been done in multicasting on core network, several challenges manifest during its implementation at Radio Access Network (RAN) side. Multicasting in the RAN part of the network has to deal with radio specific problems including multiple access (MAC), mobility and handover and the justification of investment by the service providers. Hence, the need for a new perspective of using mobility friendly Information Centric Networks approaches RAN. The goal of this work is to propose a simple and efficient multicast design which integrates the ICN's native support of multicast, obtained by using identifiers instead of locators, and efficient wireless transmission of the information by existing cellular systems. It is well known that Information Centric Networking architecture is better suited than the IP-based networks for content delivery and mobility [11], [4], [6]. ICN in backbone, nonetheless, provides a relatively small amount of relief in terms of backbone bandwidth utilization and reduced latency when the last-hop of the content delivery is a radio access network, since, RAN is a major bottleneck and a considerable source of end-to-end latency. Therefore, a good starting point of building a cellular multicast could be to evaluate the efficacy of ICN when deployed in radio access networks.

1.2 Main Contributions

The key contributions of this work are as following:

- A novel ICN based cellular architecture, "I-MAC", is proposed in which MAC layer services are tied to name/identifier associated with an ICN packet. In other words, the idea is, therefore, to push the ICN identifiers to the MAC layer with the purpose of making cellular multicast more dynamic and efficient.

- The thesis focuses on a specific use case of pull based multicast: multicasting in a vehicular to infrastructure system where a number of vehicles could request for the same content such as the request for traffic conditions in the tunnel or intersection a mile ahead.

  The goal of this work is to study the performance in terms of metrics such as system throughput and multicast gain under an alternative cellular multicast system with ICN network deployed in the core. Detailed NS-3 simulations were conducted to study the efficacy of the proposed system and on the impact of parameters such as window size for aggregation, system load and Zipf Distribution exponent, on system performance were carried out.

1.3 Related Work

Information Centric Networking emerged recently as a future internet technology which focuses on named data and adopts pull based or receiver oriented communication with built in features of efficient multimedia delivery, mobility friendly and content driven protocol. However, a significant amount of work has been focused on efficient content delivery in fixed or unlicensed WiFi wireless networks. With such features, ICN can bring noteworthy benefits when deployed in cellular networks.

A significant amount of research study is related to how caching in cellular networks could reduce the server load, reduce end-to-end latency and assist in scaling of such systems [12–17] . Reference [15] proposes ICN based edge caching system involving both EPC and RAN caching. The authors, through trace-driven evaluation results, show that RAN caching largely contributes to the reduction in both delay and traffic load as popular contents are mostly cached at last hop. Reference [18] proposes a similar system architecture by deploying additional components such as ICN proxys and HTTP proxy cache inside eNodeB. The IP traffic from either legacy device or ICN supported device is intercepted at eNodeB and sent to ICN Proxy in case it is an interest packet. This is further sent to ICN routers which are co-located with eNodeB. If the content is already cached in the co-located ICN router, then it is sent back immediately, otherwise, it is sent to the host server to retrieve the content object. The results in [18] show that the additional processing time introduced due to ICN caching is compensated by the performance enhancement achieved by ICN integration with LTE.

Authors in [17] introduce the concept of FemtoCaching in which they analyse the optimum way of assigning files to the small base stations (sBS) with low-rate backhaul and high storage cache in order to minimize the download time for contents. [14] studies the similar optimization problem of placing popular contents in small base stations such that maximum amount of data can be fetched from such sBS. The authors, assuming unknown popularity at sBS, formulates such a problem as a multi-armed bandit problem and study the performance of learning based algorithms such as Modified Combinatorial Upper Confidence Bounds (MCUCB) and Informed Upper Bound (IUB). Similarly, [19] formulates the problem of selecting the users with best channel conditions to be multicast as a multi-armed bandit problem and propose ϵ-greedy and pursuit algorithms to solve such a problem. With user generated content based applications gaining popularity, [20] addresses challenges introduced in core and radio access networks by such applications and proposes a framework exploiting Service Function Chaining as Virtual Network Functions (VNF), enforcing a dynamic offload along with the use of multicast mode of delivery to locally distribute the shared content among users.

While these studies show that in-network caching can be very effective in reducing server load, such an approach still doesn't completely solve the problem of efficiently "delivering" the content over the last-hop wireless channel. The rudimentary approach to delivering the content from the server is still unicast. As the popularity of the content increases, more and more requests arrive at the cache location which further adds to the cache load. Reference [16] points out that caching is effective when there is content *reuse* while multicast is effective when there is content request concurrency i.e many users request for a same content around same time, especially observed in scenarios of crowded places. In the era where most of the internet traffic is video based [2], the leveraging of multicast as a core principle for content distribution in the case where large number of users request for the same content can be resource efficient. The idea of using multicast is not new and as such several studies have been conducted on using multicast for content delivery. Reference [16] proposes to use multicast-aware caching solution when there is a massive demand of delay tolerant content and show that combining multicast and caching can provide significant energy gains in the case of dynamic user demands. In order to jointly optimize the average network delay and power costs under a multiple access constraint, authors in [21] propose a stochastic content multicast scheduling algorithm and show that the problem can be modelled as infinite horizon average cost Markov Decision Process (MDP).

All these studies focus on sender oriented i.e. push-based content delivery. And yet there is a dearth of on-demand or pull-based multicast studies, the plausible reasons for which are enlisted as follows. Firstly, there are fewer use-cases where aggregating the requests provide significant multicast gain such as in [22, 23]. Secondly, even if aggregation is possible, modelling such scenarios with appropriate control and data model is difficult. The aim of this project is thus, to show that with significant temporal and spatial locality, an efficient data and control model can be constructed supporting the notion for pull-based as well as push-based multicast.

It seems that a considerable work has been done in order to make unicast transmissions better and faster, however, we feel that the potential gains obtained by using multicast are highly underutilized in radio access networks.

Chapter 2

Overview of Current Cellular Multicast

3GPP in Release 9 introduced eMBMS architecture to support broadcast delivery of the content to a large number of users. The following sub-section describes in detail how LTE supports group communications over radio access networks.

2.1 Group Communication in LTE Radio Access Networks

To study the group communication in cellular networks, it is imperative to understand how a common data requested by multiple end points is disseminated by an eNodeB. LTE fundamentally uses two types of physical channels to actually transport the data: Physical Downlink Shared CHannel (PDSCH) (used for unicast data) and Physical Multicast CHannel (PMCH) (PMCH is designed for MBSFN data transmission). LTE uses PDCCH to transport the control data : they support data channel PDSCH by indicating the time-frequency resources to which the data is mapped [1]. The eMBMS standard can be predominately described by two deployment technologies: MBSFN (Multicast-broadcast single frequency network) and SC-PTM (Single Cell-Point-to-Multipoint). These are further discussed in detail in following subsections.

2.2 MBSFN Overview and Architecture

Figure 2.1 shows the eMBMS reference architecture. For the existing cellular systems, 3GPP Rel 9 specified evolved Multimedia Broadcast Services (eMBMS) for broadcasting/multicasting services over LTE networks with radio and mobile TV broadcasting, emergency alerts, live video streaming when users are spatially correlated as the target

[1]Control information of PMCH is carried in PDSCH

applications. eMBMS in LTE network uses Multicast-broadcast single-frequency network (MBSFN), which is a multicell transmission over a synchronized ‘single frequency network’, to support broadcasting to cell edge users. The MBSFN deployment consists of participating eNodeBs (i.e. eNodeBs having subscribed multicast clients) transmitting the same signal in a time-frequency synchronized manner. This benefits *cell edge* users as this synchronized transmission from multiple cells interferes constructively and the effective received signal strength is enhanced. Further, the Modulation and Coding (MCS) value is selected according to the user experiencing worst channel conditions. eMBMS was introduced as an efficient multicast system over the traditional broadcast technologies such as DVB-H, DMB-H or former MediaFLO. The advantages over these technologies were that no additional infrastructure is required, no additional spectrum is needed and user interaction is also possible.

In order to allow both unicast transmission and MBSFN transmission, LTE/LTE-A reserves special MBSFN subframes which are scheduled for such broadcast transmissions. The goal was to reuse existing LTE spectrum for broadcast with same resources divided among unicast and broadcast (up to 60% of sub-frames can be allocated to the MBMS traffic). For supporting LTE broadcast, new nodes and interfaces were introduced in addition to the existing nodes. BM-SC (Broadcast Multicast Service Center) connects the Content Provider with EPC. MBMS-GW allocates IP multicast address to which participating eNodeBs should join to receive MBMS data and forwards the user data packets to these eNodeBs. MCE (Multi-cell Coordination Entity) deals with session control and manages radio transmissions from eNodeBs within a semi-statically configured area such that exactly same configuration is used by each participating cell.

However, eMBMS suffers from considerable drawbacks. The UEs have to perform separate channel estimate for the MBSFN reception in addition to that performed for single-cell reception and so MBSFN uses a more dense reference signal pattern compared to unicast. Therefore, time division multiplexing of unicast and MBSFN data is used which means that there are designated subframes, known as MBSFN subframes, used exclusively for the transmission of broadcast/multicast data. In other words, the use of Multiple Input Multiple Output (MIMO) is not defined for MBSFN transmissions and

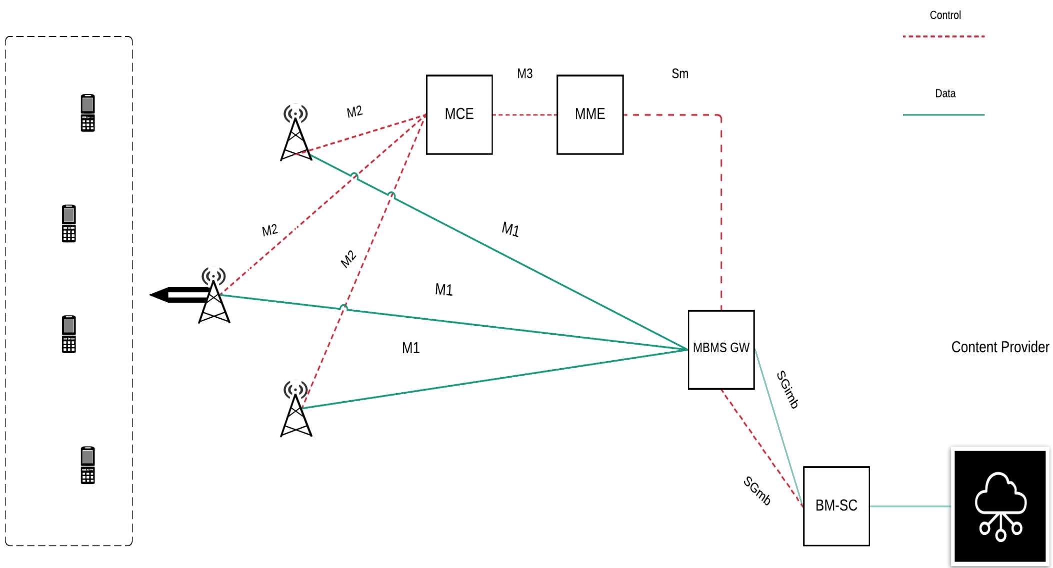


Figure 2.1: 3GPP eMBMS Architecture

thus it loses out on peak spectral efficiency that could have been obtained. Furthermore, MBSFN transmissions don't support advanced link adaptation schemes due to lack of uplink feedback and thus, HARQ re-transmissions aren't supported.

The question, then, is how can RAN multiplex unicast and multicast data across frequency domain? For that reason, it is important to understand how each data is transferred in case of unicast. Figure 2.2 describes the unicast transmissions in the physical control and data channels.

LTE uses PDSCH subframe to transmit data to a specific User Equipment (UE). After random access procedure, when any UE goes into *RRC_Connected* state, eNodeB assigns that UE a Cell Radio Network Temporary Identifier (C-RNTI), which is unique to a particular UE in that cell. Downlink Control Information (DCI) which is transmitted through the Physical Downlink Control CHannel (PDCCH), carries the control information about the OFDMA resource blocks that are allocated for the specific data for a particular UE in PDSCH. The cyclic redundancy check (CRC) part of DCI in PDCCH is scrambled based on C-RNTI of the receiving UE and thus only the UE with that particular C-RNTI could decode the DCI and further the data in PDSCH.

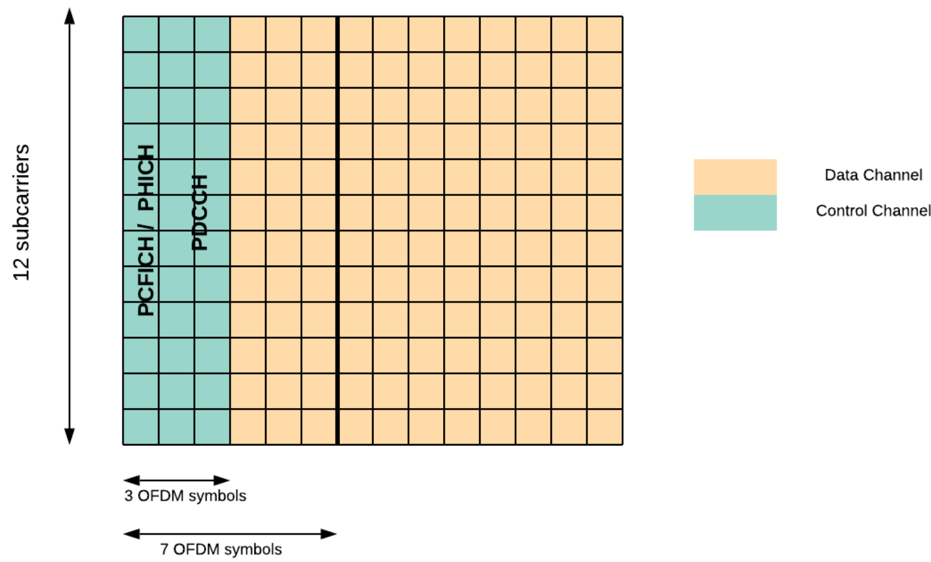


Figure 2.2: Physical Control and Data Channels (PDCCH and PDSCH)

Therefore, for group communication, LTE requires a mechanism to define the same RNTI for UEs involved in that group communication so that all of those UEs could decode the data in PDSCH. To provide a fundamental solution for such problem, 3GPP in Release 13 introduced Single Cell Point-To-Multipoint (SC-PTM) as an alternative deployment technology deploying existing eMBMS core architecture and uses PDSCH for group data transmission.

2.3 Radio Access Protocol Architecture and Signalling for SC-PTM

In Release 13 specification on SC-PTM, two new types of logical channels were introduced: Single Cell Multicast Control CHannel (SC-MCCH) and Single Cell Multicast Traffic CHannel (SC-MTCH) as control and data logical channels respectively. One SC-MCCH and one or more SC-MTCH(s) are mapped on LTE transport channel DL-SCH. Figure 2.3 shows the high level signalling protocol in SC-PTM. SC-MCCH and SC-MTCH transmissions are each indicated by a logical channel specific RNTI on PDCCH (there is a one-to-one mapping between TMGI and G-RNTI used for the reception of the DL-SCH to which a SC-MTCH is mapped). Note that in the figure MCE stands for Multicast Coordination Entity and CN stands for Core Network. The SC-MCCH provides the list of all MBMS services with ongoing sessions transmitted on SC-MTCH(s), including for each MBMS service its TMGI and optional session ID, associated G-RNTI and scheduling information. SC-MCCH is transmitted by RRC (SIB) every SC-MCCH repetition period. eMBMS architecture encounters a "chicken-and-egg" problem since both new logical channels SC-MCCH and SC-MTCH are mapped to a new transport channel MCH which in turn is mapped to PDSCH. Hence, to solve such a problem, 3GPP introduced a new System Information Block, SIB13 to include eMBMS related control information. SIB13 is broadcasted periodically on Physical Broadcast Channel (PBCH) and it carries information about the scheduling of the MCCH (i.e. the radio frames mapped in PDSCH correpsonding to MCCH), MCCH modification period and repetition period and subframe offset. At this point, it has to be noted that MBMS doesn't use PDCCH for dynamic scheduling and MCCH is actually mapped to PDSCH, the physical data channel in unicast. MCCH itself carries the one-to-one mapping between TMGI [2] and G-RNTI (Group RNTI) as well as scheduling information of SC-MTCH which carries the downlink SC-PTM data and the downlink SC-PTM data in MTCH is identified by G-RNTI. Another important point to note here is that SC-PTM requires two special types of additional RNTIs: Single Cell-RNTI (SC-RNTI), to decode the MCCH data and Group-RNTI (G-RNTI) to decode the MTCH data.

In terms of spectral efficiency, SC-PTM performs better than unicast and MBSFN transmission as it uses less resources compared to unicast on a per cell basis as well as less resources in the network than MBSFN.

---

[2]The BM-SC allocates a globally unique TMGI per MBMS bearer service

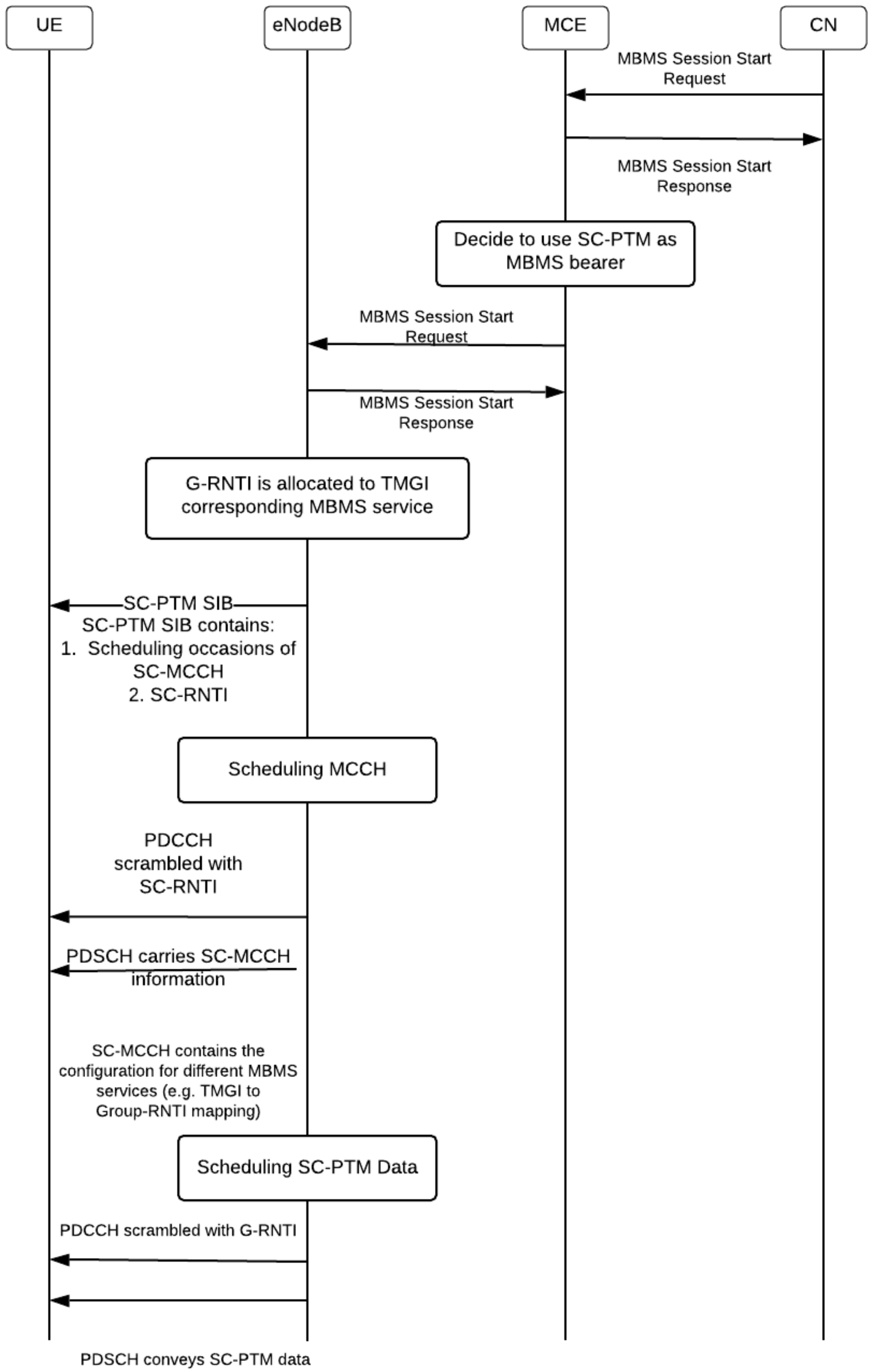


Figure 2.3: High level procedures for SC-PTM [1]

Even after multiplexing unicast and multicast data in PDSCH, SC-PTM is still not resource efficient as this protocol requires control information to be carried in shared data channel, effectively reducing the resources for both unicast and multicast data transmission. Further, SC-PTM requires continuous monitoring of PDSCH and is on a per-service basis. This current specification is relatively applicable to static media delivery and is ill suited for content based dynamic architectures such as ICN.

The aim of this work is two-fold : firstly, to enable ICN semantics to be identified at the MAC layer in cellular technology for seamless integration with name-based core network routing associated with ICN and secondly, extend this proof-of-concept to show it's effectiveness in support of fine-grained dynamic multicast services. Most of the literature on SC-PTM focuses its use in mission critical communication scenarios as in [1, 24]. We believe that multicast potential in cellular communications is under-utilized and, thus, we would like to explore more multicast use cases other than mission critical communications.

Chapter 3

ICN Multicast

ICN on cellular layer 2 protocols (PDCP, RLC, MAC, PHY) or deployed along with IP stack can help bring benefits of inherent capability of multicast, multihoming, seamless mobility support and optimized data delivery using local cache at the cellular edge. However, ICN, if deployed in cellular network architecture, faces several key challenges. First, producer mobility is an open challenge as uplink resources are utilized in updating the Name Resolution System (NRS) [25] while on the other hand subscriber mobility is easy to handle by virtue of its connection-less receiver oriented model. Second, the IP-based cellular networks are connection oriented enforcing QoS policies and deploy packet gateways (PGW/SGW) based data forwarding as most of the traffic is heading towards Internet. On the contrary, connection focused communication in ICN is incongruous with content oriented communication as the content can be accessed from different cache locations.

In order to utilize full benefits that ICN inherently provides, a complete overhaul of LTE/5G network is required, both in the core network as well as in radio access network. It is important to understand why the future cellular systems need a complete overhaul and that can be understood by looking at the control and data plane signalling in the present LTE-A networks.

Reference [26] proposes a distributed and "flat" core network which employs identifier-based protocol extensions on eNodeB and routers without the need of using centralized gateways. In this core architecture, MME is replaced by distributed name resolution or mapping system which maps identifiers associated with devices, context, services etc, to a set of locators such as the address of corresponding eNodeB. The architecture also does away the use of gateway nodes PGW/SGW and allows eNodeB to handle policy

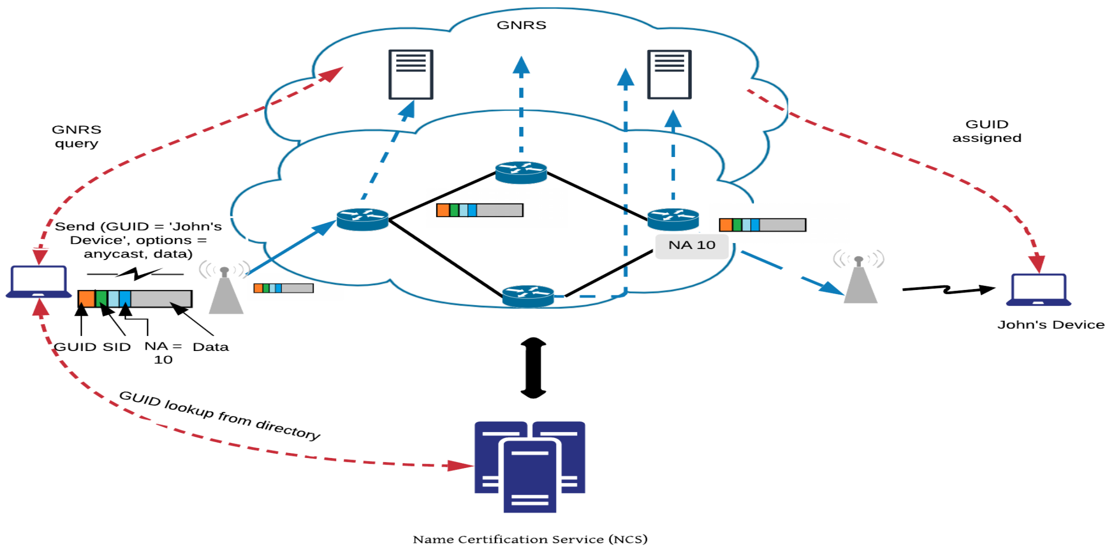


Figure 3.1: MobilityFirst Architecture Concept

and QoS enforcement. It has to be noted that every node in the physical path between client and the publisher is a router which can forward the interests intelligently as well as provides in-network caching functionality. Here, each eNodeB maintains a local name resolution database and each mobile device has a unique identifier associated with it. Since, we are using MobilityFirst as the ICN architecture, these mobile devices will be identified by unique GUIDs which are flat cryptographically secured identifiers. The work assumes that both the radio access network and core network implement the MobilityFirst stack. We use the same core network architecture in our system and assume that GTP tunneling is not considered between PGW and eNodeB. In the legacy LTE systems, uplink data is captured at PDCP layer at eNodeB and passed to user plane GTP (GTP-U) tunnel which carries this uplink data to S-GW. In our proposed work, the captured data at PDCP layer is sent to the ICN router co-located with eNodeB without the need of any tunneling.

For ICN to be integrated with current IP-based cellular architectures, we need a mechanism to support ICN identifiers with the current MAC layer identifiers. The idea is, therefore, to push the ICN identifiers to the MAC layer to support dynamic and efficient multicast. We specifically focus on use cases of pull-based multicast for two reasons. Firstly, we believe that there are special scenarios where using pull based multicast can assist the LTE system in saving considerable bandwidth resources. Secondly, the communication in ICN is consumer oriented and is inherently pull based, driven by content requests from the user. Therefore, this work aims at studying the performance of pull-based multicast in current and future cellular networks supported by ICN. We propose “I-MAC”, an ICN supported cellular architecture. For this work, we consider MobilityFirst as the ICN architecture deployed in the core network, although any other ICN architecture such as NDN/DONA can be deployed as well [27].

MobilityFirst is a clean slate future internet architecture with an emphasis on handling mobility in a trustworthy manner along with providing support for multicast, multi-homing and in-network cache. Figure 3.1 shows the MobilityFirst architecture centered on name based service layer. One of the major design features of MobilityFirst is the separation of names and addresses for all network attached entities (including devices, contents, contexts, a person). These network attached objects are assigned“flat” globally unique identifiers (GUID) by employing name certification services (NCSs) without a global root of trust. These GUIDs are mapped to a set of network addresses (NA) corresponding to current points of attachment using fast global name resolution service (GNRS) to dynamically bind the destination GUID to their respective current network address.

3.1 I-MAC Architecture

Figure 3.2 shows the conceptual I-MAC architecture. We assume that base stations deploy dual stack i.e. both IP and ICN stacks (for backward compatibility) and hence are ICN stack enabled and as a result they can identify MobilityFirst identifiers. In addition to this, each eNodeB maintains a Local Name Resolution Service (LNRS) database which is used in the control plane. LNRS supports in name management by maintaining a table of named objects (such as network device, content, context etc.) mapped to the corresponding UE RNTI. We also assume an ICN router co-located with the base station which can access the GNRS database and resolve the GUID to a set of network addresses using distributed hash tables. Moreover, these ICN routers

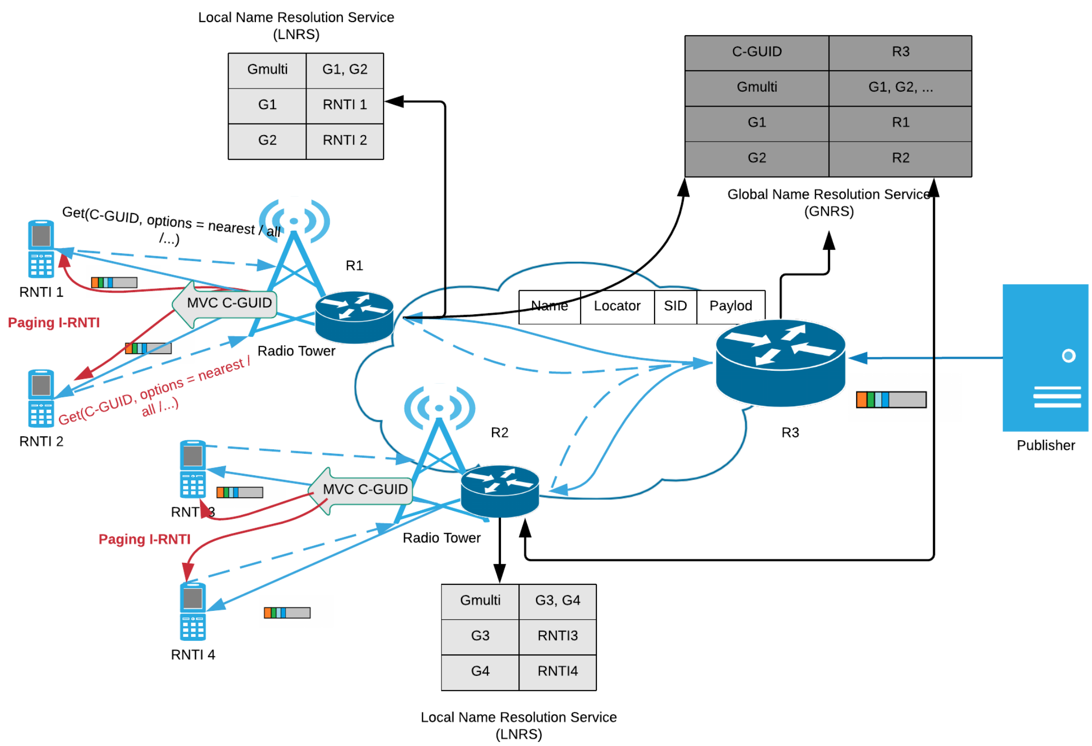


Figure 3.2: Proposed “I-MAC” Architecture

can maintain cache space in order to store and serve popular requests with a very low latency. Initially, the LNRS is populated with direct mapping entries of UE GUID and allocated UE RNTI at the completion of attach procedure which involves random access procedure. The LTE RACH (Random Access CHannel) is used to achieve uplink time synchronization for a UE which either has not yet acquired, or has lost, its uplink synchronization [28]. Figure 3.3 shows the four steps involved in contention based random access procedure. After being allocated C-RNTI at step 2, the UE sends an actual random access procedure message, such as an RRC connection request, tracking area update, or scheduling . In our proposed work, we assume that the UE GUID is tagged along with the RRC connection request message. In this way, LNRS maps the corresponding UE RNTI with UE GUID which was provided by the UE at the RRC connection set up procedure and this database is updated dynamically with each attach procedure. Let G_m be the content GUID either requested or subscribed by i^{th} UE with GUID G_i. As the eNodeB receives this content ID request, it starts populating the corresponding entry with $G_i \forall i$. When the server wants to push data to the client UEs,

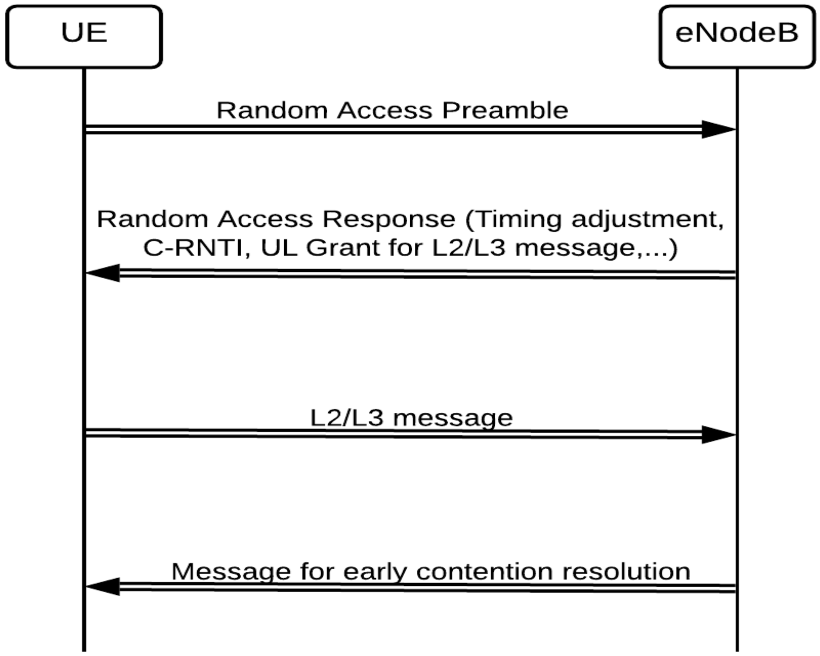


Figure 3.3: Random Access Procedure in LTE/LTE-A

after querying GNRS, it sends the data to the respective edge ICN routers which are co-located with the corresponding eNodeBs. The control part of eNodeB decapsulates the packet to get the content GUID G_m and now the eNodeB queries the LNRS, for last hop transmission, to obtain end UE RNTIs and starts the creation of data radio bearers in radio access network. The procedure of setting up of data radio bearers and radio access signalling is described in next subsection.

3.2 Radio Access Signalling

As mentioned in previous section, LTE eNodeB uses RNTI to identify a UE which is in *RRC CONNECTED* state. One drawback of SC-PTM is that UEs require two types of RNTIs - SC-RNTI (broadcasted in SIB 13) and G-RNTI (indicated by SC-MCCH control channel) in addition to periodic announcements and monitoring of SC-MCCH. This is spectrally inefficient, battery consuming and isn't suitable for dynamic scenarios. We propose a spectrally efficient and dynamic multicast protocol signalling in RAN for the proposed I-MAC architecture. In this work, rather than monitoring SC-MCCH continuously for the exact time the data will be broadcasted and any system modification changes, eNB inserts the content GUID into the paging message and sends it to the multicast UE. This is because paging is the mechanism in which the network tells a UE - “I have something for you”. The UE wakes up at every paging occasion (pre-decided) and checks PDCCH with corresponding P-RNTI (Paging RNTI). PDCCH will inform UE where to read the paging message in PDSCH. So, reading paging message in PDSCH, the UE gets the content GUID (this is a form of context that lets UE know that there is a multicast data for the corresponding content GUID) and upon receiving the content GUID, the UE initiates the RRC Connection with eNB. The response of connection set up from eNB includes the I-RNTI i.e. I-RNTI is assigned during the connection setup for decoding the multicast data. The multicast group now uses this I-RNTI to monitor the PDCCH which informs multicast UEs about the resource blocks allocated for multicast data in PDSCH. Figure 3.4 shows the proposed radio access signalling.

The drawback of using I-MAC based multicast is the extra control overhead introduced due to paging each UE individually. Nonetheless, the network takes advantage of dynamic and “on-demand” multicast.

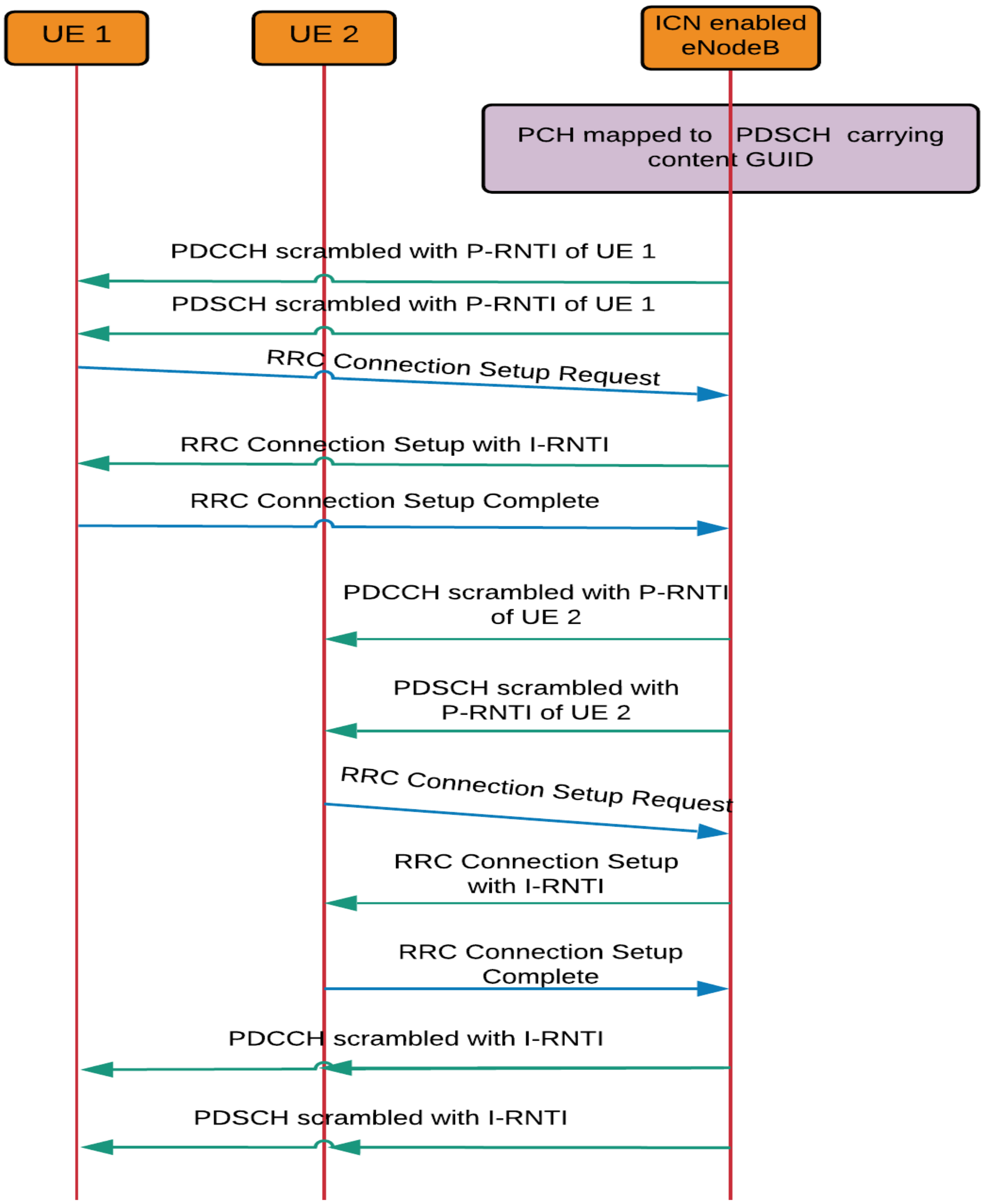


Figure 3.4: Proposed I-MAC Multicast Signalling in RAN

Chapter 4

Evaluation Results

In this section, we investigate the performance of I-MAC multicast for a realistic multicast use case. Since, we primarily study the effectiveness of aggregating the requests, we specifically focus on the scenarios that are capable of supporting multicast. We consider a pull-based multicast application in a highway - tunnel intersection as shown in figure 4.1.

To describe the probability of a content to be required by a UE, the statistics of content requests is modelled by Zipf Distribution as suggested in and [29]. The probability $p(i)$ that content i is required by a UE can be expressed as:

$$p(i) = \frac{\frac{1}{i^{\alpha}}}{\sum_{n=1}^{M} \frac{1}{n^{\alpha}}} \qquad \forall\ i = 1, ..., M \tag{4.1}$$

where $\alpha > 0$ is the Zipf parameter and M is the number of contents in the system. We assume $M = 100$ in our system. We assume that requests arrive at a Poisson rate with the mean arrival rate calculated as described. We assume that a macro eNodeB is installed at the entry of the tunnel and the vehicles in the range of 500 meters of

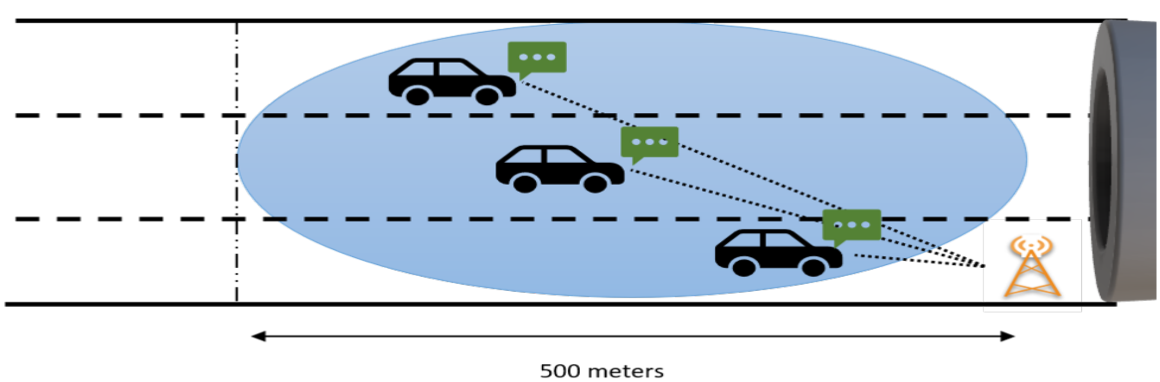


Figure 4.1: Highway-tunnel V2X pull-based multicast scenario

a 3-lane highway region request for some content, for example, status of the traffic congestion in the tunnel. The rate at which the content is requested depends on the speed of the vehicle, which in turn depends upon the density in the 3-lane highway. We obtain the speed density relationship from [30] and hence, for a given density on the highway, the corresponding average speed of the vehicles can be obtained. Now, a simple speed-distance relationship can give us the time a vehicle will take to cross the 500m distance. Since, we assume that the user, while crossing the given distance, would request for the content only once, it suffices to say that

$$\lambda = \frac{1}{t_d}$$

where, t_d is the amount of time it takes for a vehicle to cross 500m and λ is the rate of all content requests. For example, if the density on the highway is 30 vehicles/lane/500m, the approximate average speed per vehicle is $22m/s$. This gives us the time difference between two consecutive requests, which in this case, will be $0.25s$, and thus the average Poisson rate of all the content requests is 1/0.25. The requests rate for a particular GUID i can then be expressed as:

$$\lambda_i = \lambda p(i)$$

According to [29], if the probability that a content is accessed follows a Zipf distribution then the following asymptotic properties hold true:

1. The hit-ratio of a web cache grows in a log-like fashion as a function of the cache size.

2. Content requests exhibit excellent temporal locality.

The latter temporal locality property can allow to aggregate the delay-tolerant requests by waiting for a certain period of time and serve these batched requests with a single transmission. We implement such a concept for our highway-tunnel pull-based multicast use case. Using such derivations, we model the content arrivals in ns-3. We further simulate the vehicle mobility scenario by coupling NS-3 with a traffic simulator

SUMO (Simulation of Urban MObility) [31] with the aim of simulating real highway-tunnel environment smoothly. SUMO is a microscopic vehicular traffic simulator which provides interesting features such as importing maps, modelling car following models, traffic light intelligence etc. By integrating SUMO with ns-3, ns-3 accounts for change in Modulation and Coding Scheme (MCS) with the node mobility and allows us to model C-V2X scenario accurately. As a reference scenario, we have considered Lincoln tunnel connecting New Jersey and New York City.

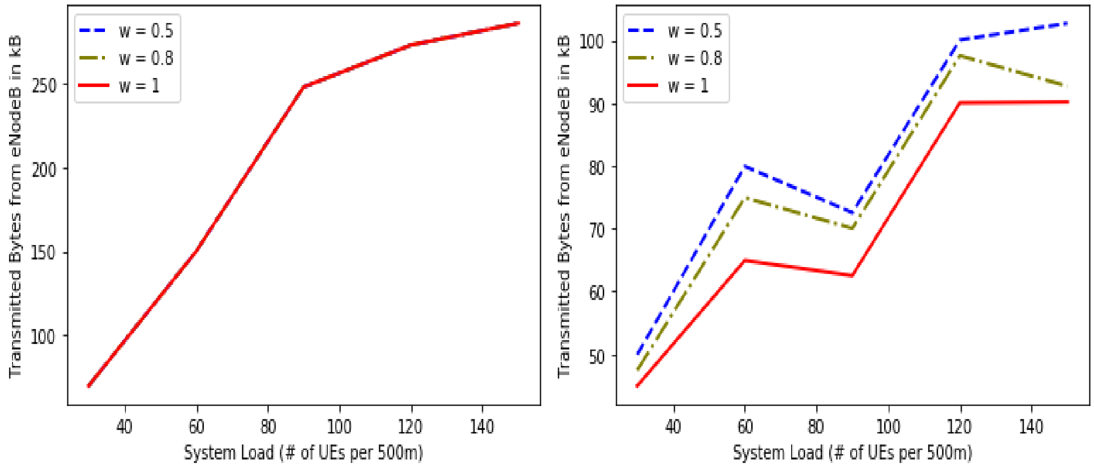
Attribute	Value
Ul/Dl Bandwidth	5MHz / 25 RBs
MAC Scheduler	Proportional Fair Scheduler
eNodeB Transmit Power	46 dBm
Pathloss Model	Friis Spectrum Propagation Model
eNodeB mobility	Constant Position Mobility Model
UE Mobility model	SUMO traces
Number of contents	100
Content Size (same for all contents)	2500 Bytes

Table 4.1: NS-3 Simulation Parameters

Table 4.1 describes the simulation parameters in NS-3. We simulate the given scenario with different window sizes, Zipf parameters and system load, and evaluate the system performance under batching requests and compare the performance of multicast and unicast modes of transmissions.

Figures 4.2 show the number of bytes transmitted from eNodeB in case of unicast and multicast mode of transmission against different batching window sizes. It can be seen from these two figures that eNodeB saves up a considerable bandwidth if it decides to multicast even after waiting at most $0.5s$. Further, it is shown that window size doesn't influence unicast transmissions.

We see a further drop in bandwidth utilization when the system load increases from 60 UEs to 90 UEs but the bandwidth utilization again increases after 90 UEs. Figure 4.3 shows the reason behind this. As the number of UEs increase, the probability of aggregating multiple UEs in a single transmission increases. However, as the system load is further increased, we predict that the spread among content requests can increase



(a) Unicast Scenario (b) Multicast Scenario

Figure 4.2: Performance comparison of unicast and multicast modes in terms of bandwidth utilization

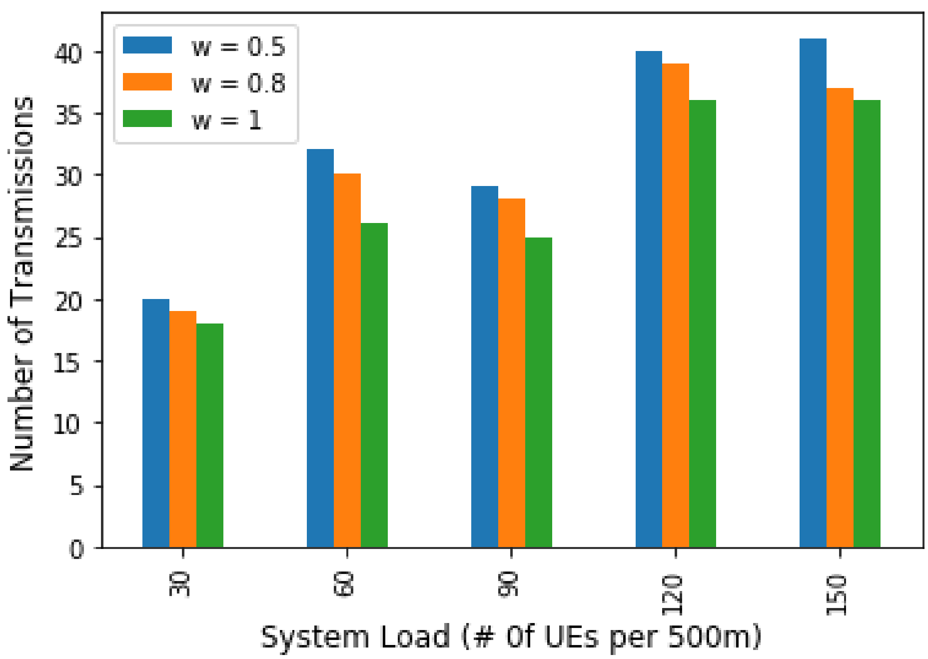


Figure 4.3: Number of transmissions from eNodeB under various window sizes in a multicast scenario

and hence, the number of UEs requesting the same content in a particular batching window will decrease and therefore the eNodeB would have to serve the users with more transmissions.

Figure 4.4 shows the impact on system bandwidth as content size increases and it can be seen that the unicast performance degrades as the size increases while the performance of multicast system is not affected to a great extent.

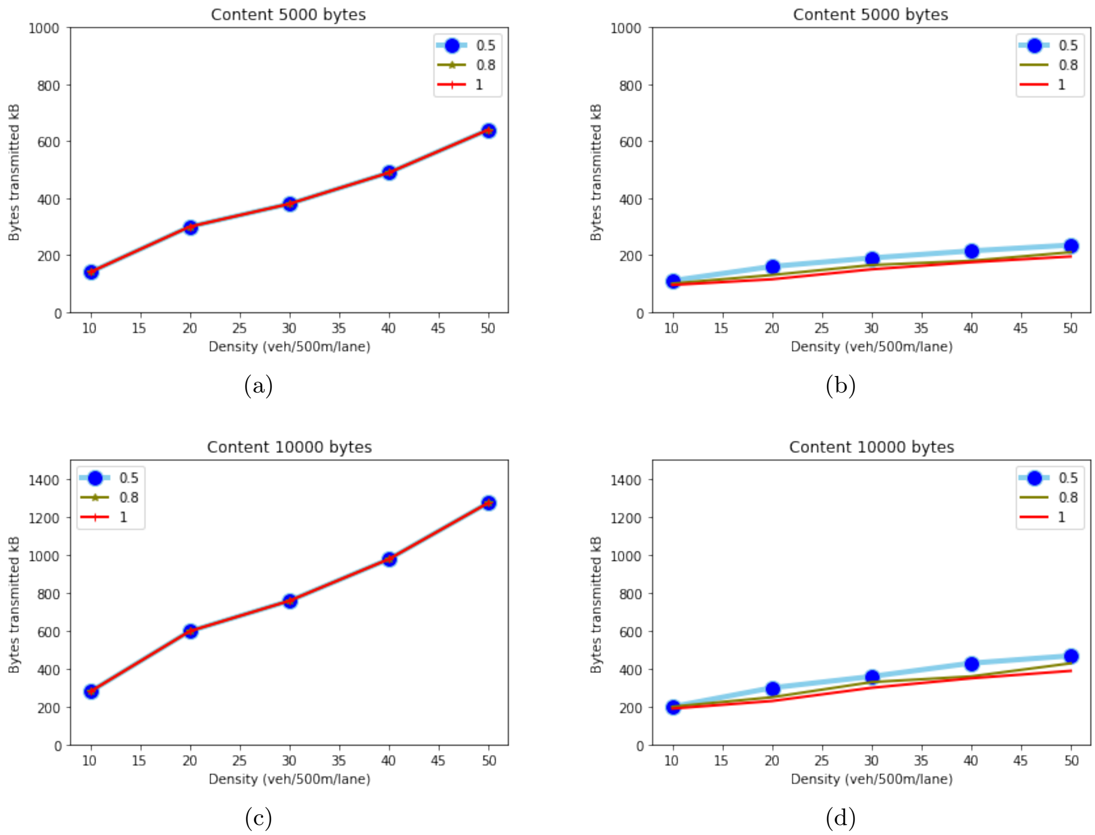


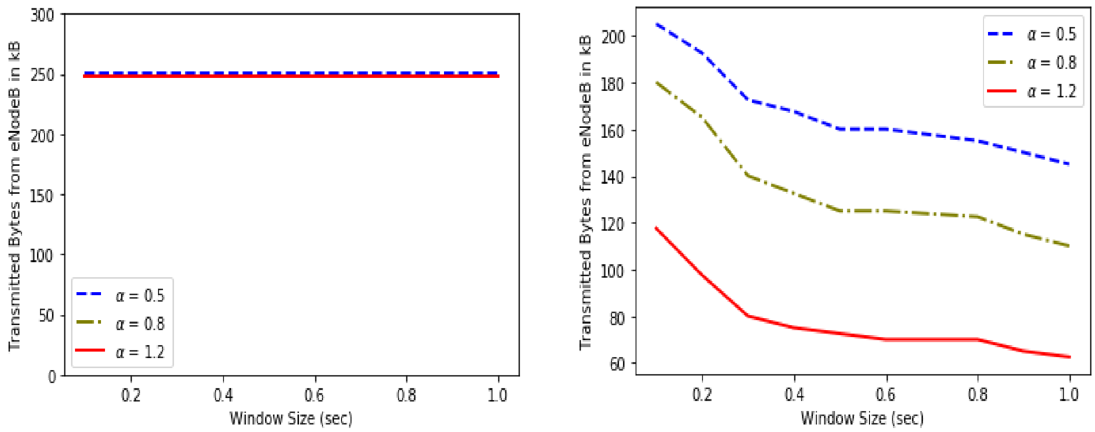
Figure 4.4: Performance Comparison between unicast and multicast: (4.4a) and (4.4b); Size = 5000 bytes, (4.4c) and (4.4d); Size = 10,000 bytes

4.0.1 Parametric Evaluation

In this section, we analyze the sensitivity of our design with respect to different parameters including different batching window sizes and Zipf Parameter α.

Impact of Batching Window Size

Figure 4.5 shows the effect of window size for different Zipf parameters when the system load is 90 UEs. As α increases, the probability of requesting the most popular content increases and this causes larger number of UEs to request for the same content, hence resulting in more request aggregation. Figure 4.5 justifies that even waiting for as low



(a) (b)

Figure 4.5: Impact of batching window size on (4.5a) for unicast; (4.5b) for multicast

as 0.2 seconds can provide considerable savings on system bandwidth utilization when dealing with delay-tolerant user requests and higher demand for the content.

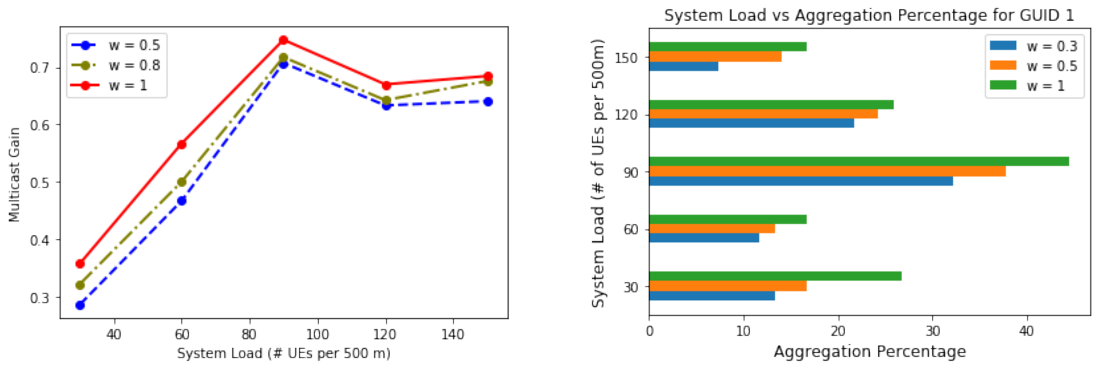
In fact, 4.6b shows that 30% aggregation can be achieved when users tolerate delay for 0.3s for $\alpha = 1.2$ and the gain increases further, around 45%, with length of batching window. A different interpretation is shown in figure 4.6a as a function of multicast gain. We define multicast gain metric γ as follows [32]:

$$\gamma = 1 - \frac{\text{Number of bytes transmitted in multicast}}{\text{Number of bytes transmitted in unicast}}$$

where, $0 \leq \gamma < 1$. When the value is close to zero, there is no use in multicasting but as the value approaches 1, the benefit of using multicast increases. Figure 4.6a underlines the better performance of batching multicast scheme in such scenarios as the system load increases.

Impact of Zipf Parameter

An interesting observation as shown in figure 4.7 when system load is 90 UEs and the most popular content is considered. As discussed before, when the Zipf exponent is high enough, for example $\alpha = 1.2$, the aggregation percentage is pretty large, and it justifies to wait so as to achieve considerable multicast-gain. However, when the



(a) (b)

Figure 4.6: Multicast gain as a function of system load for $\alpha = 1.2$

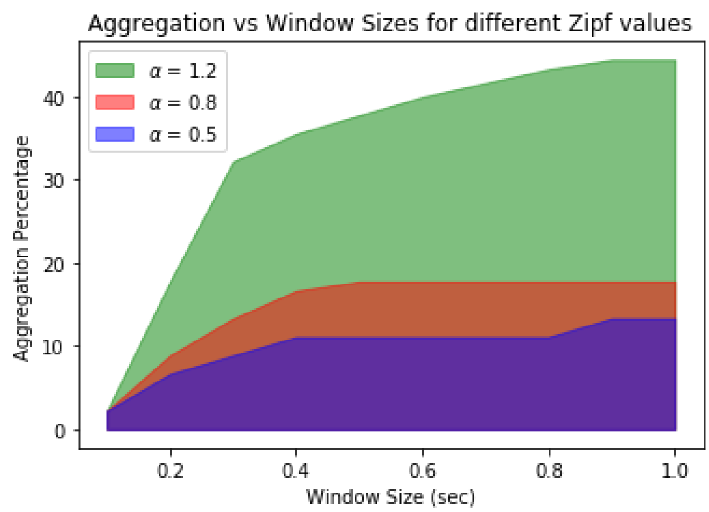


Figure 4.7: Impact of Zipf Parameter

content's popularity is not as steep, the aggregation percentage after a certain window size saturates. This behaviour can be justified by the observations in [29]. As $0 \leq \alpha < 1$, the asymptotic behaviour of temporal locality becomes less tractable as compared to when $\alpha > 1$. Thus, even a higher window size doesn't compensate for such a behaviour of Zipf-based requests.

4.0.2 Push-based multicast performance comparison between SC-PTM and I-MAC Multicast

In this section, we compare the performance of SC-PTM and I-MAC multicast in terms of PDSCH occupancy and uptime. It is important for any proposed design to support multiple use cases and determine the efficacy of the design for such scenarios. We believe that push-based multicast mostly finds application in massive Internet of Things (IoT) scenario. Thus, following evaluations are performed to determine if the proposed system is suitable for massive IoT application.

Multicast Approach	Uptime in seconds / hour
SC-PTM	47.81
I-MAC	2.81

Table 4.2: Uptime for IoT devices in seconds / hour

Table 4.2 shows the amount of time IoT devices spend receiving control information (uptime). According to [1], the scheduling occasions for SC-MCCH carried by SC-PTM SIB is every $80ms$. Thus, for a device to receive the multicast data through SC-PTM, it has to check for SC-PTM SIB every $80ms$. This is in addition to receiving paging information necessarily required by all UEs. We assume the minimum paging cycle of $1.28s$ and that the device observes both SC-MCCH and paging message for $1ms$. This gives us the amount of time an IoT device spends getting the control information in case of SC-PTM transmission. With our proposed design, the device only needs to monitor the paging message every $1.28s$ and therefore, the uptime in this case is only 2.81 $seconds/hour$. The study of uptime is important in case of IoT scenarios as it determines the energy consumption of such energy-constrained devices.

We further discuss the control data occupancy of PDSCH for both SC-PTM and I-MAC modes. For fair comparison, we assume the same amount of resources allocated for control data in both scenarios as well as the scheduling frame and paging frame are separate. Let the number of frames carrying control data for SC-MCCH scheduling be x frames/hour and let the total number of frames available in an hour be y. Then we define the PDSCH occupancy δ_{PDSCH} as:

$$\delta_{\text{PDSCH}} = \frac{x}{y} \cdot 100$$

Since, we assume the SC-MCCH scheduling period to be 80ms, SC-PTM uses 4500

Multicast Approach	PDSCH Occupancy %
SC-PTM	1.32 %
I-MAC	0.07 %

Table 4.3: PDSCH Occupancy % for control data

frames/hour carrying control data on PDSCH and furthermore, 281 frames/hour are used to carry paging message on PDSCH. Thus, δ_{PDSCH} for SC-PTM is 1.32%. Nevertheless, the I-MAC design only uses paging messages carried on PDSCH as control data and thus, only 281 frames/hour are used. Thus, δ_{PDSCH} for I-MAC is 0.07% as shown in table 4.3. These rudimentary measurements tell us that I-MAC is much spectrally efficient as compared to SC-PTM and allows for better utilization of data frames.

Chapter 5

Future Work and Conclusions

5.1 Future work

This work tried to explore few rudimentary aspects of the proposed design. However, certain features still need to be further explored.

- **Window size:** The work tried to empirically evaluate the optimum window size under other given parameters which includes system load and Zipf parameter. A study on theoretical lower bound on the window size as a function of such parameters needs to be done.
- **Large Scale Emulation:** This work evaluates the design using NS-3 simulation. There is a clear need to validate the design in both emulation and simulation in order to obtain real performance evaluations.
- **Performance evaluation in 5G systems:** We used NS-3 LTE module to simulate the proposed work and hence, the performance conforms to LTE systems. With the advent of 5G mobile communication systems, it is imperative to evaluate the proposed work with the upcoming 5G systems.

5.2 Conclusions

In this work, we proposed a cellular multicast protocol and architecture which uses Information Centric Networking identifiers and semantics. The use of proposed Local Name Resolution Table at eNodeB enables the mapping of MAC layer UE identifier and ICN identifier. Frame-level radio control signalling considering the available physical channels used by current cellular system is also presented. The goal was to validate the efficacy of using ICN at radio access network using some special multicast-based use cases. The design was validated using extensive NS-3 simulations and the results showed some promising aspects of the design.

Bibliography

[1] 3rd Generation Partnership Project, "3GPP TR 36.890 - Study on single-cell point-to-multipoint transmission for E-UTRA (Release 13)," Tech. Rep.

[2] Cisco and I. Cisco Systems, "Cisco Visual Networking Index: Forecast and Trends, 20172022 White Paper," pp. 2017–2022, 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html

[3] D. Cheriton and M. Gritter, "Triad: A new next-generation internet architecture," 2000.

[4] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," in *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM 07. New York, NY, USA: Association for Computing Machinery, 2007, p. 181192. [Online]. Available: https://doi.org/10.1145/1282380.1282402

[5] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT 09. New York, NY, USA: Association for Computing Machinery, 2009, p. 112. [Online]. Available: https://doi.org/10.1145/1658939.1658941

[6] D. Raychaudhuri, K. Nagaraja, and A. Venkataramani, "Mobilityfirst: A robust and trustworthy mobility-centric architecture for the future internet," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 16, no. 3, p. 213, Dec. 2012. [Online]. Available: https://doi.org/10.1145/2412096.2412098

[7] M. Handley, "Why the internet only just works," *BT Technology Journal*, vol. 24, no. 3, pp. 119–129, 2006.

[8] J. Choi, J. Han, E. Cho, T. Kwon, and Y. Choi, "A survey on content-oriented networking for efficient content delivery," *IEEE Communications Magazine*, vol. 49, no. 3, pp. 121–127, March 2011.

[9] J. Pan, S. Paul, and R. Jain, "A survey of the research on future internet architectures," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 26–36, July 2011.

[10] S. Mukherjee, F. Bronzino, S. Srinivasan, J. Chen, and D. Raychaudhuri, "Achieving scalable push multicast services using global name resolution," *2016 IEEE Global Communications Conference, GLOBECOM 2016 - Proceedings*, pp. 6–11, 2016.

[11] B. Ahlgren, C. Dannewitz, C. Imbrenda, and D. Kutscher, "A Survey of Information-Centric Networking," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 26–36, 2012.

[12] Y. Xu, X. Li, and J. Zhang, "Device-to-Device Content Delivery in Cellular Networks: Multicast or Unicast," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4401–4414, 2018.

[13] A. Dabirmoghaddam, M. Dehghan, and J. J. Garcia-Luna-Aceves, "Characterizing Interest aggregation in content-centric networks," *2016 IFIP Networking Conference (IFIP Networking) and Workshops, IFIP Networking 2016*, pp. 449–457, 2016.

[14] P. Blasco and D. Gunduz, "Learning-based optimization of cache content in a small cell base station," *2014 IEEE International Conference on Communications, ICC 2014*, pp. 1897–1903, 2014.

[15] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131–139, 2014.

[16] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Exploiting Caching and Multicast for 5G Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2995–3007, 2016.

[17] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, 2013.

[18] A. Gomes and T. Braun, "Feasibility of information-centric networking integration into lte mobile networks," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, ser. SAC '15. New York, NY, USA: ACM, 2015, pp. 627–633. [Online]. Available: http://doi.acm.org/10.1145/2695664.2695790

[19] F. Rebecchi, L. Valerio, R. Bruno, V. Conan, M. D. De Amorim, and A. Passarella, "A joint multicast/D2D learning-based approach to LTE traffic offloading," *Computer Communications*, vol. 72, pp. 26–37, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2015.09.025

[20] T. Taleb, A. Ksentini, M. Chen, and R. Jantti, "Coping with Emerging Mobile Social Media Applications Through Dynamic Service Function Chaining," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2859–2871, 2016.

[21] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6284–6297, 2016.

[22] V. Aggarwal, R. Caldebank, V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, and F. Yu, "The effectiveness of intelligent scheduling for multicast video-on-demand," *MM'09 - Proceedings of the 2009 ACM Multimedia Conference, with Co-located Workshops and Symposiums*, pp. 421–430, 2009.

[23] E. L. Abram-Profeta and K. G. Shin, "Scheduling video programs in near video-on-demand systems," *Proceedings of the 5th ACM International Conference on Multimedia, MULTIMEDIA 1997*, pp. 359–369, 1997.

[24] P. G. P. N. P. M. Alaa Daher, Marceau Coupechoux, "Sc-ptm or mbsfn for mission critical communications?" *IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017.

[25] W. Kim, "Beyond LTE-advance for information centric networking," *Computer Standards and Interfaces*, vol. 49, pp. 59–66, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.csi.2016.08.007

[26] S. Mukherjee, R. Ravindran, and D. Raychaudhuri, "A distributed core network architecture for 5G systems and beyond," *NEAT 2018 - Proceedings of the 2018 Workshop on Networking for Emerging Applications and Technologies, Part of SIGCOMM 2018*, pp. 33–38, 2018.

[27] D. Raychaudhuri, K. Nagaraja, and A. Venkataramani, "MobilityFirst," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, p. 2, 2012.

[28] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[29] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," *Proceedings - IEEE INFOCOM*, vol. 1, pp. 126–134, 1999.

[30] A. May, *Traffic Flow Fundamentals*. Prentice Hall, 1990. [Online]. Available: https://books.google.com/books?id=JYJPAAAAMAAJ

[31] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo – simulation of urban mobility: An overview," in *SIMUL 2011*, S. . U. of Oslo Aida Omerovic, R. I. R. T. P. D. A. Simoni, and R. I. R. T. P. G. Bobashev, Eds. ThinkMind, October 2011. [Online]. Available: https://elib.dlr.de/71460/

[32] R. C. Chalmers and K. C. Almeroth, "Developing a multicast metric," *Globecom '00 - IEEE. Global Telecommunications Conference. Conference Record (Cat. No.00CH37137)*, vol. 1, pp. 382–386 vol.1, 2000.