An accurate heuristic approach for UMTS RLC delay statistics evaluation

Michele Rossi, Michele Zorzi
{mrossi, mzorzi}@ing.unife.it
DIF, University of Ferrara, via Saragat, 44100 Ferrara, Italy

***Abstract*—This work is a preliminary investigation on the effect of a controlled amount of retransmission on the residual higher layer packet error probability in a UMTS system. In this paper, we first characterize the delivery derlay statistics regarding link layer Service Data Units (SDUs) considering in order delivery. After this we propose a very simple heuristic able to approximate such statistics with great accuracy. This heuristic is well suited to being used in energy limited mobile terminals. In the final part of the paper, we investigate how this heuristic can be effectively used at the mobile terminal to dynamically choose the right amount of link layer retransmissions while respecting the delay constraints imposed on the higher layer flows. We quantitatively evaluate the maximum achievable gain when the optimum number of retransmissions is selected.**

I. INTRODUCTION

We focus on a third generation wireless system where WCDMA is used on the air interface as the channel access technique. The configuration of the WCDMA protocol stack in order to maximize the quality perceived by the final user still remains a key research topic. The main aim of this work is to quantify the effect of the retransmission persistence at the UMTS link layer (RLC) when delay constrained flows are considered at higher layers. In particular, the real-time constraints require that a packet containing real-time data that is generated at time t_s by the transmitter host (placed, for instance somewhere in the Internet) be delivered to the receiver no later than at time t_s+d_{max}, where d_{max} is the maximum acceptable overall *end-to-end* delay. In this paper we focus on how to correctly configure the link layer of a WCDMA wireless interface in order to enhance the quality of real-time flows. In recent years many contributes studied hybrid ARQ [1] solutions in which the code strength is varied as a function of the error process. In this way, in error prone channels, a short delivery time is obtained at the cost of an increased coding redundancy. Recent reviews of the techniques used to improve the quality of real-time flows in wireless channels are reported in [2] and [3].

We consider UMTS as an example of a third generation system using the WCDMA technique. Moreover, we investigate how to configure the UMTS link layer (RLC) [4] in order to reduce, as much as possible, the residual packet error probability at higher levels (P_{res}), without exceeding delay requirements and without requiring any change in the link layer specification (i.e., without introducing additional coding at the link layer, as needed for example by hybrid ARQ techniques). In more detail, we consider a Mobile User (MH) on a dedicated channel (DCH) that is receiving a real-time flow from its serving Base Station (BS). We also consider that this data flow is originated by a fixed host (FH), placed somewhere in the Internet. We define $d_t = I_d + W_d$ as the end-to-end delay experienced by data packets belonging to that flow, where we refer with the terms I_d and W_d to the delays due to the FH $\rightarrow$ BS path and to the BS $\rightarrow$ MH wireless path, respectively.

When delay constraints are considered, rather than the residual SDU error probability (P_{pkt}), a better estimate of the SDU error rate can be obtained by $P_{res} = P_{pkt} + (1 - P_{pkt})\text{Prob}\{d_t > d_{max}\}$ where $\text{Prob}\{d_t > d_{max}\}$ is the probability that a given SDU experiences an end-to-end delay greater than d_{max}. This can be rewritten in a more convenient way by replacing the term $\text{Prob}\{d_t > d_{max}\}$ with $\text{Prob}\{W_d > d_{max} - I_d\}$, where $d_{max} - I_d$ is the maximum tolerable SDU delivery delay on the wireless link. In other words, it represents the delay constraint that must be imposed on the wireless path. In the following to evaluate the SDU error probability we consider the formula

$$P_{res} = P_{pkt} + (1 - P_{pkt})\text{Prob}\{W_d > W_d(max)\} \quad (1)$$

Let L be the maximum number of allowed retransmissions for each link layer packet. The UMTS RLC, when delay constraints are considered on the higher level flow, may be configured in Unacknowledged Mode (UM) [4], i.e., by not performing retransmissions when PDUs are detected in error ($L = 0$). However, observing Eq. (1) and that P_{pkt} is a decreasing function of L, it is clear that $L = 0$ is not necessarily the correct setting under all possible channel conditions. In particular, depending on $W_d(max) = d_{max} - I_d$ and on the behavior of $\text{Prob}\{W_d > W_d(max)\}$, there could be time for a limited number of retransmissions ($L > 0$). So, instead of setting $L = 0$, we could set L to a value greater than zero but still meeting the delay constraints imposed on the SDU flow. Therefore, for a given PDU error process and for a given $W_d(max)$, we find the optimum value of L (L_{opt}) by minimizing Eq. (1) with respect to L and considering I_d to be stationary with respect to the delivery time of a full SDU[1].

Note that the main drawback of exact analyses for the computation of delay statistics proposed so far in the literature [6][7][8][9] is that they are computationally too complex to be effectively used in a mobile terminal, i.e., their computation would be both time and energy consuming. Moreover, if the time spent for the computation of such statistics is large with respect to the stationarity period of the PDU error process, then they are useless. For this reason, in addition to energy requirements we have computational delay requirements. Therefore, obtaining low-complexity and fast heuristics able to accurately approximate such delay statistics is a key issue to enable the statistics estimate directly at the MH and to allow a real-time adaptive reconfiguration of link layer parameters. In this paper we propose a simple and accurate heuristic to approximate the delivery delay statistics and we quantify the gain achievable by an optimal selection of the maximum number of link layer retransmissions.

This work has been partially supported by ERICSSON Research.

[1] This hypothesis is confirmed, for example by the measurements in [5].

The paper is structured as follows: in Section II we report the model used here to characterize the link layer retransmission algorithm, in Section III we motivate our channel model assumption by means of accurate simulation results. In Section IV we discuss some properties of the delivery delay statistics of RLC packets, whereas in Section V we use these results to derive heuristics for the characterization of such statistics. In Section VI we investigate the achievable gain by means of an adaptive choice of the number of retransmissions and finally in Section VII some conclusions are given.

II. ARQ MODEL

We consider a UMTS unidirectional downlink data transfer, where a Dedicated Channel is assigned to the user and the Radio Link Control (RLC [4]) is used to recover from residual channel errors after coding. The RLC is configured in Acknowledged Mode (AM), i.e., link layer SDUs have to be passed in order to higher layers at the receiving terminal. To model the UMTS RLC we refer to the following simplified model of the Selective Repeat algorithm[2]. The time is slotted and the slot time corresponds to the RLC packet (PDU) transmission time. Moreover, PDUs are transmitted continuously (Heavy Traffic assumption [6][7]) as can occur in FTP file transfer or streaming flow transmissions. Packets are transmitted using a Selective Repeat algorithm. Let m be the (integer) number of PDUs transmitted in a full RLC round trip time (RTT). We consider that an acknowledgment (ACK/NACK) message indicating the *status* of a given PDU transmitted in the generic slot t is available at the sender in slot $t + m$, so that in this slot a new PDU can be transmitted or the same PDU can be retransmitted depending on its delivery *status*. With no loss in generality, we assume ACK and NACK messages to be error-free. We consider in order delivery of RLC SDUs, i.e., a given SDU can be passed to higher layers at the receiver side only if all the SDUs with lower identifier have been correctly received. The RLC SDU is considered as a higher layer packet composed by K (integer number) link layer PDUs. We focus on the characterization of the SDU Complementary Cumulative delivery delay Distribution Function (*ccdf*$[K, t]$), i.e., the probability that a RLC SDU is delivered in order in more than t slots, where t is the time elapsed between the slot in which the first PDU composing that SDU is transmitted for the first time over the channel by the transmitter and the slot where the whole SDU can be passed in order to higher layers at the receiver. Moreover we refer to the single PDU delivery delay probability $P_d[t]$ as the probability that the time elapsed between the slot where a given PDU is transmitted for the first time over the channel and the slot where it can be passed in order to higher layers by the receiver is equal to t.

III. DISCUSSION ON THE CHANNEL MODEL

In this section we evaluate by means of simulation how the PDU error process at the UMTS RLC is characterized in terms of burst length (b) and average PDU error rate (p). Both channel coding (convolutional with rate $1/2$) and interleaving are considered. The simulated scenario is composed of 9 hexagonal cells, where a base station (node B) is placed at the center of each cell and a given number of users are considered to be able to move inside the coverage area. The whole cell structure is wrapped around in order to avoid border effects. In the results presented here, we consider a population of 300 users with speed uniformly distributed in $\{0, 6, 49\}$ Km/h. As an example scenario, we consider that each user is using a downlink dedicated channel (with minimum and maximum powers, $P_{min} = -15$ dBW and $P_{max} = -5$ dBW, respectively), where the traffic source is continuous as can occur when either streaming flows or FTP file transfers are considered. The cell radius is 200 m. Path Loss, Shadowing and multi-path phenomena are considered as well. We consider a logical bit rate (i.e., the useful bit rate available after coding) at the RLC of 30 Kbps and a large interleaving depth ($TTI = 80$ ms). In Fig. 1 we report the mean PDU burst length (b) as a function of the PDU error rate (p). Vertical error bars are reported to indicate the confidence interval (95 %) of mean burst measurements. As a first observation, one can note that at low RLC bit rates the error process tends to be independent (the independent case is reported for comparison in the figure as a bold line). Hence, the independent error assumption in this case seems to be a good approximation. In Fig. 2 we report the 80-th percentile of the same burst measurements. These results also confirm that the burst length is limited and that for most of the users in the system the independent PDU error assumption holds. For higher bit rates, we verified that the burst length is also limited. As the logical RLC bit rate increases, more PDUs can be sent in each radio frame and the PDU error process is likely affected by longer bursts. We verified by simulation that the 80-th percentile of the burst length, for instance, considering a bit rate of 120 Kbps for a PDU error probability of $p = 0.1$ and a TTI (i.e., interleaving depth) of 80 ms is bounded by 4. Moreover, noting that a typical value for the RLC RTT when $TTI = 80$ ms is 220 ms and considering a typical RLC PDU size of 360 bits, we have that $m \approx 75$ PDUs are sent in this case in one RLC round trip delay. In Fig. 3 we report the SDU *ccdf* as a function of m and b considering two-state Markov chain as the PDU error model. Observing this last figure it is clear that when $b \ll m$ the independent (*iid*) assumption leads to accurate statistics approximations. Moreover, one can note that a good approximation for the bursty case can be achieved by a rigid rotation of the *iid* curve. For these reasons, in this work we consider an independent packet error process. This model, despite its simplicity, in many cases gives a good approximation for the actual packet error process as seen at the link layer of a UMTS system.

IV. SOME STATISTICAL PROPERTIES OF THE DELIVERY DELAY DISTRIBUTION

In this section we investigate the main characteristics of the single PDU delivery distribution function $P_d[t]$. We first focus on the unlimited retransmission attempts case ($L = +\infty$), where a tagged PDU can be passed in order to higher layers when all PDUs with lower identifier have been correctly received. We subdivide the time in rounds of m slots. Let round 0 be the round where the tagged packet is transmitted for the first time and consider that it is transmitted in position m of that round. Moreover, when the tagged packet is transmitted for the first time (in slot m) the only PDUs that can block its delivery at the receiver side are the ones in position 1 through

[2]We verified by simulation that despite its simplicity the model proposed here is able to capture the main statistical properties of the Selective Repeat algorithm used in UMTS.

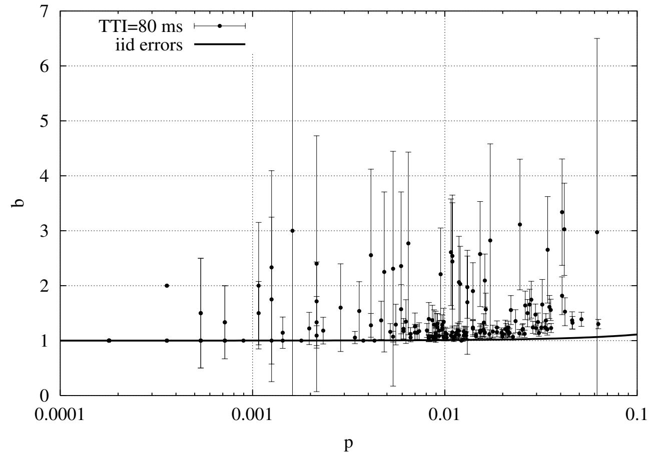

Fig. 1. Simulation points regarding mean PDU error rate p and burst lengths b for UMTS DCH downlink channels considering a RLC useful bitrate of 30 Kbps.

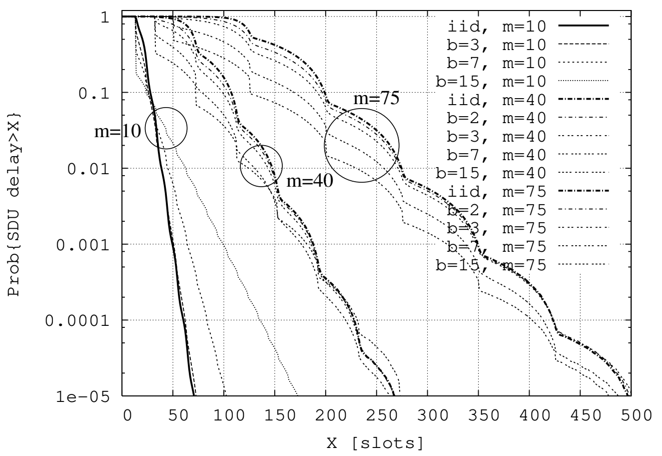

Fig. 3. Complementary cumulative delivery delay distribution by varying the round trip delay m and the burst length b for $K = 14$ and $p = 0.1$.

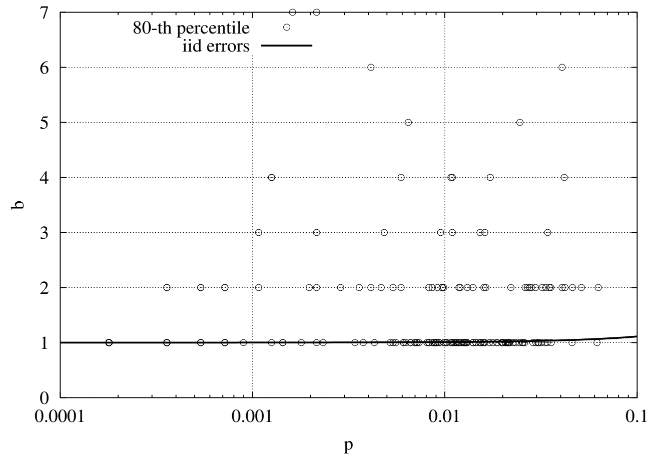

Fig. 2. 80-th percentile of the burst values measured for UMTS DCH downlink channels considering a RLC useful bitrate of 30 Kbps.

$m - 1$, i.e., they are the only ones with a lower identifier that have not been correctly delivered yet (see [6]). These packets will be referred to as *blocking packets* in the following. This follows by the fact that PDU identifiers are assigned to transmitted packets in increasing sequential order and that PDUs are transmitted round by round in the same window position until success. Let us express the delivery delay as $d = \xi m + \eta$, $\xi \geq 0$, $\eta \in \{0, 1, \ldots, m-1\}$, and let $\psi[k, 1]$ be the probability to have one erroneous PDU in a given (fixed) position $i \in \{1, 2, \ldots, m\}$ in round 0 and that k out of the remaining $m - 1$ PDUs are erroneously transmitted. The probability that an erroneous PDU in round 0 is transmitted correctly in a round up to and including round r is given by $P_c[r] = 1 - p^r$. The probability that the tagged packet is resolved in position m of round s, $s > 0$, i.e., at time $d_1 = sm$ is given by

$$P_d[d_1 = sm] = \sum_{k=0}^{m-1} \psi[k, 1] P_c[s]^k p^{s-1}(1 - p) \qquad (2)$$

where we compute the probability to have k blocking packets, that the tagged packet is transmitted in error in the slot of its first transmission (term $\psi[k, 1]$), that the k erroneous PDUs are resolved before slot d_1 (term $P_c[s]^k$) and that the tagged packet is resolved exactly in slot d_1 (term $p^{s-1}(1-p)$). The probability of resolving the tagged PDU at time $d_2 = sm + 1$, i.e., in the first slot of the following round, is

$$P_d[d_2 = sm + 1] = \sum_{k=0}^{m-1} \psi[k, 1] P_c[s]^k p^s (1 - p) \qquad (3)$$

In this case $\psi[k, 1]$ must be viewed as the probability that the PDU occupying position 1 in round 0 was in error and that exactly k packets out of the remaining $m-1$ are erroneous, $P_c[s]^k$ is the probability that PDUs in slot $\{2, 3, \ldots, m\}$ are transmitted correctly up to and including round s and $p^s(1 - p)$ represents the probability that the PDU transmitted in the first position of round 0 is resolved in slot d_2. Observing Eqs. (2) and (3), it is clear that the following relation holds

$$P_d[sm + 1] = P_d[sm]p \qquad s \geq 0 \qquad (4)$$

This is a general result that will be very useful in order to find very fast and accurate approximations of such statistics. Moreover, using again these equations, we can find the difference between the probability to resolve the tagged packet in the first slot and at the end of a given round ($s > 0$)

$$P_d[(s-1)m + 1] - P_d[sm] = \sum_{k=0}^{m-1} \psi[k, 1] p^{s-1}(1 - p)\Delta[s, k] \qquad (5)$$

where $\Delta[s, k] = P_c[s]^k - P_c[s-1]^k$. At this point, by rewriting $P_c[s]^k$ as $(1 - p^s)^k = \sum_{i=0}^{k} \binom{k}{i} (-p^s)^i$ it is easy to verify that $\Delta[s, k]$ tends to 0 as s increases. In practice this convergence is very fast and, as a consequence, $P_d[t]$, after a short transient phase, becomes almost constant inside each round. These considerations allow us to assert that the asymptotic behavior of the delivery delay statistics is described by the following equation

$$y[t] = p^{\frac{t - t_0}{m}} \qquad (6)$$

In more detail, Eq. (4) and the fact that, after a first transient phase, all points in a round have roughly the same value (Fig. 5)

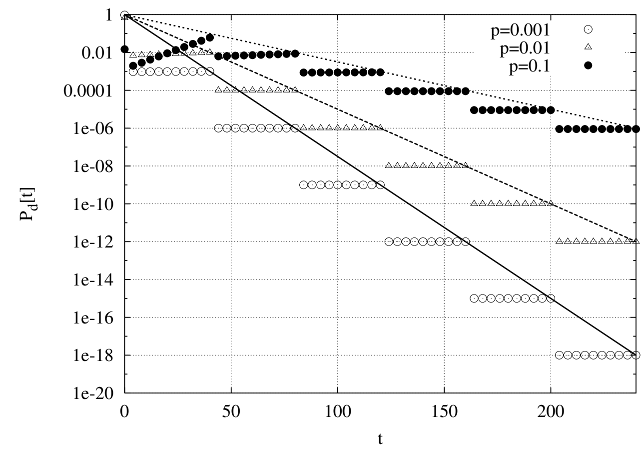

Fig. 4. Asymptotic behavior of $P_d[t]$, $m = 40$, $p = 0.1$ and $L = 7$.

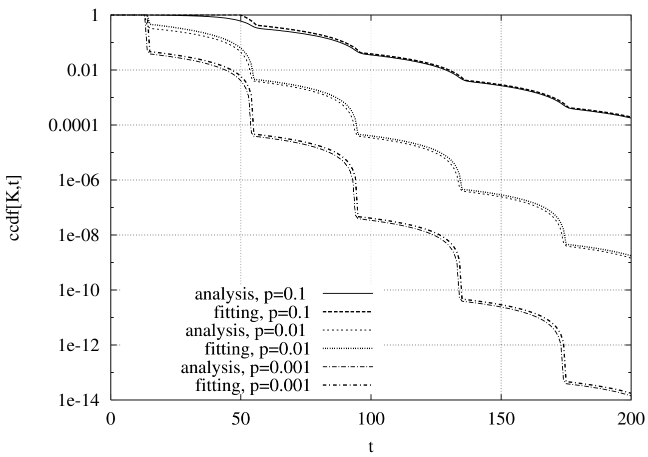

Fig. 5. Higher level packet cumulative delivery delay statistics, comparison between analysis and simulation for $m = 40$, average PDU error rate $p = 0.1$ and $L = 7$.

allow us to conclude that, in the logarithmic domain, and after a transient phase, the values of the delivery delay probability in the first (last) slot of each round are placed over a straight line. This observation allows us to accurately approximate, after a short transient phase, the behavior of $P_d[t]$ by simply using Eq. (6). Given m and p, the only parameter that needs to be computed is the value of t_0 that can be found by solving the equation $P_d[sm] = y[sm]$ for s sufficiently large[3]. t_0 can be easily tabulated.

In Fig. 5 we report $P_d[t]$ and the straight lines indicating the asymptotic behavior (Eq. (6)) by varying p and considering $m = 40$. It is worth noting that the behavior described by Eq. (6) also holds for the characterization of the SDU delivery delay statistics (see Fig. 7), i.e., also in this last case it is able to correctly track the statistics slope in the logarithmic domain.

V. Heuristics for the approximation of ARQ statistics

In this section we propose a simple heuristic for the estimate of SDU delivery delay statistics. The exact $ccdf[K,t]$ function is equal to one in $t \in \{0, 1, \ldots, K\}$, i.e., the probability that a message composed by K PDUs (a SDU) is transmitted in more than i slots, where $i \leq K$ is one. We set the origin in the slot where the first PDU composing a tagged SDU is transmitted for the first time over the channel. Moreover, we subdivide the time in rounds of m slots letting t_i, $i \geq 1$ be $t_i = K + (i-1)m + 1$. We can use the function $y[t]$ derived above to find the $ccdf[K,t]$ in the points at the beginning of each round, t_i. Moreover, we can approximate the behavior of *ccdf* between these points by means of a line whose extremes are $y[t_i]$ and $y[t_{i+1}]$. This is justified by the fact that the SDU delivery delivery delay probability (after a short transient phase) has a flat slope between the points t_i and t_{i+1}. In Fig. 5, we can observe that, after an initial transient phase the single PDU statistics is characterized, round by round, by a flat behavior. This result also holds for the SDU delivery delay statistics. In practice, *ccdf* can be well approximated by the following piecewise-linear function

$$f[K,t] = \begin{cases} 1 & t \in [0, K-1] \\ 1-(1-p)^{m-1+K} & t = K \\ y[t_i] + \dfrac{(y[t_{i+1}] - y[t_i])(t-t_i)}{m} & t \in [t_i, t_{i+1}), \\ & \forall i, \text{s.t. } i \geq 1 \end{cases} \tag{7}$$

where $(1-p)^{m-1+K}$ is the probability that all the K PDUs composing a SDU are correctly transmitted in their first transmission ($(1-p)^K$) and that no blocking packets are present at time m ($(1-p)^{m-1}$), i.e., when the first PDU is transmitted. This heuristic can be effectively used to achieve fast and accurate approximations of the exact SDU delivery delay statistics in the UMTS system, enabling the computation of delay distributions directly at the mobile terminal.

In the next we consider the limited attempt case, i.e., $L < +\infty$. When a PDU has been transmitted for $L+1$ consecutive times without success, it is discarded and a new PDU is transmitted instead. In Fig. 6, we report the $ccdf[K,t]$ by varying L. It is worth noting that when L is finite, the delivery statistics at the end of the L-th round suddenly decreases. Moreover, the slope of this last part is the same observed for the delivery delay distribution P_d regarding a RLC SDU and can be well approximated by the following function

$$y_f[t] = \left[p(1-p^L)\right]^{t-t'_o} \tag{8}$$

Hence, $y_f[t]$ can be used to approximate the behavior of $ccdf[K,t]$ after the L-th round, i.e., after t_{L+1}. The parameter t'_o can be found by imposing $y_f[t_{L+1}] = y[t_{L+1}]$ and solving for t'_o

$$t'_o = t_{L+1} \frac{\log y[t_{L+1}]}{\log \left[p(1-p^L)\right]} \tag{9}$$

VI. Selection of the optimal number of link layer retransmissions

In this section we use the heuristics derived above to compute the optimal value of L as a function of p, m and of the delay

[3]Note that for p values up to and including 0.1, $s = 1$ suffices to derive a very accurate approximation for t_0.

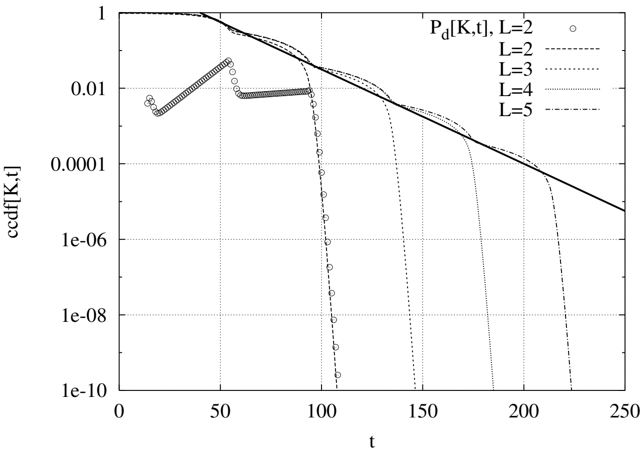

Fig. 6. $ccdf[K,t]$ by varying L and considering $K = 14$, $m = 40$ and $p = 0.1$.

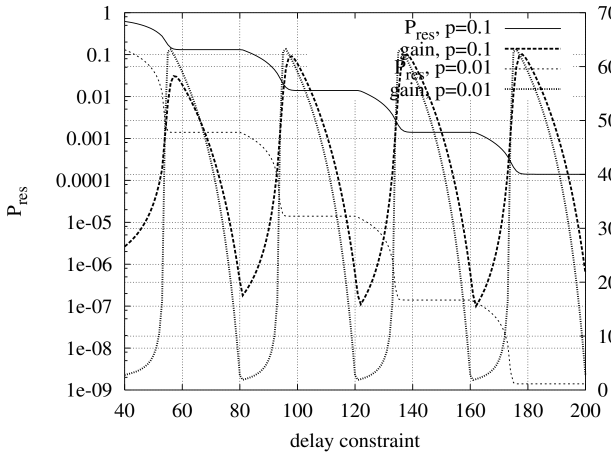

Fig. 8. $P_{res}(L_{opt})$ and maximum achievable gain.

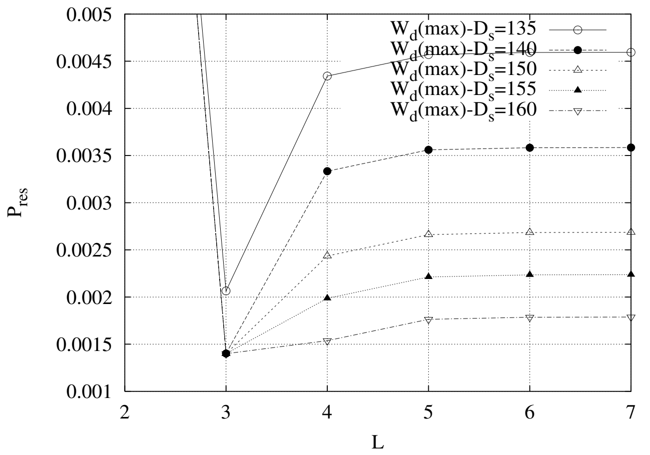

Fig. 7. Residual packet error probability, P_{res} as a function of L for $m = 40$, average PDU error rate $p = 0.1$.

constraints imposed over the wireless link $W_d(max)$. In practice, we use Eq. (1) to derive the SDU residual error probability and the heuristics to obtain the term Prob$\{W_d > W_d(max)\}$; the term P_{pkt} in Eq. (1) is obtained as $1-(1-p)^K$. Finally, we choose the value of L that minimizes the SDU error probability. In Fig. 7 we report a typical behavior of P_{res} as a function of L. From this figure it is clear that P_{res} as a function of L has a minimum (in $L = L_{opt}$). As an example, in Fig. 8 we report $P_{res}(L_{opt})$ and the maximum relative gain achievable by means of a correct configuration of L as a function of the delay constraints imposed on the wireless link $W_d(max)$, where the gain is expresses as $[P_{res}(L_{eq}) - P_{res}(L_{opt})]/P_{res}(L_{eq})$, where $L_{eq} > L_{opt}$ is the value of L after which P_{res} becomes constant ($L_{eq} = 6$ in Fig. 8). From Fig. 8 it is clear that choosing the optimum L value (L_{opt}) could lead to a considerable gain with respect to the case where L is not limited ($L > L_{opt}$ or fully reliable link layer is considered) to account for delay requirements.

VII. Conclusions

In this paper we proposed heuristics for the characterization of delivery delay statistics for the RLC level of a UMTS system when it is configured in the in order delivery mode. These heuristics are very accurate and enable the computation of such statistics directly at the mobile terminal. Moreover, we quantitatively derived the maximum achievable gain when the optimum number of retransmissions is used. We find out that a controlled amount of retransmission capability at the link layer can substantially reduce the residual packet error probability while respecting the delay constraints imposed on the SDU flow. We achieve this by means of real-time minimization of the residual SDU error probability directly at the MH and by means of simple and accurate heuristics.

References

[1] H. Liu, M. El-Zarki, “Performance of H.263 Video Transmission over Wireless Channels Using Hybrid ARQ,” *IEEE J. Select. Areas Commun.*, vol. 15, no. 9, December 1997.

[2] A. Majumdar, D.G. Sachs, I.V. Kozintsev, K. Ramchandran, M.M. Yeung, “Multicast and unicast real-time video streaming over wireless LANs,” *IEEE Trans. Circ. Syst. Video Technol.*, vol. 12, no. 6, pp. 524-534, June 2002.

[3] T. Stockhammer, H. Jenkac, C. Weiss, “Feedback and error protection strategies for wireless progressive video transmission,” *IEEE Trans. Circ. Syst. Video Technol.*, vol. 12, no. 6, pp. 465-482, June 2002.

[4] 3GPP TS 25.322, “Third Generation Partnership Project: Technical Specification Group Radio Acess Network; Radio Link Control (RLC) Specification (Release 5).”

[5] U. Horn, K. Stuhlmüller, M. Link, B. Girod, “Robust Internet Video Transmission Based on Scalable Video Coding and Unequal Error Protection,” *Image Communication, Special Issue on Real-Time Video over the Internet*, pp. 77-94, September 1999.

[6] Z. Rosberg, N. Shacham, “Resequencing delay and buffer occupancy under under the Selective Repeat ARQ,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 166-173, January 1989.

[7] J. G. Kim, M. M. Krunz, “Delay Analysis of Selective Repeat ARQ for a Markovian Source Over a Wireless Channel,” *Trans. Veh. Technol.*, vol. 45, no. 5, pp. 1968-1981, September 2000.

[8] A. G. Konheim, “A queueing analysis of two ARQ protocols,” *IEEE Trans. Commun.*, vol. 28, pp. 1004-1014, July 1980.

[9] Z. Rosberg, M. Sidi, “Selective-Repeat ARQ: The joint distribution of the transmitter and the receiver resequencing buffer occupancies,” *IEEE Trans. Commun.*, vol. 38, pp. 1430-1438, September 1990.