Dynamic Configuration of Single Frequency Networks in Mobile Streaming Systems

Saleh Almowuena
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada

Mohamed Hefeeda
School of Computing Science
Simon Fraser University
Burnaby, BC, Canada

ABSTRACT

Although the capacity of cellular networks has increased with recent generations, the growth in demand of wireless bandwidth has outpaced this increase in capacity. Not only more users are relying on wireless networks, but also the demand from each user has substantially increased. For example, it has become common for mobile users to stream full TV episodes, sports events, and movies while on the go. Further, as the capabilities of mobile devices improve, the demand for higher quality and even 3D videos will escalate, which will strain cellular networks. Therefore, efficient utilization of the expensive and limited wireless spectrum remains an important problem, especially in the context of multimedia streaming services that consume a large portion of the wireless capacity. In this paper, we introduce the idea of dynamically configuring cells in wireless networks to form single frequency networks based on the multimedia traffic demands from users in each cell. We formulate the resource allocation problem in such complex networks with the goal of maximizing the number of served multimedia streams. We prove that this problem is NP-Complete, and we propose a heuristic algorithm to solve it. Through detailed packet-level simulations, we show that the proposed algorithm can achieve substantial improvements in the number of streams served as well the energy saving of mobile devices. For example, our algorithm can serve up to 40 times more users compared to the common unicast streaming approach, and it achieves at least 80% and up to 400% improvement compared to multicast approaches that do not use single frequency networks.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Wireless communication, Network topology*

General Terms

Design, Performance, Measurement

Keywords

Mobile multimedia, single frequency network, wireless streaming

MMSys '15, March 18–20, 2015, Portland, OR, USA
Copyright 2015 ACM 978-1-4503-3351-1/15/03 ...$15.00
http://dx.doi.org/10.1145/2713168.2713179.

1. INTRODUCTION

The demand for multimedia streaming over mobile networks has been steadily increasing in the past several years. According to a recent report by Cisco [1], the data traffic over mobile networks was equivalent to 1,500 petabytes per month in 2013. It is expected that this traffic will increase 10 times to reach 15,000 petabytes per month by the end of 2018. The same report predicted that around 70% of this traffic will most likely carry videos. In order to cope with this growing demand, cellular service providers may need to rely on multicast capabilities of current and future cellular networks whenever possible. Currently, the WiMAX standard defines Multicast and Broadcast Service (MBS) in the data link layer in order to facilitate the process of initiating multicasting and broadcasting sessions [2]. Similarly, the evolved Multimedia Broadcast Multicast Services (eMBMS) allows LTE cellular networks to deliver video streams over multicast groups [3]. With these multicast-capable networks, a streaming server can substantially reduce the wireless network load by serving mobile devices interested in the same video stream using a single multicast session. For example, a major telecommunication operator in the United States (Verizon) used multicast during the 2014 Super Bowl in New Jersey to serve multimedia content to more than 30,000 customers, which consumed about 1.9 TB [4].

As defined in recent 4G standards, e.g., [5], multicast can be provided in two modes, which we refer to as independent and single frequency network (SFN). The independent mode provides multicast transmission within a single cell without any coordination or cooperation from neighboring cells. The SFN mode, however, represents a coordinated effort made by a set of base stations in order to transmit multimedia streams while minimizing the consumed wireless network resources. All base stations use the same frequency for the multicast sessions. Transmitting using SFN leads to significant improvements in the utilization of the wireless resources compared to transmitting using the independent mode. This is because in the SFN mode the coordinated cells are sending using identical radio signals, and thus receivers at the cell edges can get multiple copies of the same data but from different base stations. While these copies are considered as *inter-cell interference* in independent cells, they are translated into useful signal energy in SFN. Hence, the strength of the received signal at the cell edge is enhanced, and the interference power at the same time is largely reduced. More information on how a single frequency network manages its resources and operates in general can be found in [5, 6].

Achieving the potential gains from multicast transmissions over SFN depends on finding the optimal configuration of cells within the SFN as well as adapting this configuration to handle the dynamic nature of the multimedia traffic and the users requesting this traffic. Although several works addressed various aspects of SFNs, such as coverage and modulation schemes [7, 8], and the size of an SFN and its impact on packet delivery [9], none of the previous works considered the much more challenging problem of managing the resources of multi-cell single frequency networks in dynamic environments where the network traffic and users distribution change with time; which is the problem we address.

In this paper, we consider a general model for wireless cellular networks that support multicast services, such as LTE and WiMAX. In this model, the network is composed of multiple cells. These cells can work independently from each other so that each cell can provide unicast and multicast services to users in its range. Cells can also collaborate by forming one or more SFNs. If a subgroup of cells forms an SFN, a portion of the bandwidth is reserved for the multicast service in all cells, and the multicast service will be provided to all users within the range of this SFN. The cell membership in an SFN is dynamic, which means a cell can join or leave an SFN based on the demand from its current users. The problem we address is given user demands for different video streams in various cells, determine the optimal configuration of the wireless network that maximizes the number of users served and the energy saving for mobile devices. More specifically, decide the number (zero or more) of SFNs that should be created, and which cell should belong to which SFN. Furthermore, for each user request for a video session, decide whether it should be served using unicast, multicast within a single cell, or multicast across an SFN. This is a challenging problem; in fact, we prove that it is NP-Complete. We simplify this problem and propose a two-stage heuristic algorithm to solve it, which substantially improves the service ratio compared to current algorithms used in cellular networks.

The contributions of this paper can be summarized as follows:

- We introduce the idea of dynamically configuring cells in wireless cellular networks to form single frequency networks based the traffic demands from users in each cell.

- We formulate the resource allocation problem in multi-cell SFNs to serve multiple multimedia streams using various combinations of unicast and multicast sessions within each cell and across SFNs. We show that this problem is NP-Complete.

- We present a heuristic algorithm to solve the multi-cell SFN resource allocation problem. The algorithm has two stages. The first stage is used by each base station to independently decide whether to form an SFN or join an existing one. The second stage computes the best option to serve each multimedia stream, whether unicast, multicast, or combination thereof.

- We conduct an extensive simulation study using a detailed packet-level simulator (OPNET) to evaluate the proposed algorithm. Our results show that the proposed algorithm can serve up to 40X more users than the common unicast streaming approach. Compared to multicast approaches that do not use SFN, our algorithm can achieve up to 400% improvements in the number of users served. Even compared to approaches that do use SFN but do not configure them dynamically, our algorithm achieves up to 18% improvement in the number of users served.

The rest of this paper is organized as follows. Section 2 summarizes the related works in the literature. Section 3 describes the system model used in this paper, and Section 4 states and formulates the considered problem. Section 5 presents the proposed algorithm, and Section 6 presents our simulation results to assess the performance of our algorithm and compare it against others. Section 7 concludes the paper.

2. RELATED WORK

Several works have been introduced to assess the performance of multimedia multicast streaming over wireless cellular networks. For instance, Rong et al. [7] and Talarico and Valenti [8] present analytical models to determine the coverage of a given configuration for single frequency networks and how to utilize these models to choose the best-suitable modulation and coding scheme as well as the appropriate configuration for SFN areas. Having such knowledge prior to the network deployment helps in achieving a target bandwidth utilization. Urie et al. [10] extend this assessment and provide comprehensive evaluation of SFN performance under more realistic conditions. Alexiou et al. [9] estimate the number of neighboring cells that should be enrolled into an SFN area such that a specific average signal-to-noise ratio is achieved and a minimum communication cost is incurred. To accomplish this goal, they calculate the cost of both packet delivery and signaling procedures under a set of different network topologies and user distributions. The works [7–10] assume a static SFN configuration in which cells are registered into certain set of zones at early stages of deployment, and the enrollment of these cells do not change over time even if variations have been occurred for users distribution and network traffic. Here, we consider dynamic configuration of SFN areas.

Given a particular configuration of SFN, the available radio resources should be allocated to a mixture of unicast and multicast services to optimize network utilization. Although a number of algorithms have addressed the transmission scheduling problem for multicast services [11–14], a few research efforts have considered the hybrid unicast-multicast approach in the allocation problem. For example, Monserrat et al. [15] and Lee et al. [16] present two schemes in which both unicast and multicast connections are served with an objective of reducing the service blocking probability, whereas Rahman et al. [17] utilize the hybrid approach to minimize the average power consumption of mobile terminals. Our proposed resource allocation algorithm in this paper aims at increasing the bandwidth utilization of a cellular network, but it is different from the works in [15–17] in two main aspects: a) it utilizes an adaptive and flexible scheduling process, and b) it takes an advantage of three transmission modes. The fraction of radio resources reserved for multicast services is assumed to be constant in most existing scheduling algorithms, including those schemes in [11–17]. In our case, the resource distribution between unicast and multicast connections is done dynamically based on which served request leads to better spectral efficiency. Our proposed transmission scheduler is also an adaptive allocator because the modulation and coding scheme used for multicast services may not be suitable for those users with the worst channel conditions, especially in scenarios where their quantity is low. To offer the flexibility of resource distribution, we are exploiting three type of transmission: unicast, multicast over an SFN, and multicast within the local coverage of a cell. Wireless cellular networks have reserved sub-frames for SFN transmission, and these standards cannot easily be changed. Therefore, we increase the transmission scheduling for multicast services by delivering some multicast connections through the available resources for unicast.

3. SYSTEM MODELS

In this section, we present two models that are considered in this paper. The first describes the model for the cellular network, and

Table 1: Symbols used in this paper.

Symbol	Description
T	No. symbol columns in an allocation window
S	No. subchannel columns in an allocation window
d	Fraction of resource blocks reserved for videos
Γ	Duration of a resource allocation window
V	No. videos available for streaming within the system
r_v	The encoding rate of a particular video v
Z_v	No. segments into which a video v is divided
n_{vz}	No. mobile devices watching the segment z of a video v
M	No. Modulation and Coding Scheme (MCS) modes
c_m	Per-block capacity with a modulation and coding scheme m
$x_{v,m,z}$	Whether segment z of video v is sent MCS mode m
N	Total no. mobile devices within the coverage of a cell c
$n_{v,m,z}$	No. mobile devices watching segment z of video v with m
A_L	Upper limit of SFN areas in which a cell c can enroll
$y_{b,c}$	Whether cell c is enrolled into SFN area A_b or not
$g_{v,z}$	Multicast group in which segment z of video v is transmitted
$SG_{v,z,m}$	Subgroup prefers to watch segment z of video v with m
$A_{v,z,m}$	The weight assigned to the multicast subgroup $SG_{v,z,m}$

the second is for the multimedia streaming service. Table 1 lists all symbols and their definitions.

3.1 Wireless Network Model

We consider a wireless cellular network with base stations, mobile devices, and multi-cell coordination entities as illustrated in Figure 1. In addition to transmitting via unicast connections, our network utilizes two modes for multicast transmissions. The first mode is the single-cell point-to-multipoint transmission, which allows a feedback from mobile terminals on their channel conditions and then dynamically adjusts their suitable modulation and coding schemes based on these feedbacks. The advantage of such mode is its adaptation to any change in the current distribution of users within a cell. Multicast services using this mode can be turned-off within a particular cell in which there are no active users. The second mode of multicast transmission is the multi-cell point-to-multipoint approach, and it is called multicast over an SFN. A single frequency network represents a coordinated set of base stations in order to broadcast multimedia streams over a region of the network utilizing the same physical radio resources. To achieve such objective, a fixed modulation and coding scheme is applied to match the decoding requirements of the edge-user with the worst channel condition. Transmitting multicast services through an SFN leads to significant improvements in the total spectral efficiency as compared to multicasting within a single-cell [5]. Since the coordinated cells are sending using an identical radio signals, receivers at the cell edges are getting multiple copies of the same data but from different base stations. Hence, the strength of the received signal at the cell edge is enhanced, the interference power is largely reduced, and the overall performance remains consistent even If a user moves from one cell to another.

The mobile system shown in Figure 1 is divided into a number of SFN areas. A multi-cell coordination entity (MCE) ensures the full functionality of an SFN area by performing time synchronization as well as coordinating the usage of the same radio resources and transmission parameters across all cells belonging to a particular area. It is possible to define many SFN areas at the same time where base stations are capable of joining 1 to A_L areas. The radio resources of a base station are divided along both time, represented by sub-frame, and frequency, represented by sub-carrier, domains. Let the number of sub-carriers allocated to each cell be S. The resource allocation window has T sub-frames, indexed by t, and has a duration of Γ. The smallest resource unit (resource block) in frequency-time space is identified by (s, t), where $s \in [1, S]$ and $t \in [1, T]$. While each allocation window consists of TS resource blocks, we use d fraction of bandwidth, i.e., dTS resource blocks, for video services. The user equipment is informed about which sub-frames are assigned to its video stream via a control channel broadcasted from its nearest base station, and the allocation can be changed dynamically at specified intervals. We assume that the maximum power for each base station is P and these base stations allocate equal power to their sub-carriers [18]. Hence, the sub-carrier power allocation is $\frac{P}{S}$.

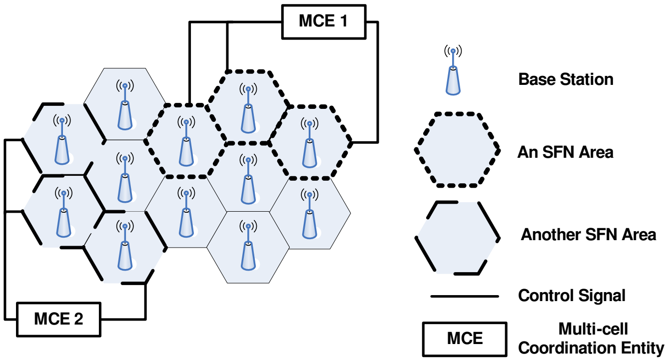


Figure 1: The considered model for a mobile network.

In this paper, we propose a novel approach to reconfigure SFN areas in a wireless cellular network based on video popularity and user distribution. To accomplish this objective, we assume that there are three possible types of cells within an SFN area, as shown in Figure 2. A cell can be either actively enrolled, un-enrolled or passively enrolled in a certain area. Both actively and passively enrolled cells transmit a video stream through an identical radio resource channel. The main difference between both cells is that there are a number of users receiving this stream in the active cell, whereas passively enrolled cells have no interested users in this video stream. Normally, passively enrolled cells are formed only in scenarios where there is low traffic within a period of time, so these cells would be merged into an active SFN area in order to enhance its quality of services and strengthen its broadcast signals. On the contrary, an un-enrolled cell does not belong to the multi-cell transmission, and it should have no impact on the signals broadcasted within its neighboring SFN cells. For this reason, an un-enrolled cell always tries to avoid causing any inter-cell interference by using the radio resources utilized for its neighboring SFN area for only unicast connections or single-cell point-to-multipoint services with a limited power. To determine the type of a certain cell in a mobile network, we assume that each base station works along with its nearest multi-cell coordination entity to make such decision, and this decision is made regularly based on certain conditions as explained in Section 5.1.

Similar to any wireless communication, our mobile network is vulnerable to environmental phenomena such as shadowing, multipath fading, and interference. The channel qualities between mobile terminals and their nearby base-station vary due to these phenomena as well as the mobility of viewers. To adapt to different channel qualities during data transmission and to allow the reception with an acceptable bit-error-rate, multiple modulation and coding schemes (MCS) are used as defined in OFDMA standards. The MCS mode is determined based on the wideband channel quality index report [19]. Higher MCS modes require good channel quality and lead to higher per-resource block capacity. On the other hand,

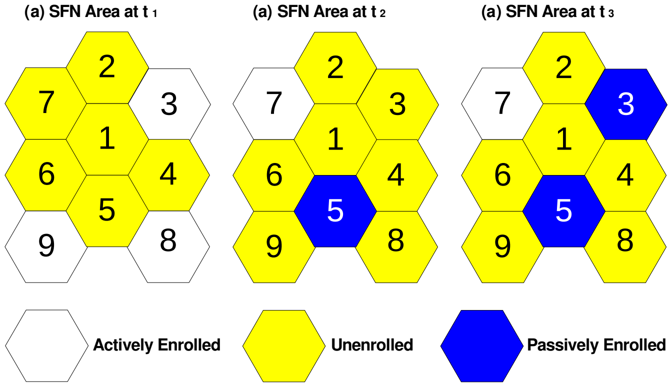


Figure 2: Three possible types of cells within an SFN area.

smaller MCS modes are more robust and usable for diverse (both good and bad) channel qualities. For our system, the MCS mode is denoted by $m \in [1, M]$. Let n_m be the number of mobile terminals in a cell that can receive the data at MCS m or better. Here, $n_1 \geq n_2 \geq ... \geq n_M$. Similarly, $n_{vzm} : v \in [1, V], m \in [1, M]$ is the number of mobile terminals in a viewers' subgroup that can receive (buffer) the video-segment (v, z) at MCS m or better, where $n_{vzm} \geq n_{vzm'}$ if $m < m'$. The per-resource block capacity (c_m) is non-decreasing quantity on $m \in [1, M]$ such that $c_1 \leq c_2 \leq ... \leq c_M$.

3.2 Streaming Model

We consider a general streaming model that can be used for live as well as on-demand streaming with some constraints. Live streaming is useful in several scenarios such as streaming sports events, live concerts, news, political debates, talks, seminars, and popular TV episodes. Live streaming is naturally suitable for multicast services as users are mostly synchronized: they are watching at the same moment in the video and functions that may disrupt this synchrony, e.g., fast forward, are not applicable. In addition, live streaming of popular events typically attracts a large number of *concurrent* users, which can put a huge load on the cellular network and may result in denying some users the streaming service because of the limited capacity. Our work in this paper optimizes the wireless resources for live streaming by carefully creating a mixture of unicast and multicast sessions, which can be within single cells or across multiple coordinated cells.

In on-demand streaming, on the other hand, users arrive asynchronously to the system. That is, users may request the same video at different times and they can be watching at different moments in the videos. This general asynchronous model for streaming is difficult to achieve using pure multicast as very few users can form a multicast session especially if they are requesting videos that are not popular. We consider a less general model, which is useful for requesting popular videos on relatively short periods of time, such as requesting news clips during morning or afternoon commute times, and streaming TV episodes during the evening peak watching times. This model is also useful for time-shifted viewing of various events and videos, where some users opt to watch such videos at different times than their original scheduled times. Even for such limited asynchronous model, multicast alone will be not be sufficient to provide true-on-demand service without imposing long waiting times on users. To solve this problem, existing delay mitigation mechanisms, such as patching videos using temporary unicast connections [20–22] may be adopted. For example, we implement an efficient patching algorithm in which a user can join any existing multicast group and receive the missing leading portions of this stream over a separate unicast connection. This temporary unicast session is shorter than a predetermined threshold; otherwise, a new multicast group for the requested video may be created.

In our streaming model, mobile terminals within a cell can request V different video streams. Mobile terminals receiving the same video streams may watch at different timestamps relative to that video. To facilitate these demands, each video stream $v \in [1, V]$ is divided into a number of segments $z \in [1, Z_v]$, and each video-segment has a play-out duration Γ, which equals to the allocation window duration. Let $r_{vz} : v \in [1, V], z \in [1, Z_v]$ be the encoding rate of the video's segment (v, z). Because of the variable bit rate nature of video streams, r_{vz} may vary across allocation windows. Since the aggregate data rates for the requested segments, $\sum_v \sum_z r_{vz}$, in an allocation window may exceed the video service capacity, dTS, we need to decide which segments to transmit for each video with the objective of maximizing the service ratio among users.

4. PROBLEM STATEMENT

In this section, we define and mathematically formulate our problem, which we divide into two sub-problems: the first determines the best configuration of single frequency networks, and the second specifies the transmission scheduling for multimedia streams using a mixture of unicast and multicast sessions.

4.1 Problem Definition

Several wireless cellular networks consider static configurations for single frequency networks [5]. However, static configurations are unaware of user distribution and video popularity across the mobile network. Therefore, these approaches may waste the radio resources of cells, especially in those scenarios where no mobile terminals are interested in a cell belonging to a predetermined multicast service. Dynamic configuration of SFNs provides more flexibility and thus can yield higher efficiency in using the radio resources. We consider an initial SFN configuration consisting of a number of hexagonal cells in which a set of mobile terminals are distributed within their transmission coverages. The information of transmitted videos in each cell along with the number of active viewers is periodically delivered to the nearest multi-cell coordination entity in the given wireless network. Our first problem in this paper can be stated as follows:

PROBLEM 1 (SFN CONFIGURATION). *For a given cell c, select the optimal subset of SFNs to join so that the: (i) obtained bandwidth utilization within this cell is maximized, (ii) total transmission cost of control signals is minimized, (iii) average single to noise ratio at a certain instant of configuration is greater than or equal to its value at a subsequent reconfiguration, and (iv) number of SFN zones in which c is enrolled does not exceed a predetermined limit A_L.*

Under any given SFN configuration, base stations and their nearest multi-cell coordination entities should cooperate to allocate the available resource blocks for both multicast and unicast connections. Here, there is another issue that still exists for the transmission scheduling over OFDMA, which is mainly related to the different data rate and quality requirements of users in the same multicast group. Generally, mobile terminals close to the base station can obtain higher data rate, while cell-edge users are forced to reduce the data rate in order to minimize the bit error rate in data reception. Conventional multicast schemes adopt a conservative approach, which restricts the rate of the multicast session to the user with the worst channel condition. This approach introduces severe inefficiencies when some users (even if they are just a few) experience poor channel conditions. The objective of our proposed transmission scheduling algorithm is to address the aforementioned inefficiency problem of multicast communications in the presence of link quality differences among users within a multicast group. Instead of transmitting the video stream to a multicast group at a very low bit rate, it could be more efficient to eliminate some users from the multicast group and serve these users with unicast streams so that they do not slow down all users in the multicast session. Thus, our second problem in this paper is to consider a joint multicast-unicast transmission scheduling for bandwidth-efficient delivery of the requested videos to mobile terminals. In this approach, a base station dynamically handles an asynchronous incoming request for a certain multimedia file by either initiating a unicast stream or extending the number of participants in a multicast session. In particular, a base station along with its assigned multi-cell coordination entity need to compute a schedule that specifies: (i) which video streams to multicast, (ii) who can be enrolled into the multicast groups, and (iii) which video streams to unicast. This joint multicast-unicast problem can be stated as follows:

PROBLEM 2 (HYBRID TRANSMISSION SCHEDULING). *Given an allocation window of T sub-frames and S sub-carriers, determine the optimal transmission schedule for each video request R_v received by mobile terminals of diverse channel conditions and using a hybrid multicast-unicast streaming approach, so that the: (i) total service ratio across all mobile terminals is maximized and (ii) no more than dTS blocks are consumed by the video service, where* d *is the fraction of resources allocated to the mobile multimedia service.*

4.2 Problem Formulation

Problems 1 and 2 have a common objective since they are aiming at maximizing the number of served users in the given mobile network. On this ground, the two problems can be formulated into a single optimization problem, and this problem formulation must consider the trade-off between the number of served users at a certain modulation and coding scheme, i.e, n_{vzm} and their resource requirements, i.e., $\frac{r_{vz}\Gamma}{c_m}$. We use the Boolean decision variable x_{vzm} ($v \in [1, V]$, $z \in [1, Z_v]$, $m \in [1, M]$) to denote whether segment z of video v transmitted using MCS mode m. That is, $x_{vzm} = 1$ if the video-segment (v, z) is transmitted using MCS mode m, and $x_{vzm} = 0$ otherwise.

We present the problem formulation in Eq. (1). The objective function of Eq. (1a) is to maximize the total number of served users within a cell. Eq. (1b) implements the resource constraint and also ensures admission control for video segments. Eq. (1c) ensures that at most one MCS mode is selected for each layer of a video segment. Eq. (1d) ensures that the obtained bandwidth utilization at a certain instant of SFN configuration (i) is greater than its value at a subsequent reconfiguration $(i + 1)$. This condition helps in avoiding ineffective reconfigurations for SFN and limits the number of signaling control. Finally, Eq. (1e) makes sure that the number of SFN zones in which c is enrolled does not exceed the maximum limit of zones A_L in which cell c is allowed to join, keeping into account that B refers to the number of active SFN areas, $y_{b,c}$ is a binary number introduced to determine whether c is enrolled in the area b or not.

Solving this optimization problem, we can find the decision matrix $\mathbf{X}$ which indicates the group sets of video segments selected for transmission and corresponding MCS modes to maximize the service rate of a video streaming service.

$$\max_{\mathbf{x}} \quad O = \sum_{v=1}^{V}\sum_{z=1}^{Z_v}\sum_{m=1}^{M} x_{vzm} n_{vzm} \tag{1a}$$

$$s.t. \qquad B = \sum_{v=1}^{V}\sum_{z=1}^{Z_v}\sum_{m=1}^{M} x_{vzm}\frac{r_{vz}\Gamma}{c_m} \le dTS \tag{1b}$$

$$\sum_{m=1}^{M} x_{vzm} \le 1 \tag{1c}$$

$$\frac{O(i)}{B(i)} > \frac{O(i+1)}{B(i+1)} \tag{1d}$$

$$\sum_{b=1}^{B} y_{b,c}^{t_i} \le A_L \tag{1e}$$

$$x_{vzlm} \in \{0,1\}, \forall\, v \in [1,V], z \in [1,Z_v], m \in [1,M] \tag{1f}$$

4.3 Problem Complexity

The formulation of Eq. (1) is an integer programming problem, which is a NP-Complete. To prove that, let us define the set of decision variables $a_{i,j,k}$ for all $i \in [1, S], j \in [1, T]$, and $k \in [1, V]$ in the transmission scheduler such that $a_{i,j,k} = 1$ if the resource block in i^{th} subchannel and j^{th} symbol is allocated to video k and 0 otherwise. Given an allocation assignment represented by the above variables, we can verify whether it is a valid assignment. We need to count the number of 1's and make sure that the count is $\le dTS$. In order to ensure that the same resource block is not allocated to more than one video, we need just to check that for any $i \in [1, S], j \in [1, T]$, and $\sum_{k=1}^{V} a_{i,j,k} \le 1$. This checking operation can be done in polynomial time.

Now, we will use the 0-1 knapsack problem in order to show that an NP-Complete instance is reducible to our problem. In the 0-1 knapsack problem, we are replacing the variables v with p and w with b to avoid confusion with our defined variables. Therefore, our instance of 0-1 knapsack is defined as following: there are n items x_l such that $l \in [1, n]$ and $x_l = 1$ if the item is chosen and 0 otherwise. The value and weight of item x_l are defined by p_l and b_l, respectively. The capacity of the knapsack is W. We assume non-negative values and weights. We have to choose among these n items to maximize the value without violating the capacity of the knapsack. Mathematically, we would like to maximize $\sum_{l=1}^{n} x_l p_l$ subject to $\sum_{l=1}^{n} x_l b_l \le W$ and $x_l \in \{0, 1\}$. To reduce this instance of 0-1 knapsack to an instance of our problem, we set $n = TSV$. We define a new variable $x'_{i,j,k}$ for each x_l for any $i \in [1, S], j \in [1, T]$, and $k \in [1, V]$ such that $x'_{i,j,k} = 1$ if the resource block in i^{th} subchannel and j^{th} symbol is allocated to video k and 0 otherwise. We introduce another variable $p'_{i,j,k}$ for each p_l such that $p'_{i,j,k}$ denotes the number of mobile devices served by allocating resource block $x'_{i,j,k}$. We replace each b_l with new $w'_{i,j,k}$ and set $w'_{i,j,k} = 1$ for all $i \in [1, S], j \in [1, T]$, and $k \in [1, V]$. Finally, we set the knapsack capacity to be $W = dTS$. This reduction can be done in polynomial time. The reduced 0-1 knapsack problem will have a solution if and only if our considered problem has a solution.

5. PROPOSED SOLUTION

We propose a heuristic algorithm to solve the problem defined in Eq. (1). The algorithm performs two main steps: 1) dynamic configuration for the single frequency network, and 2) transmission scheduling of incoming video requests received within a predefined scheduling window. The first step reconfigures a network and reconstructs its SFN areas dynamically by taking into consideration the popularity of videos and the signal-to-noise ratios of served terminals. This reconfiguration is established periodically based on a pre-determined threshold, which represents the variation of current traffic and channel conditions. The second step schedules incoming requests with an objective of maximizing the service ratio in a given system. This scheduling is done at every resource allocation window. The details of each step are described in the following subsections.

5.1 Dynamic SFN Configuration Algorithm

To reconfigure the areas within a single frequency network, it is possible to allow every multi-cell coordination entity to collect both current traffic and user distribution within its coverage and then perform a *centralized* process to search for the optimal set of areas and reconstruct itself accordingly to the obtained findings. We avoid such centralized operations and propose a *coordinated* algorithm in which a base station can dynamically help in determining whether an SFN reconfiguration is required or not. This algorithm is illustrated in Figure 3. Four different decisions are performed: a) expanding the number of areas in which cell c is enrolled to accommodate additional multicast services, b) joining an unrequested multicast session to strengthen the transmission coverage of neighboring cells, c) replacing an existing video stream with another one, and d) shrinking the number of areas in which cell c is enrolled and switching its mode for certain radio resources from SFN to single-cell mode.

In the proposed algorithm, cell c periodically sorts its ongoing multicast sessions in an ascending order and its unserved incoming video requests in a descending order. The sorting is conducted based on the estimated bandwidth utilization of each stream in the two sets. Once this phase is accomplished, cell c tries to improve the service ratio within its cell by rearranging its enrollment in the current SFN areas but without causing frequent usage of its control signals. Four types of control overhead are considered in the computation of signaling cost: C_{syn} represents the cost of conducting a synchronization process for coordinated cells, C_{poll} refers to the cost of counting interested clients for a certain multicast service, C_{init} defines the cost of initiating a new multicast session, and C_{stop} is the cost of ending an existing multicast service and releasing its allocated resources. To achieve our objective, cell c begins its examination by checking if there is enough bandwidth for multicast services over SFN (i.e. $W_M > 0$) in order to expand its offered multicast sessions. In the cases where c does not exceed the upper limit of allowed SFN areas and $W_M > 0$, cell c starts its attempts with the unserved video request whose spectral efficiency is the highest. Two possible options in this scenario can be predicted: 1) cell c joins an active area where this video is broadcasted and 2) c enrolls in a zone where there are enough resource blocks such that a new multicast session can be initiated. Enabling cell c to go with either options necessitates the use of both synchronization and initiation control signal, thereby costing the network C_{sys} and C_{init}, respectively. Polling signals are also required in the latter option to announce the new service in all enrolled cells and count how many users are interested in receiving it. This polling process is expected to cost C_{poll}. We assume these signaling values vary from an area to another based on the number of its active cells and their distances from the corresponding multi-cell coordination entity. When cell c explores all possible cases, it will select the SFN area that helps in maximizing Eq. (1a).

Sometimes, it is probable for a cell c to enjoy non-rush hours where a high number of radio resource are available and no incom-

Algorithm 1: Dynamic Configuration of SFN Areas

```
Inputs : $MS_c \leftarrow$ set of multicast services in a cell $c$
    $US_c \leftarrow$ set of unserved streams in a cell $c$
    $W_M \leftarrow$ bandwidth still available for multicasting over SFN
    $A_B \leftarrow$ set of active SFN areas in which cell $c$ can join
    $A_c \leftarrow$ set of active SFN areas in which cell $c$ is enrolled
    $\{\alpha, \lambda\} \leftarrow$ parameters for the user behavior model
    $SE() \leftarrow$ function computes for a video its ratio of served
        users to consumed bandwidth in given configuration
Output : $D \leftarrow$ the obtained decision of SFN reconfiguration

$MS_c \leftarrow sort(MS_c, SE(), Ascending)$
$US_c \leftarrow sort(US_c, SE(), Descending)$
if $(((W_M > 0) \cap (US_c \neq \emptyset)) \cap (|A_c| < A_L))$ then
    $CS_c \leftarrow US_c.pop()$
    for $x \in [1, |A_B|]$ do
        $JV_x \leftarrow SE(CS_c)$ when $(A_c \in A_x)$
        $JC_x \leftarrow C^{A_x}_{sys} + C^{A_x}_{init} + (C^{A_x}_{poll}$ only if $A_c \notin A_x)$
    end for
    $A_O \leftarrow \max_{1 \leq x \leq |A_B|} (JC_x / JV_x)$
    $W_M \leftarrow W_M - W_{CS_c}$
    $MS_c \leftarrow MS_c + CS_c$
    Expand $A_C$ to enroll $c$ into $A_o$ in order to multicast $CS_c$
return $D = \text{‘Expand’}$
else if $(((W_M > 0) \cap (US_c = \emptyset)) \cap (|A_c| < A_L))$ then
    for $x \in [1, |A_B|]$ do
        $CS_x \leftarrow MS_x.pop()$
        $JV_x \leftarrow e^{-\lambda} \times TE_{CS_x} \times (SE(CS_x)$ when $(A_c \in A_x))$
        $JC_x \leftarrow C^{A_x}_{sys} + C^{A_x}_{init} + C^{A_x}_{poll}$
    end for
    $A_O \leftarrow \max_{1 \leq x \leq |A_B|} (JC_x / JV_x)$
    $W_M \leftarrow W_M - W_{CS_x}$
    $MS_c \leftarrow MS_c + CS_x$
    Join $c$ into $A_o$ to support multicasting $CS_x$
return $D = \text{‘Join’}$
else if $((W_M < 0) \cap (US_c \neq \emptyset))$ then
    $CS_c \leftarrow US_c.pop()$
    $RS_c \leftarrow MS_c.pop()$
    while $SE(CS_c) > SE(RS_c)$ do
        $C_{CS_c} \leftarrow \frac{(C^{RS}_{stop} + C^{CS}_{sys} + C^{CS}_{init} + C^{CS}_{poll}) \times (CS_c)^\alpha}{SE(CS_c) \times TE_{CS_c}}$
        $C_{RS_c} \leftarrow \frac{(RS_c)^\alpha}{SE(RS_c) \times TE_{RS_c}}$
        if $(C_{CS_c} < C_{RS_c})$ then
            $W_M \leftarrow W_M + W_{RS_c} - W_{MS_c}$
            $MS_c \leftarrow MS_c - RS_c + CS_c$
            Replace $RS_c$ with $CS_c$
            return $D = \text{‘Replace’}$
        else if $(\frac{C^{RS}_{stop} \times CS^\alpha_c}{SE(CS_c) \times TE_{CS_c}} < \frac{RS^\alpha_c}{SE(RS_c) \times TE_{RS_c}})$ then
            $W_M \leftarrow W_M + W_{RS_c}$
            $MS_c \leftarrow MS_c - RS_c$
            Shrink $c$ from $A_x$ in which $RS_c$ is multicasted
            return $D = \text{‘Shrink’}$
        else
            $CS_c \leftarrow US_c.pop()$
        end if
    end while
end if
return $D = \text{‘None’}$
```

Figure 3: Proposed algorithm for reconfiguring an SFN.

ing requests are anticipated. For this reason, our algorithm aims at utilizing this unexploited bandwidth to improve the overall signal-to-noise ratios within the mobile system. Cell c in such scenarios would be called a passively enrolled cell, as explained in Figure 2. To ensure taking a full advantage of this passive enrollment, the multi-cell coordination entity retrieves the most spectrally efficient multicast streams within each area x of the available SFN zones, A_B, and then analyzes the gain achieved by joining cell c into the SFN area A_x, where $x \in [1, |A_B|]$. In our calculation for this gain, we are following the same formula in Eq. (1a). Yet, two additional parameters are introduced to avoid any extensive calls for reconfiguration. The first parameter depends on the arrival rate of users within the cell c. Here, we model the request arrival process in a video service using a Poisson distribution with an arrival rate λ, which is defined by: $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$. To allow c to act as a passively enrolled cell, it should have no outstanding traffic within that period of time. In other words, k should be equal to 0, leading $P(k = 0) = e^{-\lambda}$. The second parameter is related to the remaining duration of the most spectrally efficient stream CS within the SFN area x. We denote this time by TE_{CS_x}, and use this parameter to give multicast streams with longer estimated playing time higher priorities than those short video sessions.

When the allocated bandwidth for multicast services over SFN is not sufficient to serve a set of outstanding video streams, our algorithm enables cell c to assess both active multicast sessions and incoming requests. It then observes its gains and losses from the perspective of achieved service ratio within its coverage. For instance, in the occasions where an incoming request CS_c gives better bandwidth utilization than an existing multicast stream RS_c, cell c should examine the benefit of broadcasting CS_c rather than RS_c by employing the ratio of control signaling cost to the projected spectral efficiency of each operation. Substituting a multicast service over SFN with another video stream requires initiating, synchronizing and announcing the new video CS_c, and it also needs stopping the ongoing transmission of RS_c. Meanwhile, dividing the number of interested users in CS_c by its required resource blocks gives the estimated spectral efficiently of the new video. For a realistic comparison between the two streams, the remaining playing time as well as the popularity of both videos are also taken into account. To model the video popularity in a system, Zipf distribution is often used to characterize the access of viewers. If the video popularity is sorted in a decreasing order, we can assume that among the available titles, the stream CS_c has an access probability given by: $1/(CS_c)^\alpha$, where α is the skew factor of the Zipf distribution. Once the cell c finds that it is not economical to replace RS_c with CS_c, it will test another alternative choice in which the radio resources allocated for RS_c is released and switched to a single-cell mode. Changing the transmission mode from SFN to single-cell involves less amount of control signals, and it is most likely not going to increase the inter-symbol interference if these radio resources are used with low power. We call this operation of switching mode a shrinking process. If both replacing and shrinking approaches are still not cost-effective, cell c is going to discard CS_c and look for another outstanding stream.

5.2 Proposed Transmission Scheduling Algorithm

The transmission scheduler works at radio base stations, and it is responsible of assigning portions of the shared bandwidth to users. At the beginning of every allocation window, the transmission scheduler determines its allocation decisions and then informs each mobile terminal which resource blocks it has been assigned for its multimedia streaming. The duration of this allocation window should be at least equivalent to the transmission time interval (i.e., 1 millisecond as in LTE networks), and it is recommend to be equal to the size of video chunks produced by video encoders (i.e., 2 seconds as in Microsoft Silverlight).

To maximize the service ratio during a transmission scheduling window, we divide the incoming requests into subgroups based on the identification of their required videos, the time instances of the requested segments, and the best suitable modulation and coding scheme for interested users. In other words, mobile terminals asking for the same segment of a certain video and sharing identical channel status condition are clustered together into a single subgroup. If we apply the conventional multicast approach, subgroups requesting the same segment of a certain video stream would be merged into one group and then they would be served accordingly to the modulation and coding scheme suitable for those users with the worst channel condition.

To find the optimal solution that maximizes the service ratio in a scheduling window and then minimizes the service blocking in the system, we can use CPLEX [23] along with MATLAB to examine all possible combinations of these subgroups. After that, it decides which users would be served through multicast sessions, which users would be not admitted during this window, and which users would be served via unicast connections. The drawback of using this optimal solution is its exponential running time in the worst case. For this reason, it is necessary to have a fast algorithm with reasonable complexity in its running time to solve the problem of maximizing the average service ratio in video services.

The proposed greedy algorithm is presented in Figure 4. Once the transmission scheduler receives a set of incoming requests from mobile terminals within its transmission coverage, it creates a number of subgroups where each subgroup is identified by the segment number, its video identification, and its best suitable modulation and coding scheme. Each subgroup is also given a weight which is determined by two parameters: 1) the number of resource blocks needed to transmit this segment, and 2) the number of possible users who are able to receive at this modulation and coding scheme and at the same time requesting this particular segment of video. Multiplying the former with the latter parameter gives a weight which is used in prioritizing the segments and then determining which set of them are chosen to be served during the current scheduling window. After constructing the subgroups of every required segment in the scheduling window, the proposed algorithm merges these small subgroups into more confined groups where the users of every group are requesting the same segment but might have different channel conditions. Since those larger groups may have diverse users regarding the channel conditions, the video streams of these groups are transmitted accordingly to the the member with the worst channel condition. These groups are also given weights, where the weight of each group is equivalent to the weight of the subgroup whose modulation and coding scheme is the lowest among its peers.

The number of available resource blocks is usually limited, so it is most likely to have scenarios where some of the merged groups cannot be admitted during the current scheduling window. In these cases, our algorithm aims at choosing the best set of groups that maximizes the number of served users, and thereby minimizing the service blocking. To achieve such objective, our greedy algorithm selects the group with the smallest weight and then eliminates its subgroup whose modulation and coding scheme is the lowest. Once this subgroup is removed, the weight of its own group is recalculated. The scheduler tries again to accommodate the groups in hand. If the available bandwidth is still not enough, the process of removing a subgroup would be repeated until a solution is found.

Algorithm 2: Hybrid Transmission Scheduling Algorithm

```
Inputs :  ${V, Z}$ $\leftarrow$ set of requests for video streams
          $\Gamma$ $\leftarrow$ duration of the scheduling window
Output :  $X$ $\leftarrow$ set of video segments to be served
              during the current allocation window

$W_{Available} \leftarrow dTS$
$W_{Required} \leftarrow 0$
$Queue \leftarrow \{\emptyset\}$
for each required segment $(v, z)$ do
    $n_{(v,z)} \leftarrow 0$
    $g_{(v,z)} \leftarrow \{\emptyset\}$
    for $m \in [M_{MAX}, M_{MIN}]$ do
        $n_{(v,z,m)} \leftarrow$ the number of viewers interested to
                receive this segment using $MCS_m$
        $n_{(v,z)} \leftarrow n_{(v,z)} + n_{(v,z,m)}$
        // Calculating the weight for the current subgroup
        $A_{(v,z,m)} \leftarrow (n_{(v,z,m)} \times c_m \times n_{(v,z)})/(r_{(v,z)} \times \Gamma)$
        $SG_{(v,z,m)} \leftarrow \{n_{(v,z,m)}, c_m, A_{(v,z,m)}\}$
        // Merging this subgroup into its larger multicast group
        $g_{(v,z)} \leftarrow [g_{(v,z)}, SG_{(v,z,m)}]$
    end for
    $Queue \leftarrow Queue + g_{(v,z)}$
    $W_{Required} \leftarrow W_{Required} + (r_{v,z} \times \Gamma / c_{(MCS_{g_{(v,z)}(0)})})$
end for
while $W_{Available} < W_{Required}$ do
    // Retrieving the group with the smallest weight
    $Queue \leftarrow$ sort($Queue$, $A_{g_{(v,z)}(0)}$, $Ascending$)
    $g_{(v,z)} \leftarrow Queue.pop()$
    $W_{Required} \leftarrow W_{Required} - (r_{v,z} \times \Gamma / c_{(MCS_{g_{(v,z)}(0)})})$
    // Eliminating the subgroup with the lowest modulation
    // and coding scheme
    $g_{(v,z)} \leftarrow g_{(v,z)} - g_{(v,z)}(0)$
    $W_{Required} \leftarrow W_{Required} + (r_{v,z} \times \Gamma / c_{(MCS_{g_{(v,z)}(0)})})$
    $Queue \leftarrow Queue + g_{(v,z)}$
end while
$X \leftarrow Queue[1, :]$
```

Figure 4: Proposed transmission scheduling algorithm to maximize the service ratio for a video service over mobile networks.

The proposed greedy algorithm produces a feasible allocation and terminates in polynomial time: $O(N^2 Log(N))$, where N is the number of mobile terminals generating requests for video streams. The while loop in Figure 4 ensures the feasibility of the produced solution by satisfying the constraint in Eq. (1b). Moreover, in each iteration it removes the least profitable streaming ensuring that the algorithm is not trapped into an infinite loop. Hence, the algorithm is correct. The dominating computational complexity of the algorithm occurs in the third loop: (i) the while-loop there iterates at most N^2 times, and (ii) sorting the priority queue consumes $log(N)$ times in its worst-case. Combining these two computational complexity, we can easily find that the time complexity of our proposed algorithm is $O(N^2 Log(N))$.

5.3 Practical Considerations and Discussion

In our proposed algorithm, unicast and multicast sessions are initiated by coordinated effort between base stations and multi-cell coordination entities. Video content providers are assumed to cooperated or even be part of telecommunication operators, which is becoming common nowadays. For instance, two main mobile network providers in Canada have launched their video-on-demand streaming services at the end of 2014 [24,25]. Subscribers to these two streaming services should be active customers to enjoy more than 10,000 hours of content. Since video streams are being delivered simultaneously within the same network, having content caches closer to base stations helps in increasing the chances of creating multicast streams. It also helps in overcoming the probability of facing congestion issues over the delivery path.

Current cellular networks support theoretical peak data rates of 300 Mbps downlink, and future mobile networks aim at achieving a theoretical peak data rate of 3 Gbps downlink. Currently, a single active terminal could get up to almost 75 Mbps when it experiences an excellent channel quality condition, and this rate might be dropped to reach only about 1 Mbps for a cell-edge user, who usually receives a weak signal and suffers from strong interference [5,6]. Based on this ground, another example of how our algorithm might be utilized can be found in major sport games. Large events like the Super Bowl, where tens of thousands of fans gather for hours, and they can put a significant burden on any mobile network [26]. In these cases, a mobile network provider could exploit the concept of multicast over a single frequency networks within the surrounding area of a stadium during a game to deliver football-related content, such as instant replays and highlight videos. Rather than creating individual unicast connection to each incoming request, interested fans can join existing multicast sessions for popular videos.

In addition to increasing the average service ratio within the system, multicast over single frequency network also improves the network coverage and minimizes inter-cell interference. Different from unicast streams, multicast sessions are preceded by an extended cyclic prefix in order to accommodate a longer guard time, thus enabling more SFN signals from distant base stations to positively contribute into the useful signal energy [3]. The use of multicast over single frequency networks is expected to accelerate in the near future since it does not require any changes in the architecture of its underlying network or the terminal of mobile receivers. It also does not need dedicated radio resources. Instead, the spectrum can be allocated to multicast sessions at peak times and within certain regions of interest whenever it is likely to be fully utilized. Once a certain event is over, these frequencies can be easily aggregated into the radio resources used for unicast connections.

6. EVALUATION

6.1 Simulation Setup

Simulator and Algorithms: We simulated a video streaming system using OPNET modeler and its associated LTE Specialized module. OPNET is a detailed packet-level commercial simulator. It consists of a suite of protocols and technologies in order to facilitate the test and demonstration of network designs in realistic scenarios prior to the production phase. The LTE module implements all the details of realistic LTE wireless networks. To evaluate the proposed algorithms, we implemented the unicast approach and three different types of multicast policies. The unicast approach admits interested users based on their order of arrivals. In case of multicast, three policies are used to create multicast sessions: 1) FCFS (first-come first-served) which gives priority to multicast groups formed earlier, 2) Max Users which gives priority to those multicast groups with highest numbers of members, and 3) adaptive resource allocation which follows the proposed algorithm in Section 5.2. We

Table 2: LTE Network Configurations.

Parameter	Value
Physical Profile	LTE 20 MHz FDD
Maximal Transmission Power	0.01 Watt
eNodeB Antenna Gain (dBi)	15 dBi
User Equipment Antenna Gain (dBi)	-1 dBi
Modulation and Coding Scheme (MCS)	4, 8, 14, 22
Evolved Packet System Bearer for Uplink	Best Effort
Propagation Model	Free Space
Scheduling Mode	Link Adaptation
Mobility Model	Random Waypoint

compare the results obtained by our solution with these three methods from different perspectives. We also employ some heuristics to make our simulation setup more realistic. For example, when a mobile device cannot be admitted due to resource limitations, it does not leave the system immediately. Instead, it retries for the intended video up to three times with an exponential back-off waiting period starting at 2 seconds. After being rejected three times, it finally stops requesting its desired video. Another heuristic is the batching of requests such that all requests for videos within the duration of an *allocation window* are grouped together to be served at the beginning of the next allocation window. These heuristics improve the performance of the proposed solution from various aspects and are likely to be incorporated in real video streaming systems.

Wireless Network Configuration: Although the proposed algorithm is applicable to any wireless networks with multicast support, we use the LTE Release-9 standard to evaluate the performance of the proposed algorithm. In Table 2, we list the LTE configuration parameters used for the simulations, unless specified otherwise. Other parameters are set to the default values of the OPNET LTE module. We configure the LTE downlink with Evolved Packet System (EPS) bearers. We define an EPS bearer as a transmission path of defined quality, capacity, and delay [27]. The EPS bearer in LTE delivers bursty data at regular intervals, as scheduled, within Common Subframe Allocation (CSA) period and thus allows mobile devices to turn off the radio circuits between two bursts for saving energy. Moreover, the EPS bearer can be configured with specific quality of service attributes. We adjust the quality of service attributes of EPS bearers to ensure specific MCS mode and bit rate of the video for transmission. Depending on the MCS mode of the bearer, the play time of the burst varies. We choose four MCS modes, i.e., MCS 4, 8, 14, and 22, to support all possible channel qualities. We define four types of bearers with respect to these MCS modes for each of the video streams. According to the proposed algorithm, each video can be transmitted using one bearer. For the assumed bearer configurations and MCS modes, depending on the channel conditions of the mobile devices: (i) MCS 4 to MCS 7 are served by the bearer of MCS 4, (ii) MCS 8 to MCS 13 are served by the bearer of MCS 8, (iii) MCS 14 to MCS 21 are served by the bearer of MCS 14, and (iv) MCS 22 to MCS 28 are served by the bearer of MCS 22. The simulator runs the resource allocation algorithm once every allocation window of 2 seconds.

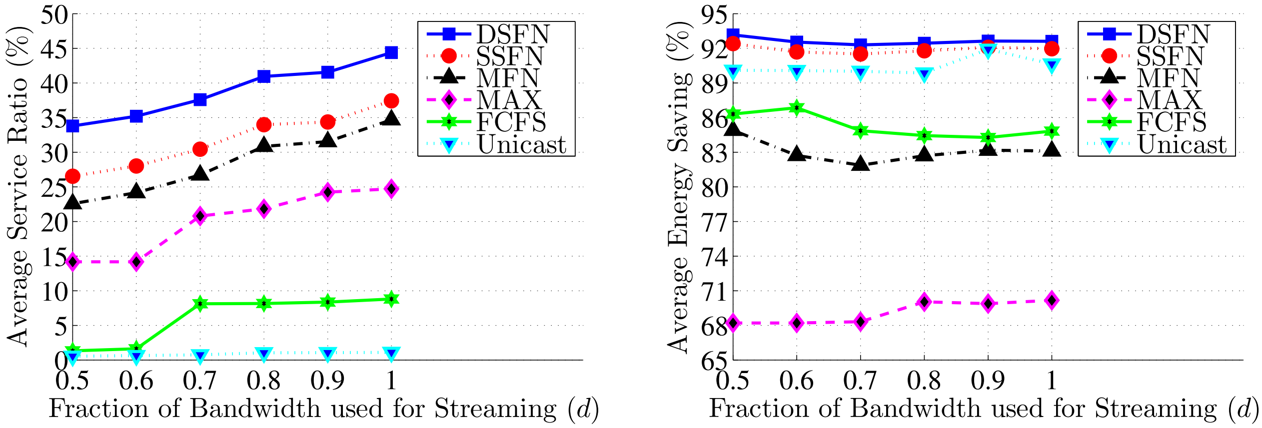
We set the cell size to be around 10 Km by 10 Km by controlling the power of the base station. Each cell is served by one non-sectorized base station, called eNodeB in the LTE standard. The video server has the capability of both multicast and unicast services. The server can be directly connected to the Evolved Packet Core (EPC) or it may be located in the Internet.

User Distribution and Mobility: We assume a population of 1000 users joining the system following a Poisson process with mean λ. λ is a simulation parameter which we set to 20 users per second by default for our simulations. We choose this value to allow users arrive over some time to cover different possible situations. We configure users to move following the random waypoint model in which mobility speed is randomly chosen between 0 and 72 km/hr. We configure the mobile devices to send a Channel Quality Indicator (CQI) report to the associated base stations every 100 ms, which allows the base stations to determine the MCS mode depending on the channel condition. We choose this reporting interval to ensure that we do not miss any channel condition changes, and at the same time we do not receive unnecessary frequent reports. Mobile users are randomly distributed within the service area of each cell such that more users, about 90% of the total number of users, are densely populated within 1/3 of cell radius and the rest of them are sparsely scattered around the rest of the cell area. This is done to mimic realistic scenarios as mobile operators usually install base stations in crowded areas to serve most users with strong signals.

Videos: We crawled YouTube to retrieve information of 1,000 videos using the YouTube dataset in [28]. Since the videos are 240p, we scaled up their bit rates to emulate videos of 720p, which are popular in modern smartphones. The popularity, bit rate and length of these videos drive our simulations. We sort videos based on their popularity. In order to imitate realistic video requesting behavior, mobile users request videos from this list following the Zipf distribution with a skewness parameter α. We set $\alpha = 1.5$ in our simulations, but this can be tuned to any desired value to cover a wide range of user behaviors.

Simulation Scenarios: We evaluated the proposed transmission scheduling algorithm in three scenarios: 1) seven independent cells serve unicast and single-cell multicast connections, 2) seven cells form a dynamic single frequency network, and 3) seven cells operate in a static single frequency mode. In the scenario applying single-cell traffic, we customize the resource allocator in eNodeB to schedule incoming requests and set up radio resources for both multicast and unicast connections. Different from the single-cell case, we designed and deployed an SFN area where eNodeBs are only responsible about scheduling incoming unicast requests, whereas the multi-cell coordination entity performs the required admission control for multicast sessions and assigns uniform radio resources among its cells in order to ensure enhanced coverage and synchronized data transmission. Our proposed algorithm follows a dynamic technique in which multicast groups as well as their assigned MCSs are configured dynamically to adapt and accommodate any changes in video popularity and channel conditions.

During our implementation, we faced a problem related to the fact that the LTE model in OPNET (which follows the 3GPP specifications for single frequency networks) considers these settings to be static during the entire running time of a simulation. In other words, the LTE configuration node cannot be adjusted dynamically, and its initial settings (including the MCS assigned for each MBMS bearer) cannot be changed after the deployment phase. To overcome this limitation, we created a virtual multicast over SFN to handle any complexity involved in the process of dynamic reconfiguration. To do that, we utilized the Physical Downlink Shared Channel (PDSCH) for multicast sessions (in addition to unicast connections) instead of using the resource blocks specifically allocated for SFN. It might be concerning that some resource blocks would not be utilized, thereby negatively impacting the achieved service ratio in the system. However, since we are not using any resource blocks of those specifically allocated for SFN, we found that the OPNET by default is freeing and releasing these resources because there is no MBMS subscribers, and it allows PDSCH to take an advantage of these inactive sub-frames and use them to extend the number of its available recourse blocks [27].



(a) Service ratio

(b) Energy saving

Figure 5: Comparison of the achieved performance by the proposed algorithm against other methods.

6.2 Comparison Against Current Algorithms

We compare our adaptive proposed transmission scheduling algorithm versus two multicast policies (i.e. Max Users and FCFS) in addition to the unicast method. The performance metrics used in this evaluation are the average service ratio and average power saving. Due to the limited radio resources in wireless networks, it may not be possible to serve all incoming video requests. For this reason, we estimate the service ratio at a certain point of time by computing the fraction of admitted mobile terminals to the number of received requests within the entire system. We define the energy saving as the percentage of time in which a served mobile device is able to turn off its network interface, thereby reducing its power consumption. The time required to switch the network interface from an active to idle is assumed to be negligible in this work. Thus, it is sufficient to utilize the time duration where a network interface is turned off as a direct representation of the energy saving for such receiver.

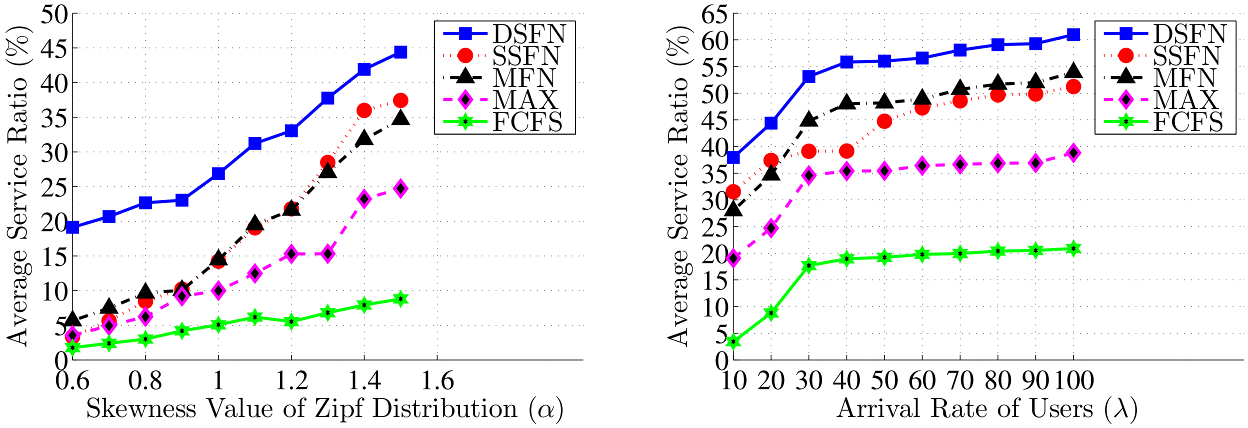
We simulate an LTE network where there are 1000 mobile terminals in each cell generating requests from a pool of 1000 possible video streams. By varying the fraction of available bandwidth for multimedia services d from 50% to 100%, we get the average achieved service ratio and energy saving shown in Figures 5(a) and 5(b), respectively. Results in these figures are the mean of 5 simulation runs with 5 different random seeds. These results indicate that our proposed transmission scheduling algorithm not only outperforms others with a significant margin regarding the achieved service ratio, but it also achieves much better energy saving. For instance, when $d = 1.0$, our proposed transmission scheduling algorithm in the seven independent cells operating under the single-cell configuration (which is denoted by MFN) admits an average of almost 35% of users at any given time, while those systems employing the Max Users and FCFS policies accept only an average of around 24% and 9% of users, respectively. This means that our algorithm in the single-cell scenario provides a service ratio which is approximately 40%, 294%, and 3070% higher than the Max Users, FCFS, and unicast methods, respectively. These numbers are based on a fully-dedicated network, where 100% of its bandwidth is allocated for video services. Similar outcome have been obtained even if d is decreased to lower values.

It can be also shown in Figure 5(a) that applying the concept of single frequency network improves the achieved service ratio by a significant gain. This can be referred to the fact that SFN services are transmitted simultaneously over the air from multiple synchronized base stations, which lead mobile terminals in the same SFN area to receive many versions of their required signals. Because these copies are broadcasted from synchronized entities, receivers are most likely to treat them as multipath components; hence, users would enjoy higher signal-to-noise ratios and then become able to decode at higher modulation and coding schemes. Figure 5(a) presents the achieved service ratio by our proposed algorithm in two types of SFN configurations: dynamic (DSFN) and static (SSFN). Applying our proposed dynamic reconfiguration helps us in reaching an average of served terminals equivalent to around 44.5% of users in fully-dedicated networks. This improvement is 18% and 28% higher than those achieved ratios in both static SFN and single-cell scenarios, while the gains over traditional techniques such as groups with Maximum Users and FCFS methods are almost 80% and 404%, respectively.

Regarding the energy saving metric, the proposed dynamic SFN algorithm outperforms all others including the unicast algorithm. The unicast algorithm usually represents the maximum energy saving possible in a single-cell configuration since individual unicast connections are served accordingly to their best-suitable modulation and coding schemes. However, both cell coverage and average signal-to-noise ratio can be enhanced when the idea of single frequency network is applied, thereby enabling our resource blocks allocator to exceed the upper limit of energy saving in single-cell configuration. For example, Figure 5(b) demonstrates that the energy saving by DSFN is approximately 3.5% more than the unicast method when the fraction of allocated bandwidth for video streaming is 50%.

6.3 The Impact of User Behavior Model

We analyze the impact of user behavior on the performance of transmission scheduling algorithms with respect to the average achieved service ratio. Two important aspects of user behavior models are considered: videos popularity and request arrival. To study the effect of video selection policy on the proposed resource scheduler, we emulate a Zipf distribution to select streams from a pool consisting of 1000 different videos. The skewness parameter α guides the selection strategy of these videos in a way that higher values of α is going to assign greater probability for most popular



(a) Video popularity distribution

(b) User arrival rate

Figure 6: Impact of the skewness in video popularity distribution on the service ratio.

videos to be chosen, and vice versa. We vary the value of α from 0.5 to 1.5 to study various policies for the video selection process. Figure 6(a) reports the impact of varying the skewness parameter on the achieved service ratio. The figure shows that the average service ratio gradually increases with the increase in the skewness parameter α. Higher values of skewness lead more users to select from the top-ranked videos, thereby resulting in more possible multicast groups. This means larger chances to serve additional mobile terminals through multicast sessions and then decrease the service blocking probability in the system.

The effect of request arrival distribution on the proposed algorithm is also examined. We utilize the Poisson process and vary its mean in order to emulate variations in the user arrivals within cells. The arrival rate λ indicates the number of incoming videos requests per second, where higher values for this parameter are offering more opportunities for the creation of multicast sessions. Figure 6(b) shows the impact of varying the arrival rate of Poisson distribution on the average achieved service ratio. We vary the value of arrival rate from 10 to 100 requests per second. In Figure 6(b), it is shown that the service ratio increases for the proposed algorithm as the arrival rate increases. This is due to the fact that higher values of the arrival rate ensure larger numbers of request arrivals per second, with the same selections of video streams since the skewness parameter is kept unchanged during these experiments. This gives a chance to merge larger number of mobile devices into multicast groups, which eventually results in higher service ratio. Figure 6(b) also indicates the effectiveness of the proposed scheme under conditions of high loads. However, we can see from the figure that the service ratio increases significantly with the increase in arrival rates until it reaches 40, after which the service ratio becomes quite steady.

Figures 6(a) and 6(b) demonstrate that our proposed resource allocation algorithm is superior in its performance results than both Max Users and FCFS schemes under any given network configurations and any chosen user behavior models. On the other hand, both Figures 6(a) and 6(b) point to a fundamental problem in the static deployment of SFN networks. Typically, the concept of single frequency network is employed to enhance the coverage and maximize the average signal-to-noise ratio such that the spectral efficiency within cells can be increased. Because the static configuration is pre-designed at an early stage of deployment, it is most probably unware of any variation in the user distributions and video requests during the operation time. As a consequence, it would inevitably waste a substantial amount of radio resources reserved for SFN, especially in those scenarios where a few number of mobile terminals are interested in the MBMS services offered by their cell. For this reason, we found in Figures 6(a) and 6(b) that the average service ratio of our proposed algorithm in the single-cell topology exceeds its counterpart in the static SFN configuration, mainly when $\alpha \leq 1.1$ or $\lambda \geq 30$. DSFN overcomes this limitation and adjusts its multicast zones according to both user distribution and video popularity. In extreme cases, cells in DSFN can shrink themselves from all SFN areas and switch their settings to the single-cell topology. In other words, our proposed transmission scheduler under the dynamic SFN configuration adapts itself in a way so that the best possible bandwidth utilization is reached as it is shown in Figures 6(a) and 6(b).

7. CONCLUSIONS

Due to the introduction of smart phones, traffic loads on mobile networks have dramatically increased during the recent decade, where a large portion of this traffic can be referred to the escalated consumption of videos. This trend of watching more multimedia content on mobile devices is expected to continue in the coming years. This creates a challenge for wireless network operators because of the constraint on their available radio resources and the substantial bandwidth requirements for each video session. This paper proposed an adaptive mobile multimedia streaming algorithm in which current user distributions and video popularities are taken into consideration during its network configurations and scheduling decisions. Different from existing works, we do not assume a static configuration of single frequency networks. Instead, we presented an approach in which cells are dynamically rearranged in SFN zones in a way that maximizes the total bandwidth utilization. We demonstrated through simulations that applying the concept of dynamic reconfiguration adds significant gain in the average service ratio, as compared to those techniques with static SFN settings, and these obtained gains are independent of the amount of available bandwidth and the model of user behaviors.

Once a proper configuration for SFNs is reached, the available radio resources are allocated for both unicast and multicast services with an objective of increasing the average service ratio. Our proposed transmission scheduler achieves this goal by utilizing a flexible allocation process in which the resource distribution between unicast and multicast connections is done dynamically. To offer the flexibility of resource distribution, this paper exploits three different types of transmission: unicast, multicast over an SFN, and multicast restricted within the coverage of a cell. According to obtained simulation results, the proposed transmission scheduling algorithm under a dynamic SFN configuration gives higher service ratios than its counterpart in static SFN and single-cell scenarios. For instance, when the fraction of bandwidth allocated for videos is 60%, the service ratio achieved by our algorithm in a dynamic SFN is 25.63% and 45.62% more than those service ratios obtained in static and single-cell configurations, respectively. Comparing with traditional scheduling policies such as: groups with maximum users and first-come first-served, the gain in service ratio by the proposed algorithm can be at least 2 and 4 times higher, respectively.

Acknowledgements

This work was supported in part by the National Science, Technology and Innovation Plan (NSTIP) of the Kingdom of Saudi Arabia and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

8. REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2013-2018, February 2014, Available: http://tiny.cc/Cisco14.

[2] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, “Quality of service support in IEEE 802.16 networks,” *IEEE Network Magazine*, vol. 20, no. 2, pp. 50–55, March 2006.

[3] 3GPP TR 26.903 V9.0.0, “Improved video support for Packet Switched Streaming (PSS) and Multimedia Broadcast/Multicast Service Services (Release 9),” March 2010.

[4] Verizon Wireless: Customers Use 1.9 Terabytes of Data in Stadium at Super Bowl, February 2014, Available: http://tiny.cc/Verizon2014.

[5] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for mobile broadband*. Academic Press, December 2013.

[6] 3GPP TS 36.300 V12.2.0, “Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Overall Description (Release 12),” June 2014.

[7] L. Rong, O. Ben Haddada, and S. Elayoubi, “Analytical analysis of the coverage of a MBSFN OFDMA network,” in *Proc. of IEEE Global Telecommunications Conference*, November 2008, pp. 1–5.

[8] S. Talarico and M. Valenti, “An accurate and efficient analysis of a MBSFN network,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2014, pp. 6994–6998.

[9] A. Alexiou, C. Bouras, V. Kokkinos, and G. Tsichritzis, “Performance evaluation of LTE for MBSFN transmissions,” *Wireless Networks*, vol. 18, no. 3, pp. 227–240, April 2012.

[10] A. Urie, A. Rudrapatna, C. Raman, and J. Hanriot, “Evolved multimedia broadcast multicast service in LTE: An assessment of system performance under realistic radio network engineering conditions,” *Bell Labs Technical Journal*, vol. 18, no. 2, pp. 57–76, September 2013.

[11] C. Suh and J. Mo, “Resource allocation for multicast services in multicarrier wireless communications,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 27–31, January 2008.

[12] J. Xu, S. Lee, W. Kang, and J. Seo, “Adaptive resource allocation for MIMO-OFDM based wireless multicast systems,” *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 98–102, March 2010.

[13] G. Araniti, M. Condoluci, L. Militano, and A. Iera, “Adaptive resource allocation to multicast services in LTE systems,” *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 658–664, December 2013.

[14] H. Won, H. Cai, D. Eun, K. Guo, A. Netravali, I. Rhee, and K. Sabnani, “Multicast scheduling in cellular data networks,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4540–4549, September 2009.

[15] J. Monserrat, J. Calabuig, A. Fernandez-Aguilella, and D. Gomez-Barquero, “Joint delivery of unicast and e-MBMS services in LTE networks,” *IEEE Transactions on Broadcasting*, vol. 58, no. 2, pp. 157–167, June 2012.

[16] J. Lee, H. Park, S. Choi, and J. Choi, “Adaptive hybrid transmission mechanism for on-demand mobile IPTV over WiMAX,” *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 468–477, June 2009.

[17] M. Rahman, C. Hsu, A. Hasib, and M. Hefeeda, “Hybrid multicast-unicast streaming over mobile networks,” in *Proc. of IFIP Networking Conference*, June 2014, pp. 1–9.

[18] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, “Distributed α-Optimal User Association and Cell Load Balancing in Wireless Networks,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 177 –190, February 2012.

[19] D. Parruca and J. Gross, “Rate selection analysis under semi-persistent scheduling in LTE networks,” in *Proc. of International Conference on Computing, Networking and Communications (ICNC 13)*, San Diego, CA, January 2013.

[20] K. Hua, Y. Cai, and S. Sheu, “Patching: A multicast technique for true video-on-demand services,” in *Proc. of the Sixth ACM International Conference on Multimedia*, Bristol, United Kingdom, September 1998, pp. 191–200.

[21] D. Eager, M. Vernon, and J. Zahorjan, “Minimizing bandwidth requirements for on-demand data delivery,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 5, pp. 742–757, September 2001.

[22] H. Hlavacs and S. Buchinger, “Hierarchical video patching with optimal server bandwidth,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, no. 1, pp. 8:1–8:23, January 2008.

[23] “IBM ILOG CPLEX,” http://www.ibm.com/software/integration/optimization/cplex-optimizer.

[24] Shomi: Video-on-demand Service, January. 2015, Available: https://shomi.rogers.com/.

[25] Crave TV: Bell Media, December. 2014, Available: http://www.cravetv.ca.

[26] J. Erman and K. Ramakrishnan, “Understanding the super-sized traffic of the super bowl,” in *In Proc. of ACM Conference on Internet Measurement Conference (IMC’13)*, Barcelona, Spain, 2013, pp. 353–360.

[27] OPNET Technologies, Inc., March. 2012, Available: http://www.riverbed.com/products/performance-management-control/opnet.html.

[28] X. Cheng, C. Dale, and J. Liu, “Statistics and social network of YouTube videos,” in *In Proc. of IEEE Workshop on Quality of Service (IWQoS’08)*, June 2008, pp. 229–238.