A Markovian Model for the Stationary Behavior of TCP

Sophie Fortin, Bruno Sericola

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Markovian Model for the Stationary Behavior of TCP

Sophie Fortin and Bruno Sericola

N˚4240

Septembre 2001

THÈME 1

ISSN 0249-6399 ISRN INRIA/RR--4240--FR+ENG

A Markovian Model for the Stationary Behavior of TCP

Sophie Fortin and Bruno Sericola *

Thème 1 — Réseaux et systèmes
Projet Armor

Rapport de recherche n ° 4240 — Septembre 2001 — 31 pages

Abstract: This paper presents a discrete-time Markov chain model for the Reno version of TCP, the transmission control protocol for reliable transport on the Internet. The purpose is the evaluation of stationary TCP flows behavior using performance measures such as the mean throughput. The model is based on previous works which are generalized by taking into account the slow start phases that appear after each time-out recovery. We consider the three different phases of the protocol : time-out, slow start and congestion avoidance and we obtain analytical expressions for the mean number of segments sent during each of these phases and their mean duration. Our model also allows us to evaluate the proportion of losses due to the timer expiry and the duplicate acknowledgments.

Key-words: TCP, Reno, congestion control, flow control, throughput, Markov chain, loss proportion.

(Résumé : tsvp)

* {Sophie.Fortin}{Bruno.Sericola}@irisa.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : 02 99 84 71 00 - International : +33 2 99 84 71 00
Télécopie : 02 99 84 71 71 - International : +33 2 99 84 71 71

Un modèle markovien pour le comportement stationnaire de TCP

Résumé : Cet article présente un modèle de chaîne de Markov à temps discret pour la version Reno de TCP, le protocole de contrôle de transmission pour le transport fiable sur l'Internet. Le but est l'évaluation du comportement stationnaire de flux TCP en utilisant des mesures de performance telles que le débit moyen. Le modèle est basé sur des travaux précédents qui sont généralisés en tenant compte des phases de démarrage lent qui apparaissent après chaque reprise de temporisation. Nous considérons les trois phases différentes du protocole : la temporisation, le démarrage lent et l'évitement de congestion et nous obtenons des expressions analytiques pour le nombre moyen de segments envoyés durant chacune de ces phases et leur durée moyenne. Notre modèle nous permet également d'évaluer la proportion de pertes dues à l'expiration de la temporisation et à la duplication des accusés de réception.

Mots-clé : TCP, Reno, contrôle de congestion, contrôle de flux, débit, chaîne de Markov, proportion de perte.

1 Introduction

Because of the great expand of the Internet, a lot of work has been done on its efficiency and on possible improvements. The apparently simple mechanism of the *Transport Control Protocol* TCP used by HTTP transfer, file transfer, email and remote access has been modeled with various stochastic tools.

First of all, a simple but interesting approximate formula for the mean throughput ρ of a TCP flow was given in [14]. Indeed, assuming a periodic window evolution, marked by random loss events separating successive congestion avoidance phases, the authors of [14] obtained $\rho = \frac{1}{RTT}\sqrt{\frac{3}{2bp}}$ segments per second, where RTT denotes the mean round trip time, p is the segment loss probability, and $b = 1$ or 2, depending on whether the TCP implementation does or does not include delayed acknowledgments. The order of $1/\sqrt{p}$ was further discussed in [14] but also in [16], in which the window size distribution function and moments were computed.

Among all other studies, such as [13], [12], [15] and [6], many are based on a fluid approach and are usually and mainly interested in getting an analytical expression for the mean throughput of a single steady-state TCP connection. In addition to these references and also considering a fluid evolution of the window size, [2], [1], [5] and [4] focus on the window size W_n just before the n-th loss, and [4] then computes the first and second moment of W_n. The case of multiple TCP connections is the subject of [3], [11], and [8]. Other tools have been explored, such as the max-plus algebra in [7] where the authors obtain expressions of the mean throughput in the case of several routers in series.

Our paper is based on previous works presented in [18], [17] and [9] which consider a discrete-time model and a discrete evolution of the window size. We propose here a discrete-time Markov chain model which aims to give analytical expressions for measures such as the mean throughput of one bulk transfer TCP-Reno flow among exogenous traffic, but which also allows us to obtain various results for the successive TCP states (exponential growth, linear growth and idle period after time-out detection) that lead to more accurate discussions about TCP behavior.

The paper is organized as follows. Section 2 reviews TCP from its implementation to its mechanisms (loss detection by *time-out* or *triple duplicate ACKs* and window control states, named *slow start* and *congestion avoidance*), in the case of the Reno version of TCP. Section 3 presents our discrete-time Markov chain model based on the notion of *rounds* while the transition probabilities of the Markov chain (MC) are evaluated in Section 4. This will be completed in Section 5 with the introduction of the *residual rounds*. Section 6 is devoted to the computation of the mean time spent in each phase and of the mean number of segments sent during each of them. In that section, we also provide expressions for the mean throughput and for the goodput. We present numerical results in Section 7 and Section 8 concludes the paper.

2 Description of TCP

The Internet Communication Protocol package called TCP/IP has been implemented and improved on Arpanet, the Internet predecessor, in the beginning of the 80's. It has then proved its efficiency and its ability to adapt to the evolution – one might say the explosion (exponential or even hyper-exponential growth) – of the number of connections, of network complexity and of users demand.

The descriptions and recommendations – rather than standards – of the TCP/IP protocols are proposed, discussed and updated in the RFCs (*Requests For Comments*), which can be found for instance in the RFC editor web page : http://www.rfc-editor.org. The main RFCs about the different kinds of TCP versions (Tahoe, Reno, Vegas, New-Reno) and their comparison are : RFC793, RFC2001, RFC2581 and RFC2582.

2.1 TCP in the TCP/IP architecture

Two transport protocols are part of TCP/IP : the *Transmission Control Protocol* TCP and the *User Datagram Protocol* UDP. As opposed to UDP, TCP is a reliable transport protocol (flow control) for connection oriented links (see [20] and [10]).

TCP/IP architecture is given in Figure 1. The TCP header contains information about source and destination, packet length, window size (see Section 2.2), acknowledgment, priority, etc. These TCP-headed packets are called *segments* (packet unit in case of a TCP connection). The IP header specifies its length, the transport protocol (TCP or UDP), the source and destination IP addresses and different kinds of options such as route recording, time stamping, etc. A segment encapsulated with an IP header is then called a *datagram*.

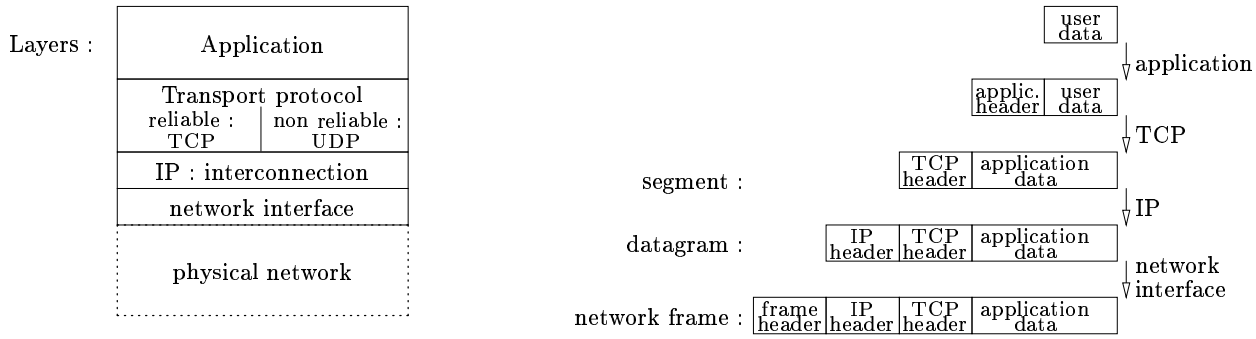

Figure 1: *The TCP/IP architecture.*

2.2 TCP dynamics

TCP is a flow control protocol that detects congestion through losses. It is based on both acknowledgment reception and sliding window.

Indeed, each successfully received packet is validated and confirmed to the source by a small packet called ACK (*ACKnowledge*). This ACK contains the sequence number of the next expected byte and a receiver's window size giving information about its buffer occupancy. But the more packets you send, the more ACKs you leave throughout the network, and the more the network might be loaded (even though these packets are small). That is why the receiver sometimes waits for more data to acknowledge before sending an ACK. Those ACKs are thus called *delayed* ACKs. The number b of segments validated per ACK is typically equal to 1 or 2 according to the implementation of TCP shared by the endpoints. However, a timer T_s will set the departure of an ACK if no new data is to be ACKed before this timer expires. Therefore, if a packet arrives and the next ones get lost, the receiver can at least acknowledge the first one.

As mentioned before, TCP is based on a sliding window dynamic. This sliding window, initialized to 1, gives the number of bytes that can be sent before receiving any ACK. Each time an ACK arrives (without any loss detection, as described later), the window slides to the right – as shown in Figure 2 – to release into the network as many bytes as the ACK validates.

However, the window size, denoted by *cwnd* in bytes or W^c in segments, is modified according to the algorithm presented in Section 2.3 and described in the RFC2001 ([21]).

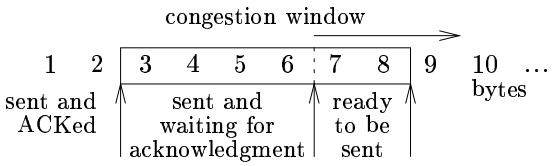

Figure 2: *Sliding window.*

Finally, there are two kinds of loss detection :

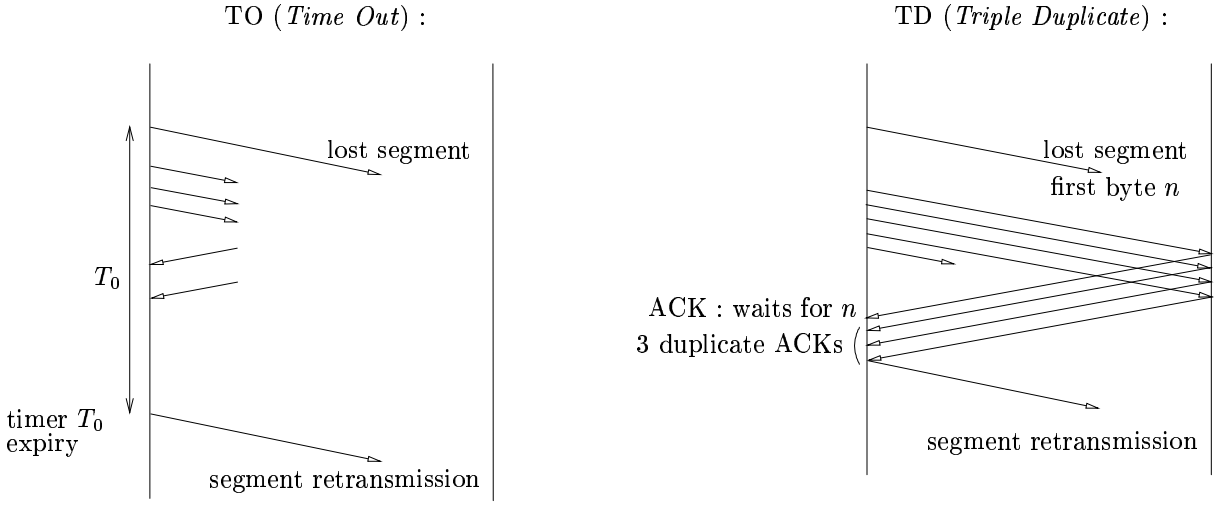

Figure 3: *Loss detection mechanisms.*

- detection by *time-out*, or TO : if no ACK is received for byte number n before the expiry of a timer T_0 (fixed or updated by round trip delay measurements), then a *time-out* occurs. The segment starting with byte n is considered lost and is thus retransmitted, and no more data is sent until byte n is ACKed (see 2.3 for further details);
- detection by the arrival of *three duplicate ACKs*, or TD : if a segment beginning with byte n is lost but some following segments are received, each of these will generate an ACK requesting byte n, that is one ACK requesting byte n and successive duplicate ACKs. The reception of the third duplicate ACK (4 similar ACKs) will halve the window and generate the segment retransmission. In fact, duplicate ACKs can be due to disordered segment reception, and the arrival of one or two duplicate ACKs is not considered as a proof of loss.

2.3 TCP congestion control algorithm

TCP flow control consists in three phases : *slow start*, *congestion avoidance* and *time-out*. Their triggering is controlled by loss events and the comparison of the congestion window size (*cwnd* or W^c) to the *slow start threshold* (*ssthresh* bytes or W^{th} segments).

- *slow start* (ss) : $W^c := W^c + 1$ each time an ACK is received (b segments ACKed). If the whole window gets successfully transmitted, then it will generate $\lceil W^c/b \rceil$ ACKs, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. For $b = 1$, a window of size W^c will thus generate W^c ACKs, so it will grow from W^c to $2W^c$. Consequently, the congestion window grows exponentially during the slow start phase;

- *congestion avoidance* (ca) : each ACK reception adds $1/W^c$ segments to the window size, so that the ACKment of the whole window increases W^c by $1/b$. Consequently, the congestion window grows linearly (of one segment every b rounds) during the congestion avoidance phase;

- *time-out* (to) : just after a TO loss detection, the apparently lost segment is retransmitted. After each retransmission failure, the timer value doubles (from T_0 to $2T_0$, $4T_0$, $8T_0$...) until $64T_0$, and then remains constant (and gets back to T_0 at the end of this time-out period, that is when the corresponding ACK arrives).

The flow control algorithm which characterizes the Reno version of TCP, described in the RFC2001 ([21]) is as follows, where $\lfloor x \rfloor$ denotes the greatest integer $\leq x$.

- initialization with $W^c := 1$ and $W^{th} := W_0^{th}$ ([21] recommending 65535 bytes); ss phase as long as $W^c < W_0^{th}$ and no loss occurs; if W^c reaches W^{th}, switch into a ca phase;

- if a TD loss occurs, then $W^{th} := \max(\lfloor W^c/2 \rfloor, 2)$ and $W^c := \max(\lfloor W^c/2 \rfloor, 1)$; switch into a ca phase;

- if a TO loss occurs, then $W^{th} := \max(\lfloor W^c/2 \rfloor, 2)$ and $W^c := 1$; switch into a to phase;

- when a to phase ends (lost segment ACKed), then enter and remain into a ss phase as long as $W^c < W^{th}$ and no loss occurs, and then switch into a ca phase.

3 The model

If the dispatch duration of all the segments and of all the ACKs held in a given window is negligible compared to the *round trip time* RTT, then we can justify the following definition of *round* given in [18], [17] and [9] : a *round* is the period of time between the departure of the first segment of the current window and the arrival of its ACK. The linear diagram in Figure 4 shows that the duration of a round is close to the round trip time when the delayed ACK timer T_s is small compared to the RTT.

We aim to model the window behavior using a homogeneous discrete-time Markov chain $X = (X_n)_{n \geq 1}$ with two components $X_n = (W_n^c, W_n^{th})$. The first component W_n^c denotes, when positive, the window size during the n-th round. The null value for W_n^c is used to represent the time-out period. The second component W_n^{th} denotes the value of the slow start threshold during the n-th round. We denote by $W_{\max}$ the maximum window size, which is the receiver's buffer capacity indicated in the ACKs. The description of the state space of this Markov chain is given, more formally, by

- $X_n = (i, j)$ with $i \in \{1, \ldots, W_{\max}\}$ and $j \in \{2, \ldots, \lfloor W_{\max}/2 \rfloor\}$ if the number of segments sent during the current round is i and the slow start threshold is j,

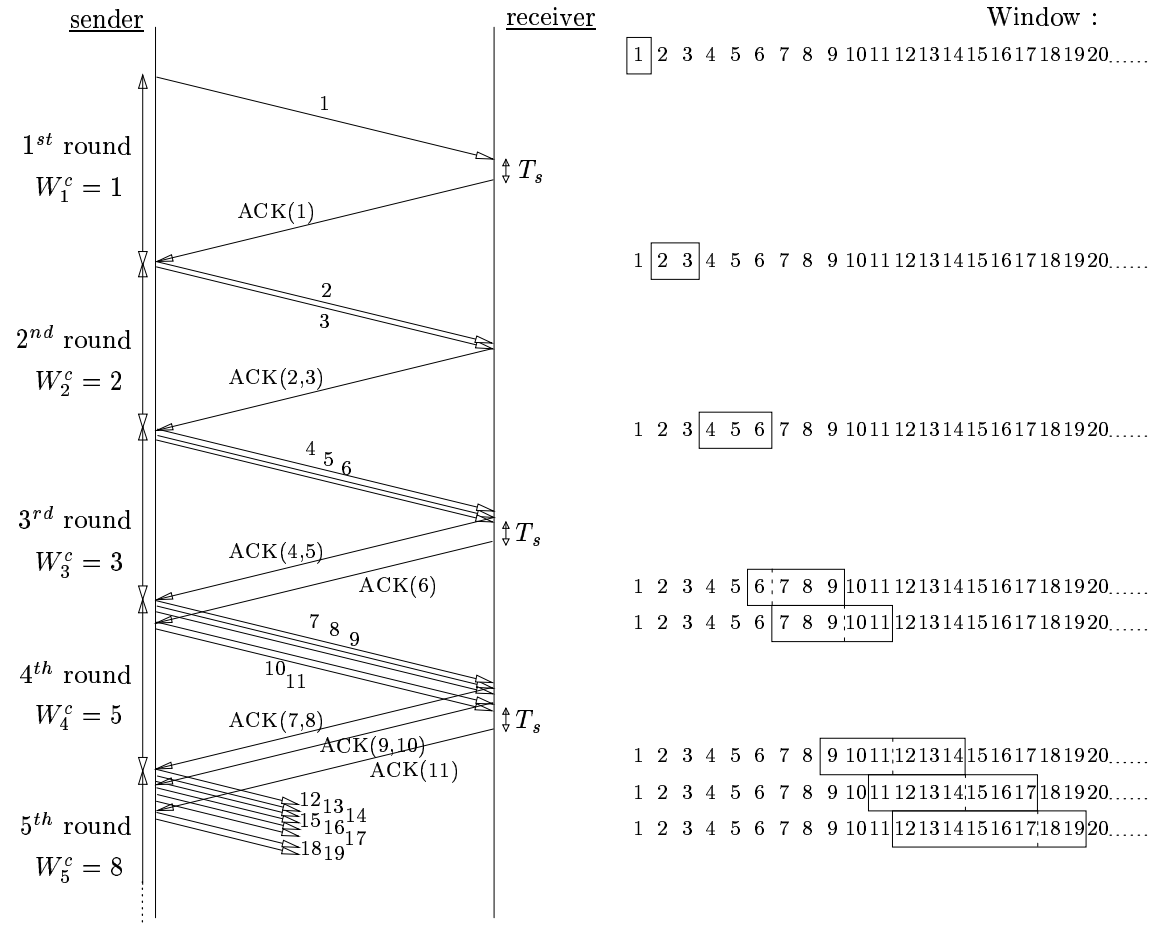

Figure 4: *Definition of a round (example in slow start phase with $b = 2$).*

- $X_n = (0, j)$ with $j \in \{2, \ldots, \lfloor W_{\max}/2 \rfloor\}$ when the connection is in a time-out period with threshold j.

As long as $W_n^c = i \geq 1$, a transition of the MC represents one round and thus lasts RTT seconds. The objective is to make the mean duration (in seconds) of a time-out period $\mathbb{E}[T_{to}]$ equal to the mean number of successive visits to the state $(0, j)$ multiplied by RTT. We thus define the two following transitions from each state $(0, j)$, for $j = 2, \ldots, \lfloor W_{\max}/2 \rfloor$:

- a transition from $(0, j)$ to $(1, j)$ with probability p_0 at the end of a time-out period,
- a transition from $(0, j)$ back to $(0, j)$ with probability $1 - p_0$ otherwise.

It follows that $\mathbb{E}[T_{to}] = RTT/p_0$. In Section 6.1, we obtain the expression of $\mathbb{E}[T_{to}]$ as a function of RTT, p, and the initial timer value T_0.

The MC state space E is a subset of $E' = \{0, \ldots, W_{\max}\} \times \{2, \ldots, \lfloor W_{\max}/2 \rfloor\}$. As we shall see below, the sets E and E' are not equal since some states of E' are never reached. However, we can already notice that for $W_{\max} = 10, 50, 100, 200$, the set E' contains respectively 44, 1224, 4949 and 19899 states.

The TCP-Reno congestion control mechanisms can then be described as follows.

- *slow start* (ss) : exponential growth of the window size as long as $1 \leq W_i^c < W_i^{th}$ and no loss occurs,

- *congestion avoidance* (ca) : linear growth of the window as long as $W_i^{th} \leq W_i^c \leq W_{\max}$ (when W^c reaches $W_{\max}$, it remains constant) and no loss occurs,
- loss detection by the arrival of three duplicate ACKs (TD) : $W_{i+1}^c = \max(\lfloor W_i^c/2 \rfloor, 1)$ and $W_{i+1}^{th} = \max(\lfloor W_i^c/2 \rfloor, 2)$, then initiate a new congestion avoidance phase,
- loss detection by *time-out* (TO) : $W_{i+1}^c = 0$ and $W_{i+1}^{th} = \max(\lfloor W_i^c/2 \rfloor, 2)$, then enter a new *time-out period*,
- *time-out* (to) : just after a TO detection, the segment is retransmitted until the reception of an acknowledgment for this segment (see Section 6.1 for further explanations), and then a new slow start phase begins with $W_i^c = 1$.

Example. A simple example of the beginning of a connection is given below to describe all the possible transitions of the MC. For simplicity, we take $W_0^{th} = 4$ segments, $W_{\max} = 8$ and the number b of segments validated per ACK is taken equal to 1.

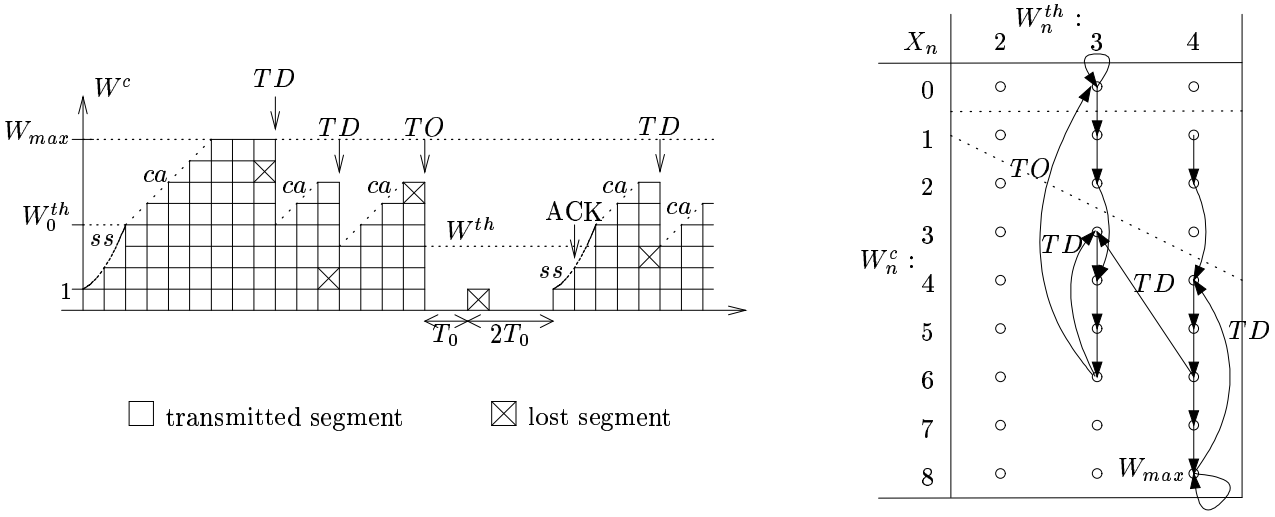

Figure 5: *Congestion window control and corresponding transitions.*

It can be noted in Figure 5, that, for instance, state $(3, 4)$ will never be reached. This is due to the fact that the window sizes reached in the slow start phase are

- for $b = 1$: $1,\ 1 + \left\lceil \frac{1}{b} \right\rceil = 2,\ 2 + \left\lceil \frac{2}{b} \right\rceil = 4,\ 8,\ 16,\ 32,\ 64,\ \ldots$
- for $b = 2$: $1,\ 1 + \left\lceil \frac{1}{b} \right\rceil = 2,\ 2 + \left\lceil \frac{2}{b} \right\rceil = 3,\ 5,\ 8,\ 12,\ 18,\ \ldots$

This example leads to the following partitioning for the state space of the MC, which is represented in Figure 6. The state space E is written as $E = E^0 \cup A \cup B$ where

- $E^0 = \{(0, j) \mid 2 \leq j \leq \lfloor W_{\max}/2 \rfloor\}$,
- $B = \{(i, j) \mid 2 \leq j \leq i \leq W_{\max} \text{ and } j \leq \lfloor W_{\max}/2 \rfloor\}$,
- $A = \left\{(i, j) \mid 1 \leq i < j \leq \lfloor W_{\max}/2 \rfloor \text{ and } \exists n \geq 0 \text{ such that } i = f^{[n]}(1)\right\}$, where $f(w) = w + \lceil w/b \rceil$, $f^{[0]}(w) = w$, and $f^{[n]} = f^{[n-1]} \circ f$, for $n \geq 1$ (see also Subsection 6.3).

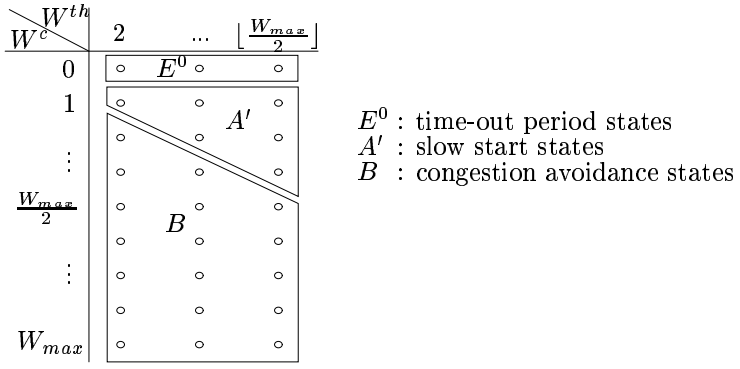

Figure 6: *Partition of the MC state space.*

The partition shown in Figure 6 is in fact a partition of the state space E' and the set A contains the reachable states of A' during the slow start phase.

This discrete-time MC is irreducible, aperiodic and has a finite state space E. It is thus ergodic and its stationary distribution π is the unique distribution verifying $\pi P = \pi$ where P is the transition probability matrix, which is given in Section 4.

In what follows, we consider the MC in stationary regime and we assume that the source behaves as a saturated source, which means that there are always packets waiting for transmission. We use the following notation illustrated in Figure 7.

- d_{to} : number of segments transmitted during a time-out period,
- $\widetilde{T}_{to}$: number of successive visits of the MC to the states of the time-out period,
- T_{to} : duration of a time-out period. We thus have $T_{to} = RTT \times \widetilde{T}_{to}$.
- d_{ss} : number of segments transmitted during a slow start phase,
- T_{ss} : duration of a slow start phase,
- d_{ca} : number of segments transmitted during a congestion avoidance phase,
- T_{ca} : duration of a congestion avoidance phase,
- N_{ca} : mean number of congestion avoidance phases in a *cycle* (i.e. between two time-out periods),
- N_{loss} : mean number of loss detections per cycle.

We study the two following flows (see [9]) :

- The *mean transmission rate* ρ, or *send rate*, takes into account all segments that have left the source, including lost segments and retransmissions. The send rate is thus the output rate seen by the network.
- The *mean effective output rate* ρ_0, or *goodput*, which does not take into account neither lost segments nor retransmissions, is the output rate seen by the receiver.

Their expression will be given in Section 6.5. We shall also be interested in the ratio $e = \rho_0/\rho$ which represents the efficiency of the connection and we shall also obtain a measure to evaluate the importance of the slow start phase.

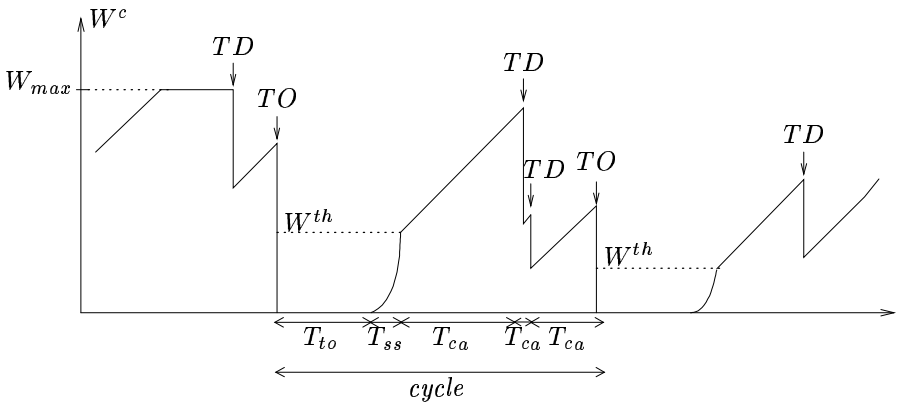

Figure 7: *Description of a cycle.*

4 The transition probabilities

We assume that losses only occur in the direction from the sender to the receiver (no loss of ACKs) and that any segment has a fixed probability p to get lost. More precisely, the random variable defined by the number of consecutive segments that are transmitted before loss has a geometric distribution with parameter $1-p$.

Let us first suppose that the connection is in slow start, i.e. $W_n^c = i < j = W_n^{th}$. As long as the MC remains in slow start, the congestion window increases by 1 segment each time an ACK is received. And because $\lceil W_n^c/b \rceil$ segments are acknowledged for the whole round, $W_{n+1}^c = W_n^c + \lceil W_n^c/b \rceil = \lceil \gamma W_n^c \rceil$ with $\gamma = 1 + 1/b$. The possible transitions from the state (i,j) are given in Figure 8. In the following propositions, we give expressions for the non-zero transition probabilities of the MC. These expressions being easy to obtain, we omit the proofs.

Proposition 1. *For each $(i,j) \in E$ such that $1 \leq i < j \leq \lfloor W_{\max}/2 \rfloor$, we get :*

- $P_{(i,j)(\lceil \gamma i \rceil, j)} = (1-p)^i$ *: no loss occurs,*
- $P_{(i,j)(0,\max(\lfloor i/2 \rfloor,2))} = \left(1-(1-p)^i\right) q_i$ *: a TO loss occurs,*
- $P_{(i,j)(\max(\lfloor i/2 \rfloor,1),\max(\lfloor i/2 \rfloor,2))} = \left(1-(1-p)^i\right)(1-q_i)$ *: a TD loss occurs,*

where q_i (computed in Section 5) denotes the probability that a loss is due to time-out when $W^c = i$.

Suppose now that the transmission is in congestion avoidance in state (i,j), i.e. $W_n^c = i \geq j = W_n^{th}$. The possible transitions of the MC from such a state are also illustrated in Figure 8.

Proposition 2. *Observing that congestion avoidance globally raises the window size by $1/b$, i.e. by 1 segment every b rounds, then for every (i,j) such that $1 \leq j \leq i < W_{\max}$:*

- $P_{(i,j)(i,j)} = (1-p)^i \left(1 - \dfrac{1}{b}\right)$ *: no loss occurs,*
- $P_{(i,j)(i+1,j)} = (1-p)^i \dfrac{1}{b}$ *: no loss occurs,*
- $P_{(i,j)(0,\max(\lfloor i/2 \rfloor,2))} = \left(1-(1-p)^i\right) q_i$ *: a TO loss occurs,*

- $P_{(i,j)(\max(\lfloor i/2 \rfloor,1),\max(\lfloor i/2 \rfloor,2))} = \left(1-(1-p)^i\right)(1-q_i)$: *a TD loss occurs.*

In order to get the model more accurate about the raise of 1 segment every b rounds, we should decompose the MC state (i,j) into b new states, say $(i,j,1)$, $(i,j,2)$, ..., (i,j,b), but, first that would of course significantly increase the MC size (even for $b=2$) and secondly, that would not change the measures of interest since the stationary distribution on the state space E remains the same after such a transformation.

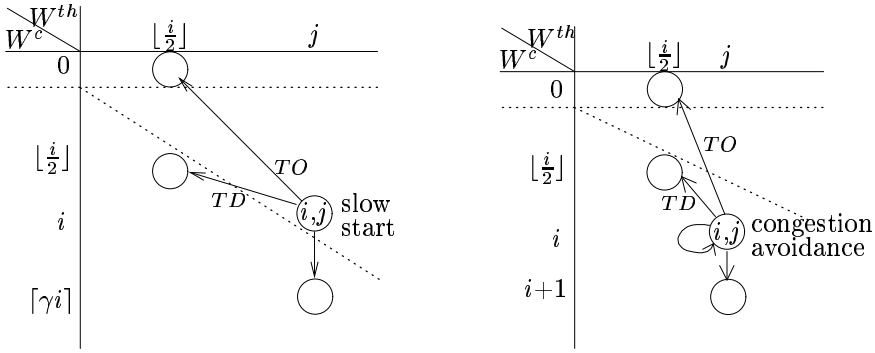

Figure 8: *Transitions of the MC during slow start and congestion avoidance.*

Proposition 3. *Similarly, for each j, we have :*

- $P_{(W_{\max},j)(W_{\max},j)} = (1-p)^{W_{\max}}$: *no loss occurs,*
- $P_{(W_{\max},j)(0,\max(\lfloor W_{\max}/2 \rfloor,2))} = \left(1-(1-p)^{W_{\max}}\right) q_{W_{\max}}$: *a TO loss occurs,*
- $P_{(W_{\max},j)(\max(\lfloor W_{\max}/2 \rfloor,1),\max(\lfloor W_{\max}/2 \rfloor,2))} = \left(1-(1-p)^{W_{\max}}\right)(1-q_{W_{\max}})$: *a TD loss occurs.*

When in time-out, the MC remains in state $(0,j)$ during $\widetilde{T}_{to}$ rounds, and then goes to state $(1,j)$. If p_0 denotes the transition probability from $(0,j)$ to $(1,j)$, the mean sojourn time in state $(0,j)$ is $1/p_0$, thus we have $p_0 = 1/\mathbb{E}[\widetilde{T}_{to}]$, i.e. :

Proposition 4. *The transition probabilities in time-out are :*

- $P_{(0,j)(0,j)} = 1 - \dfrac{1}{\mathbb{E}[\widetilde{T}_{to}]}$: *the lost segment has not been acknowledged yet,*
- $P_{(0,j)(1,j)} = \dfrac{1}{\mathbb{E}[\widetilde{T}_{to}]}$: *the acknowledgment is arrived.*

An expression of $\mathbb{E}[T_{to}] = RTT \times \mathbb{E}[\widetilde{T}_{to}]$ as a function of the timer T_0 and the loss probability p is computed in Section 6.1.

The shape of the transition probability matrix P is illustrated in Figure 9 and the regions corresponding to the different types of losses are shown in Figure 10.

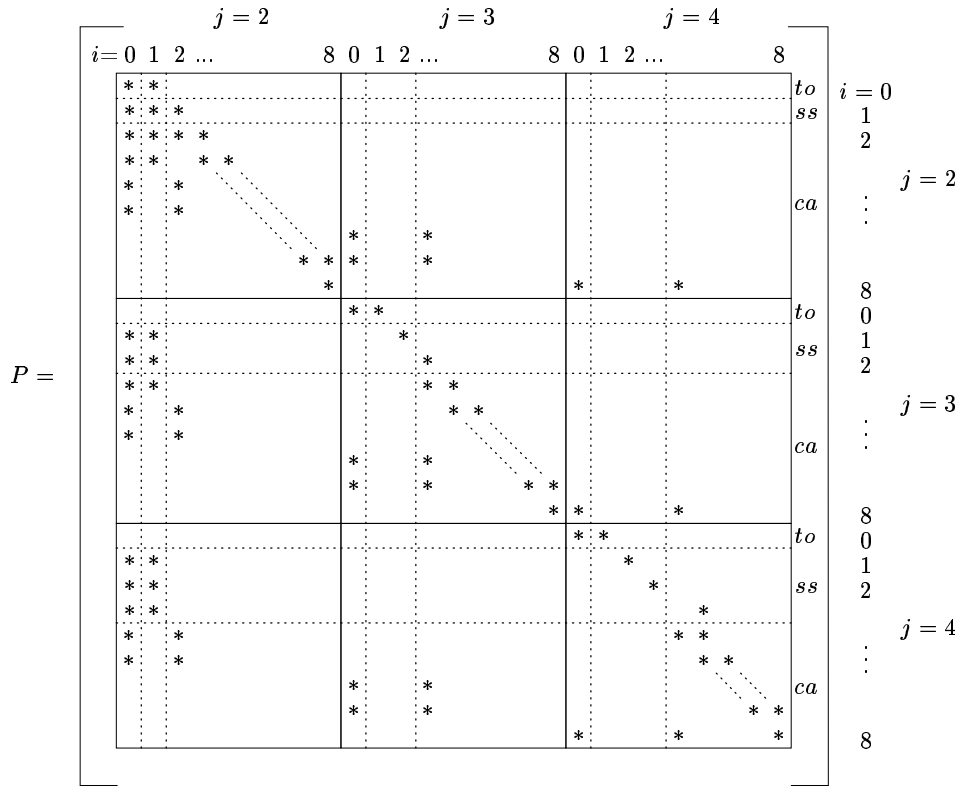

Figure 9: *Transition matrix shape for* $W_{\max} = 8$ *and* $b = 2$ *(i.e.* $\gamma = 1, 5$*).*

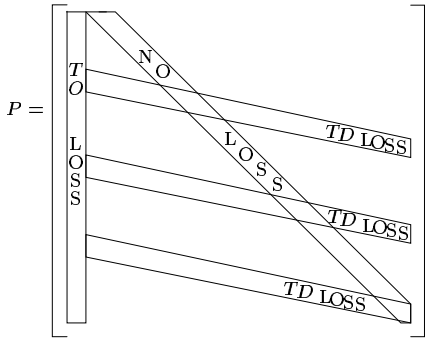

Figure 10: *Link between the transition matrix* P *and TCP.*

5 Residual rounds

Let us first emphasize the following points :

- We make the assumption that in a given round, the loss of one segment leads to the loss of the following segments (correlated losses). This should be the case in a high speed network for instance.
- In the round where the loss takes place, if k segments are however transmitted before congestion, then those segments will generate ACKs and the window will slide. This means that k new segments are transmitted in the next round, which is called the *residual round*.

This behavior is shown in Figure 11 which depicts the case where the last segment sent during the residual round is lost.

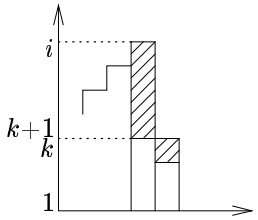

Figure 11: *The residual round.*

5.1 Computation of q_i

Proposition 5. *The probability q_i that a loss is due to TO when a loss occurs and $W^c = i$ is given by :*

$$q_i = \begin{cases} \dfrac{(1-(1-p)^{3b+1})\left(1+(1-p)^{3b+1}-(1-p)^i\right)}{1-(1-p)^i} & \text{if } i \geq 3b+2 \\ 1 & \text{otherwise.} \end{cases} \tag{1}$$

Proof. Using the notation in Figure 11, we have

- If $i \leq 3b+1$ then $k \leq 3b$ and thus no TD loss can happen (1 ACK + 3 duplicate ACKs need $b+(b+b+1) = 3b+1$ segments to be received). In this case, the loss is necessarily due to TO, i.e. $q_i = 1$.
- If $i \geq 3b+2$ then :
  - if $k \leq 3b$: for the same reason, only a TO loss can occur;
  - if $k \geq 3b+1$: there is a TO loss only when less than $3b+1$ segments from the residual round arrive at destination, i.e. the l-th segment from the residual round gets lost, with $1 \leq l \leq 3b+1$.

  Thus, if we denote by L_{k+1} the event corresponding to the loss of the $(k+1)$-th segment, we get

$$\begin{aligned} q_i &= \mathbb{P}\left(TO \mid W^c = i \ \&\ \text{loss}\right) \\ &= \sum_{k=0}^{i-1} \mathbb{P}\left(TO \mid W^c = i \ \&\ L_{k+1}\right) \mathbb{P}\left(L_{k+1} \mid W^c = i \ \&\ \text{loss}\right) \\ &= \sum_{k=0}^{3b} \frac{(1-p)^k p}{1-(1-p)^i} + \sum_{k=3b+1}^{i-1} \left(\sum_{l=1}^{3b+1} (1-p)^{l-1} p\right) \frac{(1-p)^k p}{1-(1-p)^i} \\ &= \frac{1-(1-p)^{3b+1}}{1-(1-p)^i} + \left(1-(1-p)^{3b+1}\right)(1-p)^{3b+1} \frac{1-(1-p)^{i-(3b+1)}}{1-(1-p)^i} \end{aligned}$$

□

Remark 1. The authors of [18] and [9] found, for $b = 1$,

$$q_i = \frac{(1-(1-p)^3)(1+(1-p)^3-(1-p)^i)}{1-(1-p)^i},$$

whereas Relation (1) gives

$$q_i = \frac{(1-(1-p)^4)(1+(1-p)^4-(1-p)^i)}{1-(1-p)^i}.$$

This difference is due to the fact that the segment retransmission caused by a TD loss is done after three duplicate ACKS, which means *after four identical ACKs* (the first and the three duplicate ACKs) and not three as assumed in [18] (see for instance [20]).

Proposition 6. *The mean number N_{loss} of loss detections per cycle is given by :*

$$N_{loss} = \frac{1 - \sum_{(i,j)\in E} (1-p)^i \pi(i,j)}{\sum_{(i,j)\in E} q_i \left(1-(1-p)^i\right) \pi(i,j)}. \tag{2}$$

Proof. Each cycle (see Figure 7) is composed of several TD losses and only one TO loss. Thus, we have

$$\begin{aligned}
\frac{1}{N_{loss}} &= \mathbb{P}(TO \mid \text{loss } \& \ W^c \geq 1) = \sum_{i=1}^{W_{\max}} q_i \mathbb{P}(W^c = i \mid \text{loss } \& \ W^c \geq 1) \\
&= \frac{\sum_{i=1}^{W_{\max}} q_i \mathbb{P}(\text{loss} \mid W^c = i)\mathbb{P}(W^c = i \mid W^c \geq 1)}{\mathbb{P}(\text{loss} \mid W^c \geq 1)} = \frac{\sum_{i=1}^{W_{\max}} q_i \left(1-(1-p)^i\right) \frac{\mathbb{P}(W^c=i)}{\mathbb{P}(W^c \geq 1)}}{\sum_{i=1}^{W_{\max}} \left(1-(1-p)^i\right) \frac{\mathbb{P}(W^c=i)}{\mathbb{P}(W^c \geq 1)}} \\
&= \frac{\sum_{i=1}^{W_{\max}} q_i \left(1-(1-p)^i\right) \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(i,j)}{\sum_{i=1}^{W_{\max}} \left(1-(1-p)^i\right) \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(i,j)} = \frac{\sum_{(i,j)\in E} q_i \left(1-(1-p)^i\right) \pi(i,j)}{\sum_{(i,j)\in E} \left(1-(1-p)^i\right) \pi(i,j)}.
\end{aligned}$$

□

Regarding the mean number N_{ca} of congestion avoidance phases per cycle, we have : if no loss occurs during slow start, then $N_{ca} = N_{loss}$, otherwise $N_{ca} = N_{loss} - 1$. This leads to the following obvious result.

Proposition 7.

$$N_{ca} = N_{loss} - p_{ssloss}, \tag{3}$$

where p_{ssloss} is the probability that a loss occurs during slow start (computed in Section 6.3).

5.2 The weight of residual rounds

Proposition 8. *The probability p_{rr} that a residual round appears after loss is given by*

$$p_{rr} = 1 - p\frac{1 - \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}{1 - \sum_{(i,j)\in E} (1-p)^i \pi(i,j)}. \tag{4}$$

Proof. Let K be the random variable equal to the number of segments sent before loss in the round in which that loss occurred (see Figure 11, in which we have drawn the case $K = k$). We thus have

$$\begin{aligned}
p_{rr} &= \mathbb{P}(K \neq 0 \mid \text{loss } \& \ W^c \geq 1) \\
&= \sum_{i=1}^{W_{\max}} \mathbb{P}(K \neq 0 \mid W^c = i \ \& \ \text{loss}) \mathbb{P}(W^c = i \mid \text{loss } \& \ W^c \geq 1) \\
&= \sum_{i=1}^{W_{\max}} \left(1 - \frac{p}{1-(1-p)^i}\right) \frac{\mathbb{P}(\text{loss} \mid W^c = i)\mathbb{P}(W^c = i \mid W^c \geq 1)}{\mathbb{P}(\text{loss} \mid W^c \geq 1)} \\
&= \frac{\sum_{i=1}^{W_{\max}} \left(1 - (1-p)^i - p\right) \frac{\mathbb{P}(W^c = i)}{\mathbb{P}(W^c \geq 1)}}{\sum_{i=1}^{W_{\max}} \left(1 - (1-p)^i\right) \frac{\mathbb{P}(W^c = i)}{\mathbb{P}(W^c \geq 1)}} = 1 - p\frac{\sum_{i=1}^{W_{\max}} \mathbb{P}(W^c = i)}{\sum_{i=1}^{W_{\max}} \left(1 - (1-p)^i\right) \mathbb{P}(W^c = i)} \\
&= 1 - p\frac{1 - \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}{\sum_{(i,j)\in E} \left(1 - (1-p)^i\right) \pi(i,j)} = 1 - p\frac{1 - \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}{1 - \sum_{(i,j)\in E} (1-p)^i \pi(i,j)}.
\end{aligned}$$

□

Remark 2. The probability that a residual round appears after loss in *ss* (respectively in *ca*) is easily obtained by changing in Relation (4), the state space E by the subset A (respectively B).

Because residual rounds are not considered in the MC, the duration RTT of a possible residual round and the corresponding number of segments are not taken into account in the calculation of $\mathbb{E}[T_{ca}]$, $\mathbb{E}[d_{ca}]$, $\mathbb{E}[T_{ss}]$ and $\mathbb{E}[d_{ss}]$. That is why we will need to add the following quantity in the expressions of ρ and ρ_0.

Proposition 9. *The mean number of segments $\mathbb{E}[d_r]$ and $\mathbb{E}[d_r^0]$ of a residual round that are respectively sent and successfully sent are given by*

$$\mathbb{E}[d_r] = \frac{1-p}{p} - \frac{\sum_{(i,j)\in E} i(1-p)^i \pi(i,j)}{1 - \sum_{(i,j)\in E} (1-p)^i \pi(i,j)}, \tag{5}$$

$$\mathbb{E}[d_r^0] = \frac{1-p}{p(2-p)} \left(2 - p - \frac{1 - \sum\limits_{(i,j)\in E} (1-p)^{2i} \pi(i,j)}{1 - \sum\limits_{(i,j)\in E} (1-p)^{i} \pi(i,j)} \right). \tag{6}$$

Proof. As above, we denote by K the random variable equal to the number of segments sent before loss in the round in which that loss occurred (see Figure 11). We have

$$\begin{aligned}
\mathbb{E}[d_r] &= \mathbb{E}[K \mid \text{loss } \& \ W^c \geq 1] \\
&= \sum_{i=1}^{W_{\max}} \mathbb{E}[K \mid W^c = i \ \& \ \text{loss}] \mathbb{P}(W^c = i \mid \text{loss } \& \ W^c \geq 1) \\
&= \sum_{i=1}^{W_{\max}} \left(\sum_{k=0}^{i-1} k \frac{(1-p)^k p}{1-(1-p)^i} \right) \frac{\mathbb{P}(\text{loss} \mid W^c = i) \mathbb{P}(W^c = i \mid W^c \geq 1)}{\mathbb{P}(\text{loss} \mid W^c \geq 1)} \\
&= \sum_{i=1}^{W_{\max}} \left(\frac{1-p}{p} \right) \left(\frac{1-(1-p)^{i-1}(1-p+ip)}{1-(1-p)^i} \right) \frac{\left(1-(1-p)^i\right) \sum\limits_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(i,j)}{1 - \sum\limits_{(i,j)\in E} (1-p)^i \pi(i,j)} \\
&= \left(\frac{1-p}{p} \right) \frac{\sum\limits_{(i,j)\in E} \left(1 - ip(1-p)^{i-1} - (1-p)^i\right) \pi(i,j)}{1 - \sum\limits_{(i,j)\in E} (1-p)^i \pi(i,j)} \\
&= \frac{1-p}{p} - \frac{\sum\limits_{(i,j)\in E} i(1-p)^i \pi(i,j)}{1 - \sum\limits_{(i,j)\in E} (1-p)^i \pi(i,j)}.
\end{aligned}$$

We denote by L the number of segments successfully sent before loss in the residual round. $L = l$ means that, in the residual round, l segments are received and the next one gets lost. We thus have

$$\mathbb{E}[d_r^0] = \mathbb{E}[L \mid \text{loss } \& \ W^c \geq 1] = \sum_{k=0}^{W_{\max}-1} \mathbb{E}[L \mid K = k \ \& \ \text{loss}] \mathbb{P}(K = k \mid \text{loss } \& \ W^c \geq 1).$$

The conditional expectation of L is given by

$$\mathbb{E}[L \mid K = k \ \& \ \text{loss}] = \sum_{l=0}^{k-1} l(1-p)^l p + k(1-p)^k = \frac{(1-p)\left(1-(1-p)^k\right)}{p},$$

and the conditional probability of K can be written as

$$\begin{aligned}
\mathbb{P}(K = k \mid \text{loss } \& \ W^c \geq 1) &= \sum_{i=k+1}^{W_{\max}} \mathbb{P}(K = k \mid W^c = i \ \& \ \text{loss}) \mathbb{P}(W^c = i \mid \text{loss } \& \ W^c \geq 1) \\
&= \sum_{i=k+1}^{W_{\max}} \left(\frac{(1-p)^k p}{1-(1-p)^i} \right) \frac{\left(1-(1-p)^i\right) \sum\limits_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(i,j)}{1 - \sum\limits_{(i,j)\in E} (1-p)^i \pi(i,j)}
\end{aligned}$$

Putting together these relations, we get

$$\begin{aligned}
\mathbb{E}[d_r^0] &= (1-p)\sum_{k=0}^{W_{\max}-1}\left(1-(1-p)^k\right)(1-p)^k\frac{\sum_{i=k+1}^{W_{\max}}\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\pi(i,j)}{1-\sum_{(i,j)\in E}(1-p)^i\pi(i,j)}\\
&= (1-p)\frac{\sum_{i=1}^{W_{\max}}\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\pi(i,j)\sum_{k=0}^{i-1}\left(1-(1-p)^k\right)(1-p)^k}{1-\sum_{(i,j)\in E}(1-p)^i\pi(i,j)}\\
&= (1-p)\frac{\sum_{(i,j)\in E}\pi(i,j)\left[\sum_{k=0}^{i-1}(1-p)^k-\sum_{k=0}^{i-1}(1-p)^{2k}\right]}{1-\sum_{(i,j)\in E}(1-p)^i\pi(i,j)}\\
&= \frac{1-p}{p}-\left(\frac{1-p}{p(2-p)}\right)\frac{\sum_{(i,j)\in E}\left(1-(1-p)^{2i}\right)\pi(i,j)}{1-\sum_{(i,j)\in E}(1-p)^i\pi(i,j)}
\end{aligned}$$

□

Remark 3. Similarly to Remark 2, we easily get the mean number of segments transmitted and successfully transmitted during a slow start residual round and during a congestion avoidance residual round.

6 Durations and numbers of transmissions

In this section, we obtain expressions for the mean durations $\mathbb{E}[T_{to}]$, $\mathbb{E}[T_{ss}]$, $\mathbb{E}[T_{ca}]$ and the mean numbers of segments sent $\mathbb{E}[d_{to}]$, $\mathbb{E}[d_{ss}]$, $\mathbb{E}[d_{ca}]$, respectively for the time-out, slow start and congestion avoidance phases. These expressions are given as a function of timer T_0, the mean round trip time RTT, the loss probability p, the parameter b and the stationary distribution π of the MC.

6.1 Time-out

The behavior of TCP during a time-out period is illustrated in Figure 12, where rr denotes the residual round (see also Figure 11).

The following result can be found in [18]. We give here a detailed proof of that result.

Proposition 10. *The mean number of segments sent during a time-out period and the mean duration of a time-out period are given by*

$$\mathbb{E}[d_{to}]=\frac{p}{1-p}\quad \textit{and}\quad \mathbb{E}[T_{to}]=T_0\frac{1+p+2p^2+4p^3+8p^4+16p^5+32p^6}{1-p}-RTT. \tag{7}$$

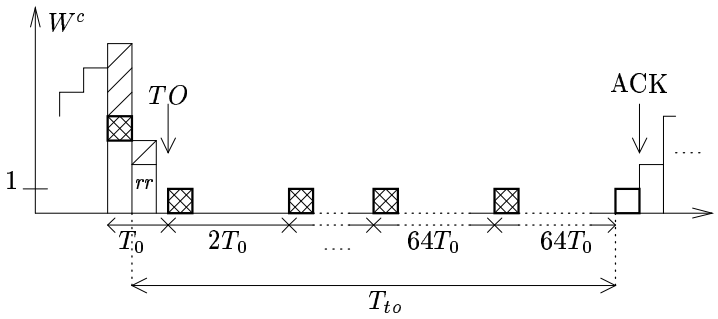

Figure 12: *Detail of a time-out period.*

Proof. The distribution of the number of segments d_{to} sent during time-out is easy to get since $d_{to} = d$ means that the retransmitted segment has been lost d times before being correctly received. Thus, d_{to} has a geometric distribution, that is

$$\mathbb{P}(d_{to} = d) = p^d(1-p) \text{ and so, } \mathbb{E}[d_{to}] = \frac{p}{1-p}.$$

Let us briefly recall the timer mechanism. After each retransmission failure, the timer doubles (T_0, $2T_0$, $4T_0$, $8T_0$, ...) until $64T_0$ and then remains equal to $64T_0$. We thus get an expression of $\mathbb{E}[T_{to}]$ by writing :

$$\begin{aligned}
\mathbb{E}[T_{to}] &= \left[\sum_{d=0}^{6}\left(\sum_{n=0}^{d} 2^n T_0\right) + \sum_{d=7}^{+\infty}\left(\sum_{n=0}^{6}(2^n T_0) + (d-6)64T_0\right)\right]\mathbb{P}(d_{to}=d) - RTT \\
&= T_0(1-p)\left(\sum_{d=0}^{6}(2^{d+1}-1) + \sum_{d=7}^{+\infty}(127+64(d-6))\right)p^d - RTT \\
&= T_0(1-p)\left(\sum_{d=0}^{6}(2^{d+1}-1)p^d + 127\sum_{d=7}^{+\infty}p^d + 64p^7\sum_{d=7}^{+\infty}(d-6)p^{d-7}\right) - RTT \\
&= T_0(1-p)\left(1+3p+7p^2+15p^3+31p^4+63p^5+127p^6+\frac{127p^7}{1-p}+\frac{64p^7}{(1-p)^2}\right) - RTT \\
&= T_0\frac{1+p+2p^2+4p^3+8p^4+16p^5+32p^6}{1-p} - RTT.
\end{aligned}$$

□

6.2 Cycle

In the following remark, we recall briefly some results on sojourn times in Markov chains. These results have been obtained in [19].

Remark 4. Consider an irreducible discrete time Markov chain with finite state space E and transition probability matrix P. We denote by π its stationary probability distribution, that is $\pi = \pi P$ and $\pi\mathbb{1} = 1$, where $\mathbb{1}$ denotes the column vector with all the entries equal to 1. Let F be a proper subset of E. We denote by F' the complementary subset $E - F$. The partition F, F' of E induces the following decomposition of P, π and $\mathbb{1}$:

$$P = \begin{pmatrix} P_F & P_{F,F'} \\ P_{F',F} & P_{F'} \end{pmatrix}, \pi = \begin{pmatrix} \pi_F & \pi_{F'} \end{pmatrix} \text{ and } \mathbb{1} = \begin{pmatrix} \mathbb{1}_F \\ \mathbb{1}_{F'} \end{pmatrix}.$$

The matrix I denotes the identity matrix of dimension given by the context. Let v_i be the stationary probability that a sojourn in F initiates in state i, $i \in F$. We denote by v the row vector composed of the v_i and by N_F the number of states of F visited during such a sojourn. From [19], we have that

$$v = \frac{\pi_F(I - P_F)}{\pi_F(I - P_F)\mathbb{1}_F} = \frac{\pi_{F'}P_{F',F}}{\pi_{F'}P_{F',F}\mathbb{1}_F}. \tag{8}$$

$$\mathbb{E}[N_F] = v(I - P_F)^{-1}\mathbb{1}_F = \frac{\pi_F\mathbb{1}_F}{\pi_{F'}P_{F',F}\mathbb{1}_F} = \frac{\pi_F\mathbb{1}_F}{\pi_F P_{F,F'}\mathbb{1}_{F'}}. \tag{9}$$

Moreover, for every $i \in F$, let $N_{i,F}$ be the number of visits to state i during a sojourn in F and let r_i be any real number. If we denote by r_F the column vector composed of the r_i and by C_F the random variable

$$C_F = \sum_{i \in F} r_i N_{i,F},$$

we easily get

$$\mathbb{E}[C_F] = v(I - P_F)^{-1}r_F = \frac{\pi_F r_F}{\pi_{F'}P_{F',F}\mathbb{1}_F} = \frac{\pi_F r_F}{\pi_F P_{F,F'}\mathbb{1}_{F'}}. \tag{10}$$

Using these results, we have the following proposition.

Proposition 11. *The mean time $\mathbb{E}[T_{E^0}^{back}]$ between the end of a time-out period (the beginning of slow start) and the next TO loss is given by*

$$\mathbb{E}[T_{E^0}^{back}] = \frac{RTT}{p_0}\left(\frac{1}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor}\pi(0,j)} - 1\right). \tag{11}$$

Proof. By definition of $\mathbb{E}[T_{E^0}^{back}]$, we have $\mathbb{E}[T_{E^0}^{back}] = RTT \times \mathbb{E}[N_{A\cup B}]$, where $N_{A\cup B}$ denotes the number of states visited during a sojourn in subset $A \cup B$. Following Remark 4, Relation (9), we have

$$\begin{aligned}\mathbb{E}[N_{A\cup B}] &= \frac{\pi_{A\cup B}\mathbb{1}_{A\cup B}}{\pi_{A\cup B}(I - P_{A\cup B})\mathbb{1}_{A\cup B}} = \frac{\sum_{(i,j)\in A\cup B}\pi(i,j)}{\pi_{E^0}P_{E^0,A\cup B}\mathbb{1}_{A\cup B}} \\ &= \frac{1 - \sum_{(i,j)\in E^0}\pi(i,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} p_0\pi(0,j)} = \frac{1 - \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor}\pi(0,j)}{p_0\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor}\pi(0,j)},\end{aligned}$$

where the last equality follows from the fact that E^0 is the subset of states $(0,j)$, $j = 2, \ldots, \lfloor W_{\max}/2 \rfloor$. □

Proposition 12. *The mean number of segments sent $\mathbb{E}[d_{E^0}^{back}]$ and the mean number of segments received $\mathbb{E}[d_{E^0}^{back,0}]$ between the end of a time-out period and the next TO loss are given by*

$$\mathbb{E}[d_{E^0}^{back}] = \frac{\sum_{(i,j)\in A\cup B} i\pi(i,j)}{p_0\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor}\pi(0,j)}, \tag{12}$$

$$\mathbb{E}[d_{E^0}^{back,0}] = \frac{(1-p)}{p} \frac{\sum_{(i,j)\in A\cup B} (1-(1-p)^i)\pi(i,j)}{p_0 \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}. \tag{13}$$

Proof. Using the result presented in Remark 4, Relation (10), we have

$$\mathbb{E}[d_{E^0}^{back}] = \frac{\pi_{A\cup B} r_{A\cup B}}{\pi_{A\cup B}(I - P_{A\cup B})\mathbb{1}_{A\cup B}},$$

where the entry (i,j) of vector $r_{A\cup B}$ is the number of segments sent when the MC is in state $(i,j) \in A \cup B$, that is $r_{(i,j)} = i$, for every $(i,j) \in A \cup B$. Relation (12) is then easily obtained using the same arguments as in the proof of Proposition 11.

In the same way, we have

$$\mathbb{E}[d_{E^0}^{back,0}] = \frac{\pi_{A\cup B} r_{A\cup B}}{\pi_{A\cup B}(I - P_{A\cup B})\mathbb{1}_{A\cup B}},$$

where the entry (i,j) of vector $r_{A\cup B}$ is now the number of segments successfully transmitted when the MC is in state $(i,j) \in A \cup B$, that is

$$r_{(i,j)} = \sum_{k=0}^{i-1} k(1-p)^k p + i(1-p)^i = \frac{1-p}{p}\left(1-(1-p)^i\right).$$

As for Relation (12), Relation (13) is then easy to get. □

6.3 Slow start

Let us denote by w_n, $n \geq 1$, the size of the congestion window during the n-th round in slow start phase and recall that $\gamma = 1 + 1/b$. If $b = 1$ we have $w_n = 2^{n-1}$, but w_n is not equal to γ^{n-1} when $b \neq 1$. Consider the integer function $f : w \mapsto \lceil \gamma w \rceil$ and define recursively $f^{[0]}(w) = w$, and $f^{[n]} = f^{[n-1]} \circ f$, for $n \geq 1$. We thus have $w_n = f^{[n-1]}(1)$. For $n \geq 1$, we denote by d_n the number of segments sent during the n first rounds of a slow start phase and we set $d_0 = 0$. We thus easily get

$$d_n = \sum_{k=1}^{n} w_k.$$

Let us consider a given slow start phase for which the threshold is equal to j and denote by n_j the maximum number of rounds in that slow start phase. Since during that phase, w_n is increasing with n, we have $n_j = \max\{n \geq 1 | w_n < j\} = \min\{n \geq 1 | w_{n+1} \geq j\}$.

Proposition 13. *The mean duration $\mathbb{E}[T_{ss}]$ of a slow start phase and the mean number $\mathbb{E}[d_{ss}]$ of segments sent during a slow start phase are given by*

$$\mathbb{E}[T_{ss}] = RTT \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \sum_{i=1}^{j-1} \pi(i,j)}{p_0 \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)} \tag{14}$$

$$\mathbb{E}[d_{ss}] = \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \sum_{n=1}^{n_j} w_n (1-p)^{d_{n-1}} \pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}. \tag{15}$$

Proof. By definition of $\mathbb{E}[T_{ss}]$, we have $\mathbb{E}[T_{ss}] = RTT \times \mathbb{E}[N_A]$, where N_A denotes the number of states visited during a sojourn in subset A, which corresponds to the slow start states. Following Remark 4, Relation (9) and as $\pi_{E^0} P_{E^0,A} + \pi_A P_A = \pi_A$, we have

$$\mathbb{E}[N_A] = \frac{\pi_A \mathbb{1}_A}{\pi_A (I - P_A)\mathbb{1}_A} = \frac{\pi_A \mathbb{1}_A}{\pi_{E^0} P_{E^0,A} \mathbb{1}_A} = \frac{\sum_{(i,j)\in A} \pi(i,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} p_0 \pi(0,j)} = \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \sum_{i=1}^{j-1} \pi(i,j)}{p_0 \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)},$$

Consider now $\mathbb{E}[d_{ss}]$. Let us denote by Z_A the state of subset A by which a sojourn in A begins. These states are necessarily the states $(1,j)$ for $j = 1, \ldots, \lfloor W_{\max}/2 \rfloor$. From Remark 4, Relation (8), $\mathbb{P}(Z_A = (1,j))$ is equal to the entry $(1,j)$ of the vector $\pi_A(I-P_A)/[\pi_A(I-P_A)\mathbb{1}_A]$, that is

$$\mathbb{P}(Z_A = (1,j)) = \frac{[\pi_A(I - P_A)](1,j)}{\pi_A(I - P_A)\mathbb{1}_A} = \frac{[\pi_{E^0} P_{E^0,A}](1,j)}{\pi_{E^0} P_{E^0,A} \mathbb{1}_A} = \frac{p_0 \pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} p_0 \pi(0,j)} = \frac{\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}.$$

Now, if the slow start phase initiates by state $(1,j)$ then the maximum number of rounds in that phase is equal to n_j. For $n < n_j$, that phase lasts n rounds (i.e. no loss occurs during the $n-1$ first rounds of this phase and a loss occurs in the n-th round) with probability $(1-p)^{d_{n-1}}(1-(1-p)^{w_n})$. It lasts n_j rounds (i.e. no loss occurs during the $n_j - 1$ first rounds of this phase) with probability $(1-p)^{d_{n_j-1}}$ since in that case the n_j-th round is the last one, a loss occurring or not. The number of segments sent during a slow start phase of n rounds is d_n. Thus, if N denotes the number of rounds in a slow start phase, we have

$$\begin{aligned}
\mathbb{E}[d_{ss} \mid Z_A = (1,j)] &= \sum_{n=1}^{n_j} d_n \mathbb{P}(N = n) \\
&= \sum_{n=1}^{n_j - 1} d_n (1-p)^{d_{n-1}} (1 - (1-p)^{w_n}) + d_{n_j} (1-p)^{d_{n_j - 1}} \\
&= \sum_{n=1}^{n_j - 1} d_n \left((1-p)^{d_{n-1}} - (1-p)^{d_n} \right) + d_{n_j} (1-p)^{d_{n_j - 1}} \\
&= \sum_{n=1}^{n_j} (d_n - d_{n-1})(1-p)^{d_{n-1}} \\
&= \sum_{n=1}^{n_j} w_n (1-p)^{d_{n-1}}.
\end{aligned}$$

The result follows by writing

$$\mathbb{E}[d_{ss}] = \sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \mathbb{E}[d_{ss} \mid Z_A = (1,j)] \mathbb{P}(Z_A = (1,j)).$$

□

Proposition 14. *If no loss occurs during a slow start phase then the mean duration* $\mathbb{E}[T_{ss}^{\max}]$ *of that phase, the mean number* $\mathbb{E}[d_{ss}^{\max}]$ *of segments sent during that phase and the mean window*

size $\mathbb{E}[W_{ss}^{\max}]$ *reached at the end of that phase are given by*

$$\mathbb{E}[d_{ss}^{\max}] = \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} d_{n_j}\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}, \mathbb{E}[T_{ss}^{\max}] = RTT\frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} n_j\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}, \mathbb{E}[W_{ss}^{\max}] = \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} w_{n_j}\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}$$

and the probability p_{ssloss} *that no loss occurs during a slow start phase is*

$$p_{ssloss} = \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (1-(1-p)^{d_{n_j}})\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}. \tag{16}$$

Proof. These results are easily obtained by using the same arguments as in the proof of Proposition 13. □

Proposition 15. *The mean number* $\mathbb{E}[d_{ss}^0]$ *of successfully transmitted segments during a slow start phase is given by*

$$\mathbb{E}[d_{ss}^0] = \frac{1-p}{p}\frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \left(1-(1-p)^{d_{n_j}}\right)\pi(0,j)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \pi(0,j)}. \tag{17}$$

Proof. We follow the steps used to obtain $\mathbb{E}[d_{ss}]$ in the proof of Proposition 13. Let us denote by K the number of segments lost during the last round of a slow start phase. We have $\mathbb{E}[d_{ss}^0] = \mathbb{E}[d_{ss}] - \mathbb{E}[K]$. If the slow start phase initiates by state $(1,j)$ then the maximum number of rounds in that phase is equal to n_j. Suppose that such a phase lasts n rounds, then we have $K \in \{0,1,\ldots,w_n\}$ and thus, if N denotes the number of rounds in a slow start phase, we have

$$\begin{aligned}
\mathbb{E}[K \mid Z_A = (1,j)] &= \sum_{n=1}^{n_j} \mathbb{E}[K1_{\{N=n\}}] = \sum_{n=1}^{n_j}\sum_{k=1}^{w_n} k\mathbb{P}(K=k, N=n) \\
&= \sum_{n=1}^{n_j}\sum_{k=1}^{w_n} k(1-p)^{d_{n-1}}(1-p)^{w_n-k}p \\
&= \sum_{n=1}^{n_j} p(1-p)^{d_{n-1}+w_n-1}\sum_{k=1}^{w_n} k(1-p)^{-k+1} \\
&= \sum_{n=1}^{n_j} p(1-p)^{d_{n-1}+w_n-1}\frac{1-(1-p)^{-w_n}\left(1+w_n\left(1-\frac{1}{1-p}\right)\right)}{\left(1-\frac{1}{1-p}\right)^2} \\
&= \sum_{n=1}^{n_j} w_n(1-p)^{d_{n-1}} - \frac{1-p}{p}\sum_{n=1}^{n_j}(1-p)^{d_{n-1}}\left(1-(1-p)^{w_n}\right) \\
&= \mathbb{E}[d_{ss} \mid Z_A = (1,j)] - \frac{1-p}{p}\sum_{n=1}^{n_j}\left((1-p)^{d_{n-1}} - (1-p)^{d_n}\right) \\
&= \mathbb{E}[d_{ss} \mid Z_A = (1,j)] - \frac{1-p}{p}\left(1-(1-p)^{d_{n_j}}\right)
\end{aligned}$$

Finally, we get

$$\mathbb{E}[d_{ss}^0 \mid Z_A = (1,j)] = \frac{1-p}{p}\left(1-(1-p)^{d_{n_j}}\right),$$

and the result follows by unconditioning. □

6.4 Congestion avoidance

Proposition 16. *The mean duration $\mathbb{E}[T_{ca}]$ of a congestion avoidance phase and the mean number $\mathbb{E}[d_{ca}]$ of segments sent during a congestion avoidance phase are given by*

$$\mathbb{E}[T_{ca}] = \frac{\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\left(X_j(\alpha_{2j}+\alpha_{2j+1}) + X_{w_{n_j+1}}\beta_j\right)}{\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}(\alpha_{2j}+\alpha_{2j+1}+\beta_j)}$$

$$\mathbb{E}[d_{ca}] = \frac{\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\left(Y_j(\alpha_{2j}+\alpha_{2j+1}) + Y_{w_{n_j+1}}\beta_j\right)}{\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}(\alpha_{2j}+\alpha_{2j+1}+\beta_j)},$$

where

$$\alpha_i = \left(1-(1-p)^i\right)(1-q_i)\sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\pi(i,j) \text{ and } \beta_j = \pi(w_{n_j},j)(1-p)^{w_{n_j}},$$

$$\begin{cases} X_i = RTT(1-p)^{-\frac{bi(i-1)}{2}}\left(\sum_{w=i}^{W_{\max}-1}\lambda_w + \mu\right) \\ Y_i = (1-p)^{-\frac{bi(i-1)}{2}}\left(\sum_{w=i}^{W_{\max}-1} w\lambda_w + \mu W_{\max}\right) \end{cases} \text{ and } \begin{cases} \lambda_w = (1-p)^{\frac{bw(w-1)}{2}}\dfrac{1-(1-p)^{bw}}{1-(1-p)^w} \\ \mu = \dfrac{(1-p)^{\frac{bW_{\max}(W_{\max}-1)}{2}}}{1-(1-p)^{W_{\max}}}. \end{cases}$$

Proof. The time T_{ca} is shown in Figure 7. Let us denote by Z_B the state by which a congestion avoidance phase begins. We must have necessarily either $Z_B = (j,j)$ or $Z_B = (w_{n_j+1}, j)$, for $j = 2,\ldots,\lfloor W_{\max}/2\rfloor$ and it is possible that for some j, we have $j = w_{n_j+1}$. So, we have

$$\mathbb{E}[T_{ca}] = \sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\left(X_jQ_{(j,j)} + X_{w_{n_j+1}}Q_{(w_{n_j+1},j)}\right)1_{\{j\neq w_{n_j+1}\}} + \sum_{j=2}^{\lfloor W_{\max}/2\rfloor} X_jQ_{(j,j)}1_{\{j=w_{n_j+1}\}}, \quad (18)$$

and

$$\mathbb{E}[d_{ca}] = \sum_{j=2}^{\lfloor W_{\max}/2\rfloor}\left(Y_jQ_{(j,j)} + Y_{w_{n_j+1}}Q_{(w_{n_j+1},j)}\right)1_{\{j\neq w_{n_j+1}\}} + \sum_{j=2}^{\lfloor W_{\max}/2\rfloor} Y_jQ_{(j,j)}1_{\{j=w_{n_j+1}\}}, \quad (19)$$

where, for $i = j$ or w_{n_j+1}, we define $X_i = \mathbb{E}[T_{ca} \mid Z_B = (i,j)]$, $Y_i = \mathbb{E}[d_{ca} \mid Z_B = (i,j)]$ and $Q_{(i,j)} = \mathbb{P}(Z_B = (i,j))$. Note that for $j = 2,\ldots,\lfloor W_{\max}/2\rfloor$, we have $w_{n_j+1} = 2,\ldots,W_{\max}$.

Let us evaluate the quantities X_i, Y_i and $Q_{(i,j)}$. Y_i represents the mean number of segments sent during a congestion avoidance phase which begins with a congestion window size $W^c = i$. During that congestion phase this window may increase from i to $W_{\max}$. Suppose that the size of this congestion window is equal to w, for $w = i,\ldots,W_{\max}$. It increases by 1 every b rounds if no loss occurs. So, when the window size is w, then the w segments of the r-th round,

$r = 1, \ldots, b$, are sent if no loss occurred during all the preceding rounds of that congestion avoidance phase, which begins with $W^c = i$. If we denote by $N(w, r)$ the total number of segments contained in all these preceding rounds, we have

$$N(w,r) = b\left(\sum_{k=i}^{w-1} k\right) + (r-1)w = \frac{b(w(w-1) - i(i-1))}{2} + (r-1)w.$$

The probability that the segments of the r-th round of size w are sent is equal to $(1-p)^{N(w,r)}$. So, we have, for $i = 2, \ldots, W_{\max}$,

$$Y_i = \sum_{w=i}^{W_{\max}-1} \sum_{r=1}^{b} w(1-p)^{N(w,r)} + \sum_{r=1}^{+\infty} W_{\max}(1-p)^{N(W_{\max},r)}.$$

Note that when the size of congestion window is $W_{\max}$, the number of rounds is unbounded. After some algebra, we easily obtain,

$$\begin{aligned}
Y_i &= \sum_{w=i}^{W_{\max}-1} w\left((1-p)^{\frac{b(w(w-1)-i(i-1))}{2}} \sum_{r=1}^{b}(1-p)^{w(r-1)}\right) \\
&\quad + W_{\max}(1-p)^{\frac{b(W_{\max}(W_{\max}-1)-i(i-1))}{2}} \sum_{r=1}^{+\infty}(1-p)^{W_{\max}(r-1)} \\
&= (1-p)^{-\frac{bi(i-1)}{2}} \sum_{w=i}^{W_{\max}-1} w\left((1-p)^{\frac{bw(w-1)}{2}} \frac{1-(1-p)^{bw}}{1-(1-p)^{w}}\right) \\
&\quad + W_{\max}(1-p)^{-\frac{bi(i-1)}{2}} \frac{(1-p)^{\frac{bW_{\max}(W_{\max}-1)}{2}}}{1-(1-p)^{W_{\max}}} \\
&= (1-p)^{-\frac{bi(i-1)}{2}} \left(\sum_{w=i}^{W_{\max}-1} w\lambda_w + \mu W_{\max}\right)
\end{aligned}$$

where

$$\lambda_w = (1-p)^{\frac{bw(w-1)}{2}} \frac{1-(1-p)^{bw}}{1-(1-p)^{w}} \text{ and } \mu = \frac{(1-p)^{\frac{b}{2}W_{\max}(W_{\max}-1)}}{1-(1-p)^{W_{\max}}}.$$

Using similar arguments, we obtain, for $i = 2, \ldots, W_{\max}$,

$$X_i = RTT \sum_{w=i}^{W_{\max}-1} \sum_{r=1}^{b}(1-p)^{N(w,r)} + RTT \sum_{r=1}^{+\infty}(1-p)^{N(W_{\max},r)},$$

which leads to

$$X_i = RTT(1-p)^{-\frac{bi(i-1)}{2}} \left(\sum_{w=i}^{W_{\max}-1} \lambda_w + \mu\right).$$

Let us now determine the probability $Q_{(i,j)}$. The only states by which a congestion avoidance phase may start are state (j, j), reached after a TD loss, and state (w_{n_j+1}, j), reached after a slow start phase without loss, for $j = 2, \ldots, \lfloor W_{\max}/2 \rfloor$. A TD loss is caused by a transition either from a state $(2j, k)$ to a state (j, j) or from a state $(2j+1, k)$ to a state (j, j). A slow start phase without loss is caused by a transition from a state (w_{n_j}, j) to state (w_{n_j+1}, j). So, following Relation (8) in Remark 4, we get

- if $j \neq w_{n_j+1}$ then

$$Q_{(j,j)} = \frac{\alpha_{2j} + \alpha_{2j+1}}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)} \text{ and } Q_{(w_{n_j+1},j)} = \frac{\beta_j}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)}$$

- if $j = w_{n_j+1}$ then

$$Q_{(j,j)} = \frac{\alpha_{2j} + \alpha_{2j+1} + \beta_j}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)},$$

where

$$\alpha_{2j} = \sum_{k=2}^{\lfloor W_{\max}/2 \rfloor} \pi(2j,k) P_{(2j,k)(j,j)} = \sum_{k=2}^{\lfloor W_{\max}/2 \rfloor} \pi(2j,k) \left(1 - (1-p)^{2j}\right) (1 - q_{2j}),$$

in the same way,

$$\alpha_{2j+1} = \sum_{k=2}^{\lfloor W_{\max}/2 \rfloor} \pi(2j+1,k) \left(1 - (1-p)^{2j+1}\right) (1 - q_{2j+1}),$$

where, we set $\pi(l,k) = 0$ for $l > W_{\max}$, and

$$\beta_j = \pi(w_{n_j}, j) P_{(w_{n_j},j)(j,j)} = \pi(w_{n_j}, j)(1-p)^{w_{n_j}}.$$

Replacing $Q_{(i,j)}$ in Relations (18) and (19), we get

$$\begin{aligned}
\mathbb{E}[T_{ca}] &= \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \left(X_j(\alpha_{2j} + \alpha_{2j+1}) + X_{w_{n_j+1}} \beta_j\right) 1_{\{j \neq w_{n_j+1}\}}}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)} \\
&\quad + \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} X_j(\alpha_{2j} + \alpha_{2j+1} + \beta_j) 1_{\{j = w_{n_j+1}\}}}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)} \\
&= \frac{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} \left(X_j(\alpha_{2j} + \alpha_{2j+1}) + X_{w_{n_j+1}} \beta_j\right)}{\sum_{j=2}^{\lfloor W_{\max}/2 \rfloor} (\alpha_{2j} + \alpha_{2j+1} + \beta_j)}
\end{aligned}$$

and, in the same way, we obtain the desired expression of $\mathbb{E}[d_{ca}]$. □

The following proposition is well-known.

Proposition 17. *The number d_{ca}^0 of successfully transmitted segments during a congestion avoidance phase has a geometric distribution with parameter $1-p$, so $\mathbb{E}[d_{ca}^0] = (1-p)/p$.*

6.5 Send rate and goodput

Proposition 18. *The send rate ρ and the goodput ρ_0 can be expressed as*

$$\rho = \frac{\mathbb{E}[d_{to}] + \mathbb{E}[d_{E^0}^{back}] + N_{loss}\mathbb{E}[d_r]}{\mathbb{E}[T_{to}] + \mathbb{E}[T_{E^0}^{back}] + RTT(N_{loss} - 1)p_{rr}}, \tag{20}$$

$$\rho_0 = \frac{1 + \mathbb{E}[d_{E^0}^{back,0}] + N_{loss}\mathbb{E}[d_r^0]}{\mathbb{E}[T_{to}] + \mathbb{E}[T_{E^0}^{back}] + RTT(N_{loss} - 1)p_{rr}}. \tag{21}$$

Proof. From Figure 7, we easily get the ratio of the first terms in expressions (20) and (21). The last terms, where N_{loss} appears, are due to the residual rounds. In counting the mean number of segments transmitted (resp. successfully transmitted) during a cycle, we also need to take into account the mean number of segments constituting the residual rounds generated by the N_{loss} loss detections. This mean number of segments is equal to $N_{loss}\mathbb{E}[d_r]$ (resp. to $N_{loss}\mathbb{E}[d_r^0]$).

For what concerns the mean cycle duration, it is increased by $p_{rr}RTT$ for each of the $(N_{loss} - 1)$ TD losses, because the TO loss residual round is taken into account in next time-out period, as shown in Figure 12. □

Remark 5. Let us denote by n_{ca} the random variable equal to the number of ca phases in a cycle, that is $N_{ca} = \mathbb{E}[n_{ca}]$. Whereas it is clear that $\mathbb{E}[T_{E^0}^{back}] = \mathbb{E}[T_{ss}] + \mathbb{E}[n_{ca}T_{ca}]$, our numerical results have shown that $\mathbb{E}[T_{E^0}^{back}]$ is very closed to $\mathbb{E}[T_{ss}] + N_{ca}\mathbb{E}[T_{ca}]$, which means that n_{ca} and T_{ca} can be considered as independent. The same results hold for variables n_{ca} and d_{ca} and for variables n_{ca} and d_{ca}^0.

7 Numerical Results

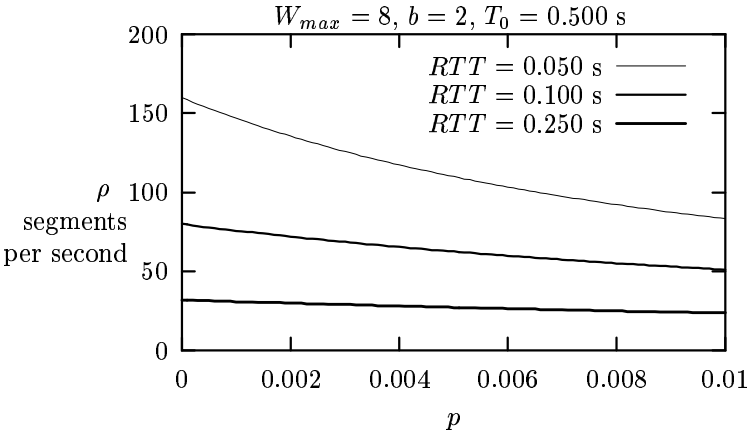

Figure 13: *Evolution of the send rate ρ for different values of RTT.*

Figure 13 : The send rate ρ gets equal to $W_{\max}$ segments per RTT ($W_{\max}/RTT$ segments per second) when loss probability p is close to zero, and converges to zero when p increases. Moreover, the shorter RTT is , the more segments are sent per second (quick acknowledgments).

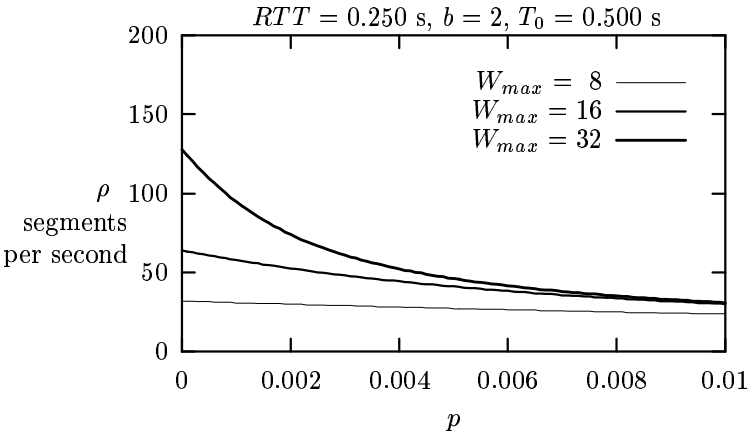

Figure 14: *Evolution of the send rate ρ for different values of $W_{\max}$.*

Figure 14 : When $W_{\max}$ increases, the window size can reach higher values and the mean throughput naturally increases too. Note that for small values of the loss probability p, ρ reaches $W_{\max}/RTT$ segments per second, and for large values of p, ρ seems to be less dependent on $W_{\max}$. Indeed, for $p = 0,01$, ρ gets close to 20 or 30 segments per second, that is around 6 segments per RTT for $W_{\max} = 8, 16, 32$.

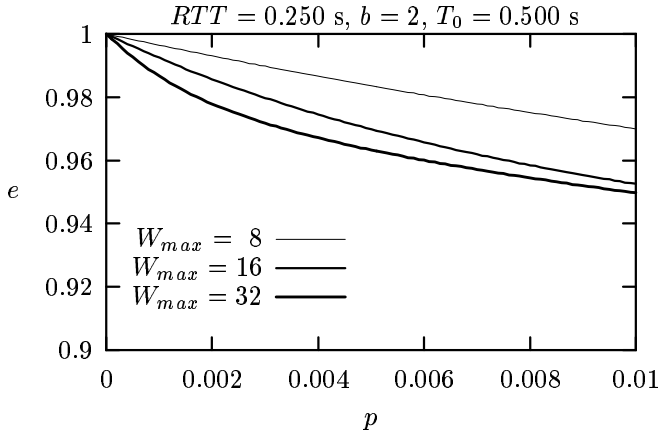

Figure 15: *Evolution of the efficiency $e = \rho_0/\rho$.*

Figure 15 : In all the examples considered, the efficiency $e = \rho_0/\rho$ appears to be quite independent of RTT, but we can see here that e decreases when $W_{\max}$ increases. In fact, the higher the bandwidth is, the faster you may transmit data, but the bigger the number of retransmissions is. Indeed, when the window size is large and one of the first segments gets lost, then all the following ones get lost too.

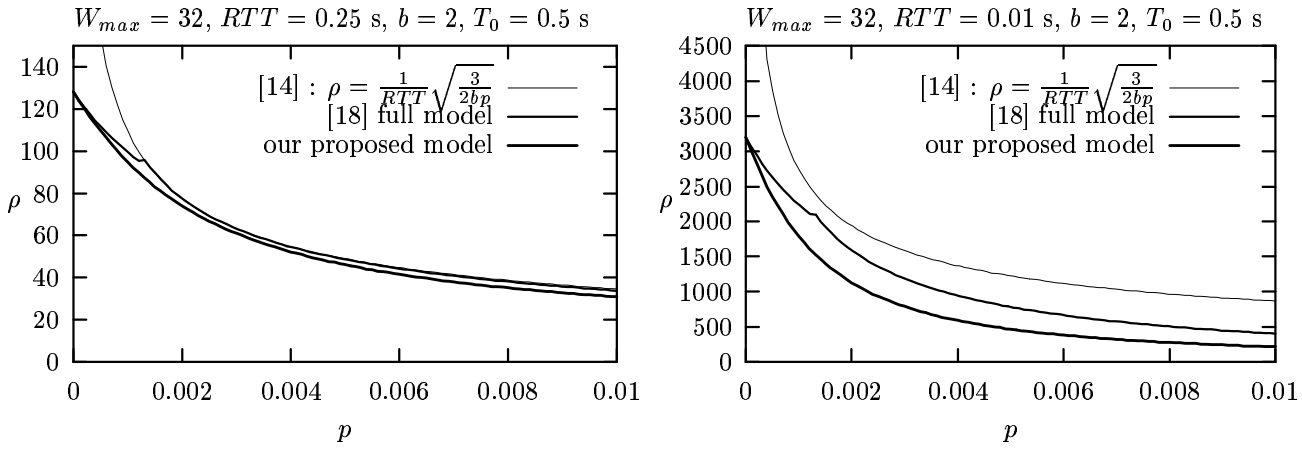

Figure 16: *Comparison to other models.*

Figure 16 : Note that the throughput of our model, evaluated with less simplifications, is lower than the one obtained by the authors of [18]. But the higher RTT is, the closer the different models are.

We show in Figures 17 and 18 other improvements obtained by using our Markov model.

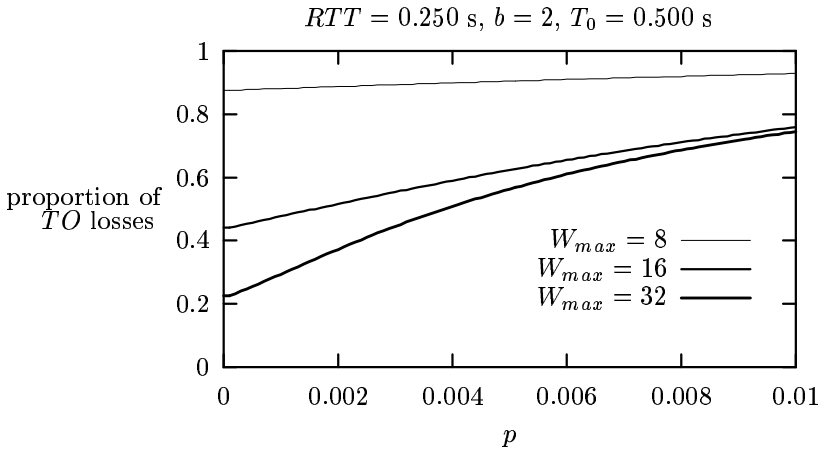

Figure 17: *Proportion of TO-type losses.*

Figure 17 : The more $W_{\max}$ gets high, the more TD events may occur, because residual rounds contain more segments and these residual rounds are the ones that generate duplicate ACKs.

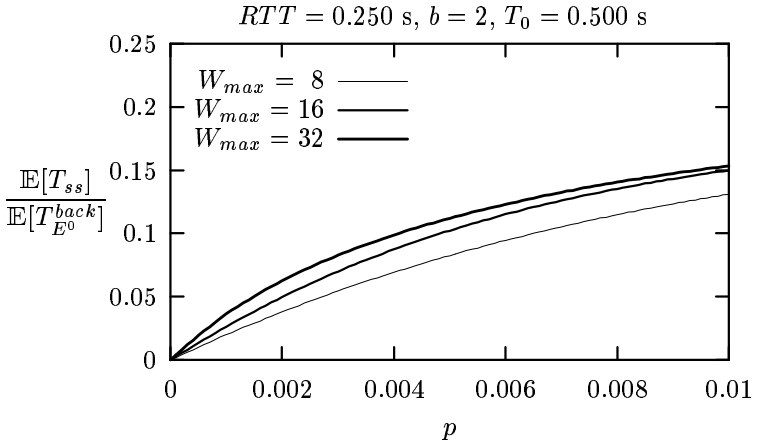

Figure 18: *Importance of slow start : proportion of time in each cycle.*

Figure 18 : Though the proportion of time spent in slow start per cycle does not appear to depend on RTT, we can notice that it seems to be sensitive to $W_{\max}$ in the sense that when $W_{\max}$ gets higher, slow start phases can reach higher thresholds and thus last longer (whereas in congestion avoidance, the bigger the window size is, the higher the probability that a loss appears, stopping thus the congestion avoidance phase). But the main remark is that the slow start phase may reach 10 or 15 per cent of a cycle duration. This shows that it must be taken into account for the performance evaluation of TCP.

8 Conclusion

The main assumption we made is that the connection is established in a high speed and wide area (large RTT) network. Indeed, the time needed to send all segments in congestion window and the time interval between ACKs must be significantly low compared to the round trip time for the identification of separated bursts, called and defined as *rounds*.

Moreover, we supposed that the loss probability p is independent to the window size, because in high capacity networks, the load of a single connection is not responsible of congestion. Concerning loss correlation (when a segment gets lost, all the following ones in the same round also get lost), we apply our model to high capacity and high speed networks with drop-tail routers, in which the connection is not the cause of congestion and packets of a given round arrive in burst in the overflowed router. And despite multiplexing, a router remains full as long as packets of the same window arrive and thus rejects all of them.

With these assumptions, we have been able to obtain analytical expressions for the send rate and goodput of a long term steady-state connection (stationary regime), for the mean duration and the mean number of segments sent and acknowledged in each time-out, slow start and congestion avoidance phase. This precise description of TCP allows an accurate study of its performance in many high speed and wide area networks.

References

[1] Alhussein A. Abouzeid, Murat Azizoglu, and Sumit Roy. Stochastic Modeling of a Single TCP/IP Session over a Random Loss Channel. In *DIMACS workshop on Mobile Networks*

and Computing, Rutgers University, New Jersey, US, March 1999.

[2] Alhussein A. Abouzeid, Sumit Roy, and Murat Azizoglu. Stochastic Modeling of TCP over Lossy Links. In *INFOCOM'00*, Tel-Aviv, Israel, March 2000.

[3] Omar Ait-Hellal, Eitan Altman, Driss Elouadghiri, Mohamed Erramdani, and Nouffisa Mikou. Performance of TCP/IP : the case of two Controlled Sources. In *ICCC'97*, Cannes, France, November 1997.

[4] Eitan Altman, Kostya Avrachenkov, and Chadi Barakat. A stochastic model of TCP/IP with stationary ergodic random losses. Technical Report RR-3824, INRIA, November 1999.

[5] Eitan Altman, Kostya Avrachenkov, and Chadi Barakat. TCP in presence of bursty losses. Technical Report RR-3142, INRIA, July 1999.

[6] Eitan Altman, Jean Bolot, Philippe Nain, Driss Elouadghiri, Mohamed Erramdani, Patrick Brown, and Dennis Collange. Performance Modeling of TCP/IP in Wide-Area Network. Technical Report RR-3142, INRIA, 1997.

[7] François Baccelli and Dohy Hong. TCP is Max-Plus Linear. Technical Report RR-3986, INRIA, 2000.

[8] Patrick Brown. Resource sharing of TCP connections with different round trip times. In *INFOCOM'00*, Tel-Aviv, Israel, March 2000.

[9] Neal Cardwell, Stefan Savage, and Thomas Anderson. Modeling TCP latency. In *INFOCOM'00*, Tel-Aviv, Israel, March 2000.

[10] Douglas Comer. *Internetworking with TCP/IP, Volume 1 : Principles, Protocols, and Architecture, 3rd edition.* Prentice-Hall, 1995.

[11] Paul Hurley, Jean-Yves Le Boudec, and Patrick Thiran. A Note on the Fairness of Additive Increase and Multiplicative Decrease. In *ITC-16*, Edinburgh, Scotland, June 1999.

[12] Anurag Kumar. Comparative Performance Analysis of Versions of TCP in a Local Networks with a Lossy Link. *IEEE/ACM Transactions on Networking*, 6(4), August 1998.

[13] T. V. Lakshman and Upamanyu Madhow. The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss. *IEEE/ACM Transactions on Networking*, 5(3), June 1997.

[14] Matthew Mathis, Jeffrey Semke, Jamshid Mahdavi, and Teunis Ott. The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm. *Computer Communications Review*, 27(3), July 1997.

[15] Vishal Misra, Wei-Bo Gong, and Don Towsley. Stochastic Differential Equation Modeling and Analysis of TCP-Windowsize Behavior. In *Performance'99*, Istanbul, Turkey, October 1999.

[16] Teunis J. Ott, J.H.B. Kemperman, and Matt Mathis. *The stationary behavior of ideal TCP congestion avoidance*, August 1996. http://www.argreenhouse.com/papers/tjo/.

[17] Jitendra Padhye, Victor Firoiu, and Don Towsley. A stochastic model of TCP Reno congestion avoidance and control. Technical Report 99-02, University of Massachussets, 1999.

[18] Jitendra Padhye, Victor Firoiu, Don Towsley, and Jim Kurose. Modeling TCP throughput : a simple model and its empirical validation. In *SIGCOMM'98*, Vancouver, Canada, September 1998.

[19] Gerardo Rubino and Bruno Sericola. Sojourn times in Markov processes. *Journal of Applied Probability*, 26, 1989.

[20] W. Richard Stevens. *TCP/IP Illustrated : Vol.1 The Protocols.* Addison-Wesley, 1994.

[21] W. Richard Stevens. *TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms*, January 1997. RFC 2001.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399