D-RMA: A Dynamic Reservation Multiple-Access Protocol for Third Generation Cellular Systems

Antonio Iera and Salvatore Marano

Abstract—In this correspondence, a dynamic reservation multiple access (D-RMA) protocol for third generation cellular mobile radio systems is proposed and its behavior is investigated under variable multimedia traffic conditions. D-RMA is a protocol which is explicitly designed to support multimedia traffic. Its structure is based on traditional PRMA protocol, but in addition to what PRMA provides, it introduces a flexible dynamic approach in the choice of the percentage of bandwidth to be used for reservation. Separation between reservation and information channels, along with dynamic adaptation of the percentage of reservation bandwidth within a frame to traffic condition, guarantee the required QoS to multimedia services. Results obtained show that the performance of D-RMA are superior when compared to a traditional “nondynamic” protocol, in terms of both the offered quality of service (QoS) and number of connections which can be activated in a microcell at one time.

Index Terms—Multiple access protocol, personal communications, quality of service (QoS) to multimedia services, third generation cellular systems.

I. INTRODUCTION

In the near future, wireless communication is expected to undergo enormous growth. In fact, Personal Communication Systems (PCS), based on wireless technologies, are evolving toward the support of a wider range of applications including voice, video, data, and multimedia [1]. The ideal future scenario in the field of telecommunications will be characterized by several Mobile Audio Visual Terminals by which users, via a base station connecting wired and wireless networks, have access to multimedia services scattered all over a high-speed communications backbone [2].

To this aim, many researchers have focused their attention on the integration between enhanced wireless systems and broadband networks, such as B-ISDN using ATM protocol. This enables the resulting wireless-PCS system to embrace a wide range of communication capabilities, besides those offered by current analog cellular mobile radio systems (substantially voice services).

The above-mentioned integration process requires the design of multiple access protocol showing high flexibility and efficiency to support future integrated services.

The protocol, called dynamic reservation multiple access (D-RMA), which is presented in this paper well meets the wireless multimedia application requirements by coupling the advantage deriving from a complete separation between Reservation and Information channels with a dynamic approach in the choice of the amount of reservation bandwidth.

II. MULTIPLE ACCESS PROTOCOLS

During the first phase of migration toward the idea of wireless PCS systems, multiple access protocols, initially proposed for satellite systems or wireless LAN [3]–[5] were reconsidered for cellular mobile radio systems. An interesting output of this activity is represented by the well-known packet reservation multiple access (PRMA) [6], [7].

In PRMA, the frame rate coincides with the arrival rate of voice packets so that a user (voice terminal) requires exactly one slot per frame. Within the frame, the terminal detects an available or reserved slot according to the feedback information stream broadcasted on the down-link from the base station.

An active terminal, which has packets to transmit, contends in order to access the channel (as in S-ALOHA) on the available slots. Following a successful attempt, there the transmission begins and the terminal holds a reservation for a slot in subsequent frames until the end of its talkspurt.

Every up-link slot, according to PRMA protocol, can be used for reservation, thus an acknowledgment at the end of each slot is required on the down-link. With PRMA, an increase in the traffic load causes a decrease in the probability of finding free slots in the frame and an uncontrolled increase in the access delay.

An adaptive protocol, named multiple access (MA) protocol, which attempts to solve the above mentioned problem, has been carried out by Mitrou et al. [8]. This protocol allows a reservation only on the so-called R-slots, and the transport of information on the I-slots.

In the MA protocol, a minimum number of R-slots is fixed to guarantee good performance under high loading conditions. When the traffic load is low, every other free I-slot can be used for reservation.

The presence of a partial separation between control and information channels is the main inconvenience of the MA protocol. In fact, while the reservation packets, transmitted over the R-slots, are strongly protected against interference; conversely this protection is absent on I-slots and the reservation packets can be corrupted by channel impairments.

Recently, a new version of PRMA, named PRMA++, has been proposed and analyzed [9]. In this protocol, the separation of control and information slots permits guarantees of different degrees of quality to control and information channels [10] and, unlike PRMA, a lower number of acknowledgment slots are required on the down-link.
In PRMA++, the number of R-slots as well as their positions within the frame are fixed. This has the effect of causing mobile terminals to experience access delay even at low load [9]. Moreover, in the presence of multiple slot assignment (i.e., multimedia traffic), the experienced access delay may become a critical factor.

III. THE D-RMA PROTOCOL

D-RMA can be regarded as belonging to the family of protocols, which dynamically adapt the traditional Reservation Multiple Access protocols to future multimedia traffic needs [15].

Its up-link frame format is constituted by a sequence of I-slot and R-slot, whose total number is set to \( N \). In order to assure the best performance to multimedia (integrated video/voice) services, the number of R-slots \( N_r \) varies dynamically according to the offered load. Thus, the condition \( (N_{r_{\text{min}}} \leq N_r \leq N_{r_{\text{max}}} < N) \) holds, where \( N_{r_{\text{min}}} (N_{r_{\text{max}}}) \) is the optimum number of the R-slots for the highest (lowest) load.

All “nonactivated” R-slots are used for transporting information, like I-slots. The \( N_{r_{\text{max}}} \) positions of R-slot within the frame are fixed. They are scattered in a homogeneous manner along the frame, so that new likely bursts do not have to wait a long time to intercept an available R-slot [8].

When a terminal has some packets to transmit (i.e., a start of a burst packet is detected), with probability \( p \) (called permission probability, as in the S-Aloha protocol), it sends a reservation packet on the first available R-slot.

If the reservation packet collides or is corrupted by channel impairments, a negative acknowledgment is transmitted from the base station on the down-link. Consequently, the contending terminals try to transmit their packets on the next free R-slot, again with probability \( p \).

On the contrary, if a success occurs, the base station sends an acknowledgment on the down-link and assigns the requested number of I-slots to the calling terminal. If that number is not available, the reservation packet is buffered in a reservation queue within the base station, and waits for I-slots to become available.

The described contention mechanism is similar to other reservation protocols. What really makes the D-RMA protocol different from previous proposals, is the dynamic management of the reservation bandwidth. In fact, in response to traffic load fluctuations, due to either originating/terminating calls or hand-off events, the base station is allowed to modify the number of active R-slots within the transmission frame.

Better performances, compared to a “fixed \( N_r \)” policy, can be obtained by correctly varying the number of active R-slots (\( N_r \)). It will be shown that a reduction in \( N_r \) is preferred when the load increases. On the contrary, the use of more R-slots helps in dealing with a low offered traffic load. The optimum number of active R-slots corresponding to each given load is computed according to an ad-hoc defined QoS metric. A block diagram of D-RMA is given in Fig. 1.

At the end of each frame, a Reservation Bandwidth Controller examines the channel load by controlling the number of terminals within the microcell which have a connection established and determines a suitable \( N_r \) value. Bandwidth requests which cannot be served are buffered and handled according to a specific rule (FIFO, Scan and Serve [11]) implemented by the Access Controller. The latter also allocates the requested Basic Bandwidth Units (BBU) to each terminal according to suitable strategies, such as Complete Sharing (CS), Complete Partitioning (CP), or Mutually Restricted Access (MRA) [11]–[13].

An efficient and simple technique to manage the variation of \( N_r \) by tracking traffic load fluctuations, is a threshold-based policy. Specifically, load-to-threshold mapping is available as a look-up table whose figures are a priori evaluated, according to the quality index \( Q \) defined in Section IV.

When many variations of traffic load around a threshold value occur, there is the persistent need for a variation of active R-slot number in subsequent frames. This phenomenon would provoke an undesired overload of signaling traffic on the down-link. In D-RMA, in order to overcome this problem, each threshold
value is provided with an hysteresis margin, which permits a change in $N_r$ only when the offered load shows an actual rising or falling trend.

IV. SIMULATION STUDY

A flexible discrete event simulator has been implemented and used to study the D-RMA protocol behavior. The achieved performance has been analyzed and compared by evaluating the following parameters:

- $D_{H_{vo}}$: (the voice holding time) the time spent from the instant in which a talkspurt is generated until its first packet is transmitted on the channel (an upper bound of 32 ms has been defined in [7]). $D_{H_{vo}}$ is actually given by $D_{vo} + D_{PHOC_{vo}}$, where the first term is the access delay (i.e., the time required to gain a channel reservation from the Base Station), and the second term is the time required to process the data before the transmission over the air interface.
- $D_{H_{vi}}$: (the video holding time) the analogous definition of $D_{H_{vo}}$ for video (an upper bound of 100 ms has been defined in [1] for time-critical traffic, like video);
- $P_{drop_{vo}}$: the average packet dropping probability for a voice terminal, computed as $\text{dropped packets} / \text{arrived packets}$ (an upper bound of 0.01 has been defined in [7]);
- $P_{drop_{vi}}$: the average packet dropping probability for a video terminal (an upper bound of $10^{-3}$ has been defined in [1]).

Furthermore, the following quality index is defined:

$$Q = \alpha \left( \frac{1 - P_{drop_{vo}}}{T_{can} + D_{vo}} \right) + (1 - \alpha) \left( \frac{1 - P_{drop_{vi}}}{T_{can} + D_{vi}} \right)$$

where $\alpha$ is the percentage of voice terminals over the total number of transmitting terminals.

For a given load, the following simple criterion is used to evaluate the optimum number of R-slots which must be activated:

“optimum $N_r$ is the one which maximizes $Q$."

Both the access delays and the packet dropping probabilities of the two examined types of traffic are considered to define the quality index $Q$.

The maximum admissible value of $Q$ (i.e., $Q = 1$) is obtained in ideal conditions when both packet dropping probabilities and access delays are zero. Performance index $Q$ considers even small additional access delays undesirable. The rationale for this can be easily understood by focusing on the concept of holding time, which takes into account the time spent within the system to cross the whole transmitting chain (contention, queuing, channel coding, interleaving, modulation, etc.). For some services the time required for data processing can be high, this reducing the value of the allowable access delay. For this reason it becomes very important to minimize $D_{vo}$ ($D_{vi}$) as much as possible to leave more time available for performing other transport layer functions. The greater the margin, the more effective the techniques can be (higher interleaving depth for example).

The speech traffic model used during the simulation tests is the Brady model [14] which considers a voice source as a sequence of talkspurts and gaps.

All talkspurts and silence periods have exponentially distributed duration with mean $T_{on}$ and $T_{off}$, respectively. If a packet is queued into a terminal buffer for a time longer than $D_{max_{can}}$, it is dropped.

Video or data terminals are modeled as bursty sources, which generate frames of variable length, exponentially distributed with mean $L$ Kb [1]. The arrival rate of a new burst is assumed equal to $\lambda_{vi}$.

Since a video service can be considered a “time critical service,” it is also assumed that video packets are dropped if their waiting time within the terminal queue exceeds the maximum delay value $D_{max_{vi}}$.

A voice source generates data at the rate of 8 Kb/s thus requiring one BBU per frame. On the contrary, considered video source generation rates are 32 Kb/s and 64 Kb/s, which require four and eight BBU’s, respectively.

Reported simulations have been run under the conditions listed in Table I.

First, we report the D-RMA performance results when in the presence of voice-only traffic; subsequently the protocol behavior analysis is extended to the case of heterogeneous traffic.

### Table I: List of Nominal Values Utilized in Simulations

<table>
<thead>
<tr>
<th>Definition</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel bit rate (Kb/s)</td>
<td>$R_c$</td>
<td>500</td>
</tr>
<tr>
<td>Frame length (ms)</td>
<td>$F$</td>
<td>6</td>
</tr>
<tr>
<td>Number of slot within frame</td>
<td>$N$</td>
<td>50</td>
</tr>
<tr>
<td>Num. of active R-slot</td>
<td>$N_r$</td>
<td>variable</td>
</tr>
<tr>
<td>Acknowledgment delay [ms]</td>
<td>$D_a$</td>
<td>1</td>
</tr>
<tr>
<td>Voice stream rate (Kb/s)</td>
<td>$R_{vo}$</td>
<td>8</td>
</tr>
<tr>
<td>Low res. video rate (Kb/s)</td>
<td>$R_{vi}$</td>
<td>32,64</td>
</tr>
<tr>
<td>new message arrival rate (meg/s/user)</td>
<td>$z_m$</td>
<td>1</td>
</tr>
<tr>
<td>Average message length (Kb)</td>
<td>$L$</td>
<td>5.12, 10.24</td>
</tr>
<tr>
<td>Average talkspurt duration (s)</td>
<td>$T_{on}$</td>
<td>1.41</td>
</tr>
<tr>
<td>Average silent period (s)</td>
<td>$T_{off}$</td>
<td>1.78</td>
</tr>
<tr>
<td>Offered channel load</td>
<td>$\rho$</td>
<td>variable</td>
</tr>
</tbody>
</table>


A. Voice Traffic Performance

Fig. 2 shows that the optimum number of active R-slots has a linearly decreasing relationship with the channel load, and consequently with the number of active terminals. This phenomenon can be explained as follows: in Reservation Multiple Access protocols two contributions affect access delay, the first is due to the access contention process (S-ALOHA), while the second is due to the waiting time of the reservation packet within the base station reservation queue

\[ D_{vo} = D_{con} + D_{que}, \quad \text{where} \quad D_{con} = \text{contention delay}, \]
\[ D_{que} = \text{queue delay}. \]

When the traffic is low, the access contention delay dominates. In such a situation, the presence in the frame of a greater number of available R-slots contributes to reduce the access delay. Conversely, under high load conditions, the delay accumulated within the reservation queue dominates. Thus, it is better to reduce the number of active R-slots in order to increase the information channel bandwidth.

Fig. 3 (Fig. 4) shows the access delay (packets dropping probability) as a function of the offered channel load \( \rho \) (defined as average number of active terminals in a frame \( N \)) for both D-RMA and a protocol which exploits a fixed number of R-slots.

The lower the access delay and the packet dropping probability, the higher the number of terminals that can be simultaneously activated.

To better understand the improvement deriving from the adoption of D-RMA instead of a “\( N_r \) fixed” solution, the reader can refer to Fig. 4 and focus on the worst case of \( P_{drop} = 10E^{-2} \). If \( P_{drop} \) has to be kept below this upper bound, then the maximum value of admissible load when
D-RMA is adopted is about 0.89. In the same condition, the values obtainable when \( N_r = 6 \) and \( N_r = 8 \) are adopted are 0.84 and 0.80, respectively. Thus, the achieved gains are about 5% and 10% (even greater margins are obtainable under different, not shown, simulated network conditions).

These percentages refer to one carrier only. As each microcell contains a set of carriers, a considerable increment in the number of simultaneously active terminals can be obtained.

B. Multimedia Traffic Performance

The following curves refer to test cases in which nominal conditions (see Table 1) and the presence of heterogeneous traffic (75% voice and 25% video [1]) are assumed. Fig. 5 (Fig. 6) show the voice access delay (the video packet dropping probability) as a function of the offered channel load \( \rho \) for both D-RMA and a “\( N_r \) fixed” protocol.

If we focus on Fig. 6, for \( P_{drop} = 10E^{-3} \) (upper bound value for the considered broadband traffic) we notice a gain in the acceptable load which is about 6.7% when compared to \( N_r = 6 \), about 11–12% when compared to \( N_r = 8 \), and about 17% when compared to \( N_r = 10 \).

These are greater margins than those achievable in the voice-only traffic case and they prove that D-RMA is particularly adequate for supporting multimedia application traffic in a heterogeneous mobile environment.

On the other hand, curves in Fig. 5 show that D-RMA can always guarantee a lower voice access delay when compared to fixed-\( N_r \) algorithms. D-RMA achieves a greater delay performance if it is compared to algorithms exploiting a lower \( N_r \).

Fig. 7 shows the improvement in the quality index \( Q \) when utilizing an “optimum \( N_r \) based” instead of a “fixed \( N_r \) based” strategy. An increase in the maximum number of simultaneously activated terminals with respect to the “\( N_r \) fixed” algorithm is always present.

Fig. 7 also reveals good \( Q \) measure performance at low loading, irrespective of the value of \( N_r \). This could be misleading if not properly explained. To better understand the
effectiveness of the D-RMA policy it must be noted that, at low loading, the adoption of a constant small $N_r$ value would correspond to an increase in the voice access delay (see Figs. 3 and 5). In the sample case of Fig. 5, this would mean that a gain of about 10 ms is achievable by substituting the $N_r = 4$ technique with D-RMA. This amount of time can be differently exploited to perform a deeper interleaving, during the transport on the air interface, while remaining within the same overall allowed transmission delay. The advantage, which cannot be evaluated through the Q index, consists in a greater robustness of the successfully transmitted information against channel errors.

A detailed test campaign has been also performed by coupling D-RMA with different bandwidth allocation strategies (Complete Sharing, Complete Partitioning, and Mutually Restricted Access) and different reservation buffer management schemes (FIFO, Scan&Serve).

Performance curves are skipped for length constraints. Nevertheless, the obtained results can be summarized by stating that the adoption of both the MRA policy and the Scan and Serve policy, allow the achievement of the best performance when compared to simple CS strategy or MRA policy with FIFO queue management.

V. CONCLUSION

The adoption of radio channel access schemes, efficient in supporting multimedia services, is a key issue in the design of future third generation cellular systems.

A dynamic reservation multiple-access protocol, called D-RMA, has been proposed, which both permits wireless networks to embrace a wider range of capabilities and contributes to the research work toward the definition of the Personal Communication Systems. Its behavior has been observed under both homogeneous and heterogeneous traffic conditions. In order to compare D-RMA behavior with that of other traditional access protocols, a new quality index has been defined. Better performance resulted from the adoption of the dynamic approach when compared to a “$N_r$ fixed” policy.

REFERENCES


**Antonio Iera**, photograph and biography not available at the time of publication.

**Salvatore Marano**, photograph and biography not available at the time of publication.