Packet Delay Modeling of Truncated Multi-Process ARQ Protocols for Parallel Communications

(Invited Paper)

Ghassane Aniba and Sonia Aïssa

INRS-EMT
University of Quebec
Montreal, QC, Canada
{ghassane, aissa}@emt.inrs.ca

***Abstract*—This paper presents a packet transmission delay modeling for networks using multiple truncated stop-and-wait (SAW) retransmission processes. The packet transmission delay includes the packet queueing delay (PQD) and the packet reordering delay (PRD). While the first type of delay is commonly known for any SAW procedure, the latter is only introduced when multiple SAW processes are considered for parallel communications. These processes work independently while sharing the same packet traffic. By modeling the queueing buffer as an M/G/∞ model, an analytical formula for the average PQD is provided. In addition, we model the PRD as a first-order Markov chain defined by a transition probability matrix, and derive the average PRD. Analytical formulae are provided considering the general case where error probabilities are different from a transmission to another, such as in automatic repeat request (ARQ) with packet combining, and for arbitrary values of the number of multi-SAW processes and maximum number of ARQ retransmissions.**

I. Introduction

Packet queueing delay (PQD) was ever the most important performance metric in wired communication systems. This is also the case and even of greater importance in wireless communications. Indeed, due to the wireless channel impairments, the packet error probability is higher, which induces more retransmissions that further result in larger delays. Most of modern wireless applications have strict requirements on the maximum or average packet transmission delay (PTD), in addition to limitations on the tolerable packet loss probability. Taking into consideration these two criteria, automatic repeat request (ARQ) protocols [1] are implemented in most wireless communication networks [2]. The first and oldest version of ARQ protocols is the stop-and-wait (SAW), which is the most used one because of its simplicity. In this scheme, at each transmission time interval (TTI) of duration T_{TTI}, the transmitter sends a packet and awaits for a corresponding response from the receiver. If the packet is lost or received in error, the receiver sends a negative acknowledgment (NACK) and the same packet is retransmitted, otherwise a positive acknowledgment (ACK) is sent and a new packet is transmitted. The main advantage of such protocol is that there is no need to assign a sequence number for each packet, which in turn reduces the NACK/ACK traffic over the feedback channel. However, there is one major problem: the transmitter has to wait during all the time between transmission of a packet and reception of its corresponding NACK or ACK. This waiting time, defined as the round-trip time, RTT, can induce a large PQD. To resolve this problem, other ARQ protocols were introduced such as go-back-N (GBN) and selective repeat (SR) [1]. Both of these protocols involve transmission of a block of packets in the same TTI and waiting for the ACK corresponding to the last correctly received packet (in GBN ARQ), or for specific packet sequence to be retransmitted if received erroneously (in SR ARQ). These two protocols reduce the PQD significantly. However, they yield a throughput loss because of retransmitting correctly received packets (in GBN ARQ), or packet reordering delay (PRD) and higher NACK/ACK traffic volume (in SR ARQ). Truncated ARQ was of course introduced in order to limit the maximum delay resulting from ARQ retransmissions [3].

Recently, and in order to combine the advantages of various ARQ protocols and reduce their drawbacks, new ARQ schemes were introduced, e.g. multi-SAW processes as in High Speed Packet Access (HSPA) networks [2], [4] and parallel-ARQ processes in multiple-input multiple-output (MIMO) systems [5]. In multi-SAW ARQ, instead of waiting for a NACK/ACK during RTT, the transmitter is allowed to send other packets assigned to other SAW processes. Each process has its own pair of buffers, at the transmitter and receiver sides, and acts independently of the other SAW processes. Moreover, each process is only responsible of the retransmission procedure for packets assigned to it. At each TTI, only one SAW process is active, i.e., if the last packet assigned to this process is lost, a retransmission is performed, otherwise, a new packet assigned to this process is transmitted. Selection among processes is done in a round-robin way, which means that considering N_{p} SAW processes, the process $\mathrm{mod}(t-1, N_{\mathrm{p}})+1$ is the one active at TTI t, where $\mathrm{mod}(\cdot,\cdot)$ denotes the modulo function. Any retransmission of the same packet is only done when the process to which it is assigned is active. Hence, when a SAW process is selected for a given TTI, the NACK/ACK corresponding to the previously transmitted packet belonging

Work supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

to this process would have already been received. In order to make sure this is the case, the number of multi-SAW processes, N_p, should be chosen such that the interval-time between activation of the same SAW process, is at least larger than RTT, i.e., $N_\mathrm{p} \geq RRT/T_\mathrm{TTI}$. Accordingly, the packet propagation delay RTT will not affect the PTD analysis. The two main advantages of this protocol are that (i) there is no waiting time at the transmitter side, since that for each TTI there is always a packet to transmit from the active SAW process, and (ii) there is no need to assign a sequence number for the packets because the SAW process preserves the packet order, and processes are also selected in a round-robin way at the receiver side.

This new protocol can be combined with other link layer techniques to reduce the number of retransmissions and, hence, transmission delays including PQD and PRD. Such techniques, like packet combining [6], affect the PTD by reducing the packet error probability from a transmission to another. In this context, we present a detailed analysis of PTD considering truncated multi-SAW processes with different packet error probabilities from a transmission to another. First, the PQD is modeled by an M/G/∞ process, and using the Pollaczek-Khintchine formula [7] we provide the average PQD expression. Secondly, we model the PRD by a first-order Markov chain defined only by the transition probability matrix, and provide the average PRD expression.

The remainder of this paper is structured as follows: Section II presents the system model with the different blocks included in the delay analysis. In Section III, the transmission state is modeled by a Markov chain and used in the following section wherein the PQD analysis is presented. The delay incurred by packet reordering is studied in Section V. Finally, concluding remarks are drawn in Section VI.

II. System Model and Definitions

Packets are assumed of fixed size, resulting from segmentation at higher layers and considering fixed modulation, and the packet transmission time is fixed and equals T_TTI. In order to analyze the PTD, including the PQD and the PRD (in terms of TTI), the transmitter (Tx) and receiver (Rx) sides are decomposed into four modules ($\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$) and the Tx/Rx chain, as illustrated in Fig. 1. The latter chain can include the use of packet combining, MIMO antenna processing, and/or other performance enhancement techniques. Given that we consider fixed modulation, the only parameter affected by this processing is the packet error probability of each transmission attempt. This general assumption makes our results applicable to different communications scenarios. Hence, hereafter we consider that each packet retransmission has a different error probability from its corresponding original transmission, whereby P_j^e denotes the packet error probability for the $(j-1)$th retransmission (or jth transmission).

Let's first describe the transmission path followed by each packet i, as illustrated in Fig. 1. We consider that the packet arrival follows a Poisson process with rate λ (packet/TTI). Let a_i be the TTI index at which packet i is buffered into module $\mathcal{A}$. The head-of-line (HOL) packet in module $\mathcal{A}$ waits until TTI b_i at which there is no packet retransmission required at the SAW process number $\mathrm{mod}(b_i-1, N_\mathrm{p})+1$, i.e., buffer number $\mathrm{mod}(b_i-1, N_\mathrm{p})+1$ in module $\mathcal{B}$ is empty. The delay between the packet arrival instant a_i and TTI b_i corresponds to the PQD of packet i. Indeed, taking into consideration the fact that only one of the N_p SAW processes is active in each TTI, the buffered packet at module $\mathcal{A}$ is transmitted at TTI t if the buffer number $k=\mathrm{mod}(t-1, N_\mathrm{p})+1$ is empty at module $\mathcal{B}$. If this is the case, packet i will be assigned to the process number k and retransmitted every N_p TTIs if necessary. Hence, each transmitted packet from module $\mathcal{B}$ to module $\mathcal{C}$ is kept into module $\mathcal{B}$ in order to be used for retransmission purposes if needed, until it is correctly received or that the maximum number of transmissions per packet, N_arq, is reached. In the latter case, the packet is deleted from module $\mathcal{B}$. Otherwise, each transmitted packet i from buffer k at module $\mathcal{B}$ proceeds to the Tx/Rx coding-decoding block, and gets forwarded to the buffer k corresponding to the active SAW process k at module $\mathcal{C}$, replacing the previously received copy of packet i. If the packet is correctly received an ACK is sent, otherwise a NACK is transmitted back to the Tx. The time needed for this processing which includes the propagation delay too is commonly much less than $N_\mathrm{p}T_\mathrm{TTI}/2$. Thus, the SAW processes are not affected by this delay because each one is active only once every N_p TTIs, and the ACK/NACK message is received between successive transmissions belonging to the same SAW process. When correctly decoded, the received packet is not forwarded to module $\mathcal{D}$ until all its predecessor packets are either correctly received or lost after N_arq failing transmissions. Thus, a packet is forwarded to module $\mathcal{D}$ if it is correctly received after N_arq possible transmissions and that all its preceding packets are correctly received or lost. The time between the admission of packet i into module $\mathcal{B}$ and its transmission to module $\mathcal{D}$ defines the PRD.

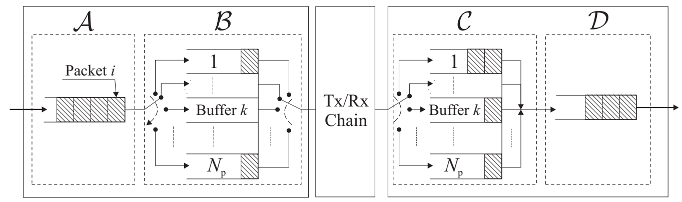

Fig. 1. System model: ($\mathcal{A}$) Queueing buffer; ($\mathcal{B}$) SAW buffers for N_p processes at Tx side and ($\mathcal{C}$) their corresponding ones at the Rx side; and ($\mathcal{D}$) buffer containing the reordered packets.

III. Transmission Steady-State Probability Vector

Delays included in the above described system model are modeled herein by different Markov chains of the first order, totally defined by transition probability matrices. The steady-state probability vectors are then deduced and used to compute the average packet delays, namely the average PQD and the average PRD.

First, we analyze the steady-state probability vector denoted by $\mathbf{P}^\mathrm{s}=\left[P_1^\mathrm{s}, P_2^\mathrm{s}, \ldots, P_{N_\mathrm{arq}}^\mathrm{s}\right]$, where P_j^s is the probability that

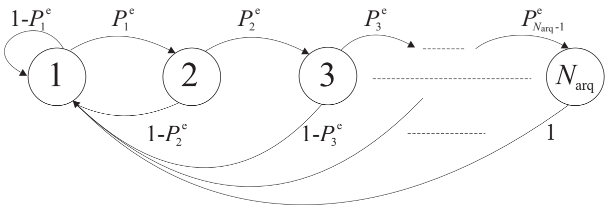

Fig. 2. Markov chain describing transition probabilities between transmission states $j = 1, \ldots, N_{\rm arq}$, where state j stands for the jth transmission of the same packet.

the actual transmission is the jth one of the same packet, with $j = 1, \ldots, N_{\rm arq}$, i.e., $j = 2$ means that the active transmission is the first retransmission of a previously sent packet. Fig. 2 shows the Markov chain representing the steady states with inter-transition probabilities which are in fact the packet error probabilities, $P_j^{\rm e}$. The transition probability matrix $\mathbf{T}^{\rm s}$ for this Markov chain can be formulated by

$$\mathbf{T}^{\rm s} = \begin{pmatrix} 1-P_1^{\rm e} & P_1^{\rm e} & 0 & \ldots & 0 \\ 1-P_2^{\rm e} & 0 & P_2^{\rm e} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-P_{N_{\rm arq}-1}^{\rm e} & 0 & 0 & \ldots & P_{N_{\rm arq}-1}^{\rm e} \\ 1 & 0 & 0 & \ldots & 0 \end{pmatrix}. \quad (1)$$

Then, the transmission steady-state probability vector is computed using equation

$$\mathbf{P}^{\rm s} \cdot \mathbf{T}^{\rm s} = \mathbf{P}^{\rm s}, \quad (2)$$

from which we deduce that[1]

$$P_j^{\rm s} = \begin{cases} \dfrac{\prod_{k=1}^{j-1} P_k^{\rm s}}{1+\sum_{l=1}^{N_{\rm arq}-1}\prod_{k=1}^{l} P_k^{\rm s}}, & \text{if } 2 \le j \le N_{\rm arq} \\ \dfrac{1}{1+\sum_{l=1}^{N_{\rm arq}-1}\prod_{k=1}^{l} P_k^{\rm s}}, & \text{if } j = 1 \text{ and } N_{\rm arq} > 1 \\ 1, & \text{if } j = 1 \text{ and } N_{\rm arq} = 1. \end{cases} \quad (3)$$

The transmission steady-state probability vector $\mathbf{P}^{\rm s}$ is a major parameter used in the delay analysis provided in the next section.

IV. Packet Queueing Delay Analysis

To determine the average PQD, denoted $D^{\rm q}$, we model the queueing buffer as an adapted version of an M/G/∞ process. A standard M/G/∞ processe is characterized by one buffer and one server, while in our case there are $N_{\rm p}$ SAW processes. However, given that only one is active at each TTI, we can model all of them as one server, represented by module $\mathcal{B}$ (Fig. 1). The server model includes definition of the *service time*, i.e., the average time needed for the server to process a packet. In our multi-SAW process configuration, this definition does not describe the server behavior exactly. Indeed, consider that packet i is transmitted at TTI t from the queueing buffer (in module $\mathcal{A}$) to module $\mathcal{B}$, and belongs to process number $\text{mod}(t-1, N_{\rm p})+1$. The next packet $(i+1)$ will not be

[1]Equation (3) can be reduced to one formula if we suppose that $\prod_{l=n}^{m} y_l = 1$ and $\sum_{l=n}^{m} y_l = 0$ for $m < n$. In order to alleviate equations we make use of this assumption in the remainder of the paper.

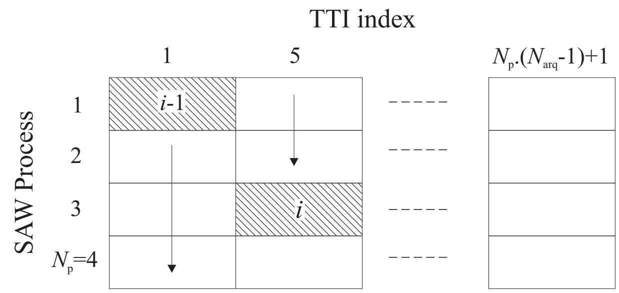

Fig. 3. An example of packet localization in time, i.e., TTI at which a packet is transmitted, and in space, i.e., the assigned process and its corresponding buffer, at module $\mathcal{B}$ for the case with $N_{\rm p} = 4$ and service time $S = 6$.

transmitted to module $\mathcal{B}$ even if packet i has been correctly received, because packet $(i+1)$ will only be transmitted to module $\mathcal{B}$ at time t' when the buffer corresponding to the process number $\text{mod}(t'-1, N_{\rm p})+1$ becomes empty, which means that the previous packet belonging to process $\text{mod}(t'-1, N_{\rm p})+1$ has been received correctly or lost. Hence, we define the *service time* as the time between admission of two consecutive packets into module $\mathcal{B}$. This new definition is also valid for a standard M/G/∞ process. Considering that the packet arrival follows a Poisson process, we can then use the Pollaczek-Khintchine formula [7] to express the average PQD according to

$$D^{\rm q} = \frac{\lambda \mathbb{E}[S^2]}{2(1-\lambda \mathbb{E}[S])}, \quad (4)$$

where S is the service time of the server expressed in TTI units, λ the packet arrival rate (packet/TTI), and $\mathbb{E}[S]$ and $\mathbb{E}[S^2]$ are the first two moments of the service time distribution.

Based on the above service time definition and the use of truncated ARQ (limit on the maximum number of transmissions $N_{\rm arq}$), the maximum service time is given by $s_{\max} = (N_{\rm arq}-1)N_{\rm p}+1$.

In order to evaluate the average PQD, $D^{\rm q}$, we need to determine the probability that the service time equals s, for $s = 1, \ldots, s_{\max}$. As aforementioned, the multi-SAW protocol behaves as $N_{\rm p}$ independent SAW processes. Thus, a packet service time equal to s TTIs is equivalent to having specific buffer states in module $\mathcal{B}$ for each SAW process during the last s TTIs, i.e., an empty or full buffer during s TTIs. Thus, the needed probability is equal to a product of probabilities corresponding to each SAW process.

To illustrate this, consider the example presented in Fig. 3, which shows the evolution of module $\mathcal{B}$ buffers in time (full or empty). The illustrated scenario is only possible when $N_{\rm arq} \ge 3$. Consider $N_{\rm p} = 4$ and let's compute the probability that $S = 6$, which is denoted by $\text{P}(S = 6)$. The first buffer is allocated to the last transmitted packet $(i-1)$ from module $\mathcal{A}$ at TTI $t = 1$. So, in order to transmit the next packet (i) at exactly TTI $t = 7$, there is only one buffer state configuration during the last $s = 6$ TTIs. This configuration is only possible if specific conditions, $C_p^{\rm q}$, $p = 1, \ldots, N_{\rm p}$, are satisfied, each pertaining to a separate SAW process or, equivalently, to its corresponding buffer state. These conditions are defined by:

- C_1^{q}: packet $(i-1)$ in buffer 1 is erroneously received at its first transmission, and must be retransmitted the next time the SAW process number 1 is active, i.e., when $t=5$.
- C_2^{q}: buffer 2 must be occupied by a packet during all 6 TTIs following the initial transmission of packet $(i-1)$, i.e., at TTIs $t=2$ and $t=6$ when the buffer takes turn according to the round-robin selection. This means that the packet belonging to process (or buffer) number 2 is received in error in its last two transmission attempts.
- C_3^{q}: buffer 3 must be occupied at TTI $t=3$ and empty at TTI $t=7$, i.e., the packet in this buffer must be received correctly or lost at TTI $t=3$, but not before this instant.
- C_4^{q}: buffer 4 must be occupied at TTI $t=4$, i.e., the packet belonging to process (or buffer) 4 is received in error in its last transmission attempt.

The probability of the first condition above, i.e. C_1^{q}, is given by

$$\mathrm{P}(C_1^{\mathrm{q}}) = P_1^{\mathrm{e}}, \quad (5)$$

whereas the probability of the condition C_2^{q} is given by

$$\mathrm{P}(C_2^{\mathrm{q}}) = P_1^s P_1^{\mathrm{e}} P_2^{\mathrm{e}} + P_2^s P_2^{\mathrm{e}} P_3^{\mathrm{e}} + \dots + P_{N_{\mathrm{arq}}-2}^s P_{N_{\mathrm{arq}}-2}^{\mathrm{e}} P_{N_{\mathrm{arq}}-1}^{\mathrm{e}}, \quad (6)$$

where P_j^s, for $j=1,\dots,N_{\mathrm{arq}}$, are defined in (3) and $N_{\mathrm{arq}} \geq 3$.

As for the probability of condition C_3^{q}, it is expressed as

$$\mathrm{P}(C_3^{\mathrm{q}}) = P_1^s P_1^{\mathrm{e}}(1-P_2^{\mathrm{e}}) + P_2^s P_2^{\mathrm{e}}(1-P_3^{\mathrm{e}}) + \cdots + P_{N_{\mathrm{arq}}-2}^s P_{N_{\mathrm{arq}}-2}^{\mathrm{e}}(1-P_{N_{\mathrm{arq}}-1}^{\mathrm{e}}) + P_{N_{\mathrm{arq}}-1}^s P_{N_{\mathrm{arq}}-1}^{\mathrm{e}}. \quad (7)$$

Finally, the probability of the last condition, C_4^{q}, is given by

$$\mathrm{P}(C_4^{\mathrm{q}}) = P_1^s P_1^{\mathrm{e}} + P_2^s P_2^{\mathrm{e}} + \cdots + P_{N_{\mathrm{arq}}-1}^s P_{N_{\mathrm{arq}}-1}^{\mathrm{e}}, \quad (8)$$

Thus, using (5), (6), (7) and (8), the probability that the service time equals $s=6$ can be expressed as

$$\mathrm{P}(S=6) = \mathrm{P}(C_1^{\mathrm{q}})\mathrm{P}(C_2^{\mathrm{q}})\mathrm{P}(C_3^{\mathrm{q}})\mathrm{P}(C_4^{\mathrm{q}}). \quad (9)$$

Considering all possible transmission scenarios, the probabilities $\mathrm{P}(S=s)$, $s=1,\dots,s_{\max}$ are provided in (10), where $\lfloor \cdot \rfloor$ denotes the floor function.

The first and second moments of the service time S can directly be computed using

$$\mathbb{E}[S] = \sum_{s=1}^{s_{\max}} s \; \mathrm{P}(S=s), \quad (11)$$

$$\mathbb{E}[S^2] = \sum_{s=1}^{s_{\max}} s^2 \; \mathrm{P}(S=s). \quad (12)$$

Finally, using (10), the average PQD, D^{q}, can straightforwardly be deduced.

V. Packet Reordering Delay analysis

As previously defined, the PRD is the time between admission of a packet into module $\mathcal{C}$ and the time it is forwarded to module $\mathcal{D}$. A correctly received packet cannot be forwarded unless each prior packet is whether correctly received, or lost after N_{arq} failing transmissions. However, even if the PRD depends on all previously transmitted packets, it can be modeled as a first-order Markov chain, which means that all the steady-state probabilities are defined by a transition probability matrix, denoted herein by $\mathbf{T}^{\mathrm{r}}$.

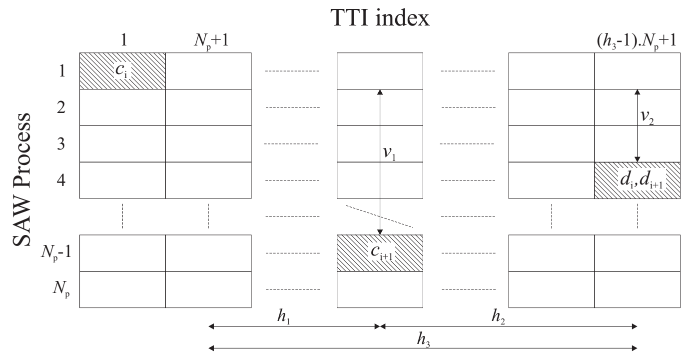

Fig. 4. An example of packet localization in time, i.e., TTI at which a packet is transmitted, and in space, i.e., the assigned process and its corresponding buffer, at module $\mathcal{C}$ for the case when $d_i = d_{i+1}$.

Consider X_{i+1} and X_i as the PRDs for packet $(i+1)$ and i, respectively, with $X_i=\infty$ indicating that the ith packet is not correctly received after N_{arq} possible transmissions. Considering N_{p} multi-SAW processes, we define the maximum PRD by $x_{\max}^{\mathrm{r}} \triangleq \max_i X_i = (N_{\mathrm{arq}}-1)N_{\mathrm{p}}$. The transition probability matrix is then defined by $\mathbf{T}^{\mathrm{r}} = [P_{x_i,x_{i+1}}^{\mathrm{r}}]_{x_i,x_{i+1}=0}^{(x_{\max}^{\mathrm{r}}+1),(x_{\max}^{\mathrm{r}}+1)}$, where

$$P_{x_i,x_{i+1}}^{\mathrm{r}} = \begin{cases} \mathrm{P}(X_{i+1}=x_{i+1}|X_i=x_i), & \text{if } 0 \leq x_i, x_{i+1} \leq x_{\max}^{\mathrm{r}} \\ \mathrm{P}(X_{i+1}=x_{i+1}|X_i=\infty), & \text{if } x_i = x_{\max}^{\mathrm{r}}+1 \\ \mathrm{P}(X_{i+1}=\infty|X_i=x_i), & \text{if } x_{i+1} = x_{\max}^{\mathrm{r}}+1. \end{cases} \quad (13)$$

Fig. 4 shows the parameters used in the PRD analysis. Specifically, c_i and d_i denote the TTI at which the ith packet arrives to module $\mathcal{C}$ and module $\mathcal{D}$, respectively. Hence, considering the reordering procedure, $d_i \leq d_{i+1}$. In order to define TTIs at which packets i and $(i+1)$ arrive in module $\mathcal{C}$ and their respective assigned processes (or buffers), we use three horizontal coordinates h_1, h_2 and h_3, and two vertical coordinates v_1 and v_2 (Fig. 4). The value h_1 defines the number of times the ith packet is transmitted before the $(i+1)$th packet is received at module $\mathcal{C}$. The values h_2 and h_3 define the number of times the $(i+1)$th and ith packets are transmitted before being forwarded to module $\mathcal{D}$, respectively. These parameter values can easily be computed using the following expressions:

$$v_1 = \mathrm{mod}(x_i - x_{i+1} - 1, N_{\mathrm{p}}), \;\; v_2 = \mathrm{mod}(x_i - 1, N_{\mathrm{p}}),$$
$$h_1 = \left\lfloor \frac{x_i - x_{i+1} - 1}{N_{\mathrm{p}}} \right\rfloor, \qquad h_2 = \left\lfloor \frac{x_{i+1}}{N_{\mathrm{p}}} \right\rfloor + 1, \quad (14)$$
$$h_3 = \left\lfloor \frac{x_i}{N_{\mathrm{p}}} \right\rfloor + 1.$$

Note that in the case when $x_i < x_{i+1}$, only parameter h_2 is used for computing the transition probability matrix ($\mathbf{T}^{\mathrm{r}}$).

First, in order to compute the transition probability matrix $\mathbf{T}^{\mathrm{r}}$, all (X_i, X_{i+1}) pairs need to be considered. There are two major cases, the first one corresponds to $d_i = d_{i+1}$ (Fig. 4)

$$
\mathrm{P}(S=s)=\begin{cases}\left(\prod_{i=1}^{\lfloor \frac{s}{N_\mathrm{p}}\rfloor} P_i^\mathrm{e}\right) P^s_{\lfloor \frac{s}{N_\mathrm{p}}\rfloor+1}\left(1-\sum_{j=1}^{\lfloor \frac{s}{N_\mathrm{p}}\rfloor+1} P_j^s\right)^{\mathrm{mod}(s-1,N_\mathrm{p})}\left(1-\sum_{j=1}^{\lfloor \frac{s}{N_\mathrm{p}}\rfloor} P_j^s\right)^{N_\mathrm{p}-\mathrm{mod}(s-1,N_\mathrm{p})-2}, & \text{if } 1\le s\le s_{\max}-2 \text{ and } \mathrm{mod}(s,N_\mathrm{p})\neq 0\\ \left(\prod_{i=1}^{\lfloor \frac{s}{N_\mathrm{p}}\rfloor-1} P_i^\mathrm{e}\right)\left(1-P^\mathrm{e}_{\lfloor \frac{s}{N_\mathrm{p}}\rfloor}\right)\left(1-\sum_{j=1}^{\lfloor \frac{s}{N_\mathrm{p}}\rfloor} P_j^s\right)^{N_\mathrm{p}-1}, & \text{if } N_\mathrm{p}\le s\le s_{\max}-1 \text{ and } \mathrm{mod}(s,N_\mathrm{p})=0\\ \left(\prod_{i=1}^{N_\mathrm{arq}-1} P_i^\mathrm{e}\right)\left(1-\sum_{j=1}^{N_\mathrm{arq}-1} P_j^s\right)^{N_\mathrm{p}-1}, & \text{if } s=s_{\max}\end{cases} \tag{10}
$$

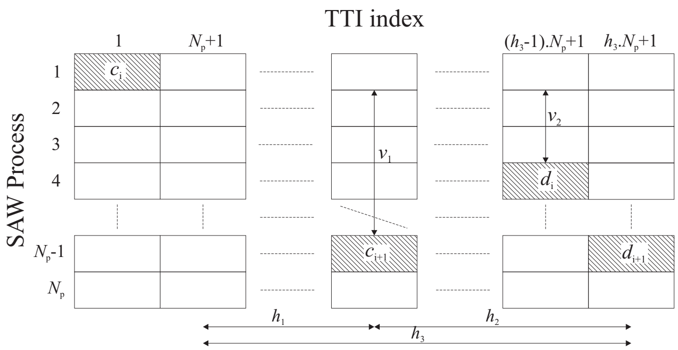

Fig. 5. An example of packet localization in time, i.e., TTI at which a packet is transmitted, and in space, i.e., the assigned process and its corresponding buffer, at module $\mathcal{C}$ for the case when $d_i < d_{i+1}$.

and the second to $d_i < d_{i+1}$ (Fig. 5). In the second case, X_{i+1} can only take values $x_{i+1} = h_1 N_\mathrm{p}$.

The same methodology is used to compute the transition probabilities $\mathrm{P}(X_{i+1} = x_{i+1}|X_i = x_i)$. First, let's consider the case when $d_{i+1} = d_i$ and $v_2 < v_1 < N_\mathrm{p} - 1$, as shown in Fig. 4. As previously mentioned, each transition probability is equal to a product of probabilities corresponding to each SAW process or its corresponding buffer. These individual probabilities can de divided into four groups. Hence, using Fig. 4 and in order to have $X_{i+1} = x_{i+1}$ while $X_i = x_i$, there are four conditions which can be formulated as follows:

- C_1^r: buffers $k = 2, \ldots, v_2+1, v_2+3, \ldots, v_1+1$ must be occupied during the last (h_1+1) transmissions.
- C_2^r: buffer (v_1+2) must be empty in the last transmission and be occupied during the previous (h_1-1) transmissions of the last transmission related to process (v_1+2). In addition, packet $(i+1)$ must be received correctly during its first h_2 transmissions.
- C_3^r: buffers $k = v_1+3, \ldots, N_\mathrm{p}$ must be occupied during the last h_1 transmissions.
- C_4^r: packet i must be erroneously received during the first h_1 transmissions.

Using the system of linear equations in (2), the probability of conditions C_p^r, $p = 1, \ldots, 4$, are given by

$$
\mathrm{P}(C_1^\mathrm{r}) = \left(1-\sum_{j=1}^{h_1+1} P_j^\mathrm{s}\right)^{v_1-1}, \tag{15}
$$

$$
\mathrm{P}(C_2^\mathrm{r}) = P^\mathrm{s}_{h_1+1}\left(1-\prod_{j=1}^{h_2} P_j^\mathrm{e}\right), \tag{16}
$$

$$
\mathrm{P}(C_3^\mathrm{r}) = \left(1-\sum_{j=1}^{h_1} P_j^\mathrm{s}\right)^{N_\mathrm{p}-v_1-2}, \tag{17}
$$

$$
\mathrm{P}(C_4^\mathrm{r}) = \prod_{j=1}^{h_1} P_j^\mathrm{e}. \tag{18}
$$

Hence, the transition probability $\mathrm{P}(X_{i+1} = x_{i+1}|X_i = x_i)$ when $d_{i+1} = d_i$ and $v_2 < v_1 < N_\mathrm{p} - 1$, is given by

$$
\mathrm{P}(X_{i+1} = x_{i+1}|X_i = x_i) = \mathrm{P}(C_1^\mathrm{r})\mathrm{P}(C_2^\mathrm{r})\mathrm{P}(C_3^\mathrm{r})\mathrm{P}(C_4^\mathrm{r}). \tag{19}
$$

Following the same approach, the transition probability matrix $\mathbf{T}^\mathrm{r}$ can be computed using (20), wherein $x_i \ge x_{i+1}$ unless otherwise stated.

In order to compute the steady-state probability vector of the PRD, denoted by $\mathbf{P}^\mathrm{r} = \left[P_0^\mathrm{r}, P_1^\mathrm{r}, \ldots, P^\mathrm{r}_{x^\mathrm{r}_{\max}+1}\right]$, we use the equation

$$
\mathbf{P}^\mathrm{r} \cdot \mathbf{T}^\mathrm{r} = \mathbf{P}^\mathrm{r}, \tag{21}
$$

where $P^\mathrm{r}_{x_i} = \mathrm{P}(X_i = x_i)$, for $x_i = 0, \ldots, x^r_{\max}$, and $P^\mathrm{r}_{x^\mathrm{r}_{\max}+1} = \mathrm{P}(X_i = \infty)$.

The solution of the linear system (21) is presented in (22). Then, the average packet reordering delay can be determined using

$$
D^\mathrm{r} = \sum_{x_i=1}^{x^\mathrm{r}_{\max}} x_i \; P^\mathrm{r}_{x_i}. \tag{23}
$$

Finally, the average transmission delay D^t, including the average PQD and the average PRD, can be computed using (4) and (23) in the following expression:

$$
D^\mathrm{t} = D^\mathrm{q} + D^\mathrm{r}. \tag{24}
$$

$$
\mathrm{P}(X_{i+1}=x_{i+1}|X_i=x_i)=
\begin{cases}
\left(1-\sum_{j=1}^{h_1+1}P_j^{\mathrm{s}}\right)^{v_1-1}P_{h_1+1}^{\mathrm{s}}\left(1-\prod_{j=1}^{h_2}P_j^{\mathrm{e}}\right)\left(1-\sum_{j=1}^{h_1}P_j^{\mathrm{s}}\right)^{N_{\mathrm{p}}-v_1-2}\left(\prod_{j=1}^{h_1}P_j^{\mathrm{e}}\right), & \text{if } v_2<v_1<N_{\mathrm{p}}-1\\
\left(1-\sum_{j=1}^{h_1+1}P_j^{\mathrm{s}}\right)^{N_{\mathrm{p}}-2}\left(\prod_{j=1}^{h_1}P_j^{\mathrm{e}}\right)\left(1-P_{h_1}^{\mathrm{e}}\right)\left(1-\prod_{j=1}^{h_2}P_j^{\mathrm{e}}\right), & \text{if } v_2<v_1=N_{\mathrm{p}}-1\\
\left(1-\sum_{j=1}^{h_1+1}P_j^{\mathrm{s}}\right)^{v_1}P_{h_1+1}^{\mathrm{s}}\left(1-\prod_{j=1}^{h_2}P_j^{\mathrm{e}}\right)\left(1-\sum_{j=1}^{h_1}P_j^{\mathrm{s}}\right)^{N_{\mathrm{p}}-v_1-3}\left(\prod_{j=1}^{h_1}P_j^{\mathrm{e}}\right), & \text{if } v_1<v_2<N_{\mathrm{p}}-1\\
\left(1-\sum_{j=1}^{h_3-h_2+1}P_j^{\mathrm{s}}\right)^{v_2}\left(\prod_{j=1}^{h_2-1}P_j^{\mathrm{e}}\right)\left(1-P_{h_2}^{\mathrm{e}}\right)\left(1-\sum_{j=1}^{h_3-h_2}P_j^{\mathrm{s}}\right)^{N_{\mathrm{p}}-v_2-2}\left(\prod_{j=1}^{h_3-h_2}P_j^{\mathrm{e}}\right), & \text{if } \mathrm{mod}(x_{i+1},N_{\mathrm{p}})=0 \text{ and } \mathrm{mod}(x_i,N_{\mathrm{p}})\neq 0\\
\left(\prod_{j=1}^{h_2-1}P_j^{\mathrm{e}}\right)\left(1-P_{h_2}\right)\left(1-\sum_{j=1}^{h_3-h_2}P_j^{\mathrm{s}}\right)^{N_{\mathrm{p}}-1}, & \text{if } \mathrm{mod}(x_{i+1},N_{\mathrm{p}})=0 \text{ and } \mathrm{mod}(x_i,N_{\mathrm{p}})=0\\
\left(\prod_{j=1}^{h_2-1}P_j^{\mathrm{e}}\right)\left(1-P_{h_2}\right), & \text{if } \mathrm{mod}(x_{i+1},N_{\mathrm{p}})=0 \text{ and } x_i<x_{i+1}\\
\mathrm{P}(X_{i+1}=x_{i+1}|X_i=x_{\max}^{r}), & \text{if } x_i=\infty\\
\prod_{j=1}^{N_{\mathrm{arq}}}P_j^{\mathrm{e}}, & \text{if } x_{i+1}=\infty\\
0, & \text{otherwise.}
\end{cases}
\tag{20}
$$

$$
P_{x_i}^{\mathrm{r}}=
\begin{cases}
\prod_{j=1}^{N_{\mathrm{arq}}}P_j^{\mathrm{e}}, & \text{if } x_i=x_{\max}^{\mathrm{r}}+1\\
\left(\prod_{j=1}^{N_{\mathrm{arq}}-1}P_j^{\mathrm{e}}\right)\left(1-P_{N_{\mathrm{arq}}}^{\mathrm{e}}\right), & \text{if } x_i=x_{\max}^{\mathrm{r}}\\
\sum_{u=x_i+1}^{x_{\max}^{\mathrm{r}}+1}P_u^{\mathrm{r}}P_{u,x_i}^{\mathrm{r}}+\left(\prod_{j=1}^{\lfloor\frac{x_i}{N_{\mathrm{p}}}\rfloor}P_j^{\mathrm{e}}\right)\left(1-P_{\lfloor\frac{x_i}{N_{\mathrm{p}}}\rfloor+1}^{\mathrm{e}}\right)\left(1-\sum_{u=x_i+1}^{x_{\max}^{\mathrm{r}}+1}P_u^{\mathrm{r}}\right), & \text{if } \mathrm{mod}(x_i,N_{\mathrm{p}})=0\\
\sum_{u=x_i+1}^{x_{\max}^{\mathrm{r}}+1}P_u^{\mathrm{r}}P_{u,x_i}^{\mathrm{r}}, & \text{otherwise.}
\end{cases}
\tag{22}
$$

VI. Conclusions

We presented analytical modeling of the packet transmission delay for networks using truncated multi-SAW ARQ protocols in conjunction with link layer performance enhancement techniques such as packet combining. The packet transmission delay includes the packet queueing delay and the packet reordering delay, and was formulated considering that error probabilities are different among multiple transmissions of a packet, which makes our analysis valid for many network scenarios and easily applicable in the design and performance evaluation and dimensioning of different kinds of wired and wireless networks.

References

[1] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request error-control schemes," *IEEE Commun. Mag.*, vol. 22, no. 12, pp. 5–17, 1984.

[2] G. Aniba and S. Aïssa, "Adaptive scheduling for MIMO wireless networks: cross-layer approach and application to HSDPA," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 259–268, Jan. 2007.

[3] E. Malkamaki and H. Leib, "Performance of truncated type-II hybrid ARQ schemes with noisy feedback over block fading channels," *IEEE Trans. Commun.*, vol. 48, pp. 1477–1487, Sep. 2000.

[4] L.-C. Wang, C.-W. Chang, and C.-J. Chang, "On the performance of an indicator-based stall avoidance mechanism for high-speed downlink packet access systems," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 691–703, Mar. 2006.

[5] C. Li and X. Wang, "Throughput analysis for parallel ARQ over correlated MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 1322–1332, Sep. 2007.

[6] S. Aïssa and G. Aniba, "BER analysis of M-QAM with packet combining over space-time block coded MIMO fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 799–805, 2008.

[7] S. M. Ross, *Introduction to Probability Models*, 10th ed. Academic Press, 2009.