VoIP over cdma2000 1xEV-DO Revision A

Mehmet Yavuz, Serafin Diaz, Rohit Kapoor, Matt Grob, Peter Black, Yeliz Tokgoz, and Chris Lott, QUALCOMM Inc.

ABSTRACT

In this article we analyze performance of VoIP services over 1xEVDO-Revision A (DO-Rev A) networks and show that high-quality VoIP with unconstrained mobility and high capacity can be achieved. Together with quality of service (QoS) requirements, we emphasize practical issues such as mobility, degradation of feedback-channel quality, and packet overheads. Novel techniques are presented for voice processing such as smart blanking and adaptive dejitter playback buffer with time warping. These techniques help to meet QoS constraints to achieve a circuitlike voice quality while improving overall capacity. Detailed end-to-end simulations are presented and system capacity is analyzed under the QoS and system stability constraints. We claim that DO-Rev A can provide VoIP capacity comparable to circuit-switched cellular CDMA technologies (e.g., IS-2000) and *simultaneously* carry significant amount of other types of traffic such as non-delay sensitive applications and downlink multicast.

INTRODUCTION

Integration of cellular voice services and high-speed data applications over wireless IP wide area networks has the potential to provide operators with better network manageability, higher network bandwidth efficiency, and greater cost savings. Such wireless networks can also provide operators with the ability to offer richer and more integrated applications together with voice.

Voice communication has stringent quality requirements on end-to-end packet delay and packet loss. Thus, quality of service (QoS) is of vital importance for VoIP services. A common misconception is that VoIP is synonymous with voice over Internet. However, to provide toll-quality voice services, operators would need to utilize reliable networks to carry voice traffic over the backhaul. In this article, we assume voice packets can be transferred over the core backhaul networks under stringent QoS requirements. Thus, our focus is on wireless access networks.

Wireless channels can degrade received signal quality due to various factors such as path loss, shadowing, and small-scale (short-term) fading. Ignoring the impacts of these impairments can yield to overly optimistic assessments of VoIP performance [1]. Prior research in analyzing VoIP performance over wireless channels has mostly focused on wireless local area networks (e.g., IEEE 802.11 WLAN) that provide only limited mobility [1]. In this article, we focus on wide area cellular networks, which are subject to wireless channel impairments but also require seamless handoffs for mobile users.

Traditional cellular systems provide voice communication services via circuit-switched networks where airlink resources are dedicated throughout the voice call. This method provides a reliable method for voice services, but is not efficient for carrying the bursty data traffic which is often seen on the Internet.

cdma2000 1xEV-DO (IS-856) is a spectrally efficient data-optimized cellular system which can be overlaid on to an existing cdma2000 network using the same 1.25 MHz channel bandwidth. Initial release of 1xEV-DO (referred to as DO Release 0) provides substantial improvements in downlink (i.e., forward link) capacity compared to cellular cdma2000 systems such as IS-2000. DO Release 0 is well suited for downlink-intensive Internet applications such as Web browsing and file downloading (among others), and it has experienced rapid subscriber growth.

Recently cdma2000 1xEV-DO Revision A (also referred to as DO-Rev A or IS-856A) system was standardized [2] by the Third-Generation Partnership Project 2 (3GPP2). DO-Rev A provides significant improvements over DO Release 0 that make it well suited both for downlink intensive applications and symmetric delay-sensitive applications such as VoIP. For example, on the downlink, DO-Rev A utilizes short packets and multi-user packets that enable significant delay reduction. Also, seamless adaptive server selection virtually eliminates transmission delays (i.e., service outage) due to downlink server changes. On the uplink (i.e., reverse link), performance is improved via hybrid automatic repeat request (H-ARQ), higher-order modulation, and finer rate quantization [3]. As a result, DO-Rev A can support large numbers of VoIP users together with other users with downlink intensive applications.

In this article, we discuss how high-quality VoIP service can be provided with full mobility over DO-Rev A networks. We first describe the

0163-6804/06/$20.00 © 2006 IEEE

voice quality and system capacity metrics used in our analysis. Next, we discuss packet overhead compression and describe voice-processing algorithms for VoIP over DO-Rev A. Then we describe detailed end-to-end system simulations and provide capacity estimates for VoIP in terms of maximum number of users supportable. Lastly, system capacity under mixed traffic configurations (e.g., VoIP, best effort, multicasting) is analyzed. Finally, conclusions are presented.

VOICE QUALITY AND SYSTEM CAPACITY METRICS

This section discusses the four metrics used in VoIP capacity and performance analysis.

METRIC 1: MOUTH-TO-EAR DELAY

A key factor that determines the VoIP quality is the mouth-to-ear delay, defined as the delay incurred from the moment the talker utters speech until the instant the listener hears it. Note that mouth-to-ear delay is different from packet delay. For a system with variable packet delay, voice frames cannot be played as soon as they arrive at the receiver. Thus, a dejitter buffer is utilized at the receiver that ensures timely playback of voice frames. If proper echo-cancellation methods are utilized, the main impact of large mouth-to-ear delay is the loss of interactivity during the voice conversation [4].

Different system components contribute to the end-to-end delay. At the source, the voice codec (also referred to as vocoder) collects speech samples[1] and then processes data to produce speech frames. Packet processing at various network nodes and propagation delay over the wireline links also contribute to end-to-end delay. On the wireless links (both uplink and downlink) each voice frame can experience a different amount of delay due to queuing, scheduling, and over-the-air transmission (e.g., H-ARQ[2]). Lastly, at the receiver, an adaptive dejitter buffer needs to be used to remove the delay jitter and play the voice frames. In our capacity analysis, mouth-to-ear delay is analyzed and compared with ITU recommendations [5]. Note that ITU recommendations are for circuit-switched networks and certain refinements may be required for VoIP systems that experience variable end-to-end delay.

METRIC 2: FRAME LOSS RATE

Packet losses generally reduce the intelligibility of the voice conversation. The relation of voice quality to frame losses largely depends on the particular vocoder used. In this study we focus on the cdma2000 family of vocoders (e.g., EVRC [6]).

In a VoIP system, frame losses can occur either on the communication paths or at the receiver playback (i.e., dejitter) buffer. The following section describes how an adaptive dejitter buffer can be utilized to exploit the trade-off between packet loss and mouth-to-ear delay.

METRIC 3: SYSTEM STABILITY

Since the uplink of DO-Rev A is based on CDMA, certain stability conditions have to be met due to the interference-limited nature of the system. DO-Rev A utilizes direct measurement of rise-over-thermal (RoT) [3], defined as the ratio of total received power to the thermal noise floor. Although this method provides a more robust method to control reverse-link loading compared to traditional CDMA systems, RoT distribution still needs to be within specifications for overall system stability. In our study, we use 3GPP2 guidelines [7] which require that the RoT not exceed 7 dB for more than 1 percent of the time.

METRIC 4: MAC CHANNEL LIMITATION

In DO-Rev A, the total number of MAC channels available for reverse-link power control (RPC) and H-ARQ [2, 3] is limited to 114. For each of the access terminals (AT) that include a particular sector in their active set, one MAC channel is used from that particular sector [2]. This puts an upper limit on the maximum number of VoIP users that can be supported on the reverse link, depending on the active set size distribution of the mobile terminals. Based on the soft-handoff factor under the 3GPP2 network model [7], this limits maximum number of users per sector to 66.

SYSTEM OVERVIEW

This section outlines how DO-Rev A networks and mobile terminals can be operated to achieve high-quality and robust VoIP performance with full mobility. We focus on the areas that are not described in the standard [2]. For more details of the DO-Rev A system in general, the reader is referred to [2, 3].

ROBUST HEADER COMPRESSION

Typically in VoIP, voice frames are carried over the RTP, UDP, and IP layers, which add significant amounts of packet overhead. One method to improve overhead efficiency is to bundle multiple frames, that is, put multiple voice frames into an RTP packet. However, this increases end-to-end delay and introduces burst errors into the decoder in the event of a packet loss. Thus, header compression is desirable in order to achieve high capacity.

RFC 3095 [8] defines a framework for robust header compression (RoHC), along with different compression protocols (profiles). Header compression protocols such as RoHC rely on the fact that there is much redundancy in the header field values between consecutive packets. The encoding of the RTP timestamp (TS) deserves special attention since the presence of silence periods between talkspurts and the use of silence suppression techniques can cause the timestamp to increase proportionally to the size of the silence period between consecutive packets. This irregular change is easily encoded by the use of a timer-based compression algorithm, which makes use of system clocks at the two ends and an estimate of the maximum jitter from the source to the destination of the packet for efficient compression.

Our detailed RoHC simulations (on the downlink and uplink) have revealed that RoHC can compress RTP/UDP/IP headers down to a minimum of 2 bytes (when UDP checksum is disabled) and a minimum of 4 bytes (when UDP

DO-Rev A can support large numbers of VoIP users together with other users with downlink intensive applications.

[1] *The vocoder needs to collect speech samples for at least the length of one voice frame. Usually there is additional delay due to "future" samples collected for more efficient encoding (i.e., Look ahead).*

[2] *H-ARQ (i.e., hybrid FER/ARQ; also known as HARQ type II) refers to incremental transmission of parity and repetition bits of a physical layer packet over multiple slots (staggered in time allowing receiver acknowledgment) so that excess* E_b/N_0 *is minimized.*

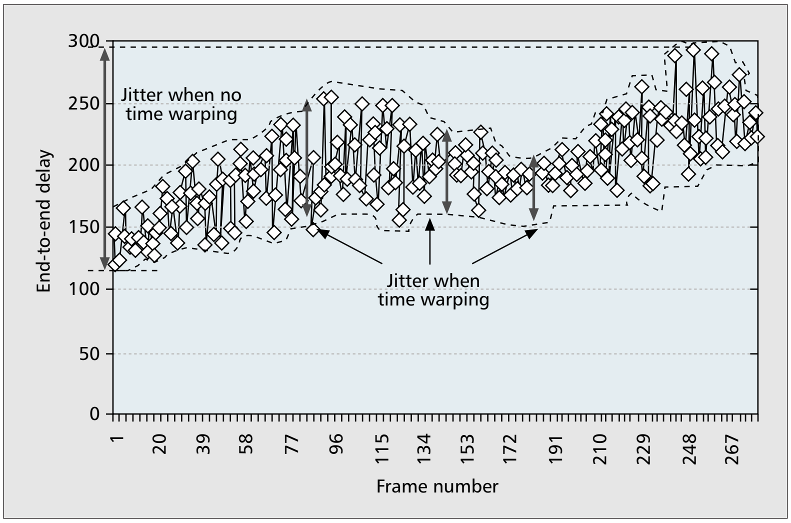

Figure 1. *Packet delay during a talk spurt.*

checksum is enabled) for most of the packets.[3] We have also found the RoHC performance to be extremely robust to packet reordering and/or drops. Only under severe frame-erasure run lengths (e.g., larger than 10), are RoHC header sizes larger than those mentioned above.

3GPP2 has chosen RoHC as the header compression protocol due to its efficiency in the face of dropped and reordered packets, both of which occur in EV-DO RevA. In IETF such support is being added by increasing the amount of resequencing that RoHC can tolerate. This can be accomplished by changing the value of the parameter p in the interpretation interval. This work is currently ongoing in the RoHC working group and is expected to be added to updated RoHC documents soon.

DO-RevA also enables Point-to-Point Protocol (PPP)-free operation via the enhanced packet application standard [9]. This eliminates several shortcomings of PPP-based transmission: Firstly, PPP results in additional header overhead (e.g., 5 bytes) and causes data-dependent packet expansion due to bit escaping. Secondly, PPP imposes high processing overhead due to byte-level processing. Thirdly, RoHC over PPP supports only one RoHC channel, which is insufficient for today's application requirements. Fourthly, with PPP the radio access network (RAN) is unaware of packet boundaries, which complicates congestion control. Lastly, features that are previously provided by or with PPP (e.g., error detection, authentication, and address assignment) are now available at the air interface or the network layer. In our simulations we assume PPP-free operation.

Voice Processing Algorithms

Dejitter Buffer Functionality — As mentioned above, the variance in end-to-end packet delay, (i.e., jitter) can lead to a situation where the playback utility is ready to play a packet,[4] but no packets are present in the buffer (i.e., underflow). Underflows can be caused by packets arriving late (delay underflows) or packets dropped on the air link. The dejitter buffer can reduce delayed underflows by increasing the amount of buffering. Thus, the underlying principle of our proposed dejitter functionality is to minimize delay (buffer size) while maintaining the packet error rate (PER) due to delayed packets below or at a predefined level, called the target PER.

A useful technique that the dejitter uses to maintain the Target PER is audio time-warping (more details on time-warping are explained below). Our studies have shown that a conservative use of time-warping leads to no perceptible loss in voice quality and can reduce delay.

When the dejitter buffer is running dry, it expands (time-warps) packets to avoid underflow, whereas when it has too many packets, it compresses them to reduce delay. This ability to expand/compress packets as necessary allows the dejitter buffer to encompass a shorter-time view of the jitter than without the use of time-warping (Fig. 1). In the absence of time-warping, the difference between the smallest delay and the largest delay would be seen as jitter. However, with the use of time-warping, as shown in Fig. 1, the dejitter buffer can adapt to the changing mean delay using time-warping and its view of jitter is only the variance around the local mean delay.

Audio Time Warping — Audio time warping is the ability to expand or compress the duration of a speech segment. It is important that the time-warping operation must not modify the pitch of the speech segment to minimize any perceived quality loss.

Encoded speech can be coarsely divided into four categories: silence, voiced, unvoiced, and transient frames.

Silence frames contain only background-noise information and can easily be time-warped (without any perceived quality loss) by simply discarding or repeating complete frames. In a VoIP system, it is preferred to execute most of the time-warping during the silent intervals of the conversation. For typical speech conversations, there is ample opportunity to exploit this technique.

Voiced frames contain speech segments with high periodicity (similar pitch periods). Examples of such speech segments are the ones containing vowels. These speech segments can be warped by repeating or deleting complete pitch periods. In order to minimize discontinuities at points where speech is added/deleted, blending techniques may be used.

Unvoiced frames contain noiselike speech. Examples of such speech segments are the ones containing "s," "ch," and "sh" sounds. Since such noiselike speech segments do not have well-defined pitch periods, they can be warped by deleting or repeating arbitrary sections of the frames. Blending techniques may also be used here to avoid speech discontinuities.

Transient frames typically contain transient speech ("t" and "p" sounds) as well as the transitions among all previous types of frames. These frames are not easily warped since any modification of these frames can potentially introduce artifacts resulting in loss in speech quality. If required, such frames may be warped by identifying the strongest pitch in the speech segment and following a procedure similar to that used for warping voiced frames.

[3] *Under typical frame erasures, average headers grow only to about 4.01.*

[4] *Note that we use voice packet and voice frame terms interchangeably.*

Silence Suppression (Discontinuous Transmission) — In a full-duplex conversation, usually one of the parties is silent. During these silence intervals, the communication channel carries only background-noise information. Proper communication of this background noise is important since it can affect the perceived voice quality.

As an example, in EVRC [6], 1/8 rate frames carrying encoded background noise information during silence are transmitted every 20 msec. Such continuous transmission of 1/8 rate frames (even when compressed) can be very expensive for VoIP due to RTP/UDP/IP packet overhead.

For accurate production of background noise at the receiver, the transmitting side needs to send a 1/8 rate frame only when the characteristics of the background noise (energy and/or frequency content) have changed "significantly." This selective transmission of 1/8 rate frames can significantly reduce overhead while still maintaining perceived voice quality.

Conversational Delay — One metric that has traditionally been used to characterize voice quality is the mouth-to-ear delay as defined above. In this section we argue that what matters from a conversational perspective is the time between when a user stopped speaking and when she first heard the other user's response. We define this delay as the conversational delay.

Figure 2 shows the timeline of a conversation between two users. The shaded segments represent talkspurts (speech segments), whereas the blank areas on the timeline represent silence periods. Once the first speech segment of User1 (blue) finishes playing at User2, User2 waits for a short interval of time before starting to speak. The beginning of User2's first speech segment (yellow) is subsequently heard by User1. The conversational round trip delay perceived by User1 is the time gap between when User1 stopped speaking to the point when User1 heard the beginning of User2's speech segment. We argue that this conversational delay (and not the one-way mouth-to-ear delay) is significant from the point of view of the users. For instance, if this conversational delay is too large for User1, it will prompt User1 to start speaking again without waiting for User2's speech segment to be played. This breaks the flow of conversation and is perceived as conversational quality degradation.

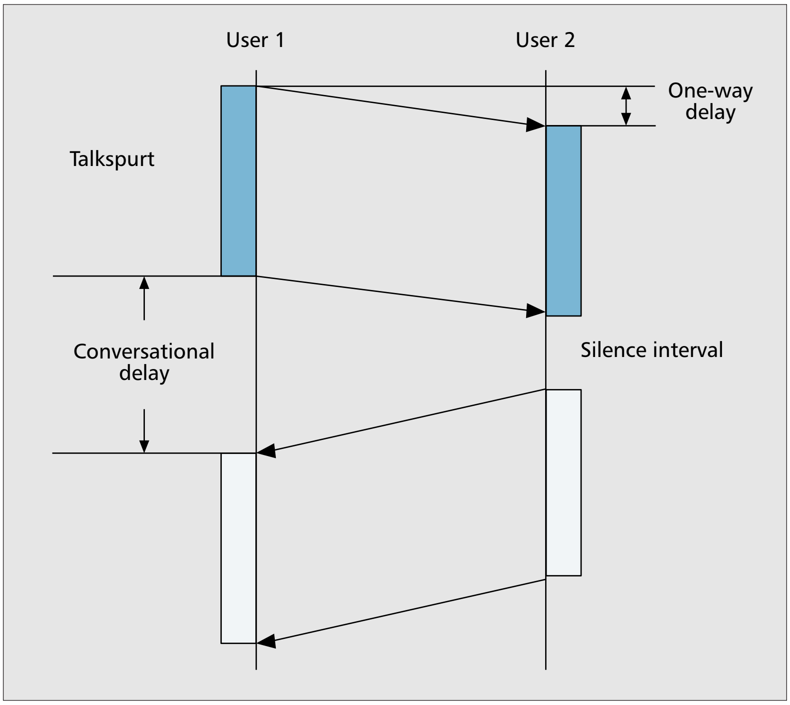

Figure 2. *Explanation of one-way delay and conversational delay.*

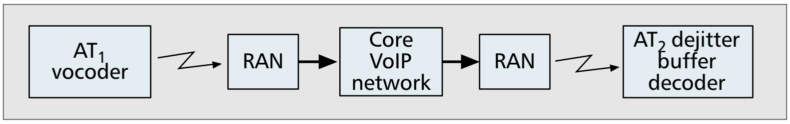

Figure 3. *Simulation model for a mobile-to-mobile call.*

System Capacity

This section describes VoIP capacity and performance simulations over 1xEV-DO Rev A. The maximum number of simultaneous users per sector is analyzed subject to the constraints described above.

Simulation Overview

We use a network level simulator based on the 3GPP2 simulation framework [7]. A few modifications to the framework were performed to align the simulator more closely to the actual implementation: firstly, based on DO-Rev A link budget analysis, the distance between base stations is set to 2.0 km and the maximum path-loss condition is 138 dB. Secondly, the AT is modeled with dual receive antennas. A 50 percent signal correlation and 4 dB gain mismatch are simulated between the receive antennas. Based on actual RAKE implementation, multiple spatial paths are combined using minimum mean square error (MMSE) and temporal paths are combined using the maximum ratio combining (MRC) method.

Our simulation platform models interference from 57 sectors in a wrap-around model [7] (i.e., there are no "edge" cells) and multiple mobiles are dropped randomly in each sector. We simulate simultaneous voice calls between randomly chosen mobiles in the network and each voice frame is simulated from source to destination. As an example, Fig. 3 shows the simulation components for a mobile-to-mobile call.

Voice traffic is simulated based on the Markov Service Option (MSO) model [7]. Every 20 ms interval a full, 1/2, 1/4, or 1/8 rate vocoder frame is generated. At the source terminal, 1/8 rate vocoder frames are blanked (i.e., not transmitted), except for the first 1/8 frame in each silence interval, which is transmitted so that the dejitter buffer will reliably detect silence intervals. The resultant distribution of frames is as follows: full rate is 29 percent, 1/2 rate is 4 percent, 1/4 rate is 7 percent, 1/8 rate is 6 percent, and blanked is 54 percent. Each vocoder frame is sent in an individual IP packet (no bundling of frames). We assume 4 bytes of packet overhead for RTP/UDP/IPv6 (assuming UDP checksum) as described above.

Delay component	Circuit–EVDO	EVDO–EVDO
Vocoder (algebraic processor)	35 ms	35 ms
Packet processing (turbo encode, demod, decode, MAC)	10 ms	20 ms
RAN (BTS-PDSN)	10 ms	20 ms
Core IP network or PSTN	15 ms	15 ms
Voice frame decoding (not including dejitter)	3 ms	3 ms
TOTAL	73 ms	93 ms

Table 1. *Constant delay components of the system.*

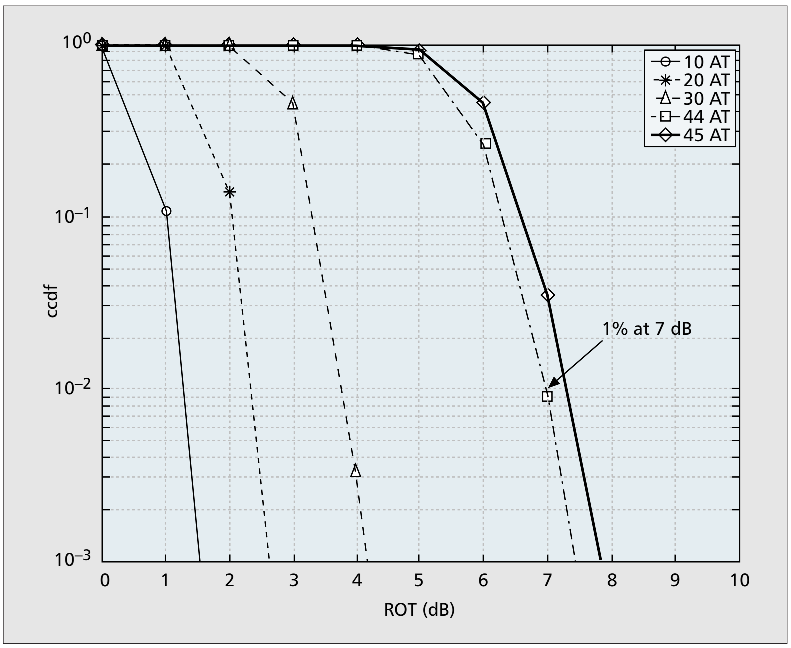

Figure 4. *Complementary cumulative distribution function of uplink rise over thermal noise for different loading levels.*

On the downlink, each IP packet is given a timestamp as it arrives to the FL transmission queue of the corresponding AT. The particular format chosen for transmission is determined by means of an access-network (i.e., base-station) scheduling algorithm. The decision is based on the information received from the individual AT on the uplink Data Rate Control (DRC) channel [2, 3], the queue states of individual users, and various other factors. If the FL queuing delay of an IP packet exceeds 200 ms, it is dropped from the queue. This value was chosen roughly to be equal to the difference between the mouth-to-ear delay requirement (e.g., 270 ms [5]) and the total constant delay term of the system (e.g., 73 ms, as detailed in Table 1). Note that if a particular packet experiences a relatively large amount of jitter on the FL or RL, the dejitter buffer may choose to drop (i.e., not play) that packet. Airlink transmission and packet decoding are modeled using short-term PER curves [7]. All control-channel overheads are modeled.[5] The medium access control (MAC) channel power allocation and performance are modeled dynamically according to channel variations.

[5] *The synchronous control channel is modeled at 76.8 kb/s.*

The AT uses a 12-slot latency target (i.e., transmit power is controlled such that 99 percent of the packets are successful within three subframes) [2, 3] and sends each IP packet in one MAC packet in the next available interlace when power headroom allows. Under the power headroom limitation, packets are allocated in sequential interlaces at the largest packet size the headroom allows. The link budget is such that all ATs can sustain the rate necessary for VoIP traffic. All uplink overhead channels are modeled dynamically in accordance with the channel variations, with standard gain settings.

The AT's channel prediction and rate-control algorithm is simulated based on the algorithm to be used in commercial terminals. A DRC erasure-mapping algorithm implemented at the access network (AN) enables uninterrupted service even when there are erasures on the DRC channel.

An adaptive dejitter buffer is simulated as described above and targets around 0.5 percent FER. Depending on the interarrival times of packets, some frames can potentially be expanded or compressed to maintain continuous playback. The rate of lost frames and mouth-to-ear delay of played packets are collected for each user during the simulation.

Table 1 displays the delay assumptions for different network components in our analysis. Unlike previous mechanisms (i.e., uplink, downlink, and dejitter buffer), these components are not simulated dynamically but modeled with constant delay. "Circuit-EVDO" column indicates the delay values for a call between a circuit phone (e.g., POTS landline phone) and a DO-Rev A phone. "EVDO-EVDO" column shows the values for a mobile-to-mobile call between two DO-Rev A terminals. Note that for the "EVDO-EVDO" case, some delay values are doubled since two airlinks are utilized in each direction of traffic flow. However, vocoder delay is not doubled since tandem vocoding is assumed to be eliminated. Changing these delay assumptions will simply shift the mouth-to-ear delay curves that we present in the next sections.

VoIP Capacity

We simulate the system with various number of simultaneously active VoIP users in a mobile-to-mobile call (i.e., EVDO-EVDO) as shown in

Fig. 3 and collect the performance metrics described above.

Figure 4 displays the complementary cumulative distribution function (ccdf) of uplink rise over thermal noise (RoT) for different number of users per sector. We observe that maximum number of users that cause the total sector load to be greater than 7 dB less than 1 percent (of the time) is 44 ATs per sector. Thus, based on the RoT metric, system capacity is 44 users.

From a delay perspective, as discussed above, we suggest conversational delay is an important metric. Our dejitter buffer implementation makes use of this fact and hence the mouth-to-ear delay of packets belonging to the beginning or end of a talkspurt is generally less than the rest. However, traditionally voice quality has been measured with one way mouth-to-ear delay of *all* packets. Thus, as a conservative approach, in this section we present mouth-to-ear delay of all packets.

Figure 5 displays the cumulative distribution function (cdf) of per-packet mouth-to-ear delay for different amounts of system loading. The mouth-to-ear delay is observed to have a distribution since some ATs have better channel conditions than others. Moreover, for a given AT, the mouth-to-ear delay can also vary slowly during the call, as described above. Even at peak capacity of 44 ATs per sector, the mouth-to-ear delay is observed to be less than 270 ms. This value is within ITU recommendations [5] for a call to be classified as satisfactory. Thus, we suggest with sector loading of 44 users, mouth-to-ear delay performance will be satisfactory.

As the last metric, we analyze per user FER as displayed in Fig. 6. A distribution among users is seen due to variations in the channel conditions. At a peak loading of 44 ATs per sector, the median FER is observed to be below 1 percent and almost all users have an FER of less than 2 percent. We believe these values are comparable or better than current cellular voice technologies for a mobile-to-mobile call.[6] In order to study other effects such as the correlation of frame erasures, we ran additional end-to-end simulations where we used real audio samples as voice sources generated by the EVRC codec. Our listening tests with actual FER events from simulations showed negligible voice-quality degradation.

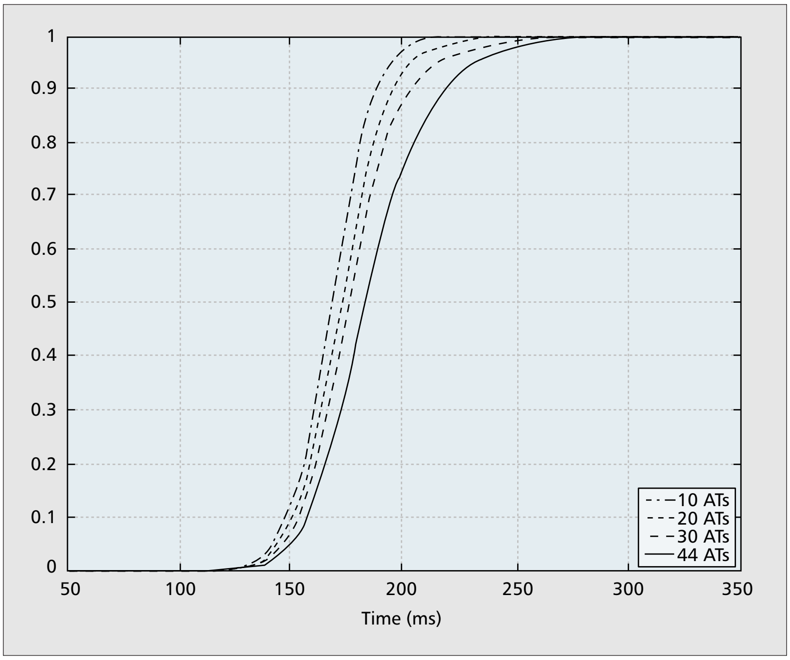

Figure 5. *Cumulative distribution function of per packet mouth-to-ear delay for different loading levels.*

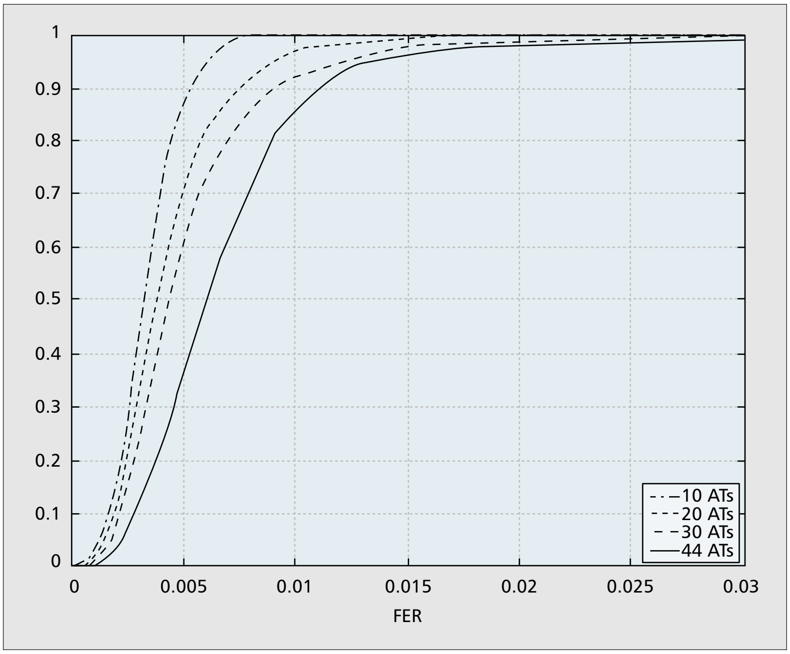

Figure 6. *Cumulative distribution function of per user FER.*

VoIP Capacity with Enhanced Features

The pilot signal transmitted from an AT is important for power control, channel estimation, and demodulation of that AT, but it is interference to others. For applications like VoIP where a large number of ATs transmit at low data rates, a significant percentage of total received power at the base station is from pilot signals transmitted by the ATs. With pilot interference cancellation (PIC) the reverse link capacity can be increased to support more VoIP users. As part of DO-Rev A systems, a PIC technique exists that can reduce significant amount of the overall pilot power arriving at the base station [10].

Another method to improve the RoT limited capacity is to utilize four-way receive diversity on the uplink. We have performed additional simulations to assess performance improvements via PIC and four-way receive diversity.

Table 2 summarizes all the metrics together with maximum number of supportable ATs for different configurations. For the mouth-to-ear delay and FER per user, we present the median and 95 percentile values. Also, a 1 percent RoT tail value and a number of used MAC channels are shown. As shown in Table 2, the default system capacity is 44 users (RoT limited). With PIC, the system is still RoT limited[7] and system capacity goes up to 50 ATs per sector. For the four-way receive diversity case, the system becomes limited by the available MAC channels at 66 ATs per sector.[8] However, note that MAC channel limitation is not likely to happen in a network with other type of traffic flows (e.g.,

[6] *Typical assumption for CDMA circuit voice call is 1 to 2 percent FER per link (e.g., a mobile-to-mobile call can experience 2–4 percent end-to-end FER).*

	Max number of ATs per sector	Mouth-to-ear delay		Frame erasure rate per user		RoT tall @ 1%	MAC channels user per sector
		50th percentile	95th percentile	50th percentile	95th percentile		
Default (EVDO-EVDO)	44	182 ms	231 ms	0.60%	1.32%	7 dB	76
With RL PIC (EVDO-EVDO)	50	187 ms	231 ms	0.69%	1.40%	7 dB	86
With 4x RL Rcv Div. (EVDO-EVDO)	66	209 ms	268 ms	0.85%	1.50%	4.0 dB	114
Note: Capacity limiting factors are underlined.							

Table 2. *Summary of capacity results for EVDO-to-EVDO calls.*

BE), which are elastic in nature or at a higher data rate than VoIP. As the capacity increases due to uplink enhancements, delay and FER are observed to degrade slightly due to downlink loading but still stay within allowable range.

Note that all the capacity results are for EVDO-to-EVDO calls. For a landline circuit to EVDO call, the delay and FER values would be even less, since there is only one airlink involved in each direction.

Mixed Traffic Capacity

In this section, we consider the DO-RevA system with other types of traffic flows in addition to VoIP. First, we consider best effort (BE) traffic (e.g., full queue). We quantify the amount of BE throughput that can be achieved when the system is also serving VoIP users. We perform end-to-end simulations where we fix the number of BE users to four (with a full-queue model) and vary the number of VoIP users. Note that we make similar assumptions for BE terminals as VoIP terminals (e.g., a dual receive antenna with 50 percent signal correlation, etc.).

We first consider the uplink performance. Figure 7 shows the normalized sector throughput of BE users on the uplink for a two-way receive antenna (i.e., default configuration) and a four-way receive antenna. We observe that as the number of VoIP users increase, the RLMAC algorithm allocates less resources to the elastic BE traffic [3], resulting in a roughly linear reduction of BE throughput. For all the cases displayed in Fig. 7, the reverse-link RoT stability metric was satisfied. The results show significant improvements in uplink BE throughput with four-way receive diversity.

The downlink sector throughput of BE users is shown in Fig. 8. The scheduler gives higher priority to VoIP users due to their delay-sensitive nature and still ensures that the QoS requirements of VoIP users are met.[9] Thus, as the number of VoIP users increase, BE capacity is reduced at a higher rate than the aggregate traffic of VoIP users. However, compared to uplink results, there is still significant BE capacity available, even for a high number of VoIP users.

The excess data capacity on the downlink suggests that other applications that are downlink intensive can be carried simultaneously with VoIP traffic. As an example, one can allocate some of the downlink time slots to the DO-Rev A Platinum Broadcast (PB) via Enhanced Broadcast-Multicast (EBM) channel [11]. The EBM traffic channel supports transmissions at a data rate of 1.8432 Mb/s across coverage. We have simulated cases with 230 and 403 kb/s broadcast channels by allocating a fraction of FL slots to EBM. The results are displayed in Fig. 8. For all simulation cases, QoS requirements of VoIP users were satisfied.[9] The results indicate that even with 403 kb/s broadcasting, significant BE capacity is available together with multiple VoIP users in the system.

Conclusion

DO-Rev A provides efficient support for both delay-sensitive and delay-tolerant applications. We have discussed how it can provide high-quality and robust VoIP services with full mobility. Novel techniques and concepts for voice processing, such as time-warping, adaptive dejitter buffer, and smart blanking, which improve QoS and capacity were presented. We analyzed the performance and capacity of VoIP individually and together with other applications. Detailed end-to-end simulations under the 3GPP2 simulation framework (with QoS and stability constraints) show that 44 to 66 simultaneous VoIP users per sector can be supported. We believe these results are comparable to circuit-switched cellular cdma technologies such as IS-2000 under similar simulation conditions [12]. Moreover, our simulations show that a significant amount of delay-tolerant traffic and downlink broadcasting can also be simultaneously supported along with VoIP.

Acknowledgments

The authors thank Mehmet Gurelli, Jean Au, Donna Ghosh, Jilei Hou, and John Smee for their contributions to the simulations used in this study and also thank the reviewers for their constructive comments that improved this article.

References

[1] O. Awoniyi and F. D. Tobagi, "Effect of Fading on the Performance of VOIP in IEEE 802.11 a WLANs," *IEEE ICC*, June 2004.
[2] 3GPP2, "cdma2000 High Rate Packet Data Air Interface Specification," C.S20024-A v. 1.0, Mar. 2004.
[3] P. J. Black *et al.*, "cdma2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview," *IEEE Commun. Mag.*, this issue.
[4] J. Janssen *et al.*, "Assessing Voice Quality in Packet-Based Telephony," *IEEE Internet Comp.*, vol. 6, no. 3, 2002, pp. 48–56.

[7] *With PIC, the system can operate based on the effective ROT as measured by the total power level after the pilot interference is cancelled. This better reflects the true interference plus noise level experienced by each user's traffic channel. Thus, the RoT values presented in Table 2 for PIC correspond to effective RoT.*

[8] *Sixty-six users per sector utilize all the available MAC channels (114) on the downlink due to average an active set value of 1.72 in the simulations.*

[9] *We analyze delay and FER statistics of VoIP users using the techniques described previously. The results are not displayed due to space considerations. For all the simulation cases (up to 40 users), these statistics were observed to be within allowable ranges (e.g., less than 270 ms).*

[5] ITU-T Rec. G.114, "One-Way Transmission Time," Feb. 1996.
[6] TIA/EIA/IS-127, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems," Jan. 1997
[7] 3GPP2 TSG-C WG3, "1xEV-DO Evaluation Methodology," 3GPP2 TSG-C Contrib. C30-20031002-004, Oct. 2004.
[8] RFC 3095, "Robust Header Compression (ROHC): Framework and Four Profiles: RTP, UDP, ESP, and Uncompressed," http://www.faqs.org/rfcs/rfc3095.html
[9] 3GPP2 C.S0063-0 v1.15, "cdma2000 High Rate Packet Data Supplemental Services," Oct. 2005.
[10] J. Hou *et al.*, "Implementing Interference Cancellation to Increase the EV-DO RevA Reverse Link Capacity," *IEEE Commun. Mag.*, this issue.
[11] H. Xu *et al.*, "Updated Harmonized Enhanced Broadcast Multicast System Proposal for 1xEV-DO," 3GPP2 Contrib. C30-20050418-039R1, Apr. 18, 2005.
[12] S. Sarkar, "Reverse Link Capacity for cdma2000," *VTC 2001 Spring*, vol. 4, 6–9 May 2001, pp. 2397–2401.

Biographies

MEHMET YAVUZ (myavuz@qualcomm.com) received a B.Sc. degree in electrical engineering from Middle East Technical University, Turkey, in 1994, and M.S. and Ph.D. degrees in electrical engineering: systems from the University of Michigan, Ann Arbor in 1996 and 1999, respectively. Currently he is with Corporate Research and Development of Qualcomm Inc., San Diego, California, where he works on system design and performance analysis of 1xEV-DO systems with a focus on VoIP. His interests include algorithms, protocols, and performance analysis of wireless networks.

SERAFIN DIAZ received his B.S. in electronic systems engineering from the ITESM, Monterrey, Mexico, in 1989 and his M.S. in electrical engineering from Southern Methodist University, Dallas, Texas, in 1994. He joined Qualcomm in 1997 and has worked at the Corporate R&D division since 1999. For the last two years, his area of work is focused on speech quality enhancements for VoIP systems over wireless networks.

ROHIT KAPOOR received his B.E. in computer science from the Indian Institute of Technology, Roorkee, India in 1999 and his M.S. and Ph.D. from the University of California, Los Angeles in 2001 and 2003, respectively. His doctoral research focused on supporting multimedia applications over the wired and wireless Internet. Since 2003 he has been working at the Corporate R&D division at Qualcomm.

MATTHEW S. GROB is senior vice president of technology for Corporate Research and Development of QUALCOMM Incorporated. He joined QUALCOMM in August 1991. He currently works on the EVDO VoIP program. His previous contributions to QUALCOMM include direct involvement in CDMA data standards IS-99, IS-657, and IS-707, the Globalstar system, as well as being awarded several patents. He holds a B.S.E.E. from Bradley University, Peoria, Illinois, as well as an M.S.E.E. from Stanford University.

PETER J. BLACK is senior vice president of technology for corporate research and development of QUALCOMM Incorporated. He joined QUALCOMM in April 1993, where he was first engaged in the system design and development of dual mode CDMA/AMPS mobile station ASICs. In 1997 he co-led the system design and prototype development of a high-speed cellular packet data system known as HDR. This system design was the framework for the cdma2000 high-rate packet data standard more commonly known as 1xEV-DO, published in 2000. He also co-led the subsequent commercialization of 1xEVDO, which has now achieved large-scale deployments in all major markets. Since 2001 he has continued to contribute to the evolution and enhancements of the EVDO standard and products. The most recent initiatives include hybrid OFDM broadcast, VOIP, and multi-carrier EVDO. He received his B.E. degree in electrical engineering from the University of Queensland, Australia, in 1985. He received his M.S.E.E. and Ph.D. degrees from Stanford University, California, in 1990 and 1993, respectively. He is a Fulbright Scholar and was awarded the University Medal by the University of Queensland in 1985.

YELIZ TOKGOZ received a B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 1998, an M.S. degree in electrical engineering from The Ohio State University, Columbus, in 2000, and a Ph.D. degree in electrical and computer engineering from the University of California at San Diego, La Jolla, in 2004. She is currently working as a system engineer at Qualcomm Inc., San Diego, California. Her research interests include multiple-antenna systems, spread spectrum communication systems, communication theory, and statistical signal processing.

CHRISTOPHER LOTT received a Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor in 2001. He is currently with the Corporate R&D Department at Qualcomm Inc., San Diego, where he is a systems designer for 1xEV-DO with a focus on MAC design. His interests include stochastic systems, resource allocation, distributed algorithms, communication theory, dynamical systems, and wireless networks.

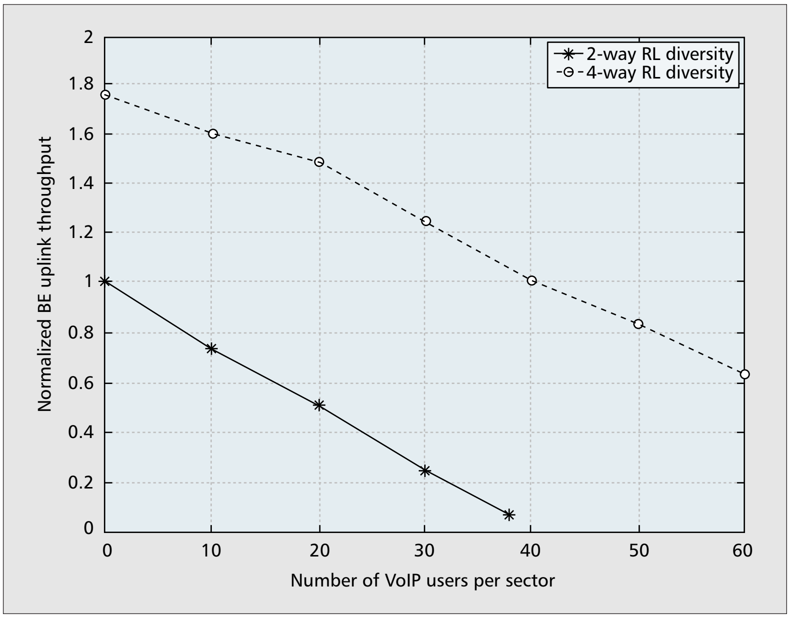

Figure 7. *Normalized sector throughput for BE users on uplink as a function of number of VoIP users in the system.*

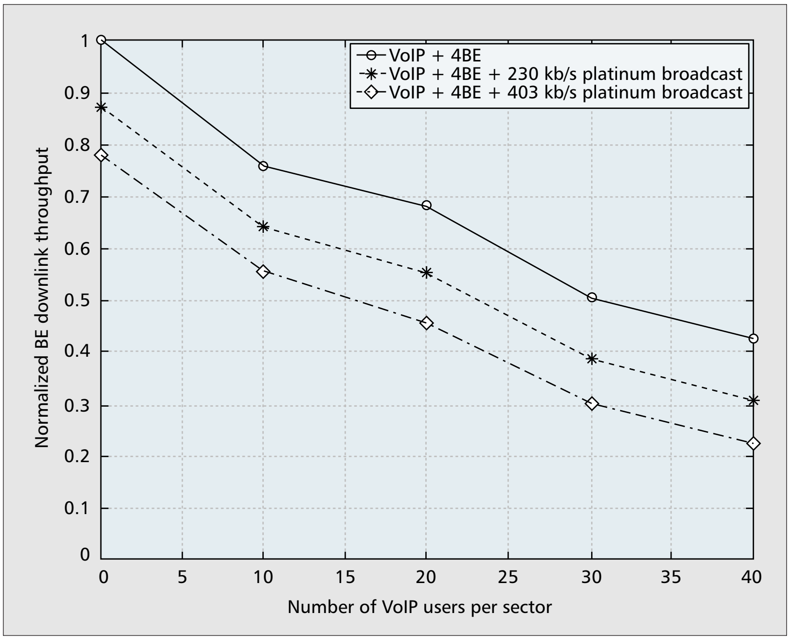

Figure 8. *Normalized sector throughput for BE users on downlink as a function of number of VoIP users in the system.*