Evaluation of the Resequencing Delay for Selective Repeat ARQ in TDD-Based Wireless Communication Systems

Francesco Chiti, *Senior Member, IEEE,* Romano Fantacci, *Fellow, IEEE,* Andrea Tassi, *Student Member, IEEE*

***Abstract*—This correspondence deals with the modelling and analysis of the resequencing delay in Time Division Duplexing communication systems which adopt the Selective Repeat Automatic Repeat-reQuest error control strategy. Under the assumption that packet mis-ordering at the receiving end is induced by channel errors, the correspondence proposes an analytical approach based on the Absorbing Markov Chain theory in order to accurately predict the impact of the resequencing delay on the quality of the provided services. Numerical results, derived by means of computer simulations, are also given in order to validate the proposed analytical model.**

***Index Terms*—SR-ARQ Schemes, Resequencing Delay, Absorbing Markov Chain.**

I. Introduction

It is well known in the literature [1] that the Selective Repeat Automatic Repeat-reQuest (SR-ARQ) approach is one of the most efficient error control protocol for packet-switched communications over lossy channels. In particular, the efficiency of the SR-ARQ approach is due to the capability of retransmitting *only* those packets received with errors [2] [3].

Even though SR-ARQ can achieve high performance in terms of throughput and delivery delay, it cannot ensure that packets are *in-order* delivered (i.e., it cannot preserve the order of packets at the receiving end). As it happens, all the upper layer protocols which require in-order packet delivery cannot process the already received information packets. Hence, all the successfully received packets have to be stored in a buffer, called *resequencing buffer*, until the entire original packet stream can be passed to the upper layers according to the original order. As a consequence, in literature the *resequencing delay* is usually defined as the delay experienced by a packet from the time of its correct reception up to its in-order delivery to the upper layers.

The analysis of the resequencing delay is a challenging topic covered in several seminal papers [4]–[6]. Usually, mis-ordering of packets at the receiving ends is caused by different communication delays [7]–[11] as happens in multipath communication systems or because of channel errors [12]–[15]. This correspondence specifically focuses on packet mis-ordering caused by channel errors as assumed in [12]–[15]. In particular, in [12] authors investigate the packet resequencing delay by providing an analytical framework based on the theory of G/M/1 queueing systems with service vacation. In [13] the in-sequence delivery delay is evaluated by considering a Radio Link Control (RLC) protocol which ensures a reliable in-sequence delivery of Service Data Units (SDUs) in wireless packet data systems. Each SDU is composed of a fixed number of RLC blocks and it is considered as a single entity (i.e., the theoretical derivation does not consider a SDU stream). Transmission of RLC blocks belonging to the next SDU is not allowed until all the RLC blocks of the current SDU are not delivered to the upper layers. In particular, the packet resequencing delay is evaluated in [13] only for the RLC blocks belonging to the same SDU.

On the other hand, [14] derives the resequencing delay by considering a multichannel SR-ARQ scheme, where the resequencing delay is reduced by opportunistically selecting the best communication channel to transmit the packet stream. Finally, a general model, based on the G/G/1 queueing system theory, for the resequencing delay evaluation is derived in [15].

Differently to [12]–[15], this correspondence focuses on a Time Division Duplexing (TDD) transmission scheme (such as the IEEE 802.16-2009 standard [16] or 3GPP's LTE-A [17]) where data are exchanged on a frame basis. As a consequence, we have to consider *bulk* service completions at the transmitting end and bulk releases from the resequencing buffer.

To the best of our knowledge, all the previous proposed approaches [12]–[15] cannot be easily extended to tackle with the aforementioned case. Moreover, the analytical approach, based on the Absorbing Markov Chain (AMC) theory [18], proposed in this correspondence is leaner than all the aforementioned alternatives ([12]–[15]) in terms of the analytical characterization. In particular, we highlight that the computational complexity of the proposed method does not depend on the value of the round trip delay. Finally, it is worth noting that the proposed derivation does not require the simplified assumption, considered in [13], of avoiding the transmission of new data packets until all packets belonging to the currently transmitted data flow have not delivered to the upper layers of the receiving end.

The correspondence is organized as follows. Section II describes the considered system model, while Section III outlines the resequencing delay analysis based on the AMC theory. Section IV presents numerical results and compares the accuracy of the analytical predictions to the performance derived by computer simulations. Finally, Section V concludes the correspondence.

II. Resequencial Delay Model

We refer here to a wireless communication system characterized by an access scheme where the time is arranged in *frames* (such as WiMAX [16] or 3GPP's LTE/LTE-A [17]).

All the authors are with the Department of Information Engineering, University of Florence, Florence, Italy (e-mail: {francesco.chiti, romano.fantacci, andrea.tassi}@unifi.it)

In particular, we assume that each frame can hold up to k information packets. We assume that traffic saturation condition holds (i.e., the transmitting nodes always has information packets to transmit to the receiving end).

Considering a stream of information packets, each of them is labelled by a Sequence Number (SN) and they are progressively transmitted according to their own SNs (starting from the lower one). In addition, we assumed that: (i) the acknowledgement process occurs over a fully reliable feedback channel, and (ii) packets transmission is always acknowledged within the end of the current transmitted frame.

For the sake of the analysis, we assumed that packet errors occur as statically independent events. This hypothesis is reasonable if we assume data packet communications under Additive White Gaussian Noise (AWGN) or slow fading Line-of-Sight propagation conditions, or the use of a suitable diversity scheme [19] - such as time, frequency or space diversity and/or their combinations - in order to avoid burst errors which impact on consecutive packets. Moreover, we assume that packets received with errors are always detected and dropped at the receiving end.

For any erroneously received packet a negative acknowledged message (NACK) is sent back to the transmitting node to request a packet retransmission during the next frame. In what follows, we will denote as P_e the packet error probability, i.e., the probability that a packet is received with errors at the receiving end. Finally, we assume that the capacity of the transmission buffer as well as the capacity of the resequencing buffer is infinite.

As soon as a packet is correctly received, the following procedure is performed:

(i) the related SN is read
(ii) if there is at least *one* packet in the resequencing buffer with a lower SN, it is stored in that buffer
(iii) otherwise,it is passed to the upper layers.

Therefore, we define the normalized resequencing delay as the time (expressed in terms of number of frames) elapsed between the correct reception of a packet and its delivering to the higher layers. Hence, the resequencing delay of the i-th information packet is null if it is directly passed to the upper layers (i.e., when it is correctly received and all the packets having a lower SN have been already passed to the upper layers).

In order to properly model the resequencing delay process of any correctly received packet, we define the i-th state s_i of that process as follows:

Definition 1: The state s_i is equal to number of packets in the resequencing buffer with a SN which is less than i.

On the basis of our assumptions, it is straightforward to note that the resequencing delay process can be modelled as a Markov chain with states $\{s_0, \ldots, s_{k-1}\}$. In particular, it is worth noting that the aforementioned model is an Absorbing Markov Chain (AMC) [18] because: (i) the state s_0 once entered, cannot be left (i.e., s_0 is the final state or *absorbing state* of the process), and (ii) any state s_i (for $i = 1, \ldots, k-1$) results to be a *transient state* (i.e., once left it is never reached again).

In order to complete the definition of the AMC model we have to derive the state transition probabilities. Let us consider a packet (say the *tagged* packet) which is correctly received and stored into the resequencing buffer in the i-position, i.e., the resequencing process associated to the packet starts from the state s_i. The transition probability $p_{i,j}$ from the state s_i to s_j (which occurs at the end of the transmission epoch) can be defined as follows:

$$p_{i,j} = \begin{cases} \binom{i}{i-j} P_e^j \left(1 - P_e\right)^{i-j} & \text{if } i \geq j \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

This means that we have a transition from the state s_i to s_j (with $i \geq j$) at the end of the next transmission epoch, if $i - j$ packets (preceding the tagged one) over i are correctly received. We remark that a packet is dequeued from the resequencing buffer only if *all* the packets previously transmitted before the considered one have been correctly received. The state transition diagram associated to the adopted AMC model is sketched in Fig. 1.

III. Absorbing Markov Chain Analysis

This Section provides the analysis of the resequencing delay model defined in Section II. In particular, by means of the AMC theory [18], we will define the *fundamental matrix* N associated to the AMC of interest. Finally, from the expression of N and the initial state probabilities of the resequencing delay process we derive the expression of the mean resequencing delay of a correctly received packet.

Let us start our analysis by proving the definition of the *fundamental matrix* N. From (1), the corresponding $k \times k$ transition matrix $\mathbf{P}$ of the AMC process can be expressed as

$$\mathbf{P} \doteq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 - P_e & P_e & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \left(1 - P_e\right)^{k-2} & \binom{k-2}{k-4} P_e^2 \left(1 - P_e\right)^{k-4} & \cdots & 0 \\ \left(1 - P_e\right)^{k-1} & \binom{k-1}{k-3} P_e^2 \left(1 - P_e\right)^{k-3} & \cdots & P_e^{k-1} \end{bmatrix}. \tag{2}$$

From (2) we note that the matrix $\mathbf{P}$ is expressed in its *canonical* form [18]. This means that $\mathbf{P}$ can be given as

$$\mathbf{P} \doteq \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{R} & \mathbf{Q} \end{array} \right], \tag{3}$$

where $\mathbf{Q}$ is a $(k-1) \times (k-1)$ transition matrix which model the behaviour of the AMC process as long as it involves only transient states; in particular, it is defined as:

$$\mathbf{Q} \doteq \begin{bmatrix} P_e & \cdots & 0 \\ \vdots & \vdots & \vdots \\ \binom{k-2}{k-4} P_e^2 \left(1 - P_e\right)^{k-4} & \cdots & 0 \\ \binom{k-1}{k-3} P_e^2 \left(1 - P_e\right)^{k-3} & \cdots & P_e^{k-1} \end{bmatrix}. \tag{4}$$

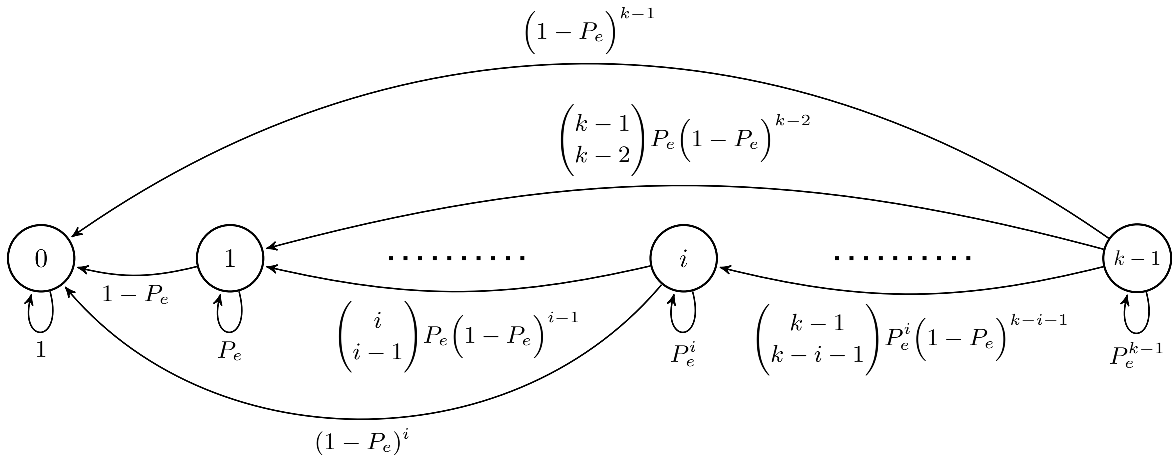

Fig. 1: State transition diagram for the considered Absorbing Markov chain model.

The term $\mathbf{R}$ is a $k-1$ dimensional column vector which lists the transition probabilities originating from a transient and directed to the absorbing state. It is given by

$$\mathbf{R} \doteq \begin{bmatrix} 1-P_e \\ \vdots \\ \left(1-P_e\right)^{k-2} \\ \left(1-P_e\right)^{k-1} \end{bmatrix}. \quad (5)$$

Finally, $\mathbf{0}$ is a $k-1$ dimensional row vector composed by null elements.

In order to derive the average resequencing delay, it is worth referring to the Proposition 1 which is a classical result in the AMC theory [18] (see Cap. III).

Proposition 1: Let $\mathbf{I}$ be the $(k-1)\times(k-1)$ identity matrix. Since the matrix $\mathbf{Q}^l$ tends to $\mathbf{O}$ (which is the $k-1\times k-1$ zero matrix) as l goes to infinity[1], the following relation holds

$$\mathbf{N} \doteq \sum_{l=0}^{\infty}\mathbf{Q}^l = \Big[\mathbf{I}-\mathbf{Q}\Big]^{-1}. \quad (6)$$

Proof: See the Appendix A. ∎

In addition, the following proposition holds:

Proposition 2: Let $N(i,j)$ be a generic element of $\mathbf{N}$. $N(i,j)$ results to be the mean value of the total number of times that the process, started in the state s_i, enters the state s_j (where both s_i and s_j are transient). Moreover, the mean number of consecutive frames $\gamma(i)$ after that the process (started from the state s_i) enters into the absorbing state s_0 (i.e., the mean value of the resequencing delay for a given packet) can be expressed as follows

$$\gamma(i) = \sum_{j=1}^{k-1}\mathbf{N}(i,j), \quad i=1,\ldots,k-1. \quad (7)$$

Proof: See the Appendix B. ∎

[1]In the correspondence with $\mathbf{Q}^l$ we will refer to the l-th power of the matrix $\mathbf{Q}$.

In order to define the mean value of the resequencing delay, we assume that any error free packet transmission (or retransmission) has occurred on any of the possible k positions of the frame with probability $1/k$. Despite this assumption seems to contrast the packet transmission/retransmission policy outlined in Section II, the good agreement between the analytical predictions and simulation results (derived according to Section II) in Figs. 2-3 validates its goodness.

For there reasons, considering a packet (transmitted in the j-th position of the frame) which is successfully received, it starts the resequencing process from the *initial* state s_i if and only if i over $j-1$ packets transmitted before it (within the current frame) have been received with errors. Hence, the probability $\pi(i,j)$ that a packet transmitted in the j-th position of the frame enters into the resequencing buffer in the i-position (i.e., the packet resequencing process starts from the s_i) is defined as

$$\pi(i,j) \doteq \begin{cases} \binom{j-1}{i}P_e^i\left(1-P_e\right)^{j-i-1} & \text{if } i\le j \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Moreover, from (8) we have that the probability that a packet starts its resequencing process from one of the possible states s_i (for $i=0,\ldots,k-1$) is given by

$$\overline{\pi}(i) \doteq \sum_{j=1}^{k}\frac{1}{k}\,\pi(i,j), \quad i=0,\ldots,k-1. \quad (9)$$

Let δ be the random variable representing the value of the resequencing delay of a packet. It follows from (9) and (7) that the mean resequencing delay $\overline{\delta}$ of a successfully received packet is

$$\overline{\delta} \doteq \sum_{i=1}^{k-1}\gamma(i)\,\overline{\pi}(i). \quad (10)$$

Finally, it is useful to define the Complementary Cumulative Density Function (CCDF), $\psi(t)$, of the random variable δ namely, i.e., the probability that the resequencing delay of a

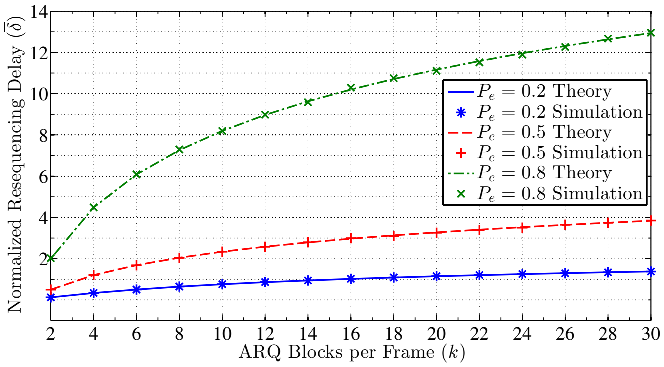

Fig. 2: Normalized resequencing delay vs. the number of ARQ blocks per frame (k) for different P_e values.

packet is greater than t (for any value of t which is integer and greater than 1).

Let $\mathbf{p}_t$ be a row vector of k items where i-th element $p_t(i)$ is the probability that the resequencing process of a packet is in the state s_i after t frame transmissions. The vector $\mathbf{p}_t$ can be expressed as follows

$$\mathbf{p}_t \doteq \Big[p_t(0), p_t(1), \ldots, p_t(k-1)\Big] = \overline{\pi} \cdot \mathbf{P}^t. \quad (11)$$

where $\overline{\pi}$ is a k-dimensional row vector where the i-th component is equal to $\overline{\pi}(i)$ (for $i = 0, \ldots, k-1$), given by (9).

From (11) the probability that a packet is passed to the higher layers (i.e., its resequencing process reaches the state s_0) after t frame transmissions (for $t \geq 1$) is the first component of the vector $\mathbf{p}_t$ (namely, $p_t(0)$). Hence, the CCDF of the resequencing process is

$$\psi(t) \doteq 1 - p_t(0), \quad t \geq 1. \quad (12)$$

IV. Numerical Results

Numerical results are provided in this Section in order to validate the analytical approach outlined in Section III. As stated before in Section II, this correspondence focuses on a TDD communication system such as that defined in the standard IEEE 802.16-2009 [16] where the time is arranged in frames which are partitioned into an uplink (UL) and downlink (DL) subframe. Within each subframe, a fixed amount of ARQ blocks are allocated to the pool of traffic flows which are transmitted [16].

We assume here that an ARQ block is equivalent to an information packet (see Section III), upon which the theoretical analysis has been developed. Furthermore, we considered a simulation setup where: (i) each flow of ARQ blocks adopts the SR-ARQ error control protocol, (ii) at the end of each frame all the transmitted ARQ blocks are acknowledged, and (iii) all the ARQ blocks which have not been correctly received are retransmitted in the next frame according to the policy described in Section II.

According to the analytical approach presented in Section III, the reported numerical results focus on the evaluation of the mean resequencing delay, normalized to the frame duration.

TABLE I: Some notable resequencing delay values reported in Fig. 2.

	$P_e = 0.2$		$P_e = 0.5$	
k	**Simulation**	**Theory**	**Simulation**	**Theory**
2	0.124	0.125	0.498	0.5
6	0.502	0.505	1.678	1.685
10	0.763	0.760	2.339	2.339
14	0.944	0.945	2.795	2.788
18	1.088	1.087	3.098	3.131
22	1.207	1.201	3.386	3.407

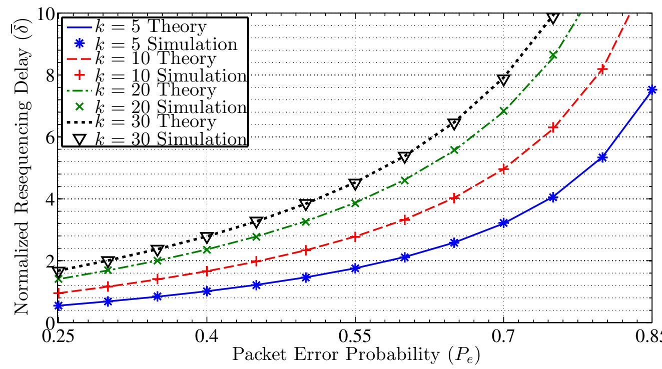

Fig. 3: Normalized resequencing delay vs. P_e, for different values of k.

Fig. 2 compares the obtained analytical predictions of the mean resequencing delay value with the simulation outcomes. In particular, Tab. I aims at highlighting the minor differences that exist between theoretical and simulation resequencing delay values (also reported in Fig. 2). We simulated the transmission of k ARQ blocks (i.e., a frame) at time while evaluating the resequencing delay associated to the reception of each information packet of that frame. It is worth noting that in the simulated system the ARQ blocks are transmitted, as stated in Section II, according to their SNs, starting from the lowest one. Furthermore, Fig. 3 shows $\overline{\delta}$ values as a function of P_e. In particular, it is worth noting that: (i) for $P_e = 0.4$, the normalized resequencing delay is greater than 1 (regardless to the values of k), and (ii) the $\overline{\delta}$ value quickly rises as either P_e or k values increase.

Figs. 2 and 3 clearly show a good agreement between the obtained analytical predictions and simulation results (derived by processing 10^4 frames of ARQ blocks). In particular, we can note that the maximum value of the difference between theoretical and simulation results shown in these figures is smaller than 0.9%. The aforementioned result comes from: (i) the accuracy of the analytical model and simulation tool, and (ii) a suitable chose of the number of simulation runs such that the convergence of the derived average values is ensured.

Fig. 4 shows the CCDF $\psi(\cdot)$ of the resequencing delay (expressed in terms of number of frames), defined by (12), in the case of k equal to 10 or 20, and P_e equal to 0.2, 0.5 or 0.8. In particular, it highlights that for $k = 10$ and $P_e = 0.5$, the 62% and 41% of the information packets which have been successfully received experience a resequencing delay greater than 1 and 2 frames, respectively.

It is worth noting from the considered figures that the

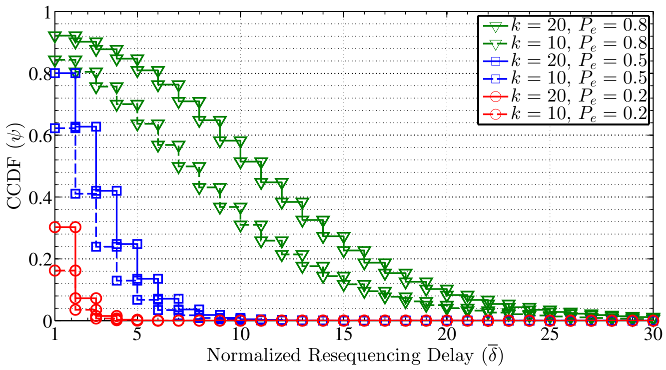

Fig. 4: CCDF of the normalized resequencing delay for different values of k and P_e.

resequencing delay can have a significant impact on data communications which are characterized by severe Quality of Service (QoS) constraints. In addition, it is straightforward to note that the resequencing delay increases at the increasing of the number of ARQ blocks per frame (k) becomes greater even though the throughput[2] increases as well. For these reasons, the theoretical model proposed in this correspondence can be efficiently used during the network design and planning phases to identify the best trade-off among the number of ARQ blocks allocated per frame, the user throughput and the mean resequencing delay of communications.

V. Conclusion

This correspondence proposes a theoretical model for the derivation of the resequencing delay caused by the SR-ARQ protocol adopted in TDD-based communication systems. Such kind of delay is experienced by all the information packets received without errors but out-of-order whenever the higher protocol layers require an in-order packet delivery. The correspondence outlines an analytical approach based on the AMC theory to derive the mean value of the resequencing delay. The proposed theoretical framework has been validated by resorting to computer simulations. Finally, it is worth noting that the proposed analytical approach is an useful tool to evaluate the impact of the resequencing delay on the overall communication delay. In particular, it can be used to properly asses the resource allocation strategies in use to guarantee the required QoS profiles.

Appendix A
Proof of the Proposition 1

Proof: From (4) the (i, j)-th entry of the matrix $\mathbf{Q}^l$ is the probability transition of entering the j-th transient state coming from the i-th one in exactly l steps (i.e., after that l frames have been transmitted). It can be proved [18] that the following relation hold:

$$\mathbf{I} - \mathbf{Q}^l = \Big(\mathbf{I} - \mathbf{Q}\Big) \cdot \Big(\mathbf{I} + \mathbf{Q} + \ldots + \mathbf{Q}^{l-1}\Big). \qquad (13)$$

[2]In this case the throughput represents the number of packet which have been successfully received (in-order or not) by a node.

Due to the fact that each element of the matrix $\mathbf{Q}^l$ tends to zero as l tends to infinity, $(\mathbf{I} - \mathbf{Q}^l)$ tends to $\mathbf{I}$. Hence, we have that the determinant of $(\mathbf{I} - \mathbf{Q}^l)$ is different than zero for sufficiently large values of l. For this reason it follows: (i) the determinant of the rightmost member of (13) is different than zero, and (ii) the determinant of $(\mathbf{I} - \mathbf{Q})$ is non-null. Thus, (13) can be rewritten as follows:

$$\Big(\mathbf{I} - \mathbf{Q}\Big)^{-1} \cdot \Big(\mathbf{I} - \mathbf{Q}^l\Big) = \mathbf{I} + \mathbf{Q} + \ldots + \mathbf{Q}^{l-1} = \sum_{l=0}^{\infty} \mathbf{Q}^l. \qquad (14)$$

Since the term $\mathbf{Q}^l$ tends to $\mathbf{O}$ as l tends to infinity, the relation (6) holds. ■

Appendix B
Proof of the Proposition 2

Proof: Let $Y_{i,j}$ be the random variable representing the total number of frame transmissions needed by the process, starting from s_i, to reach the state s_j (where both s_i and s_j are transient). Hence, the random variable defining the total number of frame transmissions needed by the process, started from the state s_i, to reach the absorbing state s_0 is

$$T_i \doteq \sum_{j=1}^{k-1} Y_{i,j}, \quad i = 1, \ldots, k-1. \qquad (15)$$

In addition, the mean value of T_i can be expressed as follows

$$\gamma(i) \doteq \sum_{j=1}^{k-1} \mathrm{E}[Y_{i,j}], \quad i = 1, \ldots, k-1. \qquad (16)$$

Let $Y_n^{(i,j)}$ be a random variable which is 1 if the process, started from s_i reaches s_j after that n frame transmission have occurred and 0, otherwise. It is straightforward to note that the following relation holds

$$Y_{i,j} = \sum_{n=0}^{\infty} Y_n^{(i,j)}, \quad i, j = 1, \ldots, k-1. \qquad (17)$$

In addition, from (4) the mean value of $Y_n^{(i,j)}$ results to be

$$\mathrm{E}[Y_n^{(i,j)}] = \mathbf{Q}^n(i,j), \quad i, j = 1, \ldots, k-1 \qquad (18)$$

where $Q^n(i,j)$ is the (i,j)-th item of the matrix $\mathbf{Q}^n$. Hence, from (17) and (18) we have that

$$\mathrm{E}[Y_{i,j}] = \sum_{n=0}^{\infty} \mathrm{E}[Y_n^{(i,j)}] = \qquad (19)$$

$$= \sum_{n=0}^{\infty} \mathbf{Q}^n(i,j), \quad i, j = 1, \ldots, k-1.$$

From (6) and (19) we have that the relation $\mathrm{E}[Y_{i,j}] = N(i,j)$ holds. As a consequence, the relation (16) can be rewritten as follows

$$\gamma(i) = \sum_{j=1}^{k-1} N(i,j), \quad i = 1, \ldots, k-1. \qquad (20)$$

This completes the proof. ■

REFERENCES

[1] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 1992.

[2] M. Anagnostou and E. Protonotarios, "Performance Analysis of the Selective Repeat ARQ Protocol," *IEEE Trans. Commun.*, vol. 34, no. 2, pp. 127–135, Feb. 1986.

[3] R. Fantacci, "Queuing analysis of the selective repeat automatic repeat request protocol wireless packet networks," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 258–264, May 1996.

[4] N. Shacham and D. Towsley, "Resequencing delay and buffer occupancy in selective repeat ARQ with multiple receivers," *IEEE Trans. Commun.*, vol. 39, no. 6, pp. 928–937, Jun. 1991.

[5] Z. Rosberg and M. Sidi, "Selective-repeat ARQ: the joint distribution of the transmitter and the receiver resequencing buffer occupancies," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1430–1438, Sep. 1990.

[6] J. G. Kim and M. Krunz, "Delay analysis of selective repeat ARQ for a Markovian source over a wireless channel," *IEEE Trans. Veh. Technol*, vol. 49, no. 5, pp. 1968–1981, Sep. 2000.

[7] G. Harrus and B. Plateau, "Queueing Analysis of a Reordering Issue," *IEEE Trans. Softw. Eng.*, vol. 8, no. 2, pp. 113–123, Mar. 1982.

[8] I. Iliadis and L.-C. Lien, "Resequencing delay for a queueing system with two heterogeneous servers under a threshold-type scheduling," *IEEE Trans. Commun.*, vol. 36, no. 6, pp. 692–702, Jun. 1988.

[9] M. Lelarge, "Packet reordering in networks with heavy-tailed delays," *Math. Meth. Oper. Res.*, vol. 67, no. 2, pp. 341–371, 2008.

[10] W. Lin and P. Kumar, "Optimal control of a queueing system with two heterogeneous servers," *IEEE Trans. Autom. Control*, vol. 29, no. 8, pp. 696–703, Aug. 1984.

[11] Z. Rosberg and N. Shacham, "Resequencing delay and buffer occupancy under the selective-repeat ARQ," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 166–173, Jan. 1989.

[12] L. Le, E. Hossain, and A. Alfa, "Delay Statistics and Throughput Performance for Multi-rate Wireless Networks Under Multiuser Diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3234–3243, Nov. 2006.

[13] W. Luo, K. Balachandran, S. Nanda, and K. Chang, "Delay analysis of selective-repeat ARQ with applications to link adaptation in wireless packet data systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1017–1029, May 2005.

[14] J. Li, Y. Zhao, Y. Zhou, and L. Lamont, "Packet delay statistics of the multichannel selective-repeat automatic-repeat-request," *Wireless Personal Communications*, vol. 66, no. 2, pp. 235–249, 2012.

[15] J. Li and Y. Q. Zhao, "Modeling and analysis of resequencing delay in selective repeat automatic repeat request," *JNW*, vol. 5, no. 7, pp. 792–799, 2010.

[16] "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface," *IEEE Std 802.16m-2011 (Amendment to IEEE Std 802.16-2009)*, pp. 1–1112, 6 2011.

[17] C. Cox, *An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications*. John Wiley & Sons, 2012.

[18] J. Kemény and J. Snell, *Finite Markov Chains*, ser. University series in undergraduate mathematics. Van Nostrand, 1960.

[19] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, Aug. 2000.