Dyaptive SYSTEMS

DMTS-8000 Assisted Data Service Rollout in CDMA2000 Networks

April 13, 2005

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

CONFIDENTIAL

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

1 Introduction

On the technical roadmap for 'any service anywhere anytime', the CDMA2000 system is the first ever manifestation of a multi-server wireless cellular network offering data rates of up to 2Mbps and supporting data transport over packet bearer mode as well as conventional circuit bearer mode. These features allow provisioning of a broad spectrum of services including high-bit rate data transfer, custom enterprise applications, massively multi-player online games, push-to-talk, multimedia messaging, video-conferencing, video-phoning and audio/video streaming in addition to conventional voice and low bit-rate services such as web-access, email and SMS. The main limitation to supporting a large number of subscribers per service per unit area, however, continues to be the scarcity of the available radio frequency spectrum. Efficient and targeted utilization of radio spectrum and other associated network resources is the key to the economical viability of these 3G networks.

Each service provider must differentiate itself from the competitors by offering new high quality services at competitive prices. The new services however shall be made available to the target demographics without seriously impacting existing services and revenue streams. Since no clear 'killer app' has emerged so far, the service providers have to treat all of their new service offerings as potential 'killer apps' and deploy them with best possible availability and quality in mind. An effective rollout strategy therefore needs to be carved out to significantly lower the 'cost per gross add' and maximize 'return on investment'. From network engineering perspectives this translates to maximizing subscriber revenue per service per unit coverage area while ensuring the QoS.

2 Understanding Quality Drivers in CDMA2000

For the packet bearer mode, the CDMA2000 standard proposes TCP and UDP as the transport protocols in the upper layer of its protocol stack. The standard allows for data traffic over common or dedicated channels. On dedicated channels, the TCP or UDP packets are segmented into RLP (Radio Link Protocol) frames by the layer-2 of the protocol stack. RLP frames are first assigned to logical dedicated traffic channels by the MAC sub-layer before being multiplexed onto the physical fundamental, supplemental or dedicated control channels by the dedicated channel multiplex sub-layer of the layer-2. The CDMA2000 standard also allows the transmission of data traffic over common control channels via upper layer signaling. From upper layer signaling these packets follow through the Signaling LAC (Link Access Control) layer, and SRBP (Signaling Radio Burst Protocol) of layer-2 before being assigned to logical signaling channels by the MAC sub-layer. The logical channels are finally multiplexed onto physical common control channels by the common control multiplex sub-layer.

The quality of data services is measured in terms of throughput, packet loss, packet delay, and delay jitter. Some data services though also specify session level QoS parameters such as session setup delay, session block rate and session drop rate. In any end-to-end multi-hop data connection traversing wireless as well as wired links, RF links are usually deemed to be the throughput bottlenecks. To address such concerns, the CDMA2000 standard offers flexibility and parameterization of almost every layer of the protocol stack. Meticulous fine tuning of all layers of the protocol stack, is needed to mitigate the affects of these unreliable and bandwidth starved RF links.

One of the QoS parameters critical to reliable data services is the packet loss rate. Packet loss rate is a direct consequence of the SINR (Signal to Interference and Noise

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

ratio) or BER (Bit Error rate) of the underlying RF channel. A TCP or UDP packet is segmented into numerous RLP frames before its transmission over the RF link. All RLP frames must be received correctly to reconstitute the IP packet on the receiver side. The segmentation of TCP/UDP packets into RLP frames therefore shall be optimized to the SINR conditions. The layer-2 Radio Link Protocol of the CDMA2000 protocol stack enhances the reliability of the data packet transfer over the RF link by retransmitting negatively acknowledged RLP frames. For real-time data services such as content delivery using RTP over UDP, retransmissions add undesirable packet delay and thus usually deteriorate rather than enhance quality of experience. The data services that employ TCP, on the other hand, usually expect reliable transmission. Given the channel conditions, appropriate RLP schemes shall be selected. The RLP schemes such as (1.2.3) in which the number of NAKs (Negative Acknowledgments) increase in each round of retransmissions are likely to be more suitable for correlated channel conditions whereas the schemes in which only one packet is sent for each retransmission may be sufficient for uncorrelated fading channels with reasonable SNR. The main responsibility of RLP therefore is to hide the noisy channel characteristics from TCP using ARQ (Automatic Repeat Request).

Various parameters associated with TCP, which itself uses window-based flow control mechanism, must be calibrated with respect to a particular RLP scheme in use. These TCP parameters include MSS (Maximum Segment Size), Initial Congestion Window and Maximum Window Size. The Maximum Window Size must be greater than or at least equal to the delay-bandwidth product of the connection otherwise the link will be underutilized. The Initial Congestion Window impacts the slow-start phase of the windows based congestion control mechanism of TCP/IP. The MSS also has an effect on the overall throughput of the connection and resource consumption. Larger the MSS, higher is the connection throughput because of better payload size to TCP/IP header size ratio and lesser amount of TCP/IP acknowledgements that the application will have to deal with in the opposite direction. However, under noisy channel conditions, a smaller MSS is more beneficial as lesser user data will be retransmitted at the TCP/IP layer level subsequent to unsuccessful attempts to deliver the data at the RLP layer level. Furthermore, the RLP scheme must deliver packets before the TCP congestion/flow control mechanism kicks-in and attempts to throttle the connection throughput. Finally, some TCP/UDP applications can enjoy higher throughput by employing VJ Header Compression or Link Layer Header Compression offered by PPP. VJ header compression when employed on TCP/IP acknowledgements, flowing in the opposite direction to the TCP/IP packets carrying the payload, can substantially reduce the bandwidth consumption. The gains in throughput however will evaporate as soon as the underlying channel becomes noisy. Applicability of Header Compression schemes should be tested and verified under noisy channel conditions before applying in real situations.

Deleted: saviour

The data services that can tolerate some packet loss, or are delay tolerant also inadvertently improve capacity. This is because unlike voice services where all users have similar BER requirements, data services do not require uniform power control. The power allocation would depend on the bit rate and thus low bit rate applications can tolerate low power assignment. Notably, the fundamental channel's target FER is always around 1% however supplemental channels can have FER ranging between 1% and 10%. Data channels therefore should be power controlled according to the target FER to save power consumption at the Mobile Station and the Base Station, as well as improve capacity.

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

The other significant QoS parameter is the packet delay. The RF link contributes to the packet delay through contention (collisions) at the MAC layer, queuing/scheduling delays at the mobile and/or base station, or lack of traffic channels or bandwidth scarcity. In a CDMA2000 system supporting a broad spectrum of services with different QoS requirements and traffic characteristics, efficient integration and multiplexing of traffic from a diverse mix of applications is quite challenging. The challenge comes from the complimentary nature of real-time services such as video and voice that are delay intolerant but can tolerate some packet-loss, and the non real-time data services that can tolerate moderate delays but not the packet loss. Moreover, disparate solutions may apply to disparate services even in the presence of similar symptoms of QoS deterioration. For example, both video conferencing and video streaming are delay and jitter sensitive, however, in the case of video streaming an acceptable quality of end-user experience could be provided by simply adding more capacity to the media play-out buffers rather than assigning more bandwidth to the connection.

Numerous options exist in CDMA2000 to choose from and to optimize the packet delay. For example, firstly, there is a choice to transmit data bursts either over signaling channels or traffic channels. In particular, long or bidirectional bursts could use dedicated control channels or dedicated traffic channels whereas short, unidirectional bursts could share reverse common channels. Short and delay sensitive bursts could be transmitted using Basic Access or Power Controlled Access of R-EACH whereas large bursts via Reservation Access channel. And in case it is decided to transport user data, or parts of it, over common signaling channels such as R-ACH or R-EACH then perhaps adding more capacity to these channels or fine-tuning of relevant parameters such as 'access probe back-off window' associated with random access shall be considered. Secondly, appropriate number of Supplemental and Fundamental Channels need to be provisioned to match the specific traffic, for example, supporting 64 kbps of a video streaming session with (8xSCH + 1x FCH) instead of using a single, large SCH. Thirdly, the duration and amount of supplemental channel assignments could be fine-tuned for bursty or variable bit rate data sessions, and trade-off between fixed supplemental channel assignment vs on-demand supplemental channel assignment could be established. Finally, various time-outs associated with the MAC layer state-machine could be optimized for specific data services. As an example, releasing dedicated traffic channels immediately after the burst is efficient from resource sharing perspectives however it adds to the setup delay when a new burst arrives and a new dedicated traffic channel needs to be acquired. A dedicated channel establishment introduces latency and signaling overhead due to renegotiation procedure including RLP resynchronization and signaling overhead etc. For certain applications that cannot tolerate session set up delays, "premium" load management features such as configuration caching could also be explored and optimized.

3 Measuring and Tuning for Quality in CDMA2000 using the DMTS-8000

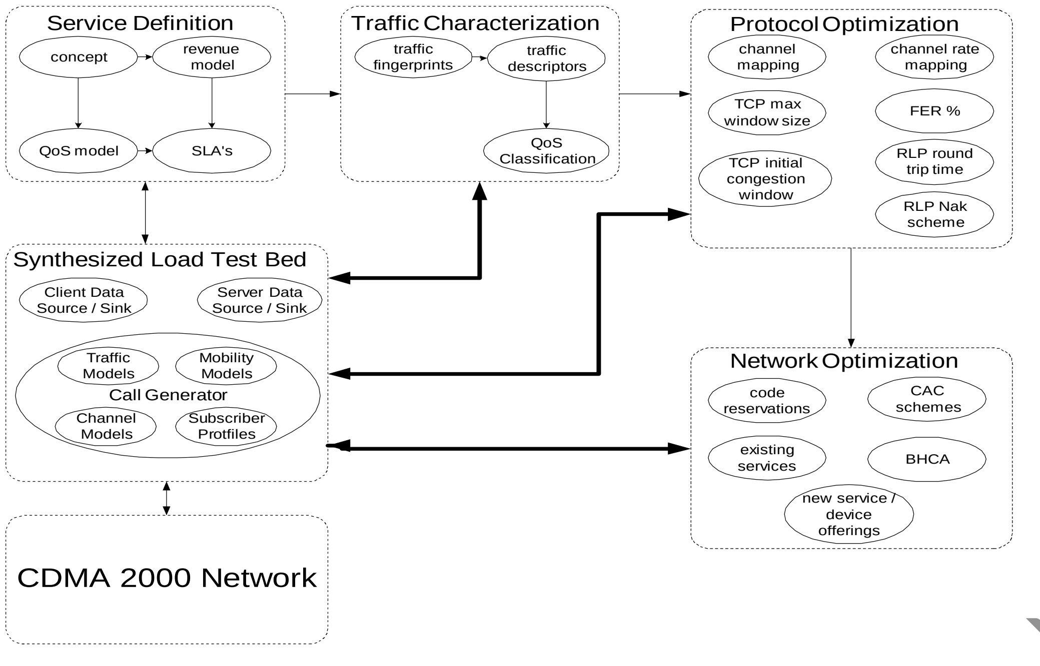
Dyaptive System's DMTS-8000 is a comprehensive system test solution that addresses the key requirements of service provisioning and optimization in CDMA2000 wireless cellular networks. It facilitates performance testing of data services without first having to deploy them on functioning CDMA2000 platforms. The DMTS-8000 exposes deficiencies in resource utilization; it helps understand delicacies of CDMA2000 protocols; and brings to the foreground intricacies among various layers of the CDMA2000 protocol stack including the application layer.

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

One of the main hurdles in the optimal provisioning of data services over wireless networks is the lack of prior knowledge of their traffic characteristics. There exist time tested traffic models for circuit-switched services however accurate traffic models for mobile, packet-switched services are still being devised. DMTS offers a unique and effective approach to this problem. It not only allows applications to be modeled using traditional Markov, Pareto, long range dependent and deterministic traffic models etc, but it can also emulate a real application session and connect it over the CDMA2000 test-bed. This is achieved by connecting real client and server components of the applications over CDMA2000 test-bed via DMTS-8000's VMTs (Virtual Mobile Terminals). A large bulk of such sessions could be created by replicating the copies of the session over multiple VMTs. Consequently, a relatively new and unknown application could be effectively provisioned and fine-tuned before its actual deployment. The DMTS approach allows the, qualitative though invaluable, QoE (Quality of User Experience) to be factored into the service provisioning process, in addition to the conventional quantitative QoS parameters.

The following use case elaborates on the key role that DMTS-8000 plays during the data service rollout. The applicability of DMTS in this process is in fact two fold. Firstly, it helps identify the most efficient strategy for deploying an application on CDMA2000 channels. A large set of CDMA2000 parameters are iteratively fine-tuned and the application performance is monitored until the throughput and QoS approaches the acceptable levels. Secondly, DMTS-8000 helps control the detrimental impact of this new offering on the existing services. Using DMTS-8000, a CDMA2000 test-bed is loaded with a mix of services and applications, including the new offering under consideration; the performance of applications is monitored; and development of efficient strategies to schedule, admission-control, multiplex and integrate the traffic from this diverse set of services and applications is facilitated.

Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.



1. An application developer, a content producer and a wireless service provider have partnered to offer a sophisticated data application involving real-time content delivery and non real-time but reliable messaging. They have developed a quality versus revenue model for deployment for the new service. The application uses IETF's SIP for signaling exchange, TCP for transporting messages and RTP over UDP for transporting content. The application is based on standard client/server architecture with SIP proxy servers, SIP registrar and content servers located in the Internet. The clients however can exchange messages among each other as peers using SIP servers for address resolution and management. In addition, CDMA2000 employs Mobile IPv4 to support IP address persistence and location management across its PDSNs. A typical session may therefore involve multiple peers exchanging messages or being involved in content exchange via content servers.
2. The network engineer has to plan an effective rollout strategy that maximizes the availability of this service to the potential user base and maintain the desired quality of user experience without impacting the existing services adversely. The quality of user experience is mapped to a set of measurable parameters that include minimum data throughput, maximum acceptable packet loss rate, maximum acceptable packet delay and delay variations. At the session (call) level as well, upper bounds on session setup delay, session block rate and session drop rate need to be guaranteed. Being a newly minted application, not enough information is available on its traffic characteristics.

Dyaptive SYSTEMS Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

3. The network engineer starts the provisioning process by first testing the application on a CDMA2000 test-bed in a lab environment. A multi-client session is established with some emulated clients communicating wirelessly whereas the rest connected via the landline Internet. The DMTS-8000 is used to emulate this session using actual client and server components of the application and using CDMA2000 test-bed for wireless connectivity. Traffic over CDMA2000 channels is monitored and a finger print of the message flows into and out of the CDMA2000 system is obtained.

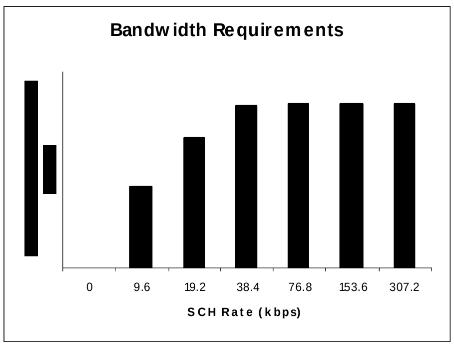
4. Using empirical data, obtained in the previous step, as well as through consultations with the application developer, the network engineer partitions the session traffic into various streams based on QoS classes, and considers assigning various combinations of CDMA2000 channels to these streams. For example, SIP or Mobile IP messages could be transported over common signaling channels or dedicated signaling channels whereas the data traffic may stay on the data traffic channels. Being crucial to connection and location management such application level signaling messages needs to be transported over reliable channels by employing appropriate RLP schemes. Similarly email notifications from email server to email clients could be exchanged as SMSs whereas the email messages themselves could be treated as SDB (Short Data Bursts). Inter-Process Communication in various gaming applications can also be provisioned as SMS or SDB with occasional use of dedicated traffic channels. Using the traffic parameters discovered in Step 3, the appropriate number of Supplemental and Fundamental Channels are estimated to match the specific traffic, for example, supporting 64 kbps of a video streaming session with (8xSCH + 1x FCH) instead of a single, large SCH etc. The numbers and rates of forward and reverse supplemental channels assigned to the application instance are varied until no further improvements in throughput are observed. The minimum amount of capacity required to maximize the session throughput is thus determined. Each channel assignment option is evaluated, possibly including new channel capabilities not presently supported in the network (e.g. EACH, EV-DO, EV_DV); data throughput, delay & jitter in each experiment are monitored using DMTS-8000; the optimal channel allocation is chosen for deployment.

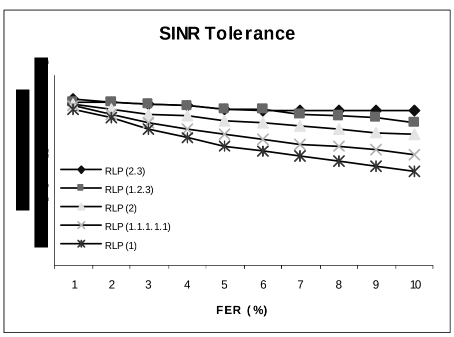
5. Next, the resilience of the application to the noisy channel conditions is evaluated. The application is optimized for underlying channel conditions by appropriately configuring the Transport Protocol, Radio Link Protocol and physical layer parameters to achieve acceptable levels of packet loss rate. The channel models appropriate for the area of interest are selected in DMTS-8000. The target FER (Frame Error Rate) of supplemental traffic channels (if chosen for transporting user data in step 4) is varied between 1% and 10% by varying the power allocation, until the throughput starts deteriorating below the acceptable levels. The optimal power allocation is selected for deployment considerations. For traffic streams requiring TCP based reliable transport, RLP is employed to hide the noisy channel conditions from the TCP layer. Aggressive RLP schemes can effectively minimize TCP retransmissions however such schemes also consume more bandwidth. DMTS-8000 is used to select a suitable RLP scheme that achieves acceptable TCP throughput while maintaining a low bandwidth overhead. Additionally, TCP parameters such as MSS (Maximum Segment Size), Initial Congestion Window and Maximum Window Size are optimized. The Maximum Window Size is increased until it approaches the delay-bandwidth

Dyaptive SYSTEMS Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

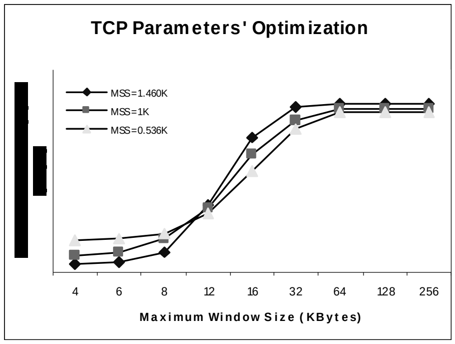
product of the connection and the connection throughput is maximized. The Initial Congestion Window is chosen according to the latency requirements of the traffic stream. Similarly, the MSS is varied until its impact on the connection throughput due to factors such as the payload size to TCP/IP header size ratio and the volume of TCP/IP acknowledgements is optimized with respect to its impact on TCP retransmissions under noisy channel conditions, as described earlier.

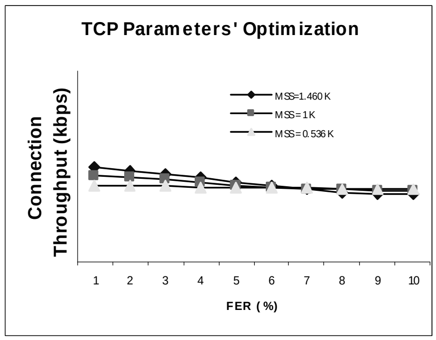
6. MAC layer is next optimized to minimize packet delay and jitter. Based on the traffic characteristics discovered in Step 3, optimization of the number, duration and capacity of supplemental channels is performed. The values of REV_SCH_DURATION, REV_SCH_NUM_BITS_IDX, FOR_SCH_DURATION, FOR_SCH_NUM_BITS_IDX in the supplemental channel request messages are varied according to the observed traffic characteristics, and the data throughput is monitored. The optimal values of the aforementioned supplemental channel parameters are selected for service provisioning. For the bursty sub-stream of the application traffic, selected in Step 4 for transportation over signaling channels, the impact of various access options on R-ACH and R-EACH e.g. basic access, power controlled access and reservation access, is observed with respect to the burst throughput. The option most suitable for the expected burst size is chosen for service provisioning.





Dyaptive SYSTEMS Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.

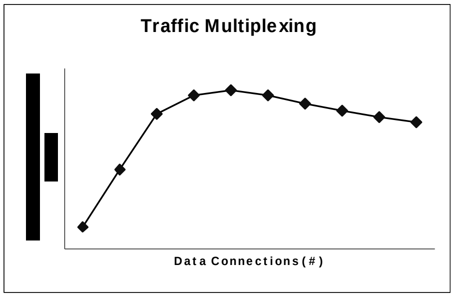


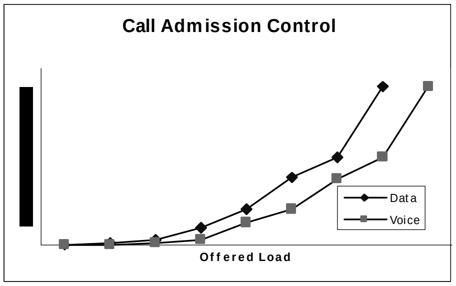


7. The network engineer emulates a sample of application sessions in the CDMA2000 test-bed using DMTS-8000 and observes the performance in the presence of background traffic such as voice BHCA. Packet data bearer modes e.g. P1, P2 & P3 and also voice modes e.g. V1 and V2 are populated. The Call Setup Delay and the Call Blocking Rate of the background voice traffic are also observed in addition to throughput, delay and jitter of the data traffic to monitor the impact of this new service on the existing services. The parameters identified in Step 6 are further optimized until the desired changes in all the aforementioned QoS parameters are observed. The 'access probe back-off window' size is altered and the numbers of R-ACH and R-EACH are increased to optimize the signaling load with respect to QoS.

8. To maximize revenue while ensuring desired quality and availability of service, the network engineer explores channel/code planning strategies. Under some CAC (Call Admission Control) policies, some codes (bandwidth) are exclusively reserved for the most revenue generating applications while the rest of the capacity is used on a first come first served basis. The number of channels reserved for each of the differentiated services is varied until desired outage probabilities (FCH request denials), service throughput and QoS is achieved.

Dyaptive SYSTEMS Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.





9. The performance of scheduling policies employed by the CDMA2000 network is monitored. For example if a fair-share based scheduling scheme is being employed then as the network load increases, all the connections must see proportional deterioration in bandwidth. The aggregate throughput of the sector as well as individual connections in monitored for such purposes.

Dyaptive SYSTEMS Copyright © 2004 Dyaptive Systems Inc.
All rights reserved.