Research Repository UCD

Title	A QoS based call admission control and resource allocation mechanism for LTE femtocell deployment
Authors(s)	Olariu, Cristian, Fitzpatrick, John, Perry, Philip, Murphy, Liam, B.E.
Publication date	2012-01-17
Publication information	Olariu, Cristian, John Fitzpatrick, Philip Perry, and Liam Murphy B.E. "A QoS Based Call Admission Control and Resource Allocation Mechanism for LTE Femtocell Deployment." IEEE, 2012.
Conference details	2012 IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, Nevada, USA, 14 - 17 January 2012
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/7524
Publisher's statement	© © 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/CCNC.2012.6181184

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

A QoS based call admission control and resource allocation mechanism for LTE femtocell deployment

Cristian Olariu
Telecommunication Software
and Systems Group
Waterford Institute of Technology,
Cork Road, Waterford, Ireland
Email: colariu@tssg.org

John Fitzpatrick
NEC Laboratories Europe,
NEC Europe Ltd.,
Kurfursten-Anlage 36, 69115
Heidelberg, Germany
Email: johnfitzpat@ieee.org

Philip Perry and Liam Murphy
School of Computer Science
and Informatics
University College Dublin
Belfield 4, Dublin, Ireland
Email: {philip.perry, liam.murphy}@ucd.ie

Abstract — Current trends show a growing number of femtocell deployments, this in turn will lead to increased volumes of voice traffic being transmitted through fixed broadband access networks such as Digital Subscriber Line. In this paper the issue of maintaining call quality through the resource constraint expedited forwarding queue of DSLAMs is investigated. A quality based Call Admission Control and resource allocation mechanism is provided to avoid resource overloading and call quality degradation. The ITU-T's E-Model is used for call quality monitoring and a message exchange interface between the mobile and fixed networks which allows dynamic adjustment to network resources is described and simulated. The results show that high voice call quality can be maintained.

Keywords: *femtocells, LTE, call admission control, QoS, E-Model, MOS, DSLAM*

I. INTRODUCTION

The Long Term Evolution (LTE) standards for cellular systems have opened many new possibilities for future mobile communications. These include concepts such as advanced Self Organizing Networks (SONs), policy based network management and further integration of femtocells and femtocell networks [1,2]. The work presented in this paper spans each of these three ideas. Specifically, this paper addresses the problem of congestion that can occur when a large number of femtocells utilise Digital Subscriber Line (DSL) connections as a backhaul link. A method to enable both Quality of Service (QoS) aware Call Admission Control (CAC) and bandwidth negotiation in the backhaul links is proposed. The algorithms developed are suitable for SON deployments and can be driven by a number of parameters that can be adjusted by operator-defined policies.

In LTE, the home base station is referred to as a Home evolved Node B (HeNB) and connects to a HeNB Gateway (HeNBGW) through the customer's fixed broadband connection, typically a DSL connection. This critical backhaul segment is often neglected in both research work and in real femtocell deployments; it is assumed to be of secondary importance to the radio link between the mobile unit and the home base station. Although the femtocells transport both voice and data, due to their real time requirements the voice connections are much more sensitive to constraints in the backhaul network where, it is often not possible to provide guaranteed network resources.

Each voice call transported via a DSL connected femtocell is essentially converted into a Voice over Internet Protocol (VoIP) call and encapsulated into a tunnel (Figure 1) for

Produced in user terminal	AMR Header	6	
	AMR Data	AMR SID	AMR 12.2
		4	32
	RTP	12	
	UDP	8	
	IP	20	
	GTP-U	8	
	UDP	8	
	IP	20	
	IPSec ESP	8	
	PPPoE	12	
	Ethernet	42	

Figure 1 - Voice over S1-U

transmission to the HeNBGW located in the operator's core network. The 3GPP have defined AMR and AMR-WB as the mandatory voice codec for LTE deployments. Each AMR VoIP call has a data rate in the region of 5 to 24 Kbps plus overhead, depending on the codec mode that is being used. In the case of IPSec secured tunnelling that is used by all femtocell deployments the overhead becomes comparable to the size of the actual payload.

Upon egress from the femtocell each VoIP packet is tagged as voice using the Differentiated Services Code Point (DSCP) field of the IP header, this allows the DSL Access Multiplexer (DSLAM) to identify it as voice and provide low-latency prioritization. In this work it is assumed that each Mobile Network Operator (MNO) operating femtocells utilizing the Internet Service Provider's (ISP) network can be individually identified on a per flow basis. This could be done by providing specific VLANs for each MNO using the ISP network, therefore allowing the DSLAM to use the VLAN tag of each packet to identify the MNO to which it belongs.

At the DSLAM all voice traffic is forwarded through the Expedited Forwarding (EF) queue. The EF queue provides the highest service level and is utilised for low latency and low bandwidth services such as voice; however in typical deployments it has a fixed and relatively limited bandwidth. As the number of HeNB deployments increases it is likely that the increased number of calls in the EF queue on DSLAMs will cause congestion and may exceed the maximum allocated bandwidth. This will lead to increased packet delays, jitter and loss at the DSLAM with a corresponding impact on the voice call quality. Therefore, the quality of the call depends on both the radio link from the customer's mobile device to their HeNB and the level of congestion on the backhaul link [3].

Figure 2 – Femtocell deployment and architecture

In this work it is assumed that each MNO will have Service Level Agreements (SLAs) in place with the ISP to provide a maximum bandwidth at each DSLAM dedicated for their VoIP traffic. Although the MNO could overprovision the dedicated resources on each DSLAM this would have increased costs. It is therefore of interest for the MNO to minimise the maximum bandwidth that is reserved on each DSLAM while maintaining high voice call quality for their customers.

The remainder of this paper is structured as follows. Section II presents other relevant work in the area; this is followed by an architecture description in Section III. The QoS based CAC and dynamic resource allocation mechanism is then described in Section IV. Section V and VI give an overview of the simulation environment and results, respectively. This paper is then concluded in Section VII.

II. RELATED WORK

There have been a number of other QoS based CAC mechanisms. In particular, an End-to-end Measurement based Admission Control (EMBAC) mechanism for VoIP was described in [4]. Here the authors use emulated VoIP streams as probes to emulate real VoIP traffic. This is also similar to previously published work on QoS based handover mechanisms for VoIP in which the authors utilise VoIP like probes to estimate call quality on candidate handoff networks [5]. Our concept differs from the one presented in [4] in two respects. Firstly, we locate the objective voice quality measurements in an intermediary node (i.e. HeNBGW) to avoid imposing the requirement for mobile clients to perform call quality measurements. Secondly, our call admission decisions utilise measurements taken from actual on-going VoIP calls rather than introducing voice probing streams to aid admission decisions, like the approach applied to WiFi CAC for example in [6]. In [6] the authors motivate the need for CAC for voice calls over Packet Switched Networks (PSN). They emphasize that there is a need for scalable and dynamic solutions to address the situation where VoIP calls are forwarded through a congested network in which no over provisioning or resource allocation exists.

The mechanism proposed in this paper is a combination of traffic policing and traffic shaping. A lot of previous work has been done on both of these concepts and an overview of each is described in [7]. Both concepts imply the existence of a SLA between the ISP and the users (e.g. MNOs). An example of network specific metrics specified in such agreements is the CIR (Committed Information Rate), Committed Burst (Bc), and Excess Burst Rate (Be).

In order to compute the Mean Opinion Score (MOS) of the VoIP calls a modified variant of the E-Model algorithm [8] is used; specifically, our implementation of the E-Model has been adapted to support AMR voice [9]. The E-Model computes a MOS for each VoIP call based on a number of network metrics including delay, loss and jitter. Accurate end-to-end network delay can be difficult to obtain unless the nodes between which it is calculated are time synchronized. In [10] the motivation and constraints for maintaining time synchronisation between femtocells and the HeNBGW is presented. It is shown that rigorous requirements for clock and frequency synchronization have to be met, NTP [11] or PTP [12] being the recommended synchronisation protocols. Based on this the HeNBGW maintains regular NTP message exchange with each femtocell to ensure high clock synchronisation accuracy. In this work we assume that such clock synchronisation exists. Indeed we analysed network traces from real femtocell deployments to confirm that this clock synchronisation is used.

III. ARCHITECTURE DESCRIPTION

The network architecture shown in Figure 2 depicts a typical femtocell deployment scenario. This was used as the reference architecture on which the proposed control mechanism and simulations were developed. It is comprised of multiple households and small offices that have HeNBs provided by the same MNO and connected through the same DSLAM to the MNO's core network. It is assumed that each HeNB is connected to a DSL broadband router.

Information gathered from multiple DSLAM vendor datasheets showed that a typical DSLAM is capable of serving approximately 1000 DSL connections. As a typical HeNB can support 4 user devices this means that potentially each DSLAM may have to transport up to 4000 simultaneous voice calls.

The other major parts of the MNO's core network are also shown but for the simulations these are assumed to have little or no impact on the call quality and, as such, are modelled with fixed low delays. The callee endpoints are also assumed to be connected via low delay links that have no impact on call quality.

Based on the current network architecture we determined two ways of creating the required interface between the MNO and the ISP as depicted in Figure 2:

- **Option 1** would be a proprietary interface developed by the ISP to allow the MNO to have limited dynamic resource allocation capabilities in the DSLAM.
- **Option 2** would utilise an interface being defined by both the Broadband Forum and 3GPP for fixed/mobile convergence [13]. This PCRF-BPCF interworking would allow the MNO to push resource allocation requests to the fixed access network. The BPCF performs the CAC and updates the resource allocation in the DSLAM.

IV. QOS BASED CALL ADMISSION CONTROL

A. Call Quality Monitor

The proposed call admission control and dynamic resource allocation mechanism is based upon the quality of on-going voice calls passing through the HeNBGW. In order to compute the call quality a *Voice monitor* application residing in the HeNBGW was developed. The role of this module is to maintain a list of on-going calls, measure the real time voice call quality (MOS), determine when problems occur and

employ the admission control and dynamic resource allocation mechanisms to restore/maintain the quality.

The *Voice monitor* uses the CIR value from the SLA to calculate a *nominal number of calls*. That is the maximum number of calls that can be supported with the default bandwidth allocation provided by the ISP to the MNO. It is computed as the ratio between the CIR [bps] and the call bit rate plus tunnelling overhead (using AMR's highest encoding rate).

B. MOS calculation

When a new packet arrives, the monitor calculates its delay, jitter and packet loss. The delay is calculated as the difference between the current time and the time when the packet was sent. The jitter is calculated using the recommended RTP jitter algorithm [14]. The RTP sequence numbers are used to compute a moving average of packet loss with a window size of 100 packets. With the formula provided by the E-model [8] the transmission factor R and the average MOS is calculated.

Keeping track of each MOS value calculated could cause scalability issues as normally around 50 values are obtained per second per voice call. Thus, we average all MOS values obtained in a time window and compute an *average MOS* value across all calls passing through the HeNBGW. We will use the term *average MOS* further in this paper to refer to this value. The *average MOS* is then used as the basis of performing call admission control and resource reservation in the DSLAM.

An alternative approach to that described in this work would be to allocate resources the moment call requests are received in the HeNBGW without requiring any call quality monitoring. However, although we assume that there is an EF queue per MNO or subscriber, this queue may be shared by other EF traffic sources in the HAN network. As such, fixed resource allocation could not determine if the allocated resources were sufficient to support high quality voice.

C. Call Admission Control algorithm

The flowchart in Figure 3 describes the decision process. The quality of all calls passing through the HeNBGW is continually monitored in real time and the *average MOS* value represents the impact of the associated DSLAM's EF queue.

There is a possibility that some calls may suffer quality degradation due to issues in the Home Area Network (HAN) in which the femtocell is installed. Our mechanism detects this situation and flags those calls with HAN problems.

The CAC and dynamic resource allocation mechanism is triggered periodically. When triggered, it is first checked whether each call's average MOS value is outside of the 95% confidence interval range. If so, those falling in that category will be filtered out and flagged as having HAN problems. A new *average MOS* value for all calls without HAN problems is then computed.

If this new *average MOS* value is less than 3.8 then no new calls will be accepted until the resource allocation in the DSLAM can be increased and the backlog of packets in the queue is cleared. If the current bandwidth used in the EF queue is lower than the committed burst then the VoIP monitor will request more bandwidth.

If the *average MOS* is higher than 4.0 and the number of calls present has decreased then the bandwidth for the EF queue is decreased, but no lower than the CIR. If the *average MOS* is higher than 3.9 then new call requests will be accepted.

The average MOS thresholds were decided upon considering Table 1 obtained from [8]. The proposed mechanism attempts to maintain all calls in the *Medium to Best* quality range.

Table 1 – Map between Quality Rating and MOS

Quality rating	MOS
Best	4.34 - 4.50
High	4.03 - 4.34
Medium	3.60 - 4.03
Low	3.10 - 3.60
Poor	1.00 - 3.10

D. Bandwidth monitoring in the EF queue and traffic shaping

The DSLAM monitors the traffic volume forwarded through the EF class [15] by each MNO. The calculation of the utilised bandwidth in the EF queue is done periodically and is based on the AMR inter-packet departure time, i.e. 20 milliseconds.

When the CIR is reached for a certain MNO, any new VoIP packets arriving in the EF queue are tagged as noncompliant. Noncompliant packets are stored until the next scheduled release time. However, the buffer has a limited storage size and thus traffic policing needs to be employed once the buffer has reached its limit. Any packets arriving when the buffer is full are therefore dropped.

V. SIMULATION SETUP

The proposed solution was implemented and validated through simulations using the Network Simulator 3 (NS3). The simulated scenario is depicted in Figure 4. It consists of pieces of User Equipment (UE) which generate VoIP traffic and are associated to Femto Access Points (FAPs). For ease, we used a wired link between the UEs and FAPs, as the radio link is not the scope of this paper. We simulated 50 FAPs, each having 4 UEs associated. An additional FAP with only one UE associated was used to create a HAN problem scenario. All FAPs are connected to a DSLAM by DSL links with a speed of 100 Mbps and a link delay of 1 ms. For the particular HAN problem scenario, the link speed was decreased to 73 kbps.

The DSLAM forwards VoIP packets through its egress EF queue to the VoIP monitor in the MNO's HeNBGW. The VoIP monitor calculates the metrics used by the mechanism described in the previous section and forwards the packets through the MNO's core network to the terminating UEs.

In our simulation, simultaneous VoIP calls are consecutively generated with an increment of 0.5 seconds plus a random variable in the range [0, 0.1] seconds with a granularity of 10 microseconds to desynchronise the VoIP sources. A simulation time of 300 seconds was chosen and we simulated 201 VoIP calls with call durations ranging from 90 to 290 seconds. The voice calls are torn down in the same randomised manner as they are created.

In order to accurately emulate DSLAM behaviour, DSLAM recommendations from Cisco [7] were used to provide a number of DSLAM parameters including the buffer size used for traffic shaping which is based on link speed. From DSLAM datasheets a typical egress link speed of 10Gbps was obtained and so this value was used, yielding a buffer size of 11,520,000 bytes.

Figure 3 – Call Admission Control and Dynamic Resource Allocation Mechanism

Figure 4 – Simulation Scenario

Given these parameters the nominal number of calls is calculated using the values from Figure 1. From this the maximum size a VoIP packet passing through the DSLAM is determined to be 176 bytes assuming that the highest AMR mode (12.20 kbps) is being used. Based on this the data rate per call is computed as:

$$176 \frac{\text{Bytes}}{\text{packet}} \times 8 \frac{\text{bits}}{\text{Byte}} \times 50 \frac{\text{packets}}{\text{Sec}} = 70.4 \text{ kbps}$$

In the results presented in this paper a CIR value of 10 Mbps and a Bc value of 15 Mbps is assumed. In order to impose the Bc value limits and perform traffic shaping, the EF queue bandwidth for each MNO is checked every 10ms which represents half of the AMR packet inter-departure time. In this way no unnecessary delaying of packets is introduced.

Using a CIR value of 10 Mbps in the EF queue, the maximum number of supported calls can be calculated as:

$$\frac{10 \text{ Mbps}}{70.4 \text{ kbps}} \approx 140 \text{ Calls}$$

For ease of implementation, only two modes of the AMR speech codec, i.e. AMR 12.20 and AMR SID were implemented. The switching between these two is specified by the speech activity parameter which is set as a random value between 30% and 80%. This emulates realistic Voice Activity Detection and silence suppression functionality of the AMR codec. The VoIP monitor in the HeNBGW computes an average MOS over all on-going voice calls every 0.5 seconds; this was chosen as a reasonable trade-off between maintaining high quality voice calls and reducing the level of overhead.

VI. RESULTS

This section presents a number of simulation results obtained using the setup described in the previous section. In order to obtain a baseline for the results, a simulation without the proposed QoS and CAC mechanisms in place was performed. The results of this simulation are depicted in Figure 5 and show the *average MOS* and the number of on-going calls over time.

As can be seen, without CAC all calls are accepted and voice packets are forwarded from the EF queue as soon as they arrive; this is unless the imposed bandwidth limitation has been reached in which case the packet is queued. This queuing means introducing extra delay and this delay increases until the buffer is full at which point the buffer will overflow resulting in dropped/lost packets. At this point the MOS of all on-going calls degrades rapidly. From Figure 5 it can be determined that the maximum number of calls that the DSLAM can accommodate is approximately 130, given our specific simulation setup and assumptions.

The fact that all voice calls are degraded draws attention to the difference between circuit switched and packet switched voice traffic. In circuit switched networks dedicated time slots are allocated for voice traffic at call setup time, while in packet switched networks voice packets from different voice calls potentially share the same packet forwarding capacity at each network router. In other words, accepting all requests overloads the forwarding queue resulting in call quality degradation of all calls utilising the same queue. This further highlights the need for dynamic CAC.

Figure 6 presents results for simulations done using the proposed QoS and CAC mechanisms enabled. It can be seen that when the overall MOS drops, the feedback mechanism rejects any new call requests and the overall MOS for all calls is restored to a high level. Figure 7 shows a plot of the bandwidth requests made by the algorithm in the HeNBGW and granted by the DSLAM.

Figure 5 – Overall MOS and number of online sessions versus time, without CAC and dynamic resource allocation

It can be seen that for each moment when the *average MOS* is below the threshold (3.8) the requested bandwidth on the DSLAM is increased up to the Bc limit. This extra bandwidth is released only when the *average MOS* has returned to a high value and the number of simultaneous calls has decreased. Since any extra reserved bandwidth would involve a cost to the MNO, increased bandwidth requests are only made when necessary, thereby maintaining a trade-off between the extra bandwidth requested and the number of accepted calls.

A third scenario was also simulated in which we introduced a femtocell whose voice traffic is being degraded in the HAN. In this case the algorithm was able to detect and remove that call from the measurement data used to make resource allocation and CAC decisions.

VII. CONCLUSION

A dynamic resource allocation and call admission control mechanism for femtocell deployments with DSL backhaul was described. The proposed mechanism used real time call quality measurements to dynamically provision DSL resources. The integration of an LTE femtocell network with existing fixed broadband DSL networks was also described. Specifically, two possible interface options to allow femtocell networks to provision resources in DSLAMs are provided.

An NS3 based simulation analysis was performed to validate the proposed solution. The simulation results demonstrate the capability of the solution to maintain high levels of quality for all voice calls and to dynamically adapt resource allocations based on changes in call quality. Thus, this approach provides a holistic and consistent solution to issues of congestion in the backhaul of femtocell deployments.

ACKNOWLEDGMENTS

This research is partially funded by Science Foundation Ireland (SFI) via grant 08/SRC/I1403 FAME SRC (Federated, Autonomic Management of End-to-End Communications Services-Strategic Research Cluster). This work was also funded by Irish Research Council for Science, Engineering and Technology, co-funded by Marie Curie Actions under FP7 and has been performed in the framework of the ICT project ICT-4-248523 BeFEMTO, which is partly funded by the European Union.

REFERENCES

[1] B. Jennings, et al., "Towards autonomic management of communications networks," *Communications Magazine, IEEE*, vol. 45, no. 10, pp. 112-121, 2007.

Figure 6 – Overall MOS and number of online calls versus time, with CAC and dynamic resource allocation

Figure 7 – DSLAM's EF queue bandwidth usage versus time

- [2] S. Davy, et al., "Policy-based architecture to enable autonomic communications - a position paper," in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, vol. 1, 2006, pp. 590-594.
- [3] C. Olariu, M. O. Foghlu, P. Perry, and L. Murphy, "VoIP quality monitoring in LTE femtocells," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, Dublin, 2011, pp. 501-508.
- [4] K. Mase and Y. Toyama, "End-to-end measurement based admission control for VoIP networks," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 2, 2002, pp. 1194-1198, vol.2.
- [5] J. Fitzpatrick, S. Murphy, and J. Murphy, "SCTP based Handover Mechanism for VoIP over IEEE 802.11b Wireless LAN with Heterogeneous Transmission Rates," in *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 5, 2006, pp. 2054-2059.
- [6] P. McGovern, P. Perry, S. Murphy, and L. Murphy, "Endpoint Based Call Admission Control and Resource Management for VoWLAN," *IEEE Transactions on Mobile Computing*, 2010.
- [7] V. Joseph and B. Chapman, *Deploying QoS for Cisco IP and Next Generation Networks: The Definitive Guide*. Morgan Kaufmann Publishers Inc., 2009.
- [8] ITU-T. G.107, "The E-Model, A Computational Model For Use In Transmission Planning," 2005.
- [9] J. Fitzpatrick, "An E-Model Based Adaptation Algorithm for AMR Voice Calls," in *IFIP Wireless Days conference (WD'11)*, 2011.
- [10] Femto Forum, "Femtocell Synchronization and Location," no. June, 2010.
- [11] D. Mills, J. Martin, J. Burbank, and W. Kasch, "RFC 5905: Network Time Protocol Version 4: Protocol and Algorithms Specification," *Internet Engineering Task Force*, 2010.
- [12] J. C. Eidson, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," *IEEE Std 1588-2002*, pp. i-144, 2002.
- [13] 3GPP and The Broadband Forum, "PCRF-BPCF Functional Split and Information Exchange," Orange, NEC, Alcatel-Lucent, 2010.
- [14] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC3550: RTP: A Transport Protocol for Real-Time Applications," *RFC Editor United States*, 2003.
- [15] S. Blake, et al., "An architecture for differentiated services," Citeseer, 1998.