UMTS Load Control with Access Class Barring

Steven Gordon, David Garbin and Denise Masi
Noblis
Falls Church, VA
Steven.Gordon@noblis.org, David.Garbin@noblis.org,
Denise.Masi@noblis.org

Patrick McGregor
Nyquetek Inc.
Millersville, MD
pat_mcgregor@msn.com

***Abstract*—Disasters can cause extraordinary service demand by the public, while concurrently causing outages that reduce network capacity to serve the surging demand. It is imperative that services supporting disaster response management perform with minimal degradation during such events. In order to provide adequate service to special users like first responders, priority treatment mechanisms have to be developed.**

The 2G CDMA and GSM technologies have been augmented with priority treatment mechanisms and provide good priority service, but these benefits will be lost in the sunset of 2G technologies in the next few years. Consequently, other priority mechanisms have to be established on UMTS/3G. One of the proposed priority-treatment concepts is Access Class Barring (ACB), which is a new feature being rolled out by carriers that sheds voice and data traffic in response to extreme overloads. However, the degree to which ACB would improve voice call completion is unknown.

Whether ACB would adequately improve priority service call completion probability is the topic of this study. A discrete-event simulation was performed to model extreme overload situations and predict the performance of ACB under various configurations. The simulation study found that ACB could drastically improve priority call completion probability from 12% to 69% in the most extreme overloads while allowing carriers to almost fully utilize their bandwidth with public traffic under extreme overloads.

Keywords—Networking; Architectures; Performance; UMTS; ACB

I. Introduction

There is a need in the wireless user community to provide priority access to commercial wireless networks, particularly Government representatives and first responders during local, regional and national emergencies. These events often cause congestion from excessive loads on wireless networks that prevent both public and priority users from reliably making voice calls. In order to provide priority service during these extreme events, the Government is currently researching the use of a new priority-treatment mechanism called Access Class Barring (ACB).

Wireless carriers previously implemented a priority service with a priority-treatment mechanism for CDMA/2G technology called the persistence delay (Psist), which provided much faster access to priority users than public users. UMTS possesses a somewhat similar mechanism, but it has minimal effect and cannot alone create an adequate priority service in UMTS. In order to sufficiently improve priority service on UMTS, carriers are considering ACB as a different, more impactful mechanism to provide the needed improvement to priority users' voice calls.

Previous modeling to assess performance of UMTS with ACB has been limited. Initial modeling based on an analysis of the 2005 Miyagi Earthquake was performed in [1], which concluded that ACB mitigates congestion more quickly than without ACB. However, this work did not examine the impact of ACB on the performance of users who are exempt from it.

A. UMTS Architecture and Background

UMTS shares a similar architecture to 2G technologies as shown in Figure 1. The radio access network (RAN) segment (wireless physical layer) is operated by the NodeBs. These have antennas that are often mounted on towers and are often in plain view of mobile users with the user equipment (UE) handsets. Simple cell phones, smart phones and fixed UMTS wireless devices are UEs. Radio network controllers (RNCs) provide most of the logic for processing user traffic including voice call and data session admittance decisions, bandwidth management, UE handoff between cells and traffic transport to the voice and data gateways. The mobile switching center (MSC) is the gateway to the public switched telephone network (PTSN) and terminates PSTN backbone circuits that serve the RAN. Similarly, the serving GPRS Support Node (SGSN) is the gateway to the packet network and provides access for wireless data traffic to the Internet as well as the carrier's other cells [2].

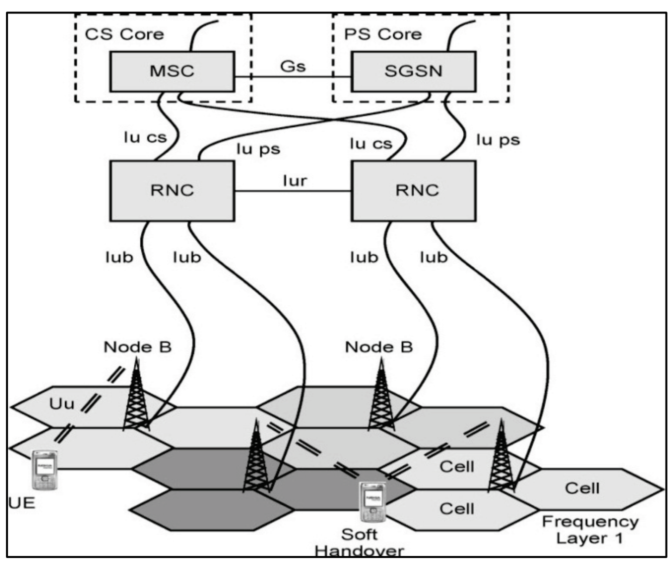

Fig. 1. UMTS Architecture

Within the Uu UE/NodeB interface in Figure 1, there are several channels with different characteristics. The reverse access channel (RACH) carries the initial signaling traffic at the beginning of a voice call or data session from the UE through the NodeB to the RNC (Figure 2). The RACH is a shared channel and is contention-based, and can experience blocking during overloads. The Uu interface also specifies dedicated control channels (DCCHs), which are set up dynamically between the UE and the RNC via the NodeB and carry signaling traffic. The initial signaling on the RACH sets up the DCCH. DCCHs are also subject to blocking, but, in contrast to the RACH, the cause of the potential blocking is unavailability of bandwidth managed by the RNC to allocate to a new DCCH. Similarly, the dedicated traffic channels (DTCHs) are dynamic point-to-point channels between the UE and the RNC via the NodeB that carry the application traffic, like a digitized voice stream or IP-based data stream.

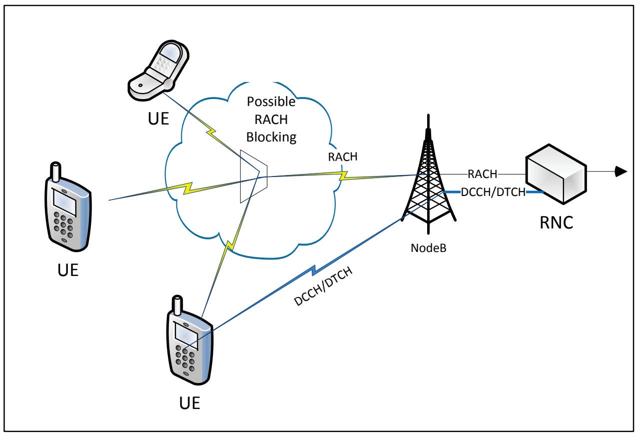

Fig. 2. RACH, DCCH and DTCH

B. *Priority Service on UMTS*

A different approach is needed for the priority service on UMTS vs. CDMA because their bottlenecks are different in extreme overloads. In CDMA, extreme overloads can congest and collapse the RACH, which carries signaling traffic from the UE to the NodeB. It is contention-based, has a (relatively) low maximum throughput and is subject to collapse under overloads, similar to Ethernets. In contrast, the UMTS RACH is much more robust and can carry much more signaling traffic, to point where it is highly unlikely to impede traffic, even under extreme overloads.

However, UMTS, even with a vastly improved RACH, can still experience unacceptable congestion during extreme overloads. Examining the voice call/data session/SMS setup process shows the blocking points in UMTS:

1. UE transmits request on the RACH [with possible RACH blocking]
2. RNC allocates DCCH [with possible DCCH blocking]; start of SMS transfer
3. Voice/Data Only: RNC allocates DTCH [with possible DTCH blocking]
4. Completed setup of voice call/data session; start of voice/data transfer.

Of these steps, only steps 2 and 3 experience significant blocking, with step 3 experiencing most of the blocking. After step 4, the bit stream for the voice call is carried on the DTCH. Because SMS traffic does not execute step 3, SMS transfer experiences much less blocking than voice and data.

In order to improve priority-service performance during extreme overloads, two general approaches exist to alleviate contention for DTCHs and increase call completion probability: (1) Provide priority users priority access to DTCH pool, and (2) eliminate or reduce the overload of public (not priority user) traffic requesting DTCHs. The first approach is implemented with an RNC feature called *traffic filtering.* The second can potentially be implemented with *Access Class Barring,* which would prohibit some of the public UEs from generating traffic under overload conditions. The analysis of ACB is the topic of the paper.

II. Access Class Barring

A. *Description*

ACB selectively limits the load offered by the public (not priority) UEs to the network under overload conditions. ACB is a highly configurable feature, but is generally configured such that (1) it engages only during extreme overloads that traffic filtering cannot effectively process, (2) it filters public traffic only, and (3) the barring process more heavily affects data traffic vs. voice.

In order to detect and react to the traffic level where the cell is experiencing an overload, ACB is often configured to trigger off failures from the Call Admission Controller (CAC). The CAC is the mechanism in the RNC that tracks the bandwidth available in the cell for allocation of DCCHs and DTCHs, and allocates it to new, incoming requests. New requests for which no bandwidth is available are rejected, and the rejections are counted in the CAC failure rate statistic. ACB is often configured to use the CAC failure rate as the trigger for engaging or disengaging.

In performing the public traffic throttling, ACB employs separate controls for the two UMTS domains, circuit switched (CS) and packet switched (PS). The CS domain encompasses voice calls and SMS, whereas PS encompasses all data traffic. The CAC is able to measure the CAC failure rates on these domains separately with the CS CAC failure rate and PS CAC failure rate. Commensurately, ACB can control the initial degree of barring in the CS and PS domains separately with the algorithmMode2CsPercentage and algorithmMode2PsPercentage ACB configuration parameters.

One significant restriction of ACB is that the barring process can only be applied in discrete steps of 10% of the public UE population, so the barring levels must be 10%, 20%, 30%, etc. No intermediate barring levels are permitted; however, these levels can be applied to CS and PS separately, so that, for example, the CS barring level can be 10% when the PS barring level is 50%. This 10% restriction exists because the whole UE population is tagged (approximately equally) with one of ten *access classes* (0-9), and a single access must be either barred or unbarred. No partial barring of an access class is possible.

B. ACB Relationship to CAC

The operation of ACB is closely tied to the CAC, which is the algorithm running in the RNC that decides if a new request from a UE is accepted or rejected. It allocates bandwidth for DCCHs and DTCHs on a first-come, first-served basis (FCFS). Its primary intent is to control ingress traffic to the limit of the available bandwidth. It is not intended to give priority for any particular traffic stream, however, because of the characteristics of the traffic streams, it gives some implicit priority to voice calls over data sessions.

DCCHs and DTCHs are allocated from a dedicated channel (DCH) pool. The CAC manages the DCH bandwidth pool mostly as FCFS with some rules and restrictions. For illustration, first, DCCHs for all traffic streams use 1.5 kbps. Second, the DTCH allocated to a voice request is 12.5 kbps whereas a DTCH for a data session is 50 kbps. Third, a (configurable) section of the bandwidth (10%) can only be used for DCCHs to prevent situations where an excessive volume of DTCHs can occupy bandwidth for DCCHs (Figure 3). Fourth, of the remaining 90% of the DCH bandwidth, 10% can only be used for **voice** DTCHs to prevent excessive data traffic from excluding voice. These allocation rules implicitly favor voice calls over data during overloads because:

- Voice DTCH can use 90% of the bandwidth, whereas data is restricted to 81% (90% of 90%)
- Voice DTCH bandwidth is much smaller than data DTCHs, which gives a voice traffic a higher probability of being successfully assigned its necessary bandwidth.

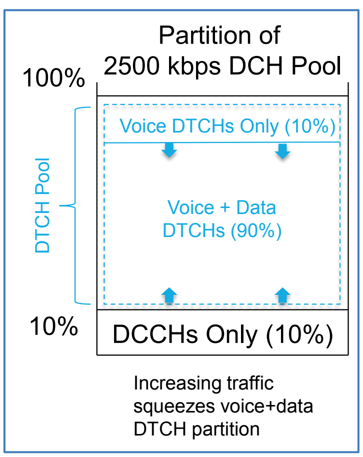

Fig. 3. Partition of 2500 kbps DCH Pool

During the process of performing the bandwidth allocation above, the CAC records the number of request arrivals to the CAC and the number of rejections by the CAC for both the CS and PS domains. From this, the CAC continuously calculates the *CS CAC failure rate* and the *PS CAC failure rate* over a 30-second sliding window. The CS CAC failure rate is used as the trigger for ACB for both the CS and PS barring processes. In addition, the CAC also calculates *CS and PS CAC arrival rate* statistics.

III. LOAD CONTROL ANALYSIS

A. Traffic Profile

A discrete event simulation model was developed that represented a large population of UEs in a single cell, a NodeB, and an RNC running the RACH access procedure, CAC and ACB as described above. The traffic that the UEs generated was a mix of public voice calls, priority voice calls, data sessions and SMS messages (Table 1). Voice sessions are assumed to have a 90 second hold time (call length) for the DTCH. In contrast, data sessions are assumed to have a much longer hold time of 300 seconds, as well as the much higher bandwidth discussed above. However, voice calls have a much higher arrival rate than data sessions. In addition, SMS messages have a much higher arrival rate than both voice and data.

TABLE I. TRAFFIC MODEL

	Voice (Public+Priority)	Data	SMS
Arrival rate	Set to 2% blocking	0.07 * voice	3 * voice
DCCH hold time	7 sec	4 sec	7 sec
DTCH hold time	90 sec	300 sec	NONE
Retries allowed?	Yes	Yes	No
Engineered Load (X)	**3,400 Public + 340 Priority calls/hour**	**240 sess./ hour**	**10,200 msgs/ hour**
Usage of DCH bandwidth	**47.6% + 4.8%**	**41.2%**	**1.2%**

B. Modeling Scenarios

We consider three scenarios of interest: (1) **ACB Off**, which shows baseline against which ACB will be compared, (2) **ACB On—Extended Surge**, which emulates the behavior during an event like an earthquake or Boston Marathon bombing, when the traffic quickly reaches a high overload and stays high for an extended period of time, and (3) **ACB On—Short Surge**, which emulates the behavior during a short event like halftime at the Super Bowl when many users rapidly initiate traffic at the beginning of halftime and rapidly (almost simultaneously) end usage at the end of halftime.

The first scenario, ACB Off, is intended to represent a baseline to which the other two scenarios will be compared. Because the CAC, even without ACB, gives some implicit priority to voice traffic (both priority and public) over data traffic, the performance of the voice traffic under extreme overloads without ACB is unknown and cannot be analytically calculated. The ACB On—Extended Surge scenario is ten hours long and is run at each of ten injected loads, ranging from engineered load (1X) to 10X. Similarly, the ACB On—Short Surge scenario is ten minutes long, and is run for the same 1X to 10X range.

The input parameters that have the most effect are the *CS CAC failure rate trigger* and the *StepIntervalTime*. The CS CAC failure rate trigger is the threshold that the CS CAC failure rate has to exceed to engage ACB or increase the barring level. If the CS CAC failure rate exceeds the CS CAC failure rate trigger and ACB is not engaged, then RNC engages ACB (both CS and PS barring) at their initial levels, otherwise the CS and PS barring levels are increased. Conversely, if the CS failure rate drops below 5% lower than the CS failure rate trigger and ACB is on, then the RNC lowers the CS and PS barring levels to their next lower level. In this process, the StepIntervalTime is a time restriction on the frequency with which the CS and PS barring levels can be changed. It specifies that the barring level cannot either increase or decrease within StepIntervalTime seconds of the last change. Higher values of the StepIntervalTime tend to slow the rate of change of ACB, which can make the process more stable but less nimble in responding to traffic spikes.

IV. Results

A. *ACB Off Scenario*

In this scenario the traffic profile in Table 1 was used as the injected traffic for public voice, public data and SMS and was scaled linearly from 1 to 10 times to represent a reasonable range of overloads. The priority voice traffic was kept at the 1X level shown in Table 1. Figure 4 shows a decrease in call served ratio between 1X and 5X that is mitigated somewhat by the CAC's tendency to displace data traffic with voice traffic. However, after 5X, little data traffic remains to be displaced by rising voice traffic, so the curve approximately flattens to a final call completion ratio of 12% at 10X.

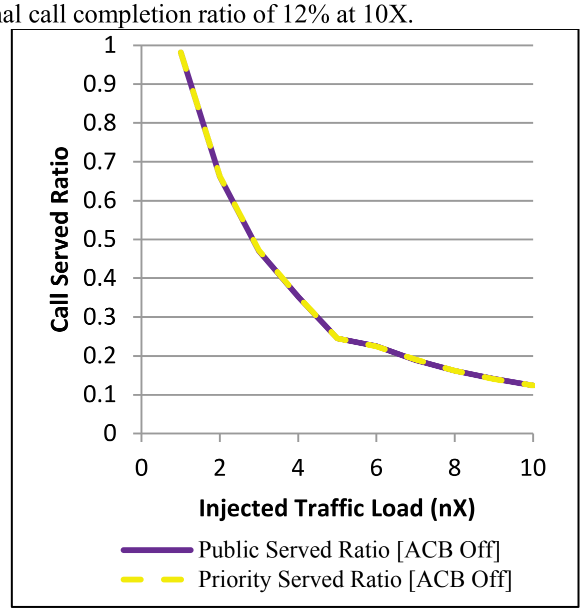

Fig. 4. Voice Call Completion With ACB Off

In addition to the voice call completion ratio, the simulation tracks and reports the utilization of the DCH bandwidth pool, from which the DCCHs and DTCHs are allocated. In Figure 5, the green trace (ACB Off) shows a DCH utilization of 86% at 1X, 98.7% at 2X and 99.7% at 10X. From this, it can be concluded that a typical expected maximum utilization of the DCH bandwidth pool is about 86% (at 1X load).

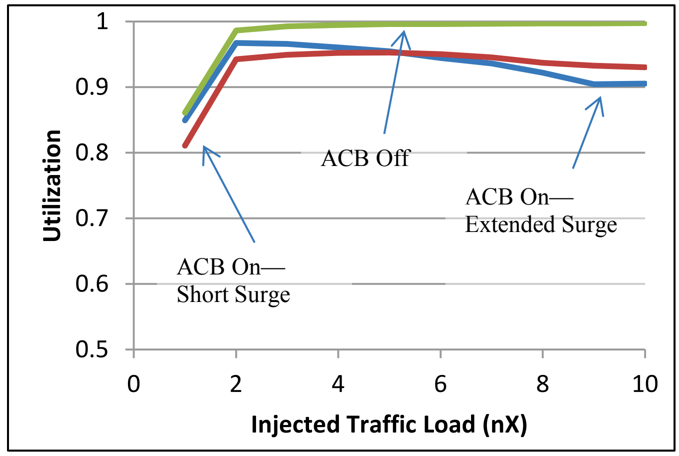

Fig. 5. DCH Bandwidth Pool Utilization for All Scenarios

B. *ACB On—Extended Surge Scenario*

An extended surge scenario was configured by setting the runtime to 10 hours, and the same range of injected traffic loads from 1X to 10X were run. In addition, the CS CAC failure rate threshold was set at 1%, 5% and 10%, and the StepIntervalTime was set at 30 and 60 seconds. A set of runs was executed for each of the six combinations of the CS CAC failure rates and StepIntervalTime parameters. Of these, the best performance is at CS CAC failure threshold of 5% and a StepIntervalTime of 30 seconds and is shown in Figure 6, which shows the improvement given to priority traffic from ACB (in the blue trace) and the slight degradation incurred by the public voice traffic.

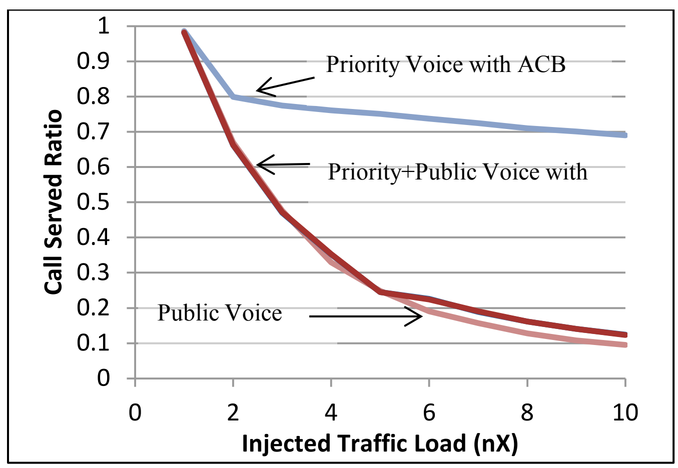

Fig. 6. Voice Call Completion for Extended Surge Scenario

In addition to the priority-traffic call completion ratio, the utilization of the DCH pool is a critical consideration. The large traffic volume disallowed by ACB raises the concern from carriers about the loss of billable traffic. This question can be assessed by examining the DCH utilization for the Extended Surge (blue trace) in Figure 5. Figure 5 shows that the DCH bandwidth utilization remains high (above 86%) even at the highest barring levels.

The ACB-enhanced priority traffic shows some important improvements:

- ACB significantly improves voice call served ratio
- ACB provides a "floor" for voice call served ratio
  - Levels out at about 69%, regardless of traffic load
  - ACB achieves this leveling by progressively increasing barring level with traffic load
- DCH utilization remains very high even at high traffic loads and ACB barring levels

C. ACB On—Short Surge Scenario

The short surge scenario was configured by setting the runtime to 10 minutes, rather than the 10 hours in the Extended Surge above. The same range of CS CAC Trigger thresholds and StepIntervalTimes were the same as the extended surge. The call completion ratios are shown in Figure 7. In addition, the DCH bandwidth utilization is shown in Figure 5 in the "ACB On-Short Surge" (red) trace.

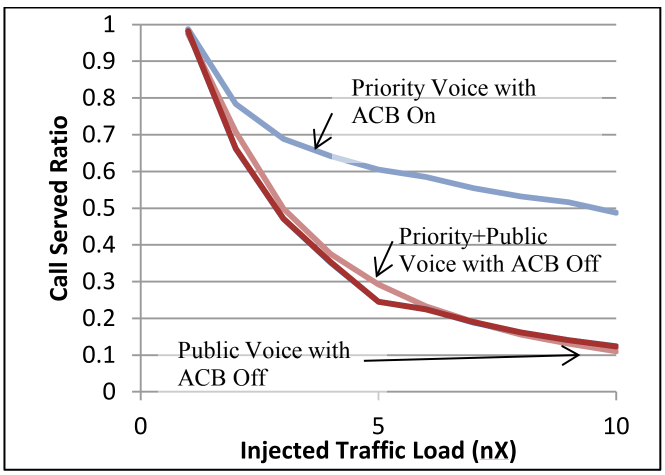

Fig. 7. Voice Call Completion for Short Surge Scenario

In comparison to Figure 6, Figure 7 shows a smaller, but significant improvement delivered by ACB in the Short Surge. The improvement is still very significant with a call completion ratio at 5X of 60% vs. 29% without ACB, but smaller than the improvement in the Extended Surge. This is due to the fact that the rampup period for ACB reaching the full effect, which can be three to four minutes for extreme surges, has a higher effect on a 10 minute scenario vs. a 10 hour scenario.

V. ACB Tuning Results

In the process of experimenting with the model, two parameters emerged as ones to which the model has high sensitivity, the CS CAC Trigger threshold and StepIntervalTime. The CS CAC Trigger threshold is the value of the CS CAC failure rate at which ACB engages or increases. During any time interval (usually 10 seconds), if the CS CAC failure rate exceeds the CS CAC Trigger threshold, then ACB will engage (if ACB was previously off) or increase in barring level (if ACB is currently on). Conversely, if ACB is on and the CS CAC failure rate drops 5% below CS CAC Trigger threshold, then ACB decreases in barring level (if previously on) or disengages (if it was at the lowest barring level).

StepIntervalTime influences the ACB engagement/ disengagement/increase level/decrease level actions by enforcing a minimum time that ACB must wait from the previous action to take the next action. StepIntevalTime is applied to each of the CS and PS barring processes separately. For example, in the situation where StepIntervalTime was set to 60 seconds, and the CS barring process had increased level at t = 1000 seconds, and the PS barring process had increased level at t = 1020 seconds, then the earliest time that the CS barring process could take an action would be t = 1060 seconds, and the earliest time that the PS barring process could take an action would be t = 1080 seconds. When StepIntervalTime is increased, the ACB process tends to be more stable, but less nimble in responding to rapid increases in traffic. When it is decreased, ACB tends to respond faster to rapid traffic increase but may cause more "overshooting" and oscillation in the barring behavior.

Figure 8 helps illustrate the relationship between the CS CAC failure rate statistic, the CS CAC Trigger threshold parameter and the StepIntervalTime parameter. This is an Extended Surge scenario where a traffic load of ⅔X begins at t = 0, and the surge of 5X begins at t = 600 seconds. The CS barring level is shown in the blue trace and the PS barring level is shown in the green trace. The CS CAC failure rate (output from the CAC) is shown in the yellow spikes. It is calculated every 10 seconds over a 30 second window. After the surge begins at 600 seconds, the RNC begins to reject requests for voice call establishments; as a result, the CS CAC failure rate rises between 620 seconds and 660 seconds. When the CS CAC failure rate rises above the 5% CS CAC Trigger threshold 630 seconds, both CS and PS barring processes engage at their initial levels of 10% and 50%, respectively. While the CS CAC failure rate stays above 5% CS CAC Trigger threshold until 810 seconds, the CS and PS barring levels continue to rise. When elevated CS and PS barring levels finally bring the CS CAC failure rate down at 810 seconds, the CS and PS barring levels begin to fall. In this process, the StepIntervalTime is the duration of the "plateau" of each step. If this parameter is reduced, the process can rise and fall more rapidly, but may risk more oscillations and therefore more instability in the dynamics of the ACB process.

A. Tuning CS CAC Trigger Threshold

Model experimentation was performed to examine the impact of the CS CAC Trigger threshold variation on the call served ratio. Lowering the CS CAC Trigger threshold should create more barring, but may do so at the expense of leaving excess capacity that could be used to complete more calls. For this set of runs, the StepIntervalTime was set at 60 seconds, which was a preliminary vendor recommendation. The CS CAC Trigger threshold was set to 1%, 5% and 10%. Figure 9 shows that significant improvement can be gained by setting the CS CAC Threshold trigger to 5% or 1%. More detailed examination (not presented) indicates that 1% probably introduces too much instability to be a usable setting, but the 5% setting appears stable. Thus, 5% is likely a better setting than the vendor's 10% preliminary recommendation.

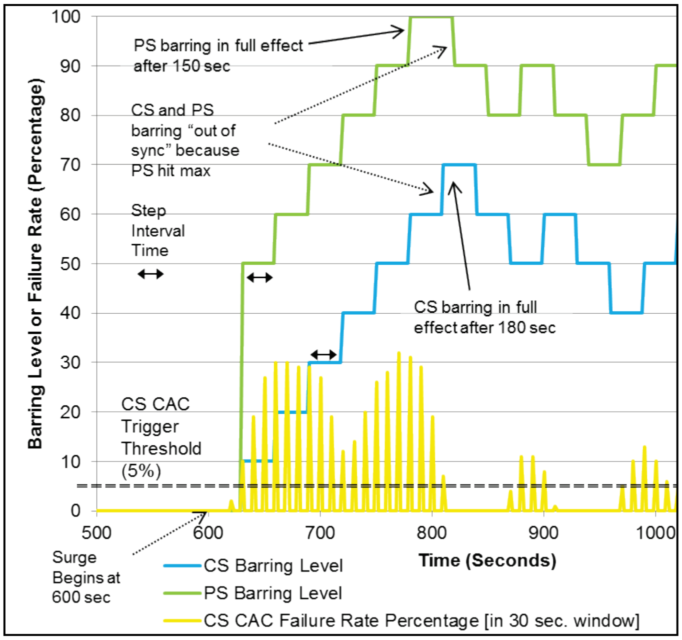

Fig. 8. ACB Barring Dynamics—Initial Engagement

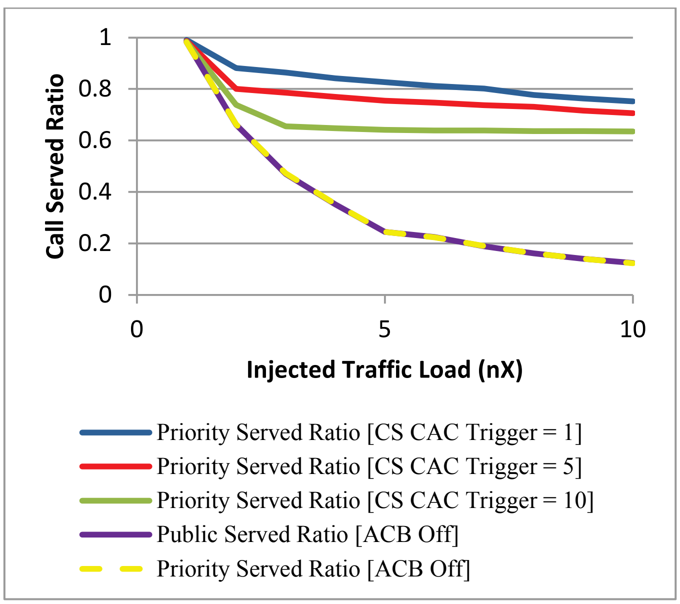

Fig. 9. CS CAC Trigger Threshold Tuning/Extended Surge

B. *Tuning StepIntervalTime*

Model experimentation was also conducted to discover the variation of call served ratio with StepIntervalTime. Lowering this can speed the reaction time of ACB to traffic spikes, but may make the barring process less stable. In addition, it primarily affects the Short Surge scenario, because in the Extended Surge scenario, a relatively short amount of time (compared to the total 10 hour run time) is spent in the ramp-up period of the traffic surge. However, in the Short Surge scenario, the rampup period is significant compared to the 10 minute runtime.

For these runs, the CS CAC Trigger threshold was set at the vendor preliminary recommendation of 10%, and the StepIntervalTime was set at 30 and 60 seconds (preliminary vendor recommendation). Figure 10 shows the results. The 30 second trace shows a significant gain in call served ratio over the 60 second trace. Other analysis (not presented) also shows that this setting is reasonably stable. In summary, 30 seconds shows to be a better setting for most cases.

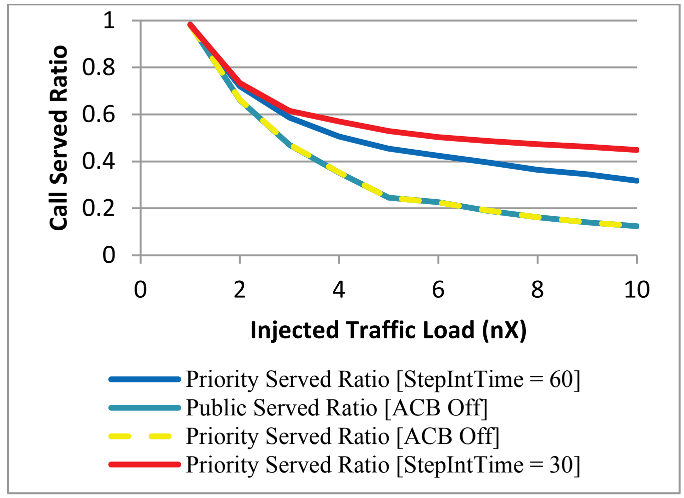

Fig. 10. StepIntervalTime Tuning/Short Surge

VI. Conclusions

In summary, Access Class Barring can be an effective method of providing priority access to commercial wireless networks with UMTS. Conclusions that can be derived from this study are the following:

- ACB gives significant improvement to the priority voice service. ACB improves priority service from less than 20% success to almost 70%.
- ACB has minimal impact on the public traffic that is transported during extreme events. There is very little difference in total served traffic with or without ACB. Also, the traffic channel capacity stays very highly utilized.
- ACB for priority voice can be improved with configuration changes. A 5% voice failure rate trigger gives 69% success at 10X while a 10% trigger gives 63% at 10X. A step timer of 30 seconds gives response time to steady state of 3-4 minutes whereas step timer of 60 seconds gives a response time of 6-8 minutes.

References

[1] EN-Universal Mobile Telecommunications System (UMTS); LTE; Study on Paging Permission with Access Control (PPAC)-V8.1.0; 3GPP TR 22.908 version 8.1.0 Release 8, Annex A: Analysis of provisioning of communication between the unauthorised users in the disaster areas.

[2] C. Johnson, Radio Access Networks for UMTS: Principles and Practice, John Wiley and Sons, 2008.

[3] J.P. Castro, The UMTS Network and Radio Access Technology: Air Interface Techniques for Future Mobile Systems, John Wiley and Sons, 2001.