Towards an LTE hybrid unicast broadcast content delivery framework

PLEASE CITE THE PUBLISHED VERSION

https://doi.org/10.1109/TBC.2017.2704429

PUBLISHER

© IEEE

VERSION

AM (Accepted Manuscript)

REPOSITORY RECORD

Christodoulou, Louis, Omar Abdul-Hameed, and Ahmet Kondoz. 2019. “Towards an LTE Hybrid Unicast Broadcast Content Delivery Framework”. figshare. https://hdl.handle.net/2134/24857.

Towards an LTE Hybrid Unicast Broadcast Content Delivery Framework

Louis Christodoulou, Omar Abdul-Hameed, Ahmet M. Kondoz

***Abstract*—The era of ubiquitous access to a rich selection of interactive and high quality multimedia has begun; with it, significant challenges in data demand have been placed on mobile network technologies. Content creators and broadcasters alike have embraced the additional capabilities offered by network delivery; diversifying content offerings and providing viewers with far greater choice. Mobile broadcast services introduced as part of the Long Term Evolution (LTE) standard, that are to be further enhanced with the release of 5G, do aid in spectrally efficient delivery of popular live multimedia to many mobile devices, but, ultimately rely on all users expressing interest in the same single stream. The research presented herein explores the development of a standards aligned, multi-stream aware framework; allowing mobile network operators the efficiency gains of broadcast whilst continuing to offer personalised experiences to subscribers. An open source, system level simulation platform is extended to support broadcast, characterised and validated. This is followed by the implementation of a Hybrid Unicast Broadcast Synchronisation (HUBS) framework able to dynamically vary broadcast resource allocation. The HUBS framework is then further expanded to make use of scalable video content.**

***Index Terms*—LTE, E-MBMS, Broadcast, Scalable Video, H.264, cellular networks.**

I. INTRODUCTION

THE last two decades have presented a continued and relentless advancement of consumer electronics. Processing power per square centimeter continues to increase exponentially, permitting cheaper, more power efficient, lighter and hence more mobile devices. The latter decade has seen a seismic transformation in the mobile device arena with the explosion in popularity of the smartphone and subsequently smart device (tablets, cars, watches etc). Whilst still a communications device, viewed from a bandwidth usage perspective, a smartphone's primary role looks very different. Whether for work or entertainment, a smartphone is most often performing "content consumption" tasks. The kind of heavy duty data consumption once limited to a stationary desktop or cumbersome laptop is now effortlessly exceeded by a device in the pockets of 1.91 billion people worldwide [1]. By 2021, mobile data traffic is expected to reach 587 exabytes annually, with video data forecast to account for over 78% of this total traffic [2].

Mobile media consumption and the associated demand it presents for mobile network bandwidth has also placed increasing pressure on the spectrum resources assigned to traditional Digital Television (DTV) services [3]. Given the forecast data trends, research in the delivery of future broadcast television over cellular networks is gaining traction. *Walker et al.* in [3] identifies that "traffic growth is far exceeding the growth in available bandwidth". Furthermore, rather than directly targeting bandwidth from DTV, the paper presents intuitive methods to share bandwidth, thus providing a greater aggregate efficiency between the two services. Moving further, work by *Shi et al.* in [4] presented a case study on DTV distribution over cellular networks. The authors made use of unicast bearers for unpopular content and showed considerable bandwidth saving over traditional DTV in urban environments. Along a similar theme, more recent work by *Lau et al.* in [5] further explores broadcast television over cellular networks, once again reinforcing the concepts linking popularity and overall spectral efficiency gains. Further more, the work used real world data and scenarios to develop an audience-driven TV scheduling framework, optimising the scheduling of broadcast resources.

The Global Mobile Suppliers Association (GSA) in [6] forecast the market for Long Term Evolution (LTE) broadcast will reach $14bn worldwide by 2020. The report goes on to explain the rising interest in LTE broadcast services predicting deployments will grow significantly during the next 5 years.

Broadcast in LTE networks is the responsibility of the enhanced Multimedia Broadcast Multicast Service (eMBMS). Implementation details for eMBMS were not specified by 3rd Generation Partnership Project (3GPP) until March 2010 with the freeze of release 9 [7]. As with Conventional Multicast Schemes (CMS), eMBMS facilitates synchronous transmission to multiple users through shared use of the same radio resources. On the radio interface of the network, this is done by establishing a Point-to-Multipoint (P-T-M) radio bearer [7]. Enhancements for eMBMS continue into release 14, targeting a June 2017 release, where 5G standardisation will also begin to be defined. Discussion in the various 3GPP Radio Access Network (RAN) meetings regarding release 14 eMBMS only serve increase the flexibility offered by the standard for broadcast resource allocation, thus strengthening the approach taken in the proposed work [8].

This research work looks beyond the imminent adoption phase; at an environment where eMBMS services are used as a delivery medium for popular content. The concept underpinning the proposed framework is formed based on the observation of two diverging trends: LTE eMBMS is expected to play a significant role in reducing the burden of delivering next generation multimedia to mobile devices. The only scenario in which broadcast technology can offer a significant spectral efficiency gain is where multiple User Equipment (UE)s are receiving the exact same data, or within the context

L.Christodoulou is with University of Surrey, Guildford, Surrey, UK (E-Mail: louis@louisc.co.uk)

O. Abdul-Hameed and A. Kondoz are with Loughborough University, London, UK (E-mail: Omar.Abdul-Hameed@outlook.com; a.kondoz@lboro.ac.uk)

of this work, the same video stream. Meanwhile, content creators and broadcasters are diverging from traditional single stream offerings, increasingly providing individual users with greater choice to personalise the way in which they consume content. These enhanced offerings are becoming increasingly popular, and can open up additional revenue streams for mobile operators and content creators who can offer 'premium services' to subscribers, thereby enhancing a live broadcast event. Example multi-stream applications include, Ultra High Definition Television (UHDTV), 3D Television, Free Viewpoint Television (FVT) and Multi/Companion Screen viewing.

The proposed work is an extension to the work first proposed in [9] and explores a hybrid delivery framework, to be defined as Hybrid Unicast Broadcast Synchronisation (HUBS), for multi stream multimedia. This area of focus is entirely inspired by the observations above. The concept allows for delivery of a popular stream via broadcast, maximising spectral efficiency. Enhancements to this broadcast stream can be delivered via a secondary unicast stream to a selected subset of subscribed users. An example application could include a scenario where the base layer of a scalable video stream encoded from a live sporting event is broadcast to all users within a cell. Users who have High Definition (HD) or Ultra High Definition (UHD) capable devices are able to enhance the base layer by requesting an enhancement layer via unicast transmission. The quality of this enhancement layer is further dynamically scaled for each subscribed user independently, since bidirectional communication exists in unicast.

II. BACKGROUND

This section presents a technical overview of the standards and technologies utilised by the proposed Hybrid Unicast Broadcast Synchronisation (HUBS) framework.

A. *Multimedia Broadcast Architecture*

The LTE eMBMS architecture is shown in Figure 1 [7]. By comparison the flow of eMBMS data through an LTE network is very different from that of unicast. Content providers will interface with the Broadcast Multicast Service Center (BM-SC) that establishes and manages the data flow configuration through the Evolved Packet Core (EPC). From here the Internet Protocol (IP) stream from the BM-SC is forwarded to the eMBMS Gateway (GW) that manages the distribution of the stream of eMBMS data packets to each participating Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access Network (E-UTRAN) Node B (eNodeB) via IP Multicast, efficiently using the backhaul network [10]. The eMBMS GW is also responsible for handling the session control signalling of each eMBMS service, which is performed via the Mobility Management Entity (MME) that keeps a record of UE properties, such as location, connected or idle status and is responsible for the setup and release of resources [11].

Connected to the MME, via the control plane, is the Multicast Coordination Entity (MCE), a key node for this research. This entity sits within the RAN and is a 'logical' entity, meaning it can be implemented as either a hardware node, or a software update in the eNodeB. The responsibilities of the MCE include the radio resource management of all eMBMS services for each of the connected eNodeBs, as well as decisions on Modulation and Coding Scheme (MCS) selection and frame allocation [11].

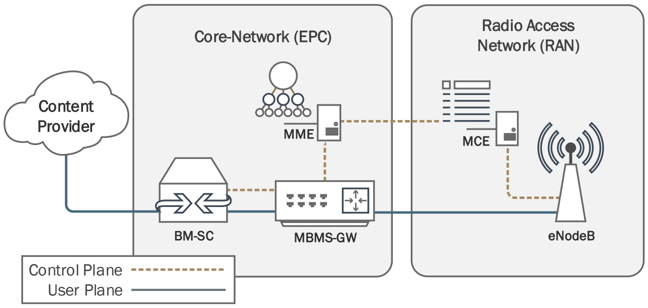


Fig. 1. LTE eMBMS logical architecture

B. *Content Processing*

Advances in video content processing techniques have been essential in the ability to successfully deliver enhanced multimedia to end-users. This section reviews the content processing and encoding techniques utilised in the HUBS framework.

The basis of compression with most modern video coding is the strong statistical correlations between consecutive video frames as well as within each frame. By exploiting these correlations, bandwidth saving can be achieved with minimal loss to visual quality. 3D or multiple viewpoint scenarios are generally shot with a pair or series of cameras at different angles, all capturing a representation of the same scene. Multiview Video Coding (MVC), an extension to the H.264/AVC standard further extends this concept through prediction between views, exploiting the redundancies and thus providing a better overall compression ratio [12] [13]. One of the outlined requirements of the extension was complete backward compatibility of the video stream by non enhanced decoders, a key feature in its implementation with the HUBS framework, giving standard users the ability to receive a broadcast stream as standard.

The development of adaptive bitrate streaming concepts has also been driven in recent years by the ever broadening range of devices on which the same multimedia is to be consumed [14]. This is the area where the development of Moving Picture Experts Group (MPEG) Dynamic Adaptive Streaming over HTTP (DASH) is positioned to offer maximal impact. Hypertext Transfer Protocol (HTTP) streaming has become an increasingly efficient protocol with which to transmit video. HTTP is implemented by nearly all Internet infrastructure and as such end to end delivery has already been streamlined [14]. Furthermore, in LTE Release 10, 3GPP released its own compatible MPEG DASH profile that was named 3GPP DASH [15].

III. OPEN SOURCE LTE EMBMS SIMULATION PLATFORM

The first logical step toward the development of a hybrid delivery framework is to establish an LTE simulation platform able to support both unicast and broadcast services. This section covers the extension and modification of the open source LTE-Sim platform to include eMBMS capability. As a stand-alone system level simulator, LTE-Sim's implementation, presented in [16], respects the layered approach of the LTE standard clearly and concisely. Furthermore, it is highly modular and makes extensive use of the object-oriented and polymorphic abilities of the c++ language. LTE-Sim has received continual support and updates from the team at *Politecnico di Bari* and continues to attract an active and engaging community of researchers. Therefore, LTE-Sim was chosen as the simulation platform on which to implement the eMBMS extension.

A. Proposed Design Considerations

1) MCE Node Unicast Broadcast Resource Allocation: Considering its central role in the management of LTE broadcast services, the starting point for design is the MCE. The 3GPP protocol definition documentation outlines the key configuration variables the MCE node will need to maintain for compliant broadcast resource allocation, these are:

- **Radio Frame Allocation Period** defining the distance, in frames, between the allocation of eMBMS enabled frames. This can be otherwise defined as the period of eMBMS frame allocation. The value of this variable must be defined as 1, 2, 4, 8, 16 or 32 frames.
- **Radio Frame Allocation Offset** that defines the offset, in frames, of the above defined allocation. This allows shifting of the allocation for this service and is useful where multiple eMBMS services are to be defined.
- **Four Frame Allocation Map enabled flag** is a boolean value that denotes the allocation mode selected. This is further explained below.
- **Sub Frame Bitmap** is either 6 or 24 bits in length based on the selection of a single or four frame allocation mode, respectively.

Both unicast and eMBMS are based on the Orthogonal Frequency Division Multiple Access (OFDMA) scheme for downlink data traffic. Despite this, the way in which each performs allocation of resources in both the frequency and time domains vary vastly. For any eNodeB that provides an eMBMS service, certain frames are periodically allocated for the transmission of the Multicast Channel (MCH) [10]. Allocations can be made in two modes, ‘oneFrame’ where a single frame is allocated each time, or ‘fourFrame’ where allocation is in sets of 4 consecutive frames [17]. Although both allocations are designed into the implementation, for the purposes of this work, only 'oneFrame' based allocation is utilised. Ordinarily, no dynamic allocation of eMBMS resources is performed, instead frame reservation is based on the ‘radioframeAllocationPeriod’ and ‘radioframeAllocationOffset’ parameters. All radio frames that satisfy:

$$SFN \bmod Ap = Ao \tag{1}$$

are reserved for the eMBMS service, where SFN is the current System Frame Number and Ap and Ao represent the chosen Allocation Period (AP) and allocation offset respectively [17].

Once a frame is reserved to contain eMBMS services, only six of the ten available sub frames within can be used for the broadcast service. This is due to synchronisation and paging that can occupy sub frames 0, 4, 5 and 9 of any LTE Type 1 frame, making them unusable for eMBMS services [18]. In order to denote which sub frames have been allocated within the reserved frame, a bitmap is used, each bit denoting true or false for an eMBMS or Unicast sub frame assignment, respectively. Only the sub frames that may be allocated to eMBMS services are represented by the bitmap; therefore, a 6-bit map would represent a 'oneFrame' allocation and a 24-bit map would be utilised for a 'fourFrame' allocation [18]. A reserved sub frame utilises the entire bandwidth allocation in the frequency domain for its duration.

In order to decode the eMBMS data, the UE must know the allocation period and offset parameters, bitmap and MCS chosen to transmit data. This control information is periodically provided by the Multicast Control Channel (MCCH), a logical channel specific for eMBMS. As such, allocation of resources for broadcast cannot be changed until an update is sent on the control channel. Strictly speaking, from a standards perspective modification of the parameters on the MCCH are currently restricted to 512 or 1024 frames via the "mcch-ModificationPeriod" parameter defined in [17]. Although the standards limit the modification period to 512 or 1024 frames, the information is transmitted repeatedly on the control channel with a more frequent interval of 32, 64, 128, and 256 frames, defined as the "mcch-RepetitionPeriod" in [17]. This allows users wishing to connect to the broadcast stream to do so without having to wait the full 512 or 1024 frames of each modification period. The motivation behind this design choice appears to be one of power saving; allowing a UE to radio sleep, assured that the broadcast scheduling parameters will not be modified more frequently than the modification period, thus ignoring the repetitions and only waking up every modification period to check the control channel (and of course perform broadcast reception). In the use case presented here, the UE is performing reception of live video, requiring a continuous connection. Furthermore, the work presented herein is based on a simultaneously established unicast connection, further restricting any possibility for radio sleep. As such, this consideration will provide little benefit in this use case. The authors propose that for such applications, the standards are updated to support modification of the MCCH parameters at a rate equivalent of the "mcch-RepetitionPeriod". Since the network is already capable of transmitting this information during the repetitions, this proposed alteration will require minimal alteration and add no control signalling overhead beyond that currently define in the standards [17]. It is the loss of the fast and dynamic ability to schedule, as well as the need for the same transmission parameters to cater for a larger user base, that can lead to lower spectral efficiency if there is little interest in the broadcast content. For the remainder of this paper, we will refer to this parameter as the "Broadcast

Scheduling Period".

2) MCE Node Broadcast Modulation and Coding: Once a sub frame is assigned to either unicast or broadcast services a call must be made to begin resource scheduling. Within the existing unicast-only platform this process of resource assignation is triggered via a call to the respective eNodeB. The proposed design has this remain true for unicast frames, but broadcast frames instead call on the MCE node. A call to either of these functions ultimately triggers the allocation of physical resources to respective unicast or broadcast bearers with data to transmit. For unicast scheduling, this is managed by the eNodeB that retrieves the selected downlink packet scheduler class and calls the scheduling function within it. With unicast allocation, the MCS is selected based on the channel quality of the given UE. Of course the limitations presented with broadcast transmission is that the MCS must be set such that the entire cell is able to receive the broadcast transmission. As such, the MCE must adopt the weakest (or potentially weakest) user's MCS.

The proposed implementation also considers Quality of Service (QoS) conditions placed upon realtime broadcast services. Thus, a further function is implemented to iterate through the broadcast radio bearers and verify that the data to be transmitted has not exceeded the maximum acceptable delay QoS parameters defined for the given application. Should the data have exceeded its defined QoS parameters, it will be dropped at the eNodeB before scheduling or transmission.

Much like unicast services, there is no 3GPP technical specification for the allocation of broadcast data to physical resources. A 'standard' algorithm was created that simply defines a set, scenario allocated MCS to each broadcast radio resource. The proposed implementation includes all of the framework allowing future resource allocation strategies to be deployed within the simulator for broadcast scenarios. This makes the platform useful for work far beyond that presented in this paper.

3) MCE Node UE Subscription Management: In order to ensure the broadcast stream is received by the correct clients, a subscription style model is implemented. This will contain eMBMS groups and a mechanism whereby UE objects can join and be tracked. This will also require the UE objects to know if they are a member of an eMBMS group. To manage this, an 'eMBMSGroup' class is created, this will maintain a subscribed user container with reference to each UE member object. Each group is referenced by an ID allocated from the simulation scenario file. UEs are subscribed to the group only upon ensuring that the given UE is not already a member. This process must also set information within the UE about their group subscription.

4) P-T-M Bearer: LTE-Sim has already established bearer classes to support Point-to-Point (P-T-P) traffic, making these a logical point at which to implement simulator support for P-T-M. For the proposed enhancements, each radio bearer instance is given a means by which it can be assigned a type. This type is given a default assignation of P-T-P to maintain compatibility. Most important of all is that a P-T-M bearer must support a group as a destination, rather than a single user object. As such the ability to define a destination 'eMBMSGroup' is established.

TABLE I
SIMULATION PARAMETERS - CHARACTERISING eMBMS AP ALLOCATION PARAMETER

Parameter	**Value**
Simulation Time	150s Per Run, 30s Warm Up, 5 Runs
Cells	Takes place in a single cell
Cell Layout	Hexagonal grid of 7. Surrounding cells generate interference
Inter Site Distance	0.5km
User distribution	Random Placement, walking in random direction
User Numbers	5 - 80 users, interval of 5
eMBMS AP	1, 2, 4, 8, 16, 32
eMBMS Allocation	Single Frame
eMBMS Bitmap	111111
eMBMS MCS	Index = 8
Frequency Reuse	Enabled (3 Clusters)
Channel Realization	Macro Cell Urban Area
Error Model	Wideband CQI Eesm Error Model
Link Adaptation	AMC Enabled
Unicast Scheduling Algorithm	Maximum-Largest Weighted Delay First (M-LWDF)

QoS Max Delay	eMBMS = 100ms
	Video = 100ms (QCI-7)
	VoIP = 100ms (QCI-1)

User Service	
Broadcast Video	100% Total Active Users Foreman H264 440Kbit

IV. eMBMS SIMULATOR CHARACTERISING

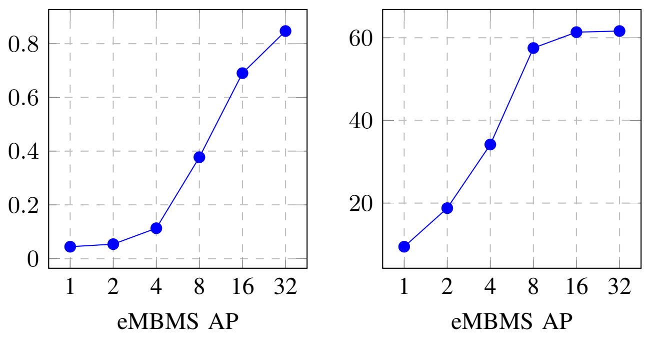
In order to assess the performance and assign credibility of the enhanced LTE-Sim simulation platform, a series of dynamic simulations were conducted and results analysed. Initially, identical simulations are performed on both the original, validated, LTE-Sim platform and the broadcast capable enhancement. Despite the broadcast architecture being in place, only unicast flows were established to facilitate a direct comparison of results. This is followed up with a characterisation of the enhanced simulation platform, where the results are examined and an analytical approach is employed through mathematical first principles.

A. eMBMS Frame Allocation

To characterise the extended functionality, a simulation scenario was established with only broadcast data present within the cell. By testing the various configuration parameters defined for eMBMS transmission and analysing the resultant output, the behaviour of the broadcast service can be characterised.

The first experiment is to both explore and validate the new Allocation Period (AP) parameter that can be defined for eMBMS transmissions. A scenario was established where every user within the cell will subscribe to only the broadcast data stream. Considering this is a test of the AP, the Sub Frame (SF) allocation map was simply set to allocate all sub frames (i.e. “111111”) to broadcast within a system frame reserved for eMBMS service. The simulation was repeated with an AP of 1,2,4,5,16 and 32. To fully test the cell from light to completely saturated, users are introduced in steps of 5 from just 5 users to 80. The remainder of the simulation parameters are listed in Table I.

Firstly, it was important to establish how both the simulator and results analysis scripts responded to increasing broadcast subscriber numbers within the cell. The tests show the simulator correctly exhibited little variation in performance with increasing broadcast subscribers.


(a) PLR

(b) Delay [ms]

(c) Throughput [Mbps]

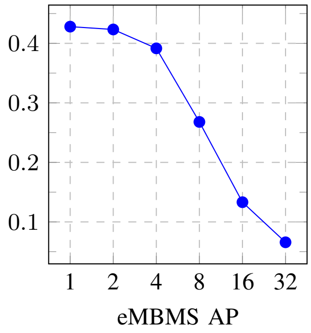


Fig. 2. eMBMS PLR, Delay and Throughput with increasing eMBMS Allocation Parameter

Figure 2 provides the average across all cell user numbers for each AP assignment. The resultant behaviour the simulator exhibits is in line with what would be expected, as fewer frames are allocated, there is a resultant increase first in delay, followed by packet loss and a drop in throughput. Of course a drop in throughput is expected due to the assignment of fewer resources, but, with QoS restricted services, a drop in throughput will also be experienced should the delay increase. The standard eMBMS scheduler mimics the tried and tested packet dropping policy of Modified Largest Weighted Delay First (M-LWDF) and Exponential Proportional Fair (EXP/PF) algorithms; implemented as a design decision to ensure delay does not build at the bearer. As the AP parameter increases, the time distance between frames reserved for broadcast also increases. There may well exist a scenario where there is sufficient bandwidth averaged over one second, yet the delay incurred by frames on a millisecond-level would exceed the 100ms QoS threshold and be subsequently dropped. This is what explains the resultant shape difference present between the delay (b) and Packet Loss Ratio (PLR) and Throughput (TP), (a) and (c) respectively in Figure 2.

Thus far, the characteristics displayed by the simulator are correct. It is also important to ensure the data produced is also valid. For a simple scenario such as this one, this can be accomplished by manually calculating the expected cell throughput for a given AP. The Transport Block Size (TBS) for a given MCS and number of Resource Block (RB)s can be derived from *Table 7.1.7.2.1-1* in the 3GPP LTE Technical Specification 36.213 [19]. Let the look up table be defined as a function $TBS(N_{rb}, I_{TBS})$, where N_{rb} is the number of RBs assigned. Let the function $I_{TBS}(I_{MCS})$ return the the row index reference derived from the chosen MCS from *Table 7.1.7.1-1* in Technical Specification 36.213 [19]. For the given scenario, the selected MCS index is 8 and at 10MHz the number of RBs available are 50 as listed in Table I. This results in:

$$I_{TBS}(8) = 8 \tag{2}$$

and:

$$TBS(50, 8) = 6968 \ bits \tag{3}$$

The TBS is the data that can be transmitted in a single Transmission Time Interval (TTI) of 1ms. Therefore, we can establish a maximum achievable bit rate using:

$$T_{Throughput}(AP) = 1000 \cdot \frac{N_{rb}}{10} \cdot TBS(N_{rb}, I_{TBS}) \cdot \frac{1}{AP} \tag{4}$$

By keying the results of Equation 2 and 3 into Equation 4, a maximum theoretical throughput can be established for each AP value. This maximum theoretical throughput has been plotted alongside the throughput of the 440Kbit video stream broadcast over the network in Figure 3. For APs of 1, 2 and 4, the stream is able to sustain its throughput, also taking into account the delay shown in Figure 2 (b) that remains relatively low below 40ms. Once the AP reaches 8, the allocation of broadcast frames are 80ms apart causing a considerable delay buildup on the bearer. It is clear that there is just about adequate theoretical bandwidth in which to transmit the data yet there is a considerable drop in throughput and subsequent rise in PLR that is now at almost 40%. This shows the importance of regular scheduling of frames when broadcasting real-time services. Furthermore, it also shows the simulator is responding in line with the theoretical calculations above based on the 3GPP technical specification.

B. *eMBMS Sub Frame Allocation*

The AP, is a somewhat course parameter to vary when allocating broadcast resources. Far more precise and granular control is achieved through allocation of sub frames to broadcast services via the SF allocation map defined by the 3GPP standards.

Once again a simple broadcast-only scenario was established to test and characterise the SF allocation functionality of

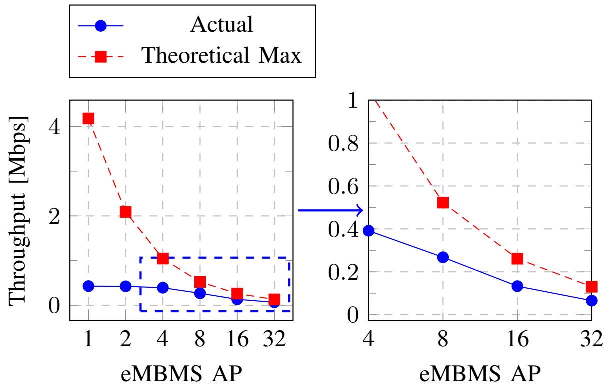


Fig. 3. Throughput of eMBMS service carrying 440kbit video stream, versus theoretical maximum of the cell given the varying AP.

TABLE II
SIMULATION PARAMETERS - SF ALLOCATION TESTING (MISSING PARAMETERS REMAIN SAME AS TABLE I)

Parameter	**Value**
Bandwidth	20 MHz (100 RBs)
User Numbers	5 and 40 users
eMBMS AP	1, 2
eMBMS Allocation	Single Frame
eMBMS Bitmaps (Static Allocation)	100000, 110000, 111000, 111100, 111110, 111111
eMBMS MCS	Index = 8
QoS Max Delay	eMBMS = 250ms
User Service **Broadcast Video**	100% Total Active Users Poznan St H.264/AVC 720p (Cam 4)

the extended simulation platform. For this test, it was decided a higher bit rate video sequence would be broadcast over the network. The Poznan Street[1] sequence was decided upon due to its familiarity amongst the research community and its availability in HD. The view from camera 4 was encoded using the H.264/AVC JM reference encoder with a QP of 27 and at a resolution of 720p (1280x720) [12]. Subsequently, a trace file was extracted taking the size, type and transmission time of each frame in the sequence. Given the larger resolution video stream, the bandwidth available to the cell downlink has been increased to 20MHz. The maximum number of cell users has been decreased to 40, since it has been proven that subscribers have little influence on the broadcast transmission. To better understand the nature of delay, the QoS maximum delay for broadcast was lifted to 250ms to avoid packets being dropped such that the response of the delay curve can be observed. The AP will also be varied but only between a value of 1 and 2 in this experiment, instead focusing on the SF bit map where assignments 100000, 110000, 111000, 111100, 111110 and 111111 will all be tested. The remainder of the simulation parameters can be found in Table II.

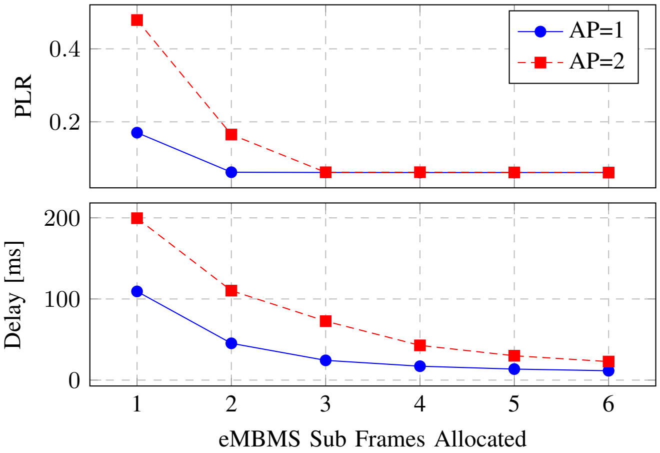


Fig. 4. eMBMS service PLR and Delay with increasing eMBMS subframe allocation shown for both an Allocation Parameter of 1 and 2.

Figure 4 illustrates the results of both the eMBMS service PLR and delay with increasing SF allocation to eMBMS services. The SFs are allocated using the SF maps shown in the simulation parameters Table II, where 1 sub frame is allocated as ‘100000’, 2 as ‘110000’ and so on. It is clear that allocation of a single frame is inadequate for this transmission even with an AP of 1 set. Doubling this to 2 has a dramatic effect with an AP of 1, dropping the PLR to a level where it shows no significant improvement with increased resource allocation. Furthermore, comparing the PLR and delay graphs in Figure 4, it is clear to see the effect the QoS packet dropping functionality in the broadcast scheduler has on the stream. As this test has had the maximum delay restriction within the QoS relaxed to 250ms, there is a clear drop in PLR at a SF allocation of 2 and 3 for an AP of 1 and 2, respectively. This serves as a verification that both the SF allocation is responding as it should be as are the QoS parameters.

Just like the last experiment, the throughput is numerically analysed to verify the functionality of the extended simulation platform. Figure 5 shows the throughput of the simulated scenario along with the theoretical maximums calculated as shown in Equation 4 but with parameters adapted for 100 RBs.

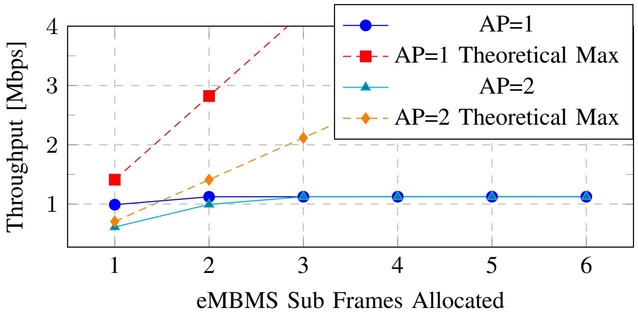


Fig. 5. eMBMS service throughput along with theoretical maximum throughput with increasing sub frame allocation for an Allocation Parameter of 1 and 2.

It seems that with broadcast as the only service in the cell, the 20MHz of spectrum allocation is somewhat under utilised. Nevertheless the platform shows, certainly clearly for an AP of 2 that the throughput falls in line with the maximum theoretical

---

[1] Production: Poznan University of Technology

throughput resulting in a bandwidth limited scenario for SF allocation of only 1 or 2 per frame. With an AP of 1 the cell is close to its limit with a SF allocation of only 1 per frame but as seen above, the throughput required for the Poznan St. sequence is reached once an allocation of 2 SFs per frame is set.

V. HUBS DYNAMIC BROADCAST RESOURCE ALLOCATION ALGORITHM

This section introduces the proposed design for the Hybrid Unicast Broadcast Synchronisation (HUBS) framework as well as its dynamic broadcast resource allocation strategy. The HUBS framework's primary objective is to, despite varying cell conditions and loading, minimise the time offset between related streams delivered jointly by LTE unicast and broadcast services. By considering the stream offset in addition to delay, the streams may adapt, together, to varying load conditions. Since bandwidth must be split between eMBMS and unicast services, this subsequently offers benefits to the entire cell.

Consider a scenario where, due to insufficient unicast resources for the requested traffic, the stream begins to see an increased bearer queue building at the eNB. In this instance, the unicast stream will show a lag versus the broadcast, the HUBS framework will then consider whether some resources from the broadcast stream can be freed and re-allocated to the unicast pool. The framework must also ensure resources are not over-provisioned to less opportunistic broadcast services as this quickly has a detrimental effect on the cell's efficiency.

A. Proposed System Design and Implementation

The LTE eMBMS architecture (presented in Section II-A), shows the MCE is uniquely positioned to gather the required user data for calculating the offset between the streams. It is also the entity that holds responsibility for radio resource management of all eMBMS services. It is therefore chosen as the key node in which to implement the HUBS framework management. As an additional challenge, this research is designed in such a way as to allow implementation into a real world LTE test bed network via a software update.

As observed in Section IV, the magnitude of the effect of varying the broadcast resource allocation parameter AP was far too course an adjustment and once a stream is established it is unlikely to require changing; therefore, the active varying of resources is performed utilising the SF map.

1) HUBS Processor: To keep this in line with the current LTE design structure, it was decided to implement HUBS entirely as a module named the 'HUBS Processor'. The 'HUBS Processor' is responsible for managing the HUBS framework. This will perform all of the required processing to keep the framework information up to date. Designed with extendability in mind, the 'HUBS Processor' separates and manages statistics and properties of each hybrid unicast broadcast service via the use of a 'HUBS group' entity.

Where the HUBS service is active during a simulation, the processor will prompt each group to refresh the statistics of each member periodically at settable intervals. This is also true for the processing of group statistics and finally decision making on the reallocation of resources via the eMBMS SF map. These periodic intervals are defined as P_{msr} and P_{gsr} system frames for member and group statistics refresh, respectively. P_{dad} is the period, in system frames, between dynamic allocation decisions as demonstrated in Algorithm 1.

Algorithm 1 HUBS Processor *process()* function. Called for every system frame.

```
Require: Variables retrieved from running simulation.
    $SFN = LTESystemFrameNumber$
    $P_{msr} = PeriodforMemberStatsRefresh$
    $P_{gsr} = PeriodforGroupStatsRefresh$
    $P_{dad} = PeriodforDynamicAllocationDecision$

function PROCESS(SFN)                ▷ where $SFN \in \mathbb{Z}^+$
    if $SFN$ mod $P_{msr} == 0$ then
        MemberStatsRefresh()
    end if
    if $SFN$ mod $P_{gsr} == 0$ then
        GroupStatsRefresh()
    end if
    if $SFN$ mod $P_{dad} == 0$ then
        DynamicBcastAllocationDecision()
    end if
end function
```

Unlike unicast services, which have a dedicated infrastructure for rapidly adapting resource allocation, LTE does not support the changing of parameters for a given eMBMS service without first issuing an update on the control channel as described in Section III-A1. With the small change proposed, updates are periodic and defined by the MCE. Due to this limitation, each group will only define an allocation decision period satisfying $P_{dad} > BroadcastSchedulingPeriod$, where $P_{dad} \in \{32, 64, 128, 256\}$. Given the frequent variability of cell conditions as well as the instantaneous nature of video stream bit rate (i.e., size of an I versus a P frame), it is inadequate to make an assessment of stream offset and subsequent dynamic allocation decision based on only a single instantaneous time sample. Therefore, an average is maintained within each HUBS group by sampling the offset with greater frequency between each dynamic allocation decision. This is the purpose of individually assigning P_{msr} that defines this more frequent sampling period, where $P_{msr} \in \{4, 8, 16, 32\}$. An example is more clearly illustrated in Figure 6. Here the member statistics sampling period is assigned as 4 system frames and both the 'Group Statistics' and 'Allocation Decision' have a period of 32 system frames assigned.

2) Deriving Stream Offset: Within a group, each member user k has the time offset δ_k between their corresponding unicast and broadcast streams calculated by probing the bearer queues at the eNodeB; this is defined as the Inter Arrival Difference (IAD). A positive or negative δ_k implies the broadcast stream is leading or lagging, respectively. This is since the broadcast stream is defined as the anchor to which the offset of each unicast stream will be measured. In the case where the broadcast and unicast streams are transmitting the same

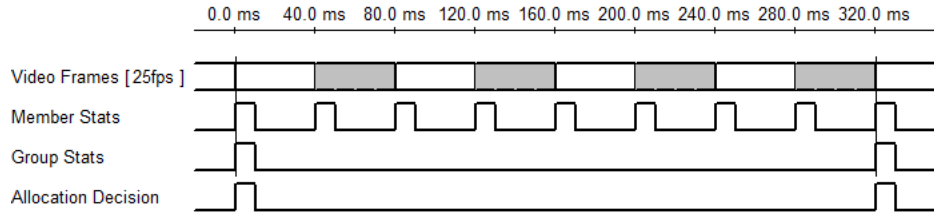


Fig. 6. Example Timing diagram where the member statistics sampling period (P_{msr}) is 4 system frames and both the Group Statistics (P_{gsr}) and Allocation Decision (P_{dad}) have a period of 32 system frames.

video frame, δ_k is set to zero.

To improve accuracy in the case where the broadcast stream leads the unicast, a record of the previous 10 video frame numbers, along with their transmission times is kept for the broadcast stream, resulting in:

$$\delta_k = t_{now} - t_{bcast,f} \tag{5}$$

where t_{now} is the current simulator time and $t_{bcast,f}$ equals the time of transmission within the broadcast stream of frame number f, kept for the preceding 10 frames. For example, where F_b is the current video frame being broadcast the following is true:

$$t_{bcast,f}\ \forall\ f \in \left\{\ x \in \mathbb{Z}^+ \mid F_b - 10 < x \le F_b\ \right\} \tag{6}$$

Maintaining a similar updated record for each unicast user in the group would prove computationally expensive. Therefore, only the current video frame $F_{u,k}$ of the unicast stream for user k is retrieved. Should a unicast user lead the broadcast stream, the stream offset will be calculated by establishing how many frames the lead consists of, multiplied by the frame duration j in milliseconds:

$$\delta_k = j(F_b - F_{u,k}) \tag{7}$$

At this point δ_k is representative of the instantaneous stream offset for user k at simulator time t_{now}. Although this instantaneous value is stored, the decision making within the HUBS framework is performed on the mean of the exponential moving average $\delta'_{k,i}$ of each user k, where i is the measurement index at system frame $Fsys$ making $i-1$ the prior measurement at $Fsys - P_{msr}$ (since updates only occur every P_{msr} system frames). The calculation is performed with the following equation:

$$\delta'_{k,i} = (\alpha\delta_k) + (1-\alpha)\delta'_{k,i-1} \tag{8}$$

The coefficient α serves to provide a factor by which the weight of older observations fall off and is assigned between 0 and 1. This weighting factor will decrease exponentially for each historic datum as a new reading is taken. Where α is closer to 1, the result will more quickly discount older observations. This implementation was chosen due to the flexibility offered without the system overheads of storing and processing historical buffer of results for each user.

3) HUBS Group Mean and Dynamic Resource Reallocation: The previous section outlined how the exponential moving average stream offset for each user within the HUBS group is derived and maintained at a frequency of P_{msr} system frames. The dynamic reallocation of resources within each group (and hence eMBMS service) is based upon a calculation of the mean of the exponential moving average across all users within the group. The calculation of this group mean is performed every group statistical refresh period of P_{gsr} system frames. Expanding on this architecture, let each HUBS group G_{ID} maintain a set of users $K = \{x \in \mathbb{Z}^+ \mid x \le N\}$ where, N is the total number of member users. The mean exponential moving average, $\triangle_{Grp}$ can thus be calculated as such:

$$\triangle_{Grp} = \frac{1}{N}\sum_{k=1}^{N}\delta'_{k,i} \tag{9}$$

At this point the group statistics have been updated and are ready to be read by the Dynamic Sub Frame (DSF) allocation algorithm to make a decision on resource reallocation. The objective of the algorithm is to minimise the time offset between the broadcast (anchor) stream and each of the unicast services whilst respecting the QoS parameters of each. The calculation is called for each group from the HUBS processor every P_{dad} system frames. The decision is based on the group's average offset $\triangle_{Grp}$ value. This is more clearly and concisely described in Algorithm 2 that expands on the functions first introduced in Algorithm 1.

There are 3 outcomes from which the decision making algorithm can select; to increase, maintain or decrease eMBMS resource allocation for the given group anchor stream. The bounds of these decisions are determined on the QoS max delay parameter, τ_{max}, defined for the broadcast application along with two thresholds that are derived by applying scaling factors to the τ_{max} delay parameter. The first threshold scaler is named the 'HUBS Delta Threshold', κ_{delta}, and can be defined between 0 and 1. This determines the area analogous to a dead band where, should $\triangle_{Grp}$ remain above $-\kappa_{delta}\tau_{max}$ but below $\kappa_{delta}\tau_{max}$ the decision is made to maintain the current allocation. Should $\triangle_{Grp}$ drift below $-\kappa_{delta}\tau_{max}$ or above $\kappa_{delta}\tau_{max}$, the algorithm will increase or decrease resources reserved for broadcast accordingly. The HUBS algorithm will also honour the broadcast bearers QoS delay constraints, guaranteeing service conditions for broadcast only users who are not members of the HUBS group. This is where the second threshold scaler appears, named the "keep within" threshold. Once again defined between 0 and 1, κ_{within} multiplied by τ_{max} defines the upper threshold to which the eMBMS broadcast stream delay may reach before the HUBS group is unable to continue reducing its own allocated resources. This

Algorithm 2 HUBS dynamic allocation algorithms for delta calculations and member and group statistics

```
Require: Retrieved from running simulation.
    $\alpha \leftarrow FalloffWeightFactor$
    $j = 1000 \times (1/framerate)$ ▷ Video frame duration (ms)

function MemberStatsRefresh(void)
    for each user $k$ where $k \in K$ do
        if user $k$ bearer queue $> 0$ then ▷ Packets Queued on Bearer
            $F_{u,k} \leftarrow$ Frame No. of Next Queued Transmission
            $\delta_k \leftarrow$ getDeltaFromAnchor($F_{u,k}$)
        else ▷ No Packets Queued on Bearer
            $F_{u,k} \leftarrow$ Frame No of Last Transmission
            $\delta_k \leftarrow$ getDeltaFromAnchor($F_{u,k}$)
        end if
        $\delta'_{k,i} \leftarrow (\alpha\delta_k) + (1-\alpha)\delta'_{k,i-1}$ ▷ Calculate Exponential Moving Average
    end for
end function

function GroupStatsRefresh(void)
    $Count \leftarrow 0$
    $\sigma \leftarrow 0$
    for each user $k$ where $k \in K$ do
        $\sigma \leftarrow \sigma + \delta'_{k,i}$ ▷ Sum Exponential Moving Average
        $Count++$
    end for
    $\triangle_{Grp} \leftarrow \frac{1}{Count} \cdot \sigma$
    return $\triangle_{Grp}$
end function

function getDeltaFromAnchor($F_{u,k}$)
    if $F_{u,k} == F_b$ then
        $\delta_k = 0$
    else if $F_{u,k} < F_b$ then
        if $F_{u,k} > (F_b - 10)$ then
            $\delta_k = t_{now} - t_{bcast,F_{u,k}}$
        else
            $\delta_k = j(F_b - F_{u,k})$
        end if
    else if $F_{u,k} > F_b$ then
        $\delta_k = j(F_b - F_{u,k})$
    end if
    return $\delta_k$
end function
```

also defines the delay threshold for the broadcast stream at which the HUBS framework will allocate further resources to the broadcast regardless of its own objectives to satisfy the QoS conditions. By setting these bounds lower than the maximum delay itself, there is less chance the delay will momentarily exceed the maximum delay (τ_{max}). This function is more concisely described in Algorithm 3.

The management of resource re-allocation is performed by a separate entity named the Dynamic Sub Frame Helper (DSFH) that is assigned to each HUBS group. This tracks the current sub frame allocation and manages requests to increase and decrease resource allocation from the HUBS Processor, translating this to a SF bitmap for the MCE. This is done by defining a 'sub frame index' where each index references a sub frame map. When an increase or decrease in eMBMS assigned resources is requested, the map index is incremented, or decremented, respectively. By default, the maps are defined with linearly increasing allocation, for example, an index of 0 would be map '10000' and 1 would be '110000' etc. This enables future expansion of the system, certainly with 24 bit maps, to offer non linear, profiled allocation of resources within given scenarios. Should a request to increment or decrement the index when at either the end or start of the available range, the allocation will remain static.

Algorithm 3 HUBS - Dynamic Broadcast Allocation Decision Function

```
Require: Retrieved from running simulation.
    $\tau_{max} \leftarrow QoSMaxDelay$
    $k_{delta} \leftarrow HUBSDeltaThreshold$
    $k_{within} \leftarrow KeepWithinThreshold$ Dynamic Threshold
    $d_{hol} \leftarrow HeadOfLineDelayforBcastBearer$
    $SFmapLength \leftarrow 6$ $or$ $24$ ▷ Length of bit map for 1 or 4 frame allocation.

function DynamicBcastAllocationDecision(void)
    if $\triangle_{Grp} > k_{delta}\tau_{max}$ and $d_{hol} < k_{within}\tau_{max}$ then
        if $index < SFmapLength - 1$ then
            $index++$
        end if
    else if $\triangle_{Grp} < -k_{delta}\tau_{max}$ or $d_{hol} > k_{within}\tau_{max}$ then
        if $index > 0$ then
            $index--$
        end if
    end if
    $bitmap \leftarrow mapFromIndex(index)$ ▷ Retrieve map for index
    $MCE- > setEMBMS_SFBitmap(bitmap)$ ▷ Command MCE to use new map
end function
```

B. Initial Performance Evaluation

To evaluate and assess the proposed HUBS DSF allocation algorithm, the design is implemented within the extended LTE-Sim platform and a mixed traffic scenario is established and simulated. For this scenario, identical simulations are run with and without the HUBS DSF allocation framework enabled. For the control simulations where no DSF allocation is utilised, the simulation is run for all possible SF bitmaps for the given AP of 1. The core parameters of each simulation remain fixed, these are shown in Table III along with assignment percentages of user services across the cell.

For this initial analysis, once again the Poznan Street[2] sequence is utilised. Two views are transmitted, jointly encoded

---

[2] Production: Poznan University of Technology

TABLE III

HUBS DYNAMIC SF INITIAL PERFORMANCE EVALUATION SIMULATION PARAMETERS (MISSING PARAMETERS REMAIN SAME AS TABLE II)

Parameter	**Value**
Simulation Time	150s Per Run, 30s Warm Up, 10 Runs
User Numbers	10 - 60 users, interval of 5
eMBMS AP	1 frame
eMBMS Allocation	single frame
eMBMS Bitmaps (Static Allocation)	111111, 111110 , 111100, 111000, 110000, 100000 When HUBS Dynamic SF Allocation DISABLED

QoS Max Delay	eMBMS = 250ms
	Video = 250ms

HUBS Parameters	When HUBS Dynamic SF ENABLED
P_{msr}	4 system frames
P_{gsr}	32 system frames
P_{dad}	32 system frames
a **coefficient**	0.2
User Service	
Broadcast	100% Total Active Users
Video	Poznan St CAM4 (720p QP27)
Enhancement	60% Total Active Users
Video	Poznan St CAM3 (720p QP27)
Voice Calls	30% Total Active Users
Internet Browsing	10% Total Active Users

as an MVC stream with the JMVC reference encoder [20]. Camera 4 of the sequence is chosen as the popular stream that is transmitted via broadcast to end users and is encoded as the MVC 'independent' stream. Camera 3 of the sequence is encoded as the enhancement to the broadcast stream and thus encoded as the MVC 'dependant' stream. This is transmitted to a subset of users via unicast. Since the unicast enhancement stream depends on the broadcast primary stream for decoding, in this scenario, any users receiving the secondary stream are subscribed to the HUBS group. Users who successfully decode both streams are able to enjoy stereoscopic 3D video. Where DSF allocation is disabled, static allocation maps are assigned for the duration of the simulation. The eMBMS AP is 1 frame, implying that every frame in the simulation will have some allocation of resources to broadcast.

Firstly a look at the key metric results of Inter Arrival Difference (IAD), measured in milliseconds. The DSF IAD against that of all the assignable static SF maps is shown in Figure 7. It is clear that the minimum allocation of just a single SF in every frame is completely insufficient to sustain the broadcast throughput. As such the plot representing the '100000' allocation has been excluded from the results as it falls completely out of range and appears below -200ms in a lightly loaded cell.

It is immediately clear that the DSF allocation mechanism introduced performs desirably at both reducing and maintaining the IAD in the given scenario. Furthermore, it illustrates how stable the DSF allocation algorithm performs even at the two extremes of such a wide variation in cell loading. The average of the DSF allocation IAD plot is 36.5ms with a standard deviation of just 4.3ms. This implies for the most part the IAD is reduced to less than the duration of a single video frame, significantly reducing the delay and complexity introduced through the requirement of buffering techniques.

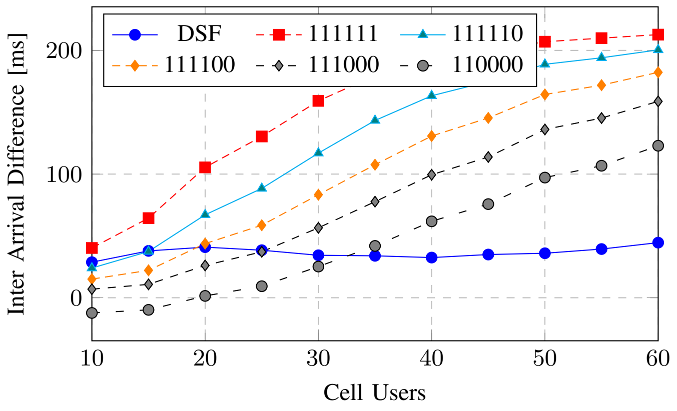


Fig. 7. Results for Inter Arrival Difference with increasing cell users; illustrating the time difference in arrival of the unicast and broadcast streams to a given HUBS subscriber with DSF or Static SF map allocation.

VI. HUBS DYNAMIC CONTENT ADAPTATION ALGORITHM

The work in this section is a continuation in the development of the HUBS framework. The motivation and objectives underpinning the original research remain identical: facilitating the delivery of high quality interactive multimedia utilising a hybrid broadcast unicast approach. Furthermore this must be achieved in a spectrally-efficient manner, without compromising end-user's experience all whilst remaining compliant with LTE 3GPP technical specifications. Whilst the DSF allocation algorithm was a cell centric approach, making changes at cell-level based on cell-level statistics, the Dynamic Content Adaptation (DCA) algorithm proposed here works at a member user level, adapting each member's content stream based on a hybrid of cell and user statistics.

The functionality on which the HUBS Dynamic Content Adaptation (DCA) algorithm builds is based upon existing concepts of dynamic adaptive video streaming protocols; namely the MPEG DASH protocol already widely accepted as the standard in adaptive video streaming for LTE. The proposed algorithm is presented with a selection of video streams at varying quality levels. As the video quality deteriorates through higher compression or reduced spatial resolution, so does the resultant bitrate. By varying between these levels, and thus bitrates, streams can be brought into synchronisation. The DCA algorithm should always seek to provide the user with the best possible quality stream in any given scenario.

A. Proposed HUBS DCA design and integration

In order to implement the support for dynamic adaptive video within the LTE-Sim platform, a new unicast application was generated named the 'Trace Based Enhanced' application.

This application is able to read in a trace file that contains multiple quality levels. The mechanics and calculations for the DCA algorithm are contained within a separate class named the 'DCA Helper'. This stores and tracks information such as the number of available layers, the currently selected layer index and the time elapsed since the last segment. Finally, the framework also defines a DCA profile, containing an array of settable parameters to tune the behaviour of the algorithm.

Once again modelled from the MPEG DASH framework, a minimum duration for each quality layer selection is defined by the 'segment length' variable. With the HUBS implementation, when the segment length duration has elapsed, the 'DCA Helper' makes a decision on whether the current layer index is suitable or should be reallocated in either direction. Unlike the most popular, "client driven" form of DASH style content delivery, where the client will request segments of a particular quality level, this implementation will be driven by the network itself. Of course any implementation of this nature will induce some control overhead. The feedback required for this implementation would amount to nothing more than a unique identifier for each user along with a quality level to serve to this user. Furthermore, by only transmitting updates on a quality level transition, the frequency of these feedback transmissions is further reduced.

B. The HUBS DCA Algorithm

The algorithm is primarily governed by the Head Of Line (HOL) delay, $D_{HOL,k}$, for each user k as part of the HUBS group G_{ID}. The Head Of Line (HOL) delay is a measure of the total delay of the packet at the head of the bearer queue, amounting to the total delay of the bearer. This is built upon the architecture in Section V-A3 where each G_{ID} maintains a set of users $K = \{x \in \mathbb{Z}^+ \mid x \leq N\}$ where, N is the total number of member users. To ensure the video stream packets are not lost due to delay-based QoS restrictions. This is also where the two threshold parameters, τ_1 and τ_2 come into play, defining a lower and upper bound, respectively. Together, these thresholds define three segments of a user's HOL delay, each of which will see the algorithm employ a different behaviour. The complete DCA algorithm is shown in Algorithm 4, where behaviours break down as follows:

Where $D_{HOL,k} \leq \tau_1$ the algorithm will perform a check on the average group HOL delay, $D_{HOL,G_{ID}}$, maintained by the HUBS processor for group G_{ID}. Only if this should meet the condition $D_{HOL,G_{ID}} \leq \tau_1$, is the quality index for user u incremented by one. When these conditions are met simultaneously, favorable conditions are experienced by both the UE in question as well as other member users within the group. This second condition check prevents the quality of a given, particularly fortunate user increasing the burden on what may be an otherwise saturated cell, impeding the quality of other cell users and services. This segment fulfills the algorithm's condition to ensure that the greatest quality stream is provided to the user where conditions allow.

The next segment lies between the defined thresholds, satisfying $\tau_1 < D_{HOL,k} \leq \tau_2$. Here the algorithm assumes more typical HUBS behaviour, comparing the exponential moving average IAD, δ'_k of each user, introduced in Section V-A2, to the HUBS threshold τ_{HUBS}. This results in users who are ahead of the broadcast by a time difference greater than that of τ_{HUBS} having their quality increased. The opposite is true for users who are behind the broadcast stream by greater than τ_{HUBS}; their quality is reduced. This has the effect of tightening each user individually around the broadcast stream, but only where the user in question is within the two threshold values.

The final segment is where the condition $D_{HOL,k} > \tau_2$ is satisfied, showing a user's delay approaching the QoS maximum delay value, τ_{max}. This is a critical area, ensuring users do not suffer a complete loss of the unicast stream as well as alleviating cell congestion. For this, a novel back-off function was developed, factoring in the upper threshold as well as the number of quality layers available within the service.

First, a quality scaling factor, φ, is calculated based on the number of available quality layers, L within the service. This calculation breaks down as follows:

$$\varphi = \frac{(\tau_{max} - \tau_2)^3}{(L \cdot 10^3)} \tag{10}$$

Once φ is calculated, it does not require re-calculation unless the DCA profile changes. At this point a calculation is performed to establish how fast to reduce the video quality by means of, $\lambda(D_{HOL,k})$, the value dictating how many layers to jump back. This is done based on the current value of $D_{HOL,k}$ for user k. The complete function takes the form:

$$\lambda(D_{HOL,k}) = -\rho \frac{\left(\frac{D_{HOL,k} - \tau_2}{10}\right)^3}{\varphi} \tag{11}$$

Where ρ serves as a quality back-off sensitivity value settable as part of the HUBS DCA profile. To get a better understanding of the construction of the backoff function, along with the calculated number of levels by which to scale back the quality of the stream, Figure 8 illustrates several variations of possible profiles and parameters in (A) through (D).

Figure 8 (a) varies the 'Quality Backoff Sensitivity' variable ρ. Where $\rho = 1$, when the HOL delay reaches τ_{max} the algorithm will have dropped the streams index L quality levels, essentially ensuring that the minimum quality stream is now being transmitted. In most cases, the most desirable result is to ensure the quality is scaled back before τ_{max} is reached. Thus, increasing ρ has the effect of increasing the sensitivity of the algorithm. A value of $\rho = 2$ will result in a drop of quality index of L levels at 90ms where $\tau_{max} = 100ms$. Figure 8 (b) shows the function scaling to scenarios with different numbers of available stream quality levels. Once again, where $\rho = 1$ and HOL delay reaching τ_{max}, it is ensured that the minimum quality layer is selected. Figure 8 (c) shows the function becoming more aggressive with scaling back the layers as the segment between τ_2 and τ_{max} is reduced. Lastly, Figure 8 (d) confirms the behaviour of the function scaling to accommodate different profile values for τ_{max}.

Algorithm 4 HUBS DCA Algorithm implementation pseudocode

```
function DCA_DECISION_MAKER($D_{HOL,k}$)
    if $D_{HOL,k} \leq \tau_1$ then                         ▷ User HOL Below Threshold 1
        if $D_{HOL,G_{ID}} \leq \tau_1$ then                ▷ Group Below Threshold 1
            LayerIndexChange(1)                             ▷ Increment Quality index by 1
        end if
    else if $\tau_1 < D_{HOL,k} \leq \tau_2$ then           ▷ User HOL between Threshold 1 and 2
        if $\delta'_k \leq -\tau_{HUBS}$ then               ▷ User ahead of broadcast by > HUBS Threshold
            LayerIndexChange(1)                             ▷ Increment Quality index by 1
        else if $\delta'_k \geq \tau_{HUBS}$ then           ▷ User behind broadcast by > HUBS Threshold
            LayerIndexChange(-1)                            ▷ Decrement Quality index by 1
        end if
    else if $D_{HOL,k} > \tau_2$ then                       ▷ User HOL Greater than Threshold 2
        $\varphi \leftarrow (\tau_{max} - \tau_2)^3/(L \cdot 10^3)$       ▷ Calculate Quality Layers Scaling Factor
        $\lambda \leftarrow -\rho[(D_{HOL,k} - \tau_2)/10]^3/\varphi$     ▷ Calculate Change in Quality
        LayerIndexChange(round($\lambda$))                  ▷ Adjust Quality index
    end if
end function
```

VII. PERFORMANCE EVALUATION

A. 3D Stereoscopic Evaluation Scenario

In order to more systematically evaluate the proposed HUBS framework, a more detailed scenario has been established. Designed based on trend forecasts in both research and industry to establish how the framework may be used in a real network delivering next generation multimedia.

This scenario explores the delivery of a popular live television event with stereoscopic 3D coverage (i.e., a football game or a motorsport event). In this case, users wishing to view the broadcast in stereoscopic 3D must be receiving the left and right views simultaneously. The left view is broadcast to all subscribing users within the cell. Since it is likely only a subset of users within the cell will be capable of, or choose to watch the coverage in 3D, these users are catered for using unicast transmissions.

This scenario provides a particular focus on how effectively the DSF allocation of the HUBS framework responds to variations of video content during the broadcast. Given the content has a significant impact on the video encoder and resultant data rate, in order to best test the proposed model, a video sequence with properties true to a typical live broadcast was created. This sequence was formed of multiple test clips compiled from sequences available and familiar to the research community. Each of these has been chosen to provide a mix of both spatial and temporal information more representative of a live video broadcast. This included fixed and panning camera shots as well as scene cuts. Table IV lists, in order, the name and duration of each clip used to compile the final sequence totalling 120 seconds run time.

For use with the DCA algorithm, several quality levels are required. The video stream was encoded using H.264/AVC with the open source x264 based encoding library, at a range of spatial resolutions and quantization parameters [21]. The final encoded sequence properties are shown in Table V.

Once again the right view is unicast at half the spatial resolution, as it will be displayed simultaneously with the left view on the end-user's device, as the current industry standard side by side mechanism does.

TABLE IV
BREAKDOWN OF COMPILED VIDEO SEQUENCE

	Sequence Frames		**Duration**
New Clip Name	**Start**	**Total**	**(Sec)@25fps**
24h Clip 1	1320	500	20
24h Clip 2	2500	500	20
Big Buck Bunny	10500	1000	40
Café Cam 4+5	0	250	10
Poznan St Cam 4+5	0	250	10
Shark Cam 4+5	0	250	10
Micro World 1+2	0	250	10
Total	-	**3000**	**120**

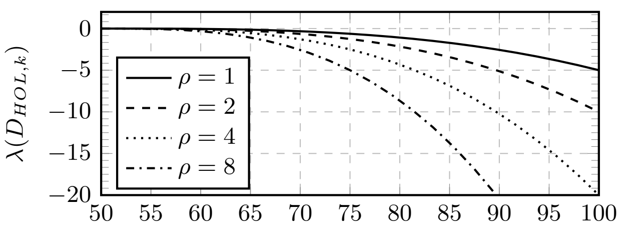
TABLE V
ENCODED SEQUENCE PROPERTIES FOR BROADCAST (LEFT) VIEW AND MULTI STREAM UNICAST (RIGHT) VIEW. THE QUALITY INDEX (QI) IS THAT SELECTED BY THE DCA ALGORITHM.

QI	**View**	**Resolution**	**CRF**	**PSNR**	**Bit rate**
-	Left	1280x720	27	37.12	1180 Kbps
4	Right	640x720	27	37.5	845.69 Kbps
3			30	36.6	658.94 Kbps
2			33	35.3	480.83 Kbps
1		480x540	30	36.3	447.8 Kbps
0			33	34.6	311.56 Kbps

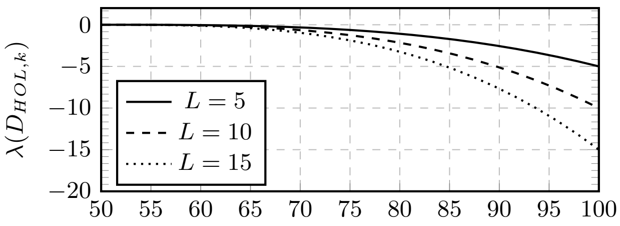
B. Simulation Parameters

Where DCA is not implemented, the simulation is carried out with the right stream encoded at 640x720 with a CRF of 30, or Quality Index (QI) number 3 in Table V. This makes the range of selectable quality levels a single QI increase and a 3 QI decrease when compared to the statically allocated map. The complete system level simulation parameters are listed in Table VI.

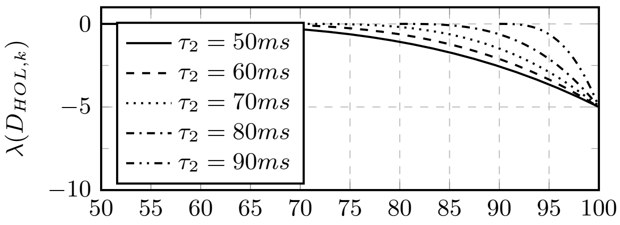
(a) $[\tau_{max} = 100ms \mid \tau_2 = 50ms \mid L = 5]$



(b) $[\tau_{max} = 100ms \mid \tau_2 = 50ms \mid \rho = 1]$



(c) $[\rho = 1 \mid L = 5 \mid \tau_{max} = 100ms]$



(d) $[\rho = 1 \mid L = 5 \mid \ \tau_2 = 50ms]$

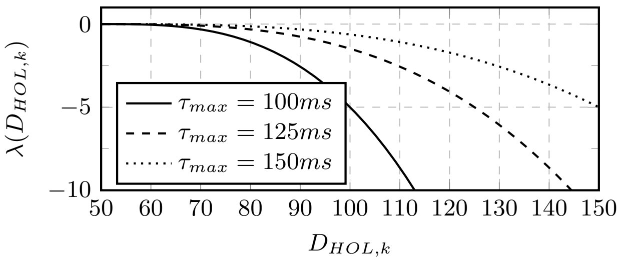


Fig. 8. Example plots for $\lambda(D_{HOL,k})$ where $D_{HOL,k} > \tau_2$. (a) Varies quality back-off sensitivity, (b) Varies number of quality layers, (c) varies upper threshold parameters, (d) shows response with increasing QoS max delay.

TABLE VI
SIMULATION SCENARIO PARAMETERS FOR DCA PERFORMANCE EVALUATION (MISSING PARAMETERS REMAIN SAME AS TABLE III)

Parameter	**Value**
Simulation Time	150s Per Run, 30s Warm Up, 15 Runs
User Numbers	5 - 60 users, interval of 5

QoS Max Delay	eMBMS = 100ms
	Enhancement Video = 100ms (QCI-7)
	Other Video = 100ms (QCI-7)
	VoIP = 100ms (QCI-1)

HUBS Parameters	When HUBS DSF Allocation ENABLED
P_{msr}	4 system frames
P_{gsr}	32 system frames
P_{dad}	32 system frames
a **coefficient**	0.8
DCA Profile	$\tau_1 = 10ms$, $\tau_2 = 50ms$, $\tau_{HUBS} = 10ms$ $\tau_{max} = 100ms$, $\rho = 2$, $L = 5$ Segment Length = 1s
User Service	
Broadcast Video	60% Total Active Users Compiled Sequence Left View 1280x720 CRF27 H.264/AVC
Enhancement Video	50% Total Broadcast Users Compiled Sequence Right View DCA Disabled: 640x720 CRF30 H.264/AVC
Other Video	10% Total Active Users Foreman H264 440Kbit
Voice Calls	20% Total Active Users
Internet Browsing	10% Total Active Users

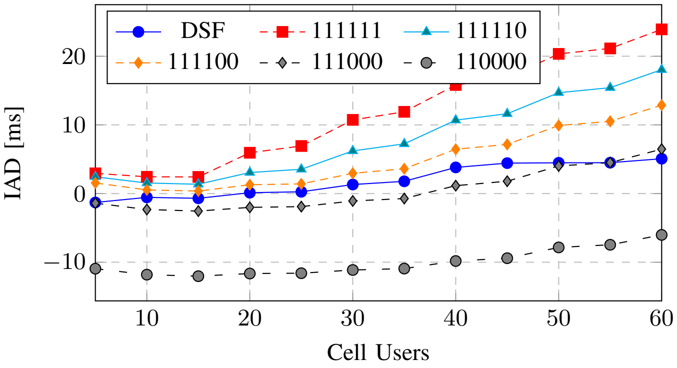


Fig. 9. Results for Inter Arrival Difference with increasing cell users; illustrating the time difference in arrival of the joint unicast and broadcast streams to a given HUBS subscriber with DSF or Static SF map allocation.

C. Results and Discussion

Initially, the simulation scenario is used to evaluate the DSF algorithm introduced in Section V. The results for IAD are shown in Figure 9. Once again an allocation map of '100000' is unable to sustain the bit rate required for the broadcast stream. The '110000' SF map is also showing some signs of elevated delay, remaining in the negative values even at 60 users.

Further understanding can be gained where the IAD is assessed over the duration of the simulation time. Figure 10 plots the IAD averaged across all runs against simulation time. The results in Figure 10 reveal the large and unstable variations experienced in IAD. This instability between the streams is due to the variation in the bit rate of the encoded sequence caused by the dynamic nature of the video content. Here the proposed HUBS algorithm reveals its real potential in offering maximum flexibility to mobile network operators whilst dealing with the difficult real world multimedia scenarios and applications an LTE network will face. Furthermore, both static allocation maps ’110000’ and ’111000’ show a great deal of variation and jitter, some of which greatly exceeds 100 milliseconds. Where the SF allocation is performed dynamically through the proposed HUBS algorithm, there is a marked improvement in stability as the algorithm ramps the resource allocation to best suite the unicast and broadcast needs. The same is seen repeated at 10 and 60 cell users, respectively. To summarise, the static SF map allocations of ’110000’ and ’111000’ are not entirely stable choices for the eMBMS delivered stream and should not be selected for static broadcast transmissions. Given these revelations, the SF maps ’111111’, ’111110’ and ’111100’ remain the focus of this analysis moving forward.

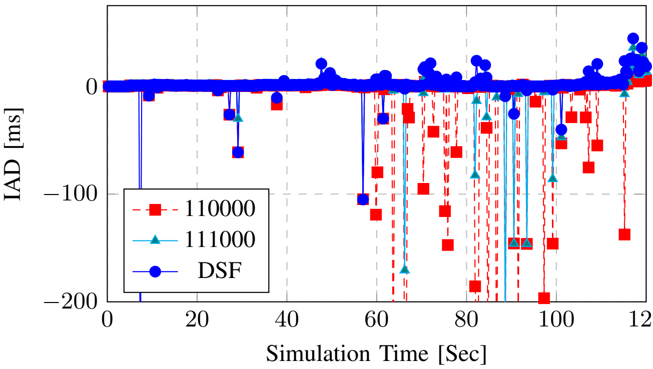


Fig. 10. IAD for SF maps 11000 and 111000 and DSF allocation against simulation time in seconds for 30 users. IAD shown is calculated from the average across all runs at the given time interval.

Now an examination of the effect of the DCA strategy on the IAD metric is performed. Figure 11 shows an array of graphs plotting IAD with increasing user numbers for a range of SF maps. In the case where the broadcast leads the unicast (Maps 111111, 111110 and 111100), DCA is clearly showing improvements in reducing the IAD between the streams as the unicast begins to struggle with greater user numbers. Since higher SF allocation maps also take greater resources from the unicast pool, the effect of DCA is most pronounced for these maps, where the algorithm is able to make the biggest impact. The DCA algorithm is primarily positioned to be most influential where broadcast is ahead due to resource limitations for unicast services within the cell. This is to complement the DSF algorithm that is most effective in the opposite light. Furthermore, given the range of quality choice the DCA algorithm has available within this scenario, it is clear to see why it is less effective in the opposing direction.

For a clearer understanding on the choices being made by the DCA algorithm, Figure 12 plots the average of the chosen quality index with increasing cell users. Also on the graph is a line representing the encoded stream utilised for the statically allocated control simulation. With fewer cell users the algorithm is comfortable to raise the quality of the end-user’s stream to the highest quality of 4, surpassing the quality offered by the static allocation. Given the graph represents an average of the selected index across each simulation run, across all users within the run and across the run duration, where the results lay between index values it is likely the algorithm has assigned alternate quality levels based on the scenario and user distribution at a given time.

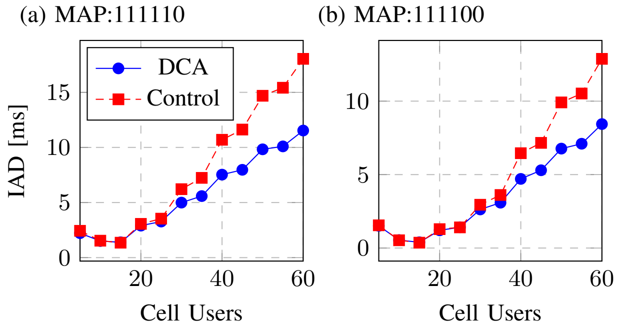


Fig. 11. Results comparing average Inter Arrival Difference with and without DCA with increasing cell users at various SF maps.

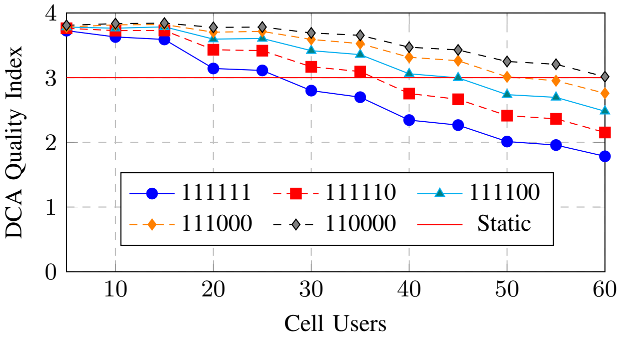


Fig. 12. Average of DCA chosen quality index with increasing users.

Following the testing of the HUBS DSF and DCA strategies independently, a more complete simulation is performed on the integrated platform. As shown, SF allocation map ’111000’ and below do not provide sufficient resources to accommodate the broadcast stream for its entirety. For this reason, maps ’111100’ and ’1111110’ are examined as these have been shown as the most plausible static allocation candidates, avoiding over or under allocation of resources.

The improvements offered by the novel DSF allocation algorithm over the default statically allocated resources have already been shown. Further to this, this section has shown the advantages offered by the DCA algorithm for stream synchronisation, particularly where the cell becomes heavily loaded. Figure 13 shows the performance of the integrated framework against standalone DSF and DCA enabled simulations. The standalone DSF algorithm reduces and stabalises IAD across the board to almost negligible levels already producing more than acceptable results for most cell conditions. The DCA algorithm shows very little change with fewer user numbers, instead helping most when the cell becomes congested and the unicast streams are adaptively scaled back to help alleviate cell load.

Figure 14 (a) and (b) present the broadcast and unicast delays, respectively. The broadcast delay for the integrated system remains well within the QoS maximum delay conditions and varies by only 6ms. As the cell users increase, the HUBS DSF algorithm will more aggressively assign resources back to unicast at opportune moments. This naturally has the

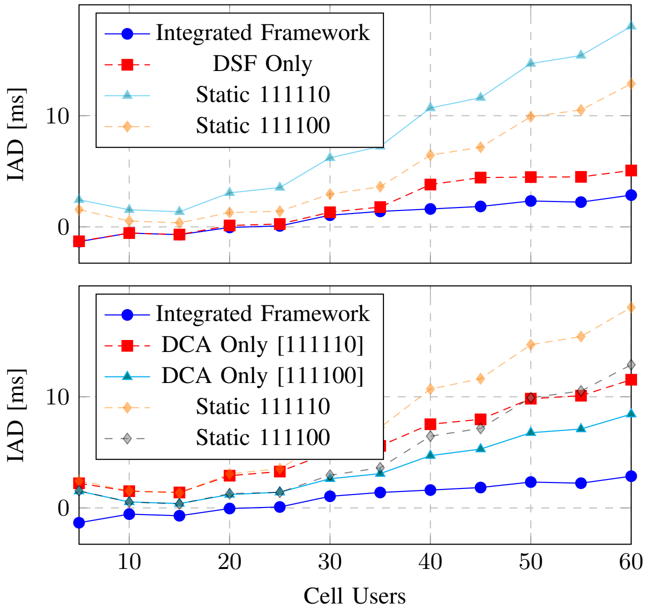


Fig. 13. Integrated HUBS framework with increasing cell load compared with standalone DSF and DCA simulations.

effect of increasing the delay slightly. This delay will only rise to approach the delay being experienced by the unicast stream. Plot (b) shows the DCA algorithm ramping down the quality and successfully reducing the delay experienced in heavy loading. (c) illustrates how releasing some broadcast resources, as well as making content allocation decisions frequently to avoid unmanageable peaks in loaded cells, the achievable throughput of the service will also increase along with the delay decrease.

Figure 14 (d) to (f) plot results from various standalone cell services, completely unlinked to the HUBS algorithm. VoIP service delay, shown in (d) sees an improvement on an already low delay. (e) graphing stand-alone video service delay also shows a reduction when making use of the integrated algorithm. Here the algorithm is opportunistically making use of lower SF allocations, all of which if statically allocated would otherwise not support the stream through its entire duration. Both the Voice Over IP (VoIP) and Video services are real-time and thus may have additional resources allocated through the unicast scheduling algorithm that is bound to attempt to meet the QoS conditions assigned. The best-effort service, with throughput graphed in (f) is not considered real-time and as such has fewer QoS boundaries and a lower scheduling priority. The traffic generation for the best-effort service is considered as an infinite buffer, meaning that it will request to get as much information through per TTI as the scheduler will assign. The instantaneously freed up resources unused by the other real-time services (or due to the fairness element of the M-LWDF scheduling algorithm in use) are also distributed among the best effort services, resulting in a clear increase in throughput.

VIII. CONCLUSION

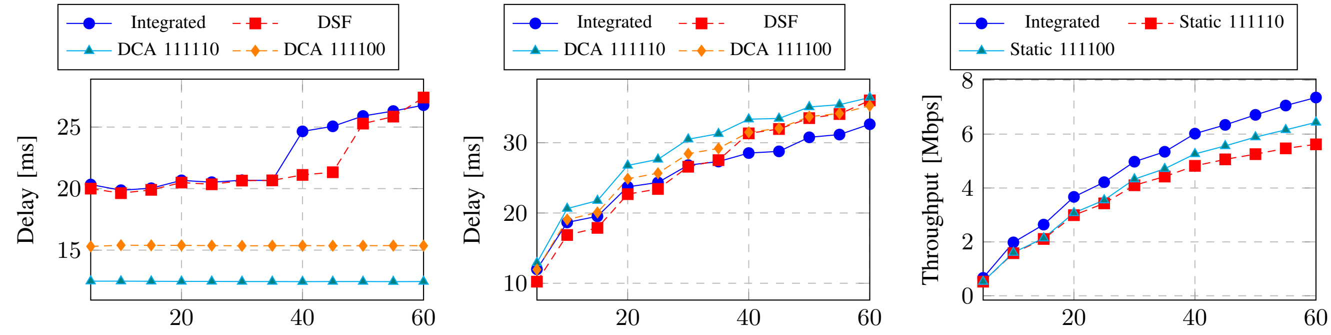
The work in this paper begins with the presentation of the design, implementation and testing of an open source LTE eMBMS simulation platform built upon the existing LTE-Sim code. The extended platform models the eMBMS Multicast Coordination Entity, a core component in eMBMS cell deployment. eMBMS services can be fully managed and configured from a standard LTE-Sim scenario file, including broadcast group Management and UE subscription handling. This fully integrates with the unicast capabilities of the simulator, obeying the resource allocation limitations outlined by the 3GPP specification documents.

A Hybrid Unicast Broadcast Synchronisation (HUBS) framework is proposed, towards fulfilling the primary objective laid out for this research: to facilitate synchronous delivery of hybrid unicast and broadcast multi-stream multimedia content. Housed within the framework is a novel Dynamic Sub Frame (DSF) allocation algorithm, which performs the task of dynamically reallocating resources based upon the Inter Arrival Difference (IAD) time between the unicast and broadcast streams. The DSF algorithm's primary objective is to minimise the IAD time between the streams whilst respecting the QoS restrictions imposed. A realistic simulation scenario challenged the DSF algorithm with a stereoscopic sequence, where the left view was broadcast and the right view delivered to HUBS subscribed users requesting 3D content. This stream was compiled of several test sequences mimicking the changing nature of real world broadcast content. The DSF algorithm demonstrated improved performance not only with the IAD metric, but also improving on the delay, throughput and PLR statistics of other cell services. Finally the algorithm also provided a "set and forget" method for mobile network operators to provide content to users in an instantaneous and straightforward way.

Finally the HUBS framework and DSF algorithm is expanded with a novel Dynamic Content Adaptation (DCA) algorithm operating on the unicast streams. The algorithm operates within three segments, each of which rely on different statistical measures. Which segment is implemented for each dynamic allocation decision is based upon the delay currently experienced by the user the decision is being made for. Should this delay be close to the maximum QoS delay conditions, the quality of the stream content will be scaled down. Alternatively, should the delay be minimal, the algorithm suggests increasing the quality. Thus, by only making quality allocation decisions based upon IAD where the cell is comfortably away from QoS limits, the algorithm is able to safely serve the HUBS objectives. Finally, a simulation with both DSF and DCA algorithms enabled and performance evaluation confirms the integrated framework is able to improve results through a range of cell users. By releasing resources in instances where they remain unused, the unicast allocation algorithms are able to fulfill more services with their opportunistic and fast changing strategies.

The entire HUBS framework implementation has been designed with integration into LTE in mind, keeping each entity and respective actions compliant with the 3GPP specification. The framework also avoids the use of complex databases and keeps calculations minimal to greatly minimise implementation overheads.

(a) Broadcast Delay (b) Enhanced Delay (c) Enhanced Throughput



(d) VoIP Delay (e) Video Delay (f) Best Effort Throughput

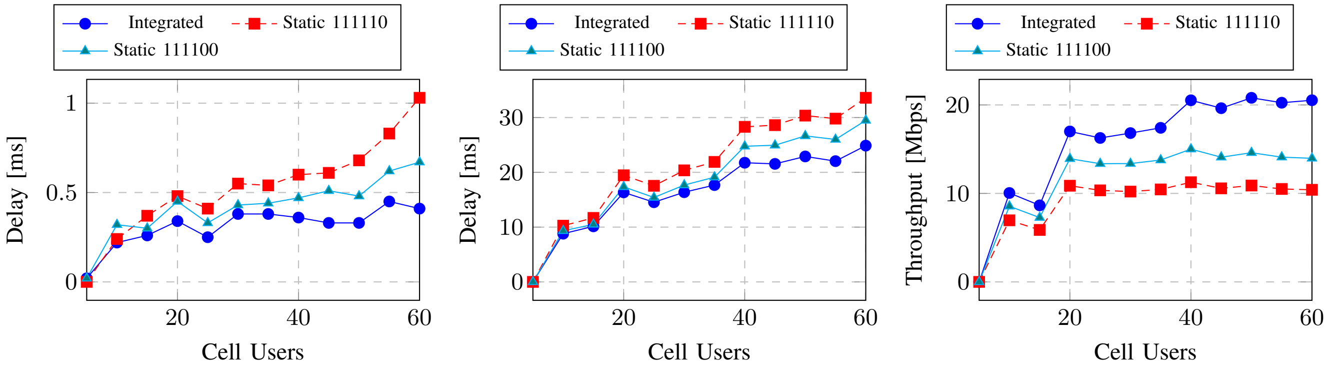


Fig. 14. Integrated HUBS Framework results for various services within the cell.

IX. FUTURE WORK

Currently the HUBS architecture proposed releases unicast resources by freeing up unrequired SFs during a particular time period. Once reallocated to unicast, these resources are assigned by the unicast scheduling strategy to any active unicast service. It would be interesting to examine whether allocating those spared resources to only HUBS unicast enhanced streams would increase the framework's performance. Furthermore, the knock on effects this has on the remainder of the services within the cell.

Given the work presented, mobile network operators are now in a position where they will have to consider a tradeoff when a cell becomes congested; reducing stream quality but servicing a greater number of users, or blocking additional users from joining the network and maintaining the stream quality of existing users. Perhaps the ideal answer lies somewhere between these choices. A logical next step for the HUBS framework would be an ability to dynamically adapt the quality of the broadcast content also. This will allow further resources to be devoted to unicast services where cell congestion is critical. This can also be linked to the allocation of the broadcast SFs through the DSF algorithm. Of course, expanding this concept further, where a cell is congested, services could also be tiered further by having two broadcast transmissions; an example would be a base layer transmission with a low order MCS and a secondary enhancement layer transmission, increasing the quality of the base layer broadcast with a higher order MCS. This will allow users with a greater signal strength, capable of receiving the more efficient transmission to benefit from a greater quality viewing experience.

REFERENCES

[1] Emarketer, "2 Billion Consumers Worldwide to Get Smart(phones) by 2016," 2014. [Online]. Available: http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694
[2] CISCO, "Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update 2016 to 2021," 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-forecast-qa.pdf
[3] G. K. Walker, J. Wang, C. Lo, X. Zhang, and G. Bao, "Relationship Between LTE Broadcast/eMBMS and Next Generation Broadcast Television," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 185–192, jun 2014.
[4] L. Shi, E. Obregon, K. Sung, J. Zander, and J. Bostrom, "CellTV-On the Benefit of TV Distribution Over Cellular Networks: A Case Study," *IEEE Transactions on Broadcasting*, vol. 60, no. 1, pp. 73–84, 2014.
[5] C. P. Lau, A. Alabbasi, and B. Shihada, "An efficient live tv scheduling system for 4g lte broadcast," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2016.
[6] GSA, "Evaluating the LTE Broadcast Opportunity," 2015.
[7] 3GPP, "ETSI TS 23.246 REL-10 - Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description," 2011.
[8] ——, "New WID: eMBMS enhancements for LTE," 2016.
[9] L. Christodoulou, O. Abdul-Hameed, A. M. Kondoz, and J. Calic, "Adaptive subframe allocation for next generation multimedia delivery over hybrid lte unicast broadcast," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 540–551, Sept 2016.
[10] D. Erik, P. Stefan, and S. Johan, *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011, no. 978-0-12-385489-6.
[11] 3GPP, "ETSI TS 36.300 REL-9 - E-UTRAN; Overall description," 2013.
[12] M. C. Dolby Laboratories Inc., Fraunhofer-Institute HHI, "H.264/MPEG-4 AVC Reference Software Manual," 2009. [Online]. Available: http://iphome.hhi.de/suehring/tml
[13] N. R. Group, "3D Video & The MVC Standard," pp. 1–2, 2009.

[14] A. Vetro and I. Sodagar, "Industry and Standards The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.

[15] 3GPP, "ETSI TS 26.247 REL-10 - Adaptive Streaming over HTTP (3GP-DASH)," 2011.

[16] G. Piro, L. Grieco, and G. Boggia, "Simulating LTE cellular systems: An open-source framework," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 498 – 513, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5634134

[17] 3GPP, "ETSI TS 36.331 REL-10 - Radio Resource Control (RRC); Protocol specification," 2013.

[18] ——, "ETSI TS 36.443 REL-9 - E-UTRAN; M2 Application Protocol (M2AP)," 2010.

[19] ——, "ETSI TS 36.213 REL-11 - E-UTRA; Physical layer procedures," 2013.

[20] J. V. Team, "JMVC reference software," 2008. [Online]. Available: garcon.ient.rwthaachen.de

[21] V. Organization, "VideoLAN - x264, the best H.264/AVC encoder," 2013. [Online]. Available: http://www.videolan.org/developers/x264.html

Dr. Louis Christodoulou received his Ph.D. degree in Electronic Engineering from the University of Surrey, United Kingdom, in 2016. His Ph.D. research explored the development of a hybrid unicast broadcast enhanced multimedia delivery framework over LTE. He also holds an MEng degree in Electronic Engineering also awarded from the University of Surrey and completed with First Class Honours in 2012. His current research interests include multipoint transmission techniques and with experience working in the television broadcast industry; future multimedia content delivery, mobile broadcast and radio resource management. He is now a 5G Research Engineer with Samsung Electronics R&D Institute UK.

Dr. Omar Abdul-Hameed received his BEng degree in Electronic and Electrical Engineering with a First Class Honours and the Ph.D. degree from the University of Surrey, Guildford, UK in 2003 and 2009, respectively. From 2007 to 2014, he was a Research Fellow at the I-Lab, Multimedia Communications Research Centre at the University of Surrey, where he was actively involved and working in several European Commission funded FP6 and FP7 collaborative research projects. He then joined the Institute for Digital Technologies (IDT) at Loughborough University London in October 2014 as a Research Associate. His main research interests include technologies for 5G mobile communications, radio resource management in mobile broadband networks, the efficient and robust video content delivery over wireline and wireless networks, and experimental test-bed design and implementation with IP-based multimedia applications. He has several publications in international journals and conferences as well as several book chapters.

Professor Ahmet Kondoz received his PhD in 1987 from the University of Surrey where he was a research fellow in the communication systems research group until 1988. He became a lecturer in 1988, reader in 1995, and in 1996 he was promoted to professor in multimedia communication systems. He was the founding head of I-LAB, a multi-disciplinary multimedia communication systems research lab at the University of Surrey. Since January 2014, Prof. Kondoz has been appointed as the founding Director of the Institute for Digital Technologies, at Loughborough University London, a post graduate teaching, research and enterprise institute. He is also Associate Dean for Research for Loughborough University London. His research interests are in the areas of digital signal processing and coding, fixed and mobile multimedia communication systems, 3D immersive media applications for the future Internet systems, smart systems such as autonomous vehicles and assistive technologies, big data analytics and visualisation and related cyber security systems. He has over 400 publications, including six books, several book chapters, and seven international patents, and graduated more than 75 PhD students. He has been a consultant for major wireless media industries and has been acting as an advisor for various international governmental departments, research councils and patent attorneys. He is a director of MulSys Ltd., a university spin-off company marketing the worlds first secure GSM communication system through the mobile voice channel. Professor Kondoz has been involved with several European Commission FP6 & FP7 research and development projects, such as NEWCOM, e-SENSE, SUIT, VISNET, MUSCADE etc. involving leading universities, research institutes and industrial organisations across Europe. He coordinated FP6 VISNET II NoE, FP7 DIOMEDES STREP and ROMEO IP projects, involving many leading organisations across Europe which deals with the hybrid delivery of high quality 3D immersive media to remote collaborating users including those with mobile terminals. He co-chaired the European networked media advisory task force, and contributed to the Future Media and 3D Internet activities to support the European Commission in the FP7 programmes.