QoS for IEEE 802.16m

Document Number: IEEE C802.16m-08/1054r1

Date Submitted:
2008-09-11

Source:

Xiangying Yang (xiangying.yang@intel.com)	Intel Corporation
Jie Hui	Intel Corporation
Muthaiah Venkatachalam	Intel Corporation
Ben Metzler	Intel Corporation
Maruti Gupta	Intel Corporation
Rath Vannithamby	Intel Corporation

Venue: IEEE 802.16m-08/033 Call for Comments and Contributions on Project 802.16m System Description Document (SDD) for session 57, on the topic of "MAC: Data Plane".

Base Contribution:

Purpose: Discussion and approval of the proposal into the IEEE 802.16m System Description Document

Existing Scheduling Services in the Reference System

- UGS:
  - periodical BW allocation to the SS. Used for CBR traffic.
  - suitable for fixed/periodic BW requirement, not efficient for ON/OFF traffic
- rtPS:
  - periodical polling to SS. Efficient for always-on real-time traffic.
  - suitable for always-on applications with periodic bandwidth requirement
- ertPS:
  - designed for VoIP with silence suppression
  - Not flexible in changing grant interval
- nrtPS:
  - Use either contention or uni-cast polls to request bandwidth.
  - Designed for delay-tolerant applications.
- BE:
  - Use contention BW request to request bandwidth.
  - Designed for delay-tolerant applications.

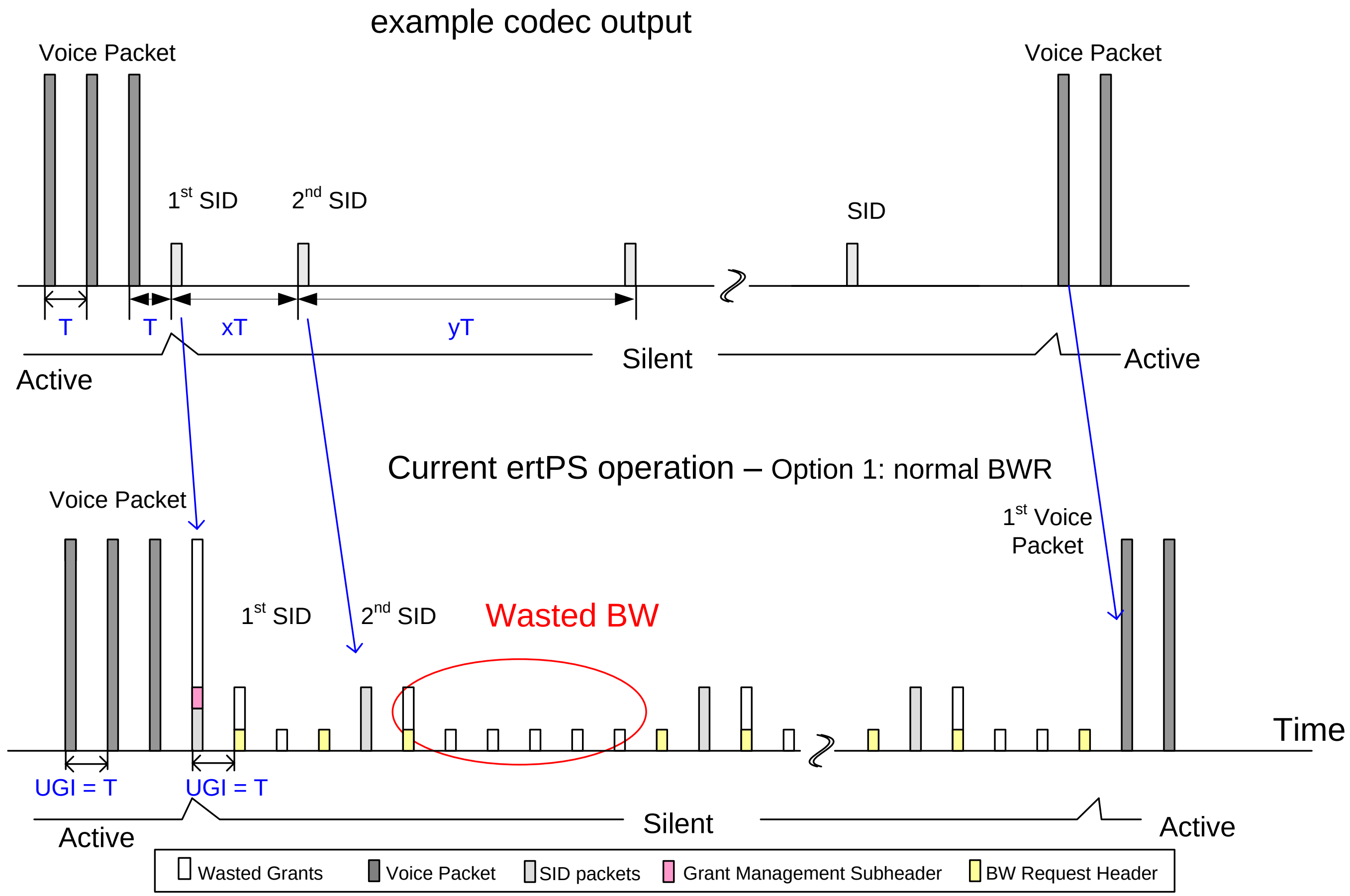
Adaptive Grant and Polling

Example: inefficiency of ertPS

when MS indicates the silence period by zero BWR

- **Option 1:** normal BWR

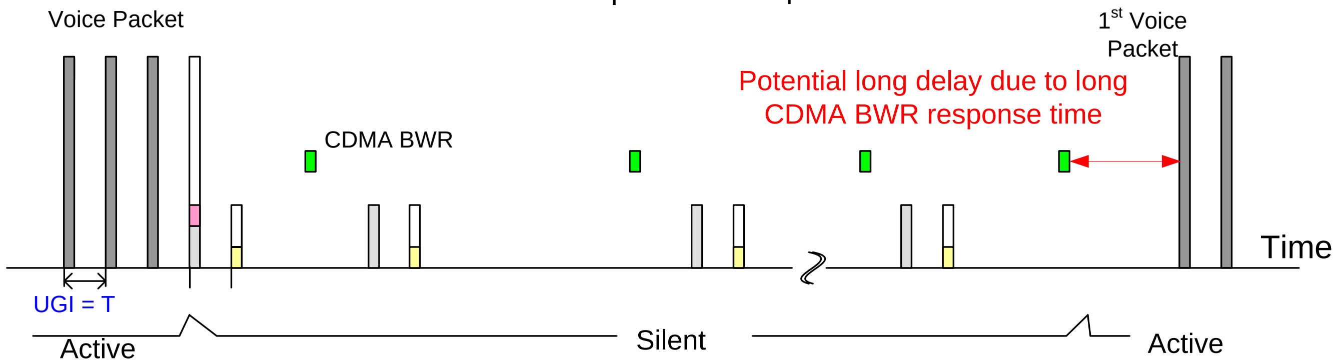
  BS allocates 6 bytes for bandwidth request header (BWR) every 20 ms. ➔ wasted UL BW



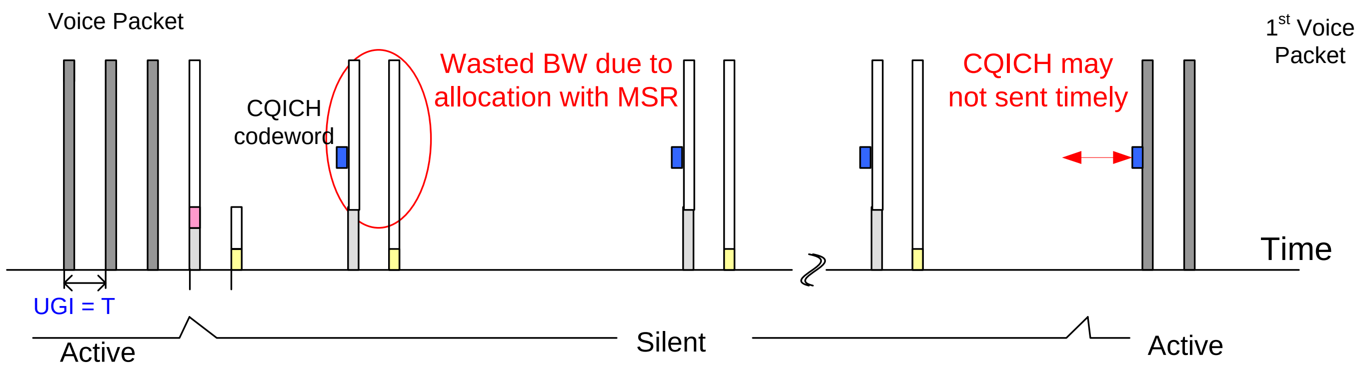
Continued: inefficiency of ertPS

- **Option 2:** CDMA BWR
  - BS provide no allocations; MS uses CDMA BWR for SID or 1st voice packet ➔ increased signaling overhead and potential longer delay
- **Option 3:** CQICH codeword
  - BS provide no allocations; MS uses CQICH codeword for SID or 1st voice packet, and BS then allocates using the maximum sustained rate ➔ wasted UL BW and potential longer delay

Current ertPS operation – Option 2: CDMA BWR



Current ertPS operation – Option 3: CQICH codeword



Requirement for QoS support

Many real-time/interactive applications have on/off bursty traffic pattern

- VoIP, Gaming, Push-2-Talk
- different delay sensitivity, for different applications

- Their traffic pattern is ON/OFF.
  - ON period:
    - Frequent packet arrival. Can be handled by UGS or regular polling.
  - OFF period:
    - Much longer packet inter-arrival time or no packet at all
    - May not know ON→OFF transition
    - require fast OFF→ON transition, in order to satisfy QoS.

We need a new scheduling service to satisfy both delay constraint and efficiency.

Adaptive Grant and Polling Service (aGPS) for IEEE 802.16m

- **Design concept:**
  - **On-period grant or polling: similar to the current ertPS or rtPS.**
  - **Off-period detection and handling, two possible solutions**
    - **Implicit:** BS itself adjusts the grant or polling configuration adaptively. The adaptive algorithm can be optimized with different functions for different applications.
    - **Explicit:** BS adapts grant or polling configuration with MS's assistance.
  - **Off→On transition:** all three methods including BWR, CDMA BWR and CQICH may be used simultaneously for quick access

- **Advantage:** significantly reduces the signaling overhead for polling, with marginal increase on latency.

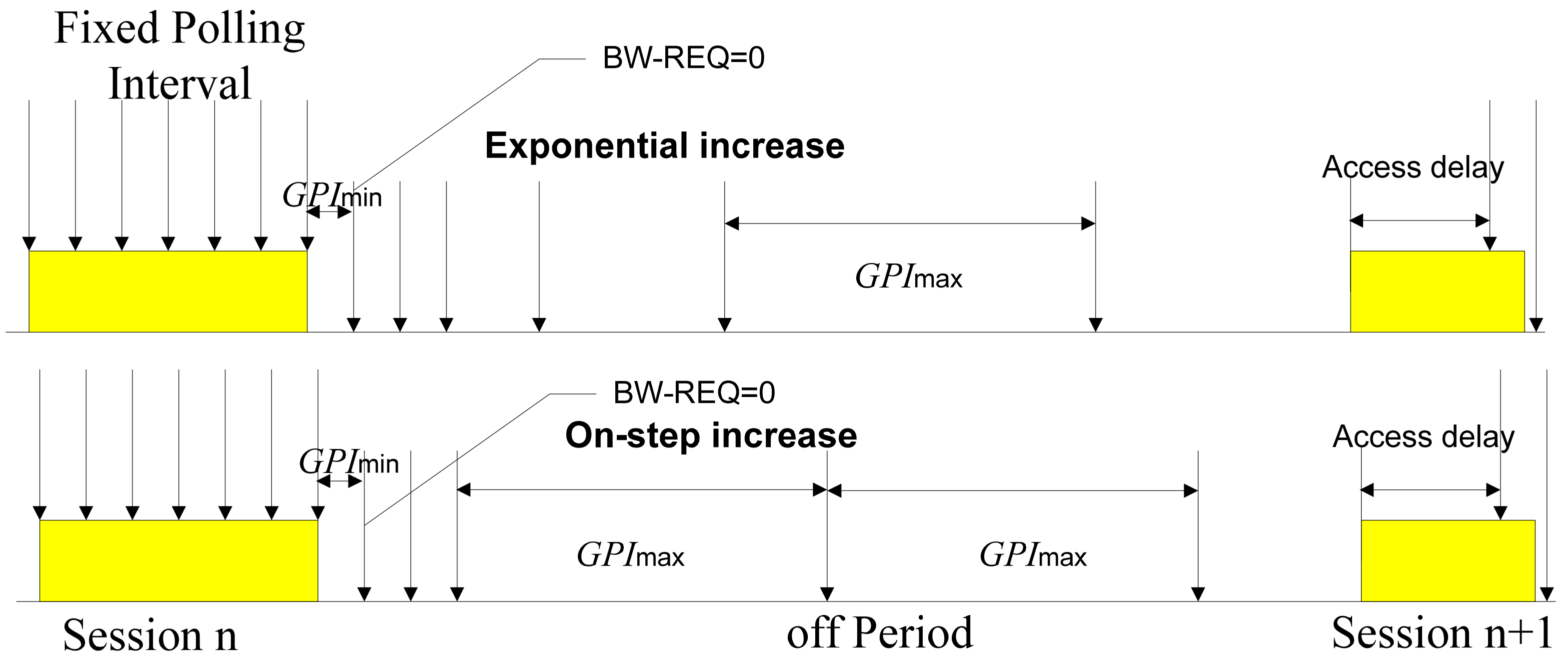
- **Parameters to negotiate for aGPS** (informative)
  - Polling or grant QoS parameters such as Minimum Grant/Polling Interval (GPI_{min}) etc.
  - implicit or explicit On→Off handling
    - Implicit:
      - Grant size: if grant size=6 bytes, it is polling, otherwise it is granting
      - Maximum Grant/Polling interval (GPI_{max})
      - Adaptation method (exponential increase, one-step increase, …)
    - Explicit:
      - Grant size: if grant size=6 bytes, it is polling, otherwise, it is granting
      - Maximum Grant/Polling interval (GPI_{max})

aGPS without MS assistance

- **Initialization:** adaptive grant or polling parameters are initialized between BS and SS based on the traffic characteristics:
  - Grant size: if grant size=6 bytes, it is polling, otherwise it is granting
  - Minimum grant or polling intervals (GPI_{min})
  - maximum grant or polling intervals (GPI_{max})
  - Adaptation method: exponential increase, one-step increase
- **Run-time operation:**
  - Upon the *N*-th zero sized BW-REQ, grant/polling adaptation begins
  - Upon non-zero sized BW-REQ, resume ON period grant or polling
- Pros
  - Can be backward compatible
  - Less MS complexity

Example: adaptive polling w/o MS assistance

- *Grant size = 6 bytes → adaptive polling*
- T_n: the polling interval of the *n*th poll.
- GPI_{min}; GPI_{max}: minimum and maximum polling interval used by the BS.
- *N*: the first number of polls in which the polling interval is fixed.
- *M:* the number of polls so that the polling interval is exponentially increased from the *N*+1 to *N*+*M* intervals.



aGPS with MS assistance

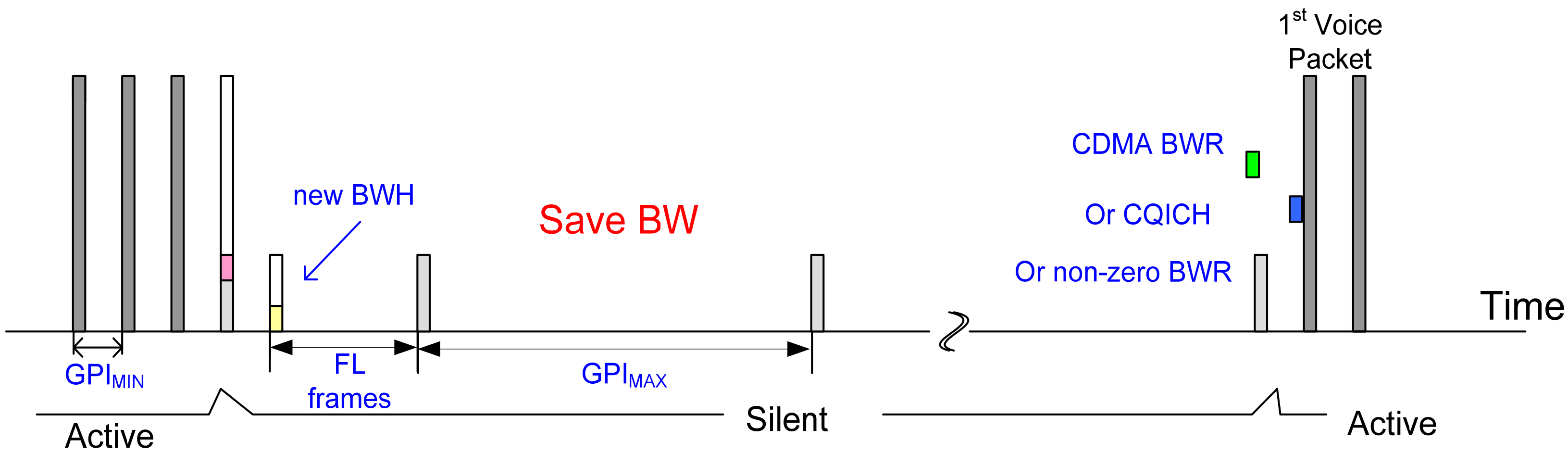
- **<u>Solution</u>:** the new grant or polling intervals and grant size to be used in OFF period are indicated in the **Bandwidth Request related MAC Headers** (BWH) explicitly and can be changed dynamically

- **<u>Runtime Operation</u>:**
  - **MS:** send a BWH indicating a new grant or polling interval (GPI_{max}) and a new grant size (if grant size=6 bytes, it is polling, otherwise, it is granting) upon detecting of state change; and this does not exclude the use of CDMA BWR or CQICH
  - **BS:** use the new grant or polling interval (GPI_{max}) and grant size indicated by MS for the future allocation.

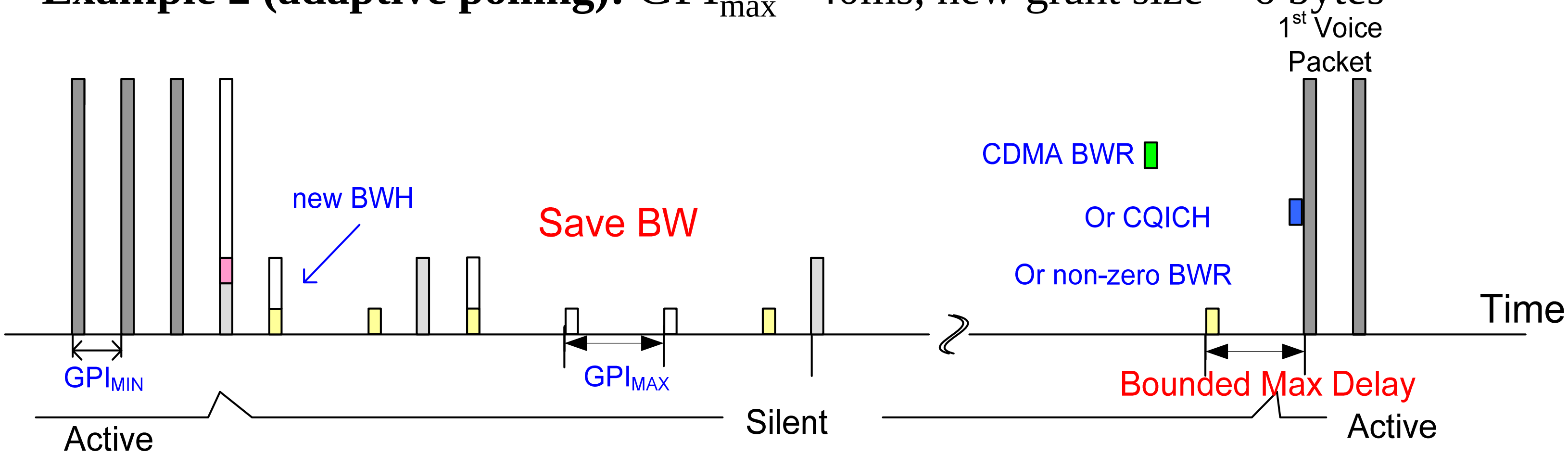
- **Pros**
  - Grant or polling interval can be adjusted dynamically with MS assistance who knows best about its application characteristics

Examples: adaptive grant or polling with MS assistance for VoIP

- **Example 1 (adaptive grant):** GPI_{max} = 160ms, New grant size = SID pkt



- **Example 2 (adaptive polling):** GPI_{max} =40ms, new grant size = 6 bytes

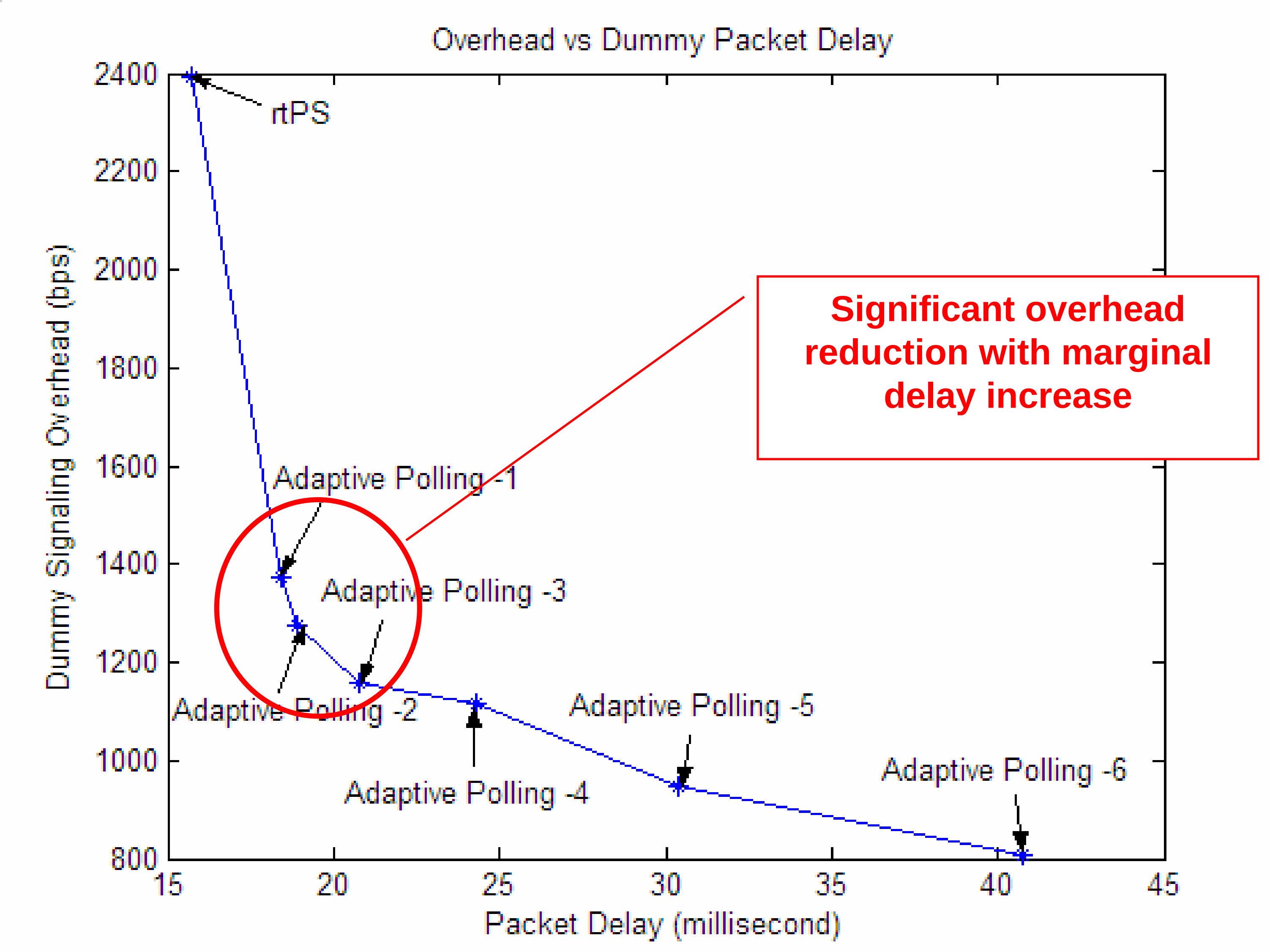


Simulation setup: aGPS vs. rtPS

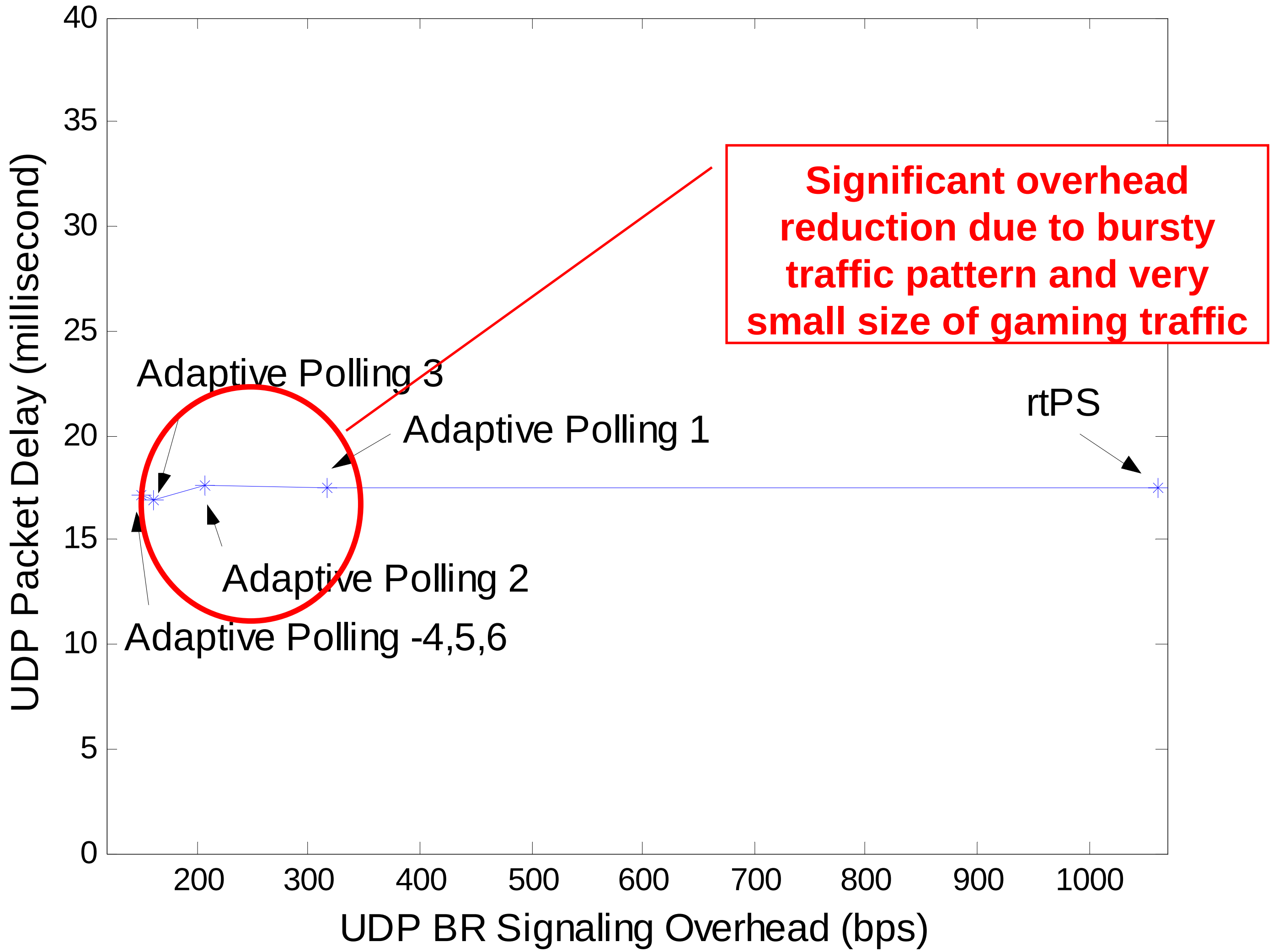
- Traffic pattern
  - VoIP Traffic Model : 2-state Markov model in [2].
  - Real gaming trace: “Age of Empires”
- aGPS:
  - Adaptive polling: assume all polling, only adapt polling intervals
  - Adaptation method: exponential

Scheduling service configurations		Minimum polling interval	Maximum polling interval	*N*
rtPS	N/A	20 ms	20 ms	N/A
Adaptive Polling	1	20 ms	40 ms	3
	2	20 ms	80 ms	3
	3	20 ms	160 ms	3
	4	20 ms	320 ms	3
	5	20 ms	64 ms	3
	6	20 ms	1280 ms	3

performance of adaptive polling w/o MS assistance
delay vs. signaling overhead



aGPS vs. rtPS: AOE game



aGPS vs. ertPS: VoIP AMR

Simulation Setup

- Traffic pattern
  - VoIP Traffic Model : 2-state Markov model in [2].
- aGPS:
  - Adaptive grant: GPI_{min} = 20ms, GPI_{max} = 160ms, N=1, new grant size in OFF period = SID pkt
  - Adaptation method: one-step increase
- ertPS: assume current ertPS operation Option 1 (refer to slide 4)

Results

- Average **UL BW savings of aGPS** comparing to current ertPS (option 1)
  - Analysis results: **11.9%**
  - Simulation results: **12.4%**

Service Class	UL_Granted_BW (Kbps)
ertPS	12.84
aGPS	11.245

Proposed Text for SDD

Insert the following text into Medium Access Control sub-layer sub-clause [1]:

----------------------------- Text Start **--**

10.x.x Adaptive Grant/Polling Service (aGPS)

IEEE 802.16m aGPS supports grant/polling interval adaptation. The BS determines the grant/polling interval to be used based on QoS parameters defined in service flow and event triggers. The MS may assist grant/polling interval adaptation at the BS with explicit signaling.

----------------------------- Text End **--**

Video QoS Improvement Proposal

Problem 1: Unintelligent Buffering

- Currently QoS parameters (traffic priority and MTL) are statically assigned.
- Video streaming applications are delay sensitive and have a jitter buffer.
- Intelligent video applications can adjust buffer size based on changing channel conditions, thus effectively increasing the delay bound.
- Use of a static MTL can result in lost packets as the size of the application buffer may exceed the MTL.

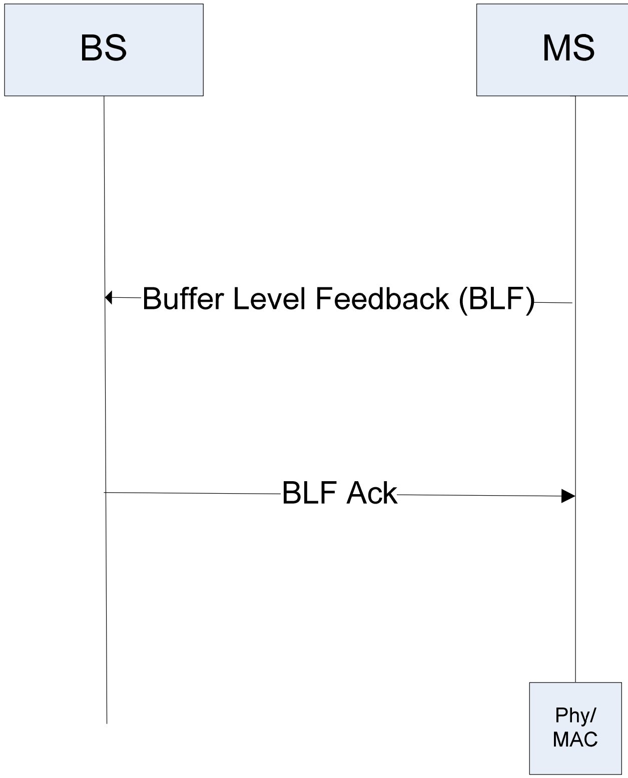
Problem 2: Undifferentiated Packet Distribution

- Two clients with the same service flow type and QoS parameter sets can generate two different levels of bandwidth demand. For example, User 1 can fast forward through a video stream and consume their jitter buffer at a higher rate than User 2 who is watching video in real time.
- However, the Base Station assigns resources based on the scheduling scheme, e.g. PF, to these two subscribers, resulting in buffer starvation at User 1.

Conclusion: We need a mechanism for conveying the application buffer information to the base station to ensure accurate scheduling and buffering.

Solution(s)

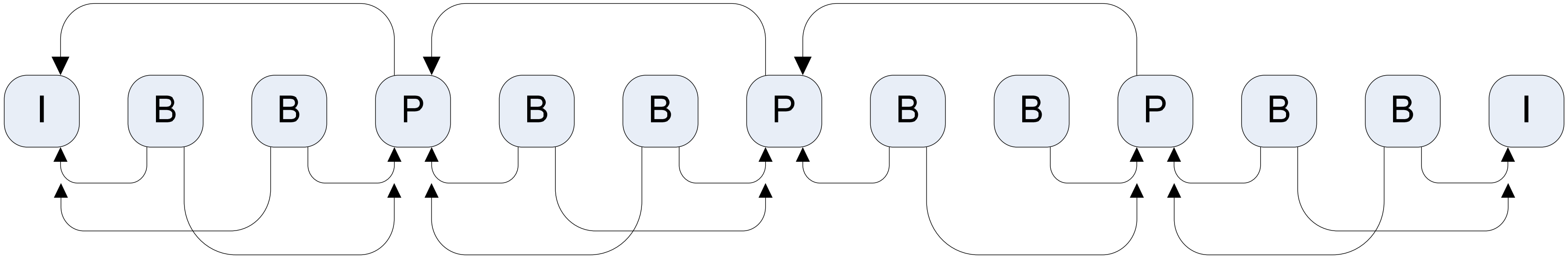
- Provide a mechanism for communicating application buffer size and occupancy to the Base Station, either event based or periodically.
- The Base Station can take appropriate actions based on scheduling implementation.
- Currently the standard has the DSC messages, but these do not provide necessary information to the BS to reflect the user behavior such as fast forwarding to realize user demand. Buffer level feedback would give a more accurate picture for better scheduling at the BS.



Unequal Error Protection Proposal

Problem: Equal Packet Handling

- Video traffic can be broken into critical (I) and non-critical (P,B) frames, as shown below in a frame dependency figure.
- The loss of a critical frame results in a significant disruption in the video stream for an extended period of time whereas the loss of a non-critical frame is less disruptive.
- However, the Base Station classifies all packets in the same stream equally.



Conclusion: The Base Station must be able to differentiate between critical and non-critical frames.

Proposed Solution – I-Frame Protection

Provide a mechanism (FFS) for the MAC layer to identify the criticality of the packets based on information provided by the video application.

Based on this information the MAC layer can increase the probability of a critical frame being delivered to the subscriber in the following techniques:

- Assign a more robust MCS to the critical packets.
- Adjust the priority of the critical packets in the scheduling queue.
- Assign critical and non-critical frames to separate service flows.

Proposed Text for SDD

Insert the following text into Medium Access Control sub-layer sub-clause [1]:

-------------------------- Text Start ---

10.x.x Video QoS Optimization

IEEE 802.16m provides an efficient signaling mechanism between the MS and the BS that communicates information such as buffer capacity and occupancy level at either a regular interval (periodic) or as an event-driven signaling scheme. This information can be used by the BS to dynamically adjust the QoS parameters (such as MTL) to improve video performance.

IEEE 802.16m provides a mechanism for detecting critical frames in video streams. This information can be used to ensure that these frames are successfully delivered with a higher probability, reducing the impact to user perceived QoS.

------------------------------ Text End ---

References

[1] C802.16m-08/003r4, Draft IEEE 802.16m System Description Document (SDD)

[2] IEEE 802.16m-08/004r2, Project 802.16m Evaluation Methodology Document (EMD)