A TCP Model for Short-Lived Flows to Validate Initial Spreading

Renaud Sallantin, Cédric Baudoin, Emmanuel Chaput, Fabrice Arnal,
Emmanuel Philippe Dubois, André-Luc Beylot

A TCP Model for Short-Lived Flows to Validate Initial Spreading

Renaud Sallantin*,
Cédric Baudoin‡, Emmanuel Chaput*, Fabrice Arnal‡, Emmanuel Dubois† and André-Luc Beylot*

* Université de Toulouse - IRIT
Email: {sallantin, chaput, beylot}@enseeiht.fr
† CNES
Email: {emmanuel.dubois}@cnes.fr
‡ Thales Alenia Space
Email: {cedric.baudoin, fabrice.arnal}@thalesaleniaspace.com

***Abstract*— With a vast majority of Internet connections shorter than 10 segments, designing a new fast start-up TCP mechanism is a major concern. While enlarging the Initial Window (IW) up to 10 segments is the fastest solution to deal with a short-lived connection in uncongested networks, numerous researchers are concerned about the impact of the large initial burst on congested networks.**

We designed Initial Spreading to remove those concerns. The initial empirical evaluation showed the potential of Initial Spreading in performing similarly to a large IW in uncongested networks without its adverse effect in congested networks. However, these conclusions were based on empirical data, and considering the implications to the TCP performance a more thorough evaluation is necessary.

In this paper, we propose a TCP model for short-lived flows to theoretically evaluate the impact of the bursts on the individual performance. The model is then used to confirm that Initial Spreading takes full advantages of the burstiness of TCP to offer significant improvements.

***Index Terms*—TCP; Fast Startup; burst; model; Initial Window; RTT; congestion**

I. Introduction

Today, a vast majority of the web objects, and thus Internet connections are shorter than 10 segments [1]. Improving the Transmission Control Protocol (TCP), in particular its behavior and efficiency at the beginning of a connection, is therefore a major concern. Many fast start-up mechanisms have been proposed to improve, circumvent or even replace the slow and conservative original slow-start stage.

In an uncongested network, enlarging the Initial Window (IW) and then sending up to 10 segments as soon as the connection is established is the fastest solution to transmit a short-lived connection. Some studies therefore encourage its global use in the Internet [2]. Nevertheless, many researchers are concerned about the consequences of releasing a large burst of segments in traffic as sporadic as TCP traffic, and support a more conservative approach [3].

Based on traffic observations, simulations and experiments, Initial Spreading [4] [5] has been designed to remove those concerns. The initial empirical evaluation showed the potential of Initial Spreading in performing similarly to a large IW in uncongested networks without its adverse effect in congested networks. However, these conclusions were based on empirical data, and considering the implications to the TCP performance a more thorough evaluation is necessary.

In this paper, we propose a TCP model for short-lived flows to theoretically evaluate the impact of the bursts, and notably of the large initial bursts, on the individual performance.

Indeed, even if the modeling of TCP behavior has received considerable attention in recent years [6] [7], firstly focusing on its steady state, then on the slow-start stage, few models have actually focused on short-lived TCP flows [8], or provided an accurate evaluation of the bursts impact on the average performance [9] [10]. This paper therefore proposes an original analytical model that focuses on describing the bursts and their impacts in order to provide an accurate estimation of the average duration of short-lived TCP connections, according to the selected start-up mechanisms.

Section II describes our work on the Initial Spreading design. Section III analyzes the bursts in detail to explain why a TCP model for short-lived flows must focus on the bursts. Sections IV and V present our analytical model and corroborate our empirical work on Initial Spreading, confirming that taking full advantage of TCP's burstiness, Initial Spreading significantly improves the short-lived flows performance.

II. Initial Spreading: a Fast Start-Up TCP mechanism

A. The original idea

Whether due to a long delay or a large queuing latency, a long Round Trip Time (RTT) deteriorates regular slow-start performance. This particularly impacts the short-lived connections.

The original idea of Initial Spreading [5] was to consider the RTT as a resource to exploit, rather than as a constant to bypass. As soon as the RTT is larger than a few milliseconds, it can therefore be used as an opportunity to safely send a large amount of data during the first RTT after the connection establishment. Spacing the data along the RTT would in fact hopefully un-correlate the sent segments and therefore

enable a high independent probability for each segment to be successfully transmitted.

B. Initial Spreading design

Initial Spreading mechanism uses the permitted upper bound value of the TCP's IW to space out a number of segments smaller or equal to this value across the first RTT before letting the TCP algorithm continue conventionally. Its simple algorithm has three steps:

1) The RTT is measured during the SYN-SYN/ACK exchange.
2) According to the RTT value, a Spreading Time ($T_{spreading}$) is computed. Depending on the number of segments to be sent, until n segments are sent every $T_{spreading}$, with n equal to the IW size.
3) After the transmission of the IW, the regular TCP algorithm is used.

$T_{spreading}$ [5] is large enough for two successive segments to be considered as un-correlated, i.e. they are not belonging to a same burst and have therefore an independent probability to be successfully delivered. Thus, no large initial burst downgrades the transmission of short-lived connections, but bursts continue to prevent an overload of the network in the case of long-lived connections. This enables Initial Spreading to not suffer from large IW or Pacing flaws [11].

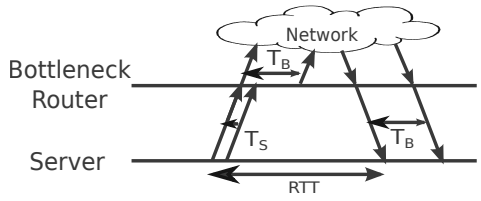

Figure 1. Timers and diagrams explanations

In the following, T_B and T_S respectively denote the time to forward a packet at the bottleneck rate and at the sender rate as illustrated on Fig.1. Both values are therefore independent of the current congestion.

Fig. 2 presents the behavior of the Initial Spreading in comparison with a large TCP's IW and also with Pacing with a large IW, when transmitting 12 segments.

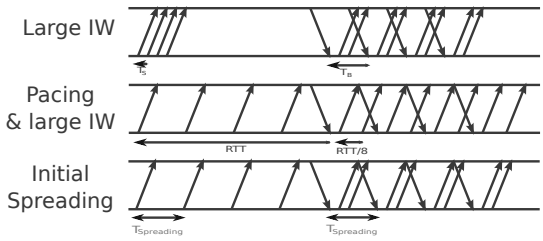

Figure 2. Time diagram illustrating the transmission of 12 segments with the three different mechanisms using an IW of four segments

An extensive set of simulations [4] and real experiments [5] have shown the benefits of using Initial Spreading instead of a regular large IW, or even Pacing. However, these conclusions were based on empirical data, and considering the implications to the TCP performance a more thorough evaluation is necessary.

In the following, we accurately study the burst phenomenon, and then introduce an analytical model that focuses on the short-lived flows and enables to take into account the burst impact on the individual connections. This model is used in section V to validate the convincing results we empirically observed when using Initial Spreading.

III. Burst Analysis

It is commonly admitted that the burstiness of TCP has a major impact on its global performance. However, the impact of an initial burst on a short-lived connection performance remains relatively unstudied.

A. Burst definition

In the literature, the " burst" definition is generally related to the "round" definition. A round begins with the transmission of a window of segments and ends with receipt of one or more ACKs. The segments sent during a same round form a burst [12] [9].

This definition, probably adapted to model the TCP steady state, does not suite an accurate modeling of the short-lived flows, notably when considering different fast Start-Up TCP mechanisms. Indeed, with such a definition, sending an IW of 10 segments, with or without Pacing would be similar in a burst point of view. Any model using this burst definition would therefore not be able to illustrate the measured and well-known differences.

For the remainder of this paper, we consider therefore that segments are belonging to a same burst because they impact the network, rather than because they have been sent in a same round. Thus, we assume that two segments that encounter independent bottleneck buffer state are not belonging to the same burst, whatever the round they belong.

B. Main types of burst

The slow start is the first stage of a TCP connection. It is used to probe an unknown network and reach a maximum bit rate as quickly as possible. During this phase, each expected ACK increases the sender CWND size by one segment, enabling the sender to transmit two new segments at its own bit rate. Since the ACKs are generated at the bottleneck bit rate, for each expected ACK received, the sender sends twice as many segments as the bottleneck router can forward (see Fig. 2).

According to our previous burst definition, the sender does therefore not transmit a burst of $CWND$ segments in one RTT but $\frac{CWND}{2}$ bursts of 2 segments. Those mini-bursts are spaced enough not to belong to a same large burst.

The IW is the second leading cause of bursts. It corresponds to the number of segments that are sent, at the sender bit rate, after the connection establishment. An IW of size n is then responsible for a burst of size n.

C. Accurate study of the burst impact

To appreciate the cost of a burst in congested networks, we performed a set of real experiments using a Dumbbell network topology (see Fig. 3) as test bed with a bottleneck bit rate equal to 10Mbps and other link bit rates equal to 100Mbps.

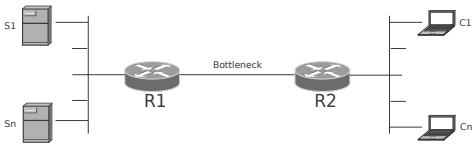

Figure 3. Dumbbell Network topology

8 parallel long-lived connections are used to generate the background traffic. Those connections are responsible for an average loss rate of 5%, and an average bottleneck buffer occupancy of 85%. We then transmitted different flow sizes in one burst, and observed the impact of the burst. For each flow size, 100 000 iterations have been done.

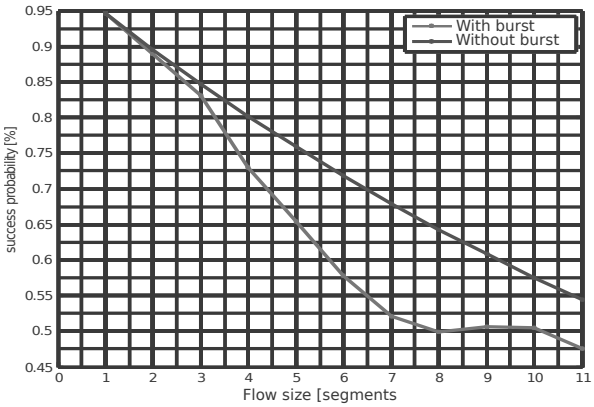

Figure 4. Probability to have no loss as a function of the flow size, with and without burst

Fig. 4 shows the average probability of successfully transmit all the segments of a burst as a function of its size, and compares it to a theoretical case in which segments losses are independent:

- for flow size lower than 4 segments, both curves are very close.
- for longer flows, sending a burst of segments significantly reduce the success probability

To further deepen our understanding of the bursts, Fig. 5 underlines the correlation of the losses between the segments sent in one burst. Fig. 5 shows the measured probability to have 1,2 or 3 segments of a burst correctly received whereas one of the previous segments of the burst has been lost. We plotted it according to the position of the first loss in a burst of 10 segments.

Two main results shall be noted:

- the probability of successfully sending a segment after a loss is different from zero.

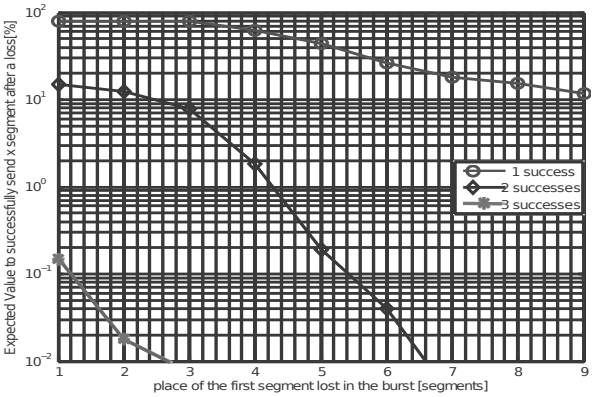

Figure 5. Expected estimation of having X successes as a function of the place of the first loss in a burst of 10 segments ($X \in 1, 2, 3$)

- whatever the position of the first loss, the probability of having 3 of the following segments of the burst successfully transmitted is almost zero. So considering the transmission of a short-sized flow in one burst, the Fast Retransmit and Fast Recovery algorithms cannot be triggered.

In conclusion, transmitting segments in a burst correlates the segments losses, increases the loss probability and reduces the possibility of using regular TCP fast recovery mechanisms. Bursts are therefore significantly downgrading the average performance of a short-lived connection.

D. Burst modeling

When it is question of burst modeling for links with dropTail queuing, the "bursty loss model" is regularly adopted [6]. It assumes that when no previous loss occurred in a round, all segments of the round have the same probability of being lost, independently of any other considerations (number of segments in the window, position in the burst,...). However, any segment lost means that all the subsequent segments of the round are lost too, while losses in one round are independent of the losses in any other round.

According to our burst definition and observations, a burst is different from a round, and several independent bursts can occurred in one round. We therefore use a slightly revised flavor of the "bursty loss model", using the same rules but applied to the bursts instead of to the rounds.

Considering the previous subsection, this burst model is inaccurate and notably in case of large bursts. Nevertheless, our TCP model mainly aims to estimate Initial Spreading impact for the short-lived flows, which is based on short-sized bursts. We assume therefore that using this burst modeling has a minor impact on the results.

IV. Short-lived TCP flow modeling

The objective of our analytical model is to estimate the average delivery delay, i.e. the average time it takes a source to successfully send i segments with an IW of size n [8]. Our model focuses on the short-lived connections in a congested

network and targets the burst impact on the transmission of segments in order to be able to distinguish but also accurately depict the TCP performance with and without Initial Spreading.

A. Assumptions

This study aims to model the short-lived connections (around 10 segments). This has numerous impacts on the assumptions we made.

We assume then that the bit rate is regulated only by the sender CWND size and not by the receiver window size.

In regards to the short size of the TCP connection, we assume that when Fast Retransmit and Fast Recovery algorithms used in association with the Selective Acknowledgement (SACK) [13] can be triggered, they are sufficient to recover from the small amount of data lost. The reception of 3 duplicated ACKs (DUP ACK) enables the fast re-transmit of the first lost segment, and then the entry into Fast Recovery. In this mode, each new acknowledged segment leads to the re-transmission of one of the lost segments.

Thus, if the number of DUP ACKs does not make it possible to enter into Fast Recovery, then a loss is recovered using a slow-start with a CWND of 1 segment after the expiration of a Retransmission Time Out (RTO), otherwise, the first lost segment is resent on receipt of the third DUP ACKs.

[14] is used to define the RTO value. The RTO is initially equal to 1s before taking the maximum value between 1s and the value computed from the RTT measurements. We assume 1s to be the maximum value in the majority of our tests. Then, for the remainder of this paper, we consider the RTO is equal to 1s. Following [14], there is a sole retransmission Timer that is set off each time an expected ACK is received.

The bit rate difference between the bottleneck link and the rest of the network has an important impact on the performance analysis. As seen in section III, we considered that n segments can only be sent in 3 different ways:

- in one burst of n segments if they correspond to the IW of size n and that Initial Spreading is not used
- independently (i.e. each segment does not belong to a burst) if the IW is sent with Initial Spreading
- by mini-bursts of 2 segments regarding all the other cases

The bursty loss model is used for segments of a same burst.

In the following, we assume that both T_S and T_B (see Fig. 1) are negligible in comparison with the RTT and the delay induced by the spreading due to Initial Spreading. The model will therefore only take the spreading value into account.

B. Model description

In its first stage, TCP can be modeled using a finite state transition system. Each state corresponds to the delivery of a certain amount of segments with a limited Window (IW or CWND) and a specific type of burst management. Solely two actions lead to a state change: the reception of all the ACKs of the previous sent window or the expiration of the Retransmission Timer. After each transition, a new state is reached that corresponds to the delivery of the segments that still have to be sent with an updated CWND and an updated burst management.

We defined two different start states:

- D_i^n: i segments are delivered with an IW of size n. Initial Spreading is not used, and so the IW is transmitted in one burst
- S_i^n: i segments are delivered with an IW of size n. Initial Spreading is used, and no burst occurred in the transmission of the IW.

The change in the slow-start burstiness, due to the difference of bit rate, has an important impact on the short-lived flows performance, and needs to be accurately depicted. Thus, for all the intermediate states, we introduce:

- B_i^n: i segments are delivered with a CWND of size n. The CWND is transmitted in $\lfloor \frac{n+1}{2} \rfloor$ bursts of 2 segments.

The final state is reached when all the segments have been delivered.

A couple of values is associated to each transition: the associated probability to change from one state to another and the required time.

Finally, we calculated $T(D_i^n)$ and $T(S_i^n)$, the average delivery delays, corresponding to the expected duration to change from an initial state to the final state.

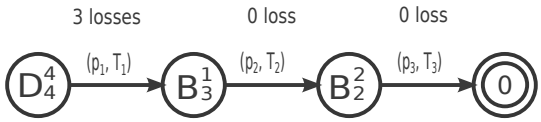

Figure 6. One of the possible scenarios to achieve D_4^4

Fig. 6 depicts one of the possible scenarios that can occurred when delivering D_4^4. The first state change is related to the expiration of the Retransmission timer, resulting from the loss of 3 segments. Then the 3 lost segments have to be re-transmitted with a CWND of 1 segment. The second transition happens when the ACK arrived and the new state corresponds to the delivery of the last 2 segments with an updated CWND of 2 segments. Then the reception of the 2 ACKs affords to enter in the final state. For this peculiar case, the delivery delay is equal to $T_1 + T_2 + T_3$ s.

To describe our model, the following variables are used:

- R: Average RTT
- T_0: Retransmission Timer after a Retransmission Time Out (RTO)
- p: data segment dropping probability
- $q = 1 - p$: success probability

C. Initialization

By definition,

$$\forall i, \begin{cases} D_i^1 = S_i^1 \\ T\left(D_i^1\right) = T\left(S_i^1\right) \end{cases}$$

A regular slow-start is used with or without Initial Spreading, so a burst of 2 segments is sent each time an ack is correctly received (see section III.b). Thus, for an IW of 1

segment, the acknowledgement of the first segment leads to the transmission of the remaining $i-1$ segments with an IW of 2:

$$\forall i, \quad T\left(D_i^1\right) = T\left(D_1^1\right) + T\left(B_{i-1}^2\right)$$

Considering that the Time Out is doubled when a loss has not been recovered at the Retransmission Timer expiration, the average time to successfully transmit one segment, depending on p, R and T_0 is equal to:

$$T\left(D_1^1\right) = R + q\sum_{i=1}^{\infty} p^i \sum_{j=1}^{i} 2^{j-1} T_0 = R + T_0 \frac{p}{1-2p}$$

When sending 2 segments with an IW of size 2, different cases have to be considered.

Fig. 7 shows that without Initial Spreading, the loss of the second segment means a re-transmission at $R+T_0$, while the loss of the first segment that implies the loss of the complete burst means the re-transmission of both segments at T_0. In the first case, the Retransmission Timer has been set off by the ACKs of the first sent segment.

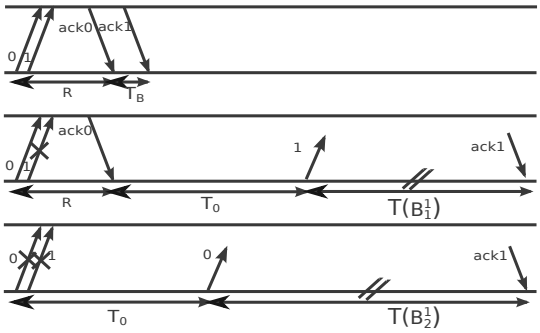

Figure 7. $T\left(D_2^2\right)$, the different scenarios

Fig. 8 gives the transition diagram for D_2^2.

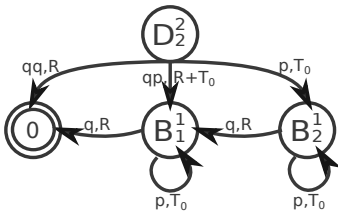

Figure 8. TCP model graph for D_2^2

$$T\left(D_2^2\right) = q^2 R + qp(R + T_0 + T\left(B_1^1\right)) + p(T_0 + T\left(B_2^1\right))$$
$$T\left(B_2^2\right) = T\left(D_2^2\right)$$

With Initial Spreading, losses are independent and so the second segment can be successfully received whereas the first segment has been lost. Fig. 9 denotes the 4 different scenarios that can occur.

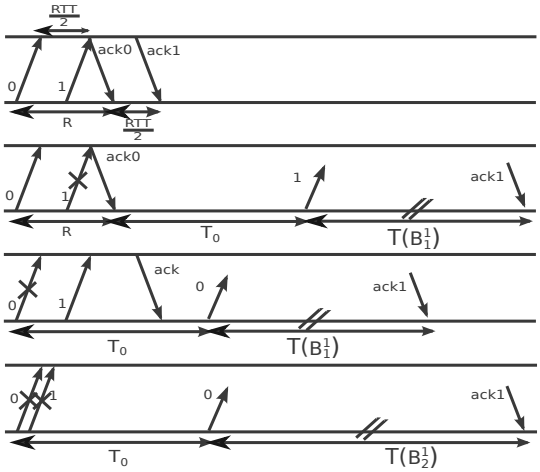

Figure 9. $T\left(S_2^2\right)$, the 4 possible scenarios

$$T\left(S_2^2\right) = q^2(R + \frac{R}{2}) + qp(R + T_0 + T\left(B_1^1\right)) + pq(T_0 + T\left(B_1^1\right)) + p^2(T_0 + T\left(B_2^1\right))$$

We then generalized our model for $n > 2$ and separately addressed each type of burst management.

D. Generalization: $\forall n \geq i \geq 2$

The main benefits of the enlarged IW, with or without Initial Spreading, lay in the transmission of short-lived flows, when the flow size is shorter or equal to the IW size. The following subsection describes the model for $i \leq n$

1) Large IW without Initial Spreading: Considering D_i^n, the bursty loss model is applied. Note that:

$$\forall n > i, \quad T\left(D_i^n\right) = T\left(D_i^i\right)$$

The first transition can decline D_n^n into $n+1$ different states representing the n different positions where the first loss can appear, and the case without any loss.

Using similar reasoning as for D_2^2, we can write:

$$T\left(D_n^n\right) = q^n R + \sum_{i=1}^{n-1} q^{n-i} p(R + T_0 + T\left(B_i^1\right)) + p(T_0 + T\left(B_n^1\right)) \quad (1)$$

2) Large IW with Initial Spreading: case S_n^n: Loss independence due to Initial Spreading makes S_n^n more complex, increasing the number of different scenarios. S_n^n can therefore be declined into 2^n different states, depending on the number of losses and their position in the burst.

The position of the first loss in the IW transmission determines when the RTO is triggered. Fig. 10 gives the example of the fifth segment lost in an IW equal to 5 segments with Initial Spreading. The RTO is then triggered at $R + \frac{3R}{5}$.

So, except when the first segment is lost, each loss adds an extra delay $\in \{R, R+\frac{R}{n}, ..., R+(n-1)\frac{R}{n}\}$ to the Time Out.

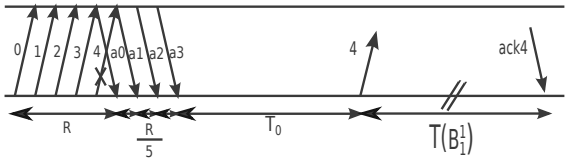

Figure 10. $T\left(S_5^5\right)$, loss of the fifth segment

$R_{i,n,j}$ sums those extra delays when j losses occur among the i segments transmitted with an IW of size n. k is the position of the first loss.

$$R_{i,n,j} = \sum_{k=2}^{i+1-j} \binom{i-k}{j-1} \left\{ R + (k-2)\frac{R}{n} \right\}$$

$T\left(S_n^n\right)$ is then equal to the sum of the terms with no loss, with all segments lost, and with j losses $\in \{1, .., n-1\}$:

$$T\left(S_n^n\right) = q^n(R + (n-1)\frac{R}{n}) + p^n T\left(B_n^1\right) \quad (2)$$
$$+ \sum_{j=1}^{n-1} q^{n-j}p^j \left\{ \binom{n}{j} \left(T_0 + T\left(B_j^1\right)\right) + R_{i,n,j}) \right\}$$

3) S_n^n, with Fast Retransmit and Recovery: Regarding our assumption (see section IV.a), B_j^1 is sent when three DUP ACKs have reported the last loss.

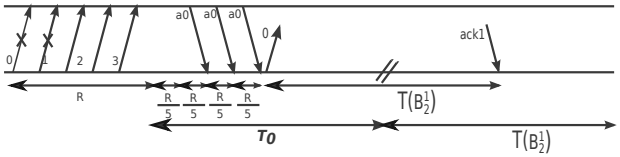

Figure 11. $T\left(S_5^5\right)$, FR triggered after 2 losses

Fig. 11 illustrates this for S_5^5 when the 2 first segments are lost. The Fast Retransmit can be triggered at $R + 4\frac{R}{5}$ then enabling a saving of $T_0 - 4\frac{R}{5}$.

Let FR be the savings gained by using Fast Retransmit, j the number of losses and x the last lost segment, then:

$$FR = \sum_{x=j}^{n-3} \binom{x-1}{j-1} \left\{ T_0 - (x+2)\frac{R}{n} \right\} \quad (3)$$

and

$$T\left(S_n^n\right) = (2) - FR$$

4) S_i^n, with $i < n$: Initial Spreading splits the RTT into n intervals, with n the IW size. So if the number of segments (i) to be sent is smaller than the IW size, only the i first intervals are used.

Using (2), we define S_i^n as :

$$T\left(S_i^n\right) = q^i(R + (n-1)\frac{R}{n}) + p^i T\left(B_i^1\right)$$
$$+ \sum_{j=1}^{i-1} q^{i-j}p^j \left\{ \binom{n}{j} (T_0 + T\left(B_j^1\right)) + R_{i,n,j}) \right\} - FR$$

5) Model for intermediate state: B_i^n: We study the case when n segments are sent by $\lfloor \frac{n+1}{2} \rfloor$ "mini-bursts" of 2 segments. Every burst is considered as independent and losses between bursts are not correlated, but the loss of the first segment triggers the loss of the second one.

Let $P_{j,n}$ be the probability of having j losses among n sent segments with this peculiar bursty traffic.

$$P_{j,n} = \begin{cases} \text{if } n = 2u, & P_{j,n} = \sum_{t=max(0,j-u)}^{\lfloor \frac{j}{2} \rfloor} \left\{ \binom{u}{t,j-2t,u-(j-t)} \right. \\ & \left. \times p^t (qp)^{j-2t} q^{2(u-(j-t))} \right\} \\ \text{if } n = 2u+1, & P_{j,n} = q \times P_{j,2u} + p \times P_{j-1,2u} \end{cases}$$

with t, $j-2t$ and $u-(j-t)$ respectively the number of burst with 2 losses, 1 loss and without loss.

We next derive $P_{j,n}$ into 3 conditional probabilities:

$$P_{j,n} = \begin{cases} Z_{j,n} & \text{given the first burst has 0 loss} \\ X_{j,n} & \text{given the first burst has 1 loss} \\ Y_{j,n} & \text{given the first burst has 2 losses} \end{cases} \quad (4)$$

Using (4) and the same reasoning as for S_n^n, B_n^n can now be defined.

Note that $\forall n > i, B_i^n = B_i^i$.

$X_{j,n} + Y_{j,n}$ is the probability of having j losses among the n sent segments given the first segment is not lost. Similarly to Fig. 7, this implies that the next states are reached at $R + T_0$. Considering the remaining $Z_{j,n}$ possibilities, next states are reached at T_0.

$T\left(B_n^n\right)$ is equal to the sum of the different delivering times, according to the number of losses, and their positions.

$$T\left(B_n^n\right) = q^n R + \sum_{j=1}^{n} \left\{ (X_{j,n} + Y_{j,n}) \times (R + T_0 + T\left(B_j^1\right)) \right.$$
$$\left. + Z_{j,n} \times (T_0 + T\left(B_j^1\right)) \right\} \quad (5)$$

E. Generalization: $\forall i > n$

Our model targets the transmission of a number of segments smaller than the IW. Based on this assumption, we model only B_i^n that both D_i^n and S_i^n use to re-transmit the lost segments of the first IW.

1) B_i^n with $i > n$: We consider the n first segments have been sent. If no loss occured, the following $i-n$ segments are sent in slow-start mode, otherwise, lost segments are retransmitted at the expiration of T_0 with a CWND equal to 1.

Fig. 12 depicts B_4^3 with the loss of the second and forth segments. When the ACK of the first segment arrives, the RTO is updated and the forth segment is sent. Then, $2 \times T_B$ seconds later (corresponding to twice the bottleneck processing time), the ACK of the third segment arrives. As the ACK of the first segment does not arrive before the expiration of this timer, B_2^1 is reached at $R + T_0$ and contains segments number 2 and 4.

Without any loss in the first RTT, CWND = 2×IW, while the position of the first loss in the IW determines the number

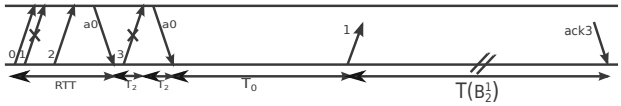

Figure 12. $T\left(B_4^3\right)$ with segments number 2 and 4 lost

of mini bursts that can be sent before waiting for the ACK of the lost segment.

Let j be the number of losses in the IW and u the number of initial success. According to the number of segments that still have to be sent, $M = min\{i-n, 2\times u\}$ is the number of segments that can be sent after the first RTT using bursts of 2 segments. We next define $T\left(B_{i,j,M}^n\right)$ as the estimated duration for sending i segments with an IW of size n, given that j losses occur and that the segment number $u+1$ is lost. This means the first u segments are acknowledged in the first RTT and that M segments are sent in bursts of 2 segments in the second RTT, when waiting for the $u+1^{th}$ ACK. When the Retransmission Timer expires, the $n-j$ segments lost in the IW transmission, the t segments potentially lost during the M re-transmissions and the $i-n$ remaining segments are sent with an IW of size 1.

$$T\left(B_{i,j,M}^n\right) = \sum_{t=0}^{M} \left\{ P_{t,M} \times \left[R + T_0 + T\left(B_{i-(n-j)-(M-t)}^1\right) \right] \right\} \tag{6}$$

We note $P_{j,n,u}$ as the probability of having j losses among n segments given the first u segments sent have succeeded.

$$P_{j,n,u} = \begin{cases} X_{j-1,n-(u+1)} & \text{if } u \text{ even} \\ Y_{j,n-u} & \text{if } u \text{ odd} \end{cases}$$

Using (6), we define $T\left(B_i^n\right)$:

$$T\left(B_i^n\right) = q^n(R + T\left(B_{i\times n}^{2\times n}\right)) + p^{\lfloor \frac{n}{2} \rfloor + 1}(T_0 + T\left(B_i^1\right))$$
$$+ \sum_{j=1}^{n-1} \sum_{u=1}^{n-j} \left\{ P_{j,n,u} \times T\left(B_{i,j,M}^n\right) + Y_{j,n}(T_0 + T\left(B_{i-(n-j)}^1\right)) \right\}$$

Using (D_i^n), (S_i^n) and (B_i^n), we can solve the previous systems, and calculate the average delivery delays to successfully transmit the IW with and without Initial Spreading.

V. Validations

The objective of this paper is twofold:

1) propose and validate an accurate model for short-lived TCP flows that enable to consider the bursts
2) validate with the model the benefits we empirically observed using Initial Spreading

A comparison between the proposed model and ns2 simulation measurements was carried out for both. The size of the connections that we study insures a good confidence in the ns2 behavior. Indeed, in most cases, the congestion avoidance algorithm is not used, and only the IW, the slow start and recovery algorithms affect the results. Our model depicts the same recovery algorithms that have been implemented in ns2, but can easily be modified to depict other recovery mechanisms when the standards will evolve.

For the following comparison, the test bed of section III-C has been used, and the results for each flow size are plotted with a 95% confidence interval, given that several thousand iterations were made.

A. Model evaluation

The model aims to predict the average delivery latency for a short-lived connection, given the mean RTT and loss probability.

1) Short-sized bursts: Our model is based on an accurate understanding of the bursts and in particular of the different impacts they have on a connection performance in congested environments. Two main types of bursts have therefore been isolated, the initial burst and the mini-burst. Our assumption is that only mini-bursts have an impact on the Initial Spreading behavior.

In order to evaluate the accuracy of the mini-burst modeling, we compared the model with the simulation measurements for short-lived connections using an IW of 1 segment. Fig. 13 presents a comparison between the measured and predicted completion times for different delays and loss probabilities.

The model accurately tracks the simulation results and provides a good prediction of the burst impacts. However, we can notice that the effects of the recovery mechanisms are greater on the simulation than on the model. In a congested network, recovery mechanisms have therefore for consequences to offer a non-monotonic increase of the average delivery latency in function of the flow size. For example, a flow of 7 segments that has 4 segments transmitted in the third RTT if its IW was equal to 1, has a higher probability to trigger recovery mechanisms and then recover from previous losses without waiting for an RTO than a flow of 6 segments. The average latency for the delivery of 7 segments is then shorter than for a flow of 6 segments [4].

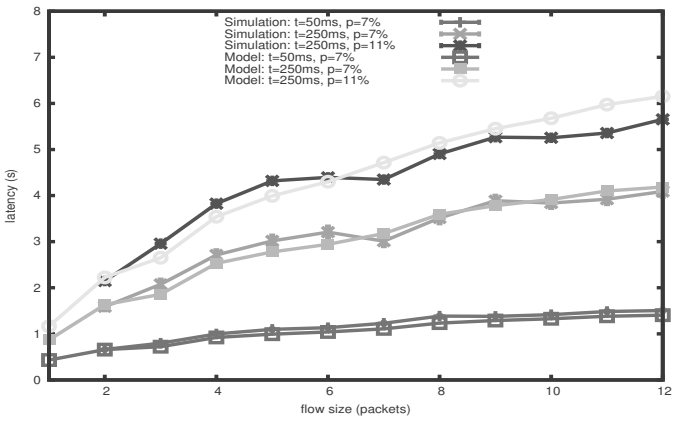

Figure 13. Comparison between the model and the simulation measurements for IW=1 and different loss probabilities and RTT

2) Long-sized bursts: In this case, the selected burst modeling is less accurate and in particular has two weaknesses, underlined in section III.D.

In Fig. 14, the comparison between the model and the simulated measurements for an IW of 10 segments without Initial Spreading illustrated the repercussions on those weaknesses on our model. First, the measured average delivery delay is higher than the model because of the rise in the probability of loss, and then becomes lower because the probability of successfully transmitting a segment after a loss is not zero.

Finally, the shortcomings of the burst modeling impact our model, which is less accurate in the case of large bursts but still enables a good approximation and results in high accuracy for short-lived flows with short-sized bursts.

B. Initial Spreading validation

We next use our model to validate the results we obtained with Initial Spreading.

Fig. 14 compares the model and the simulation measurements for an IW of size 1 and 10, with and without Initial Spreading. In this test, the average loss probability was 6.5% and the link delay was equal to 50ms. Once again, both measured and predicted results are very close, except in the case without Initial Spreading.

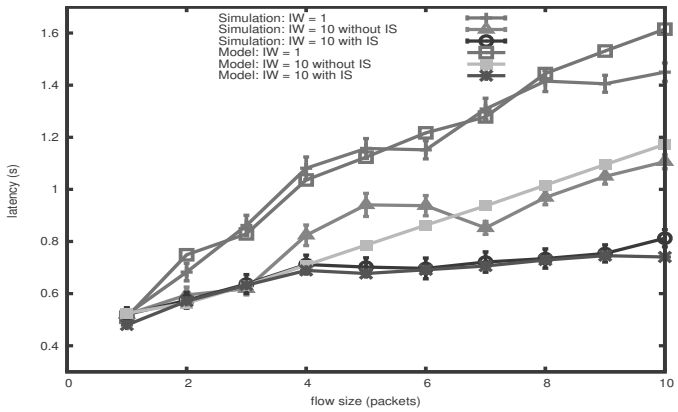

Figure 14. Comparison between the model and the simulation measurements for IW=1 & 10, with and without IS

The model validates the extensive simulations carried out. Regarding the short-lived connections, the model confirms significant savings when using Initial Spreading in association with a large IW.

Moreover, using Initial Spreading makes easier the performance prediction. This point was first illustrated by the respective standard deviation of the simulations with and without Initial Spreading, and is now confirmed by our modeling of both mechanisms.

VI. Conclusion

In this paper, we have presented a TCP model for short-lived flows in order to predict the average delivery delay. This model is adaptive and can easily be modified to stay close to the future TCP updates (e.g. Tail Loss Probe [15] and Early Retransmit [16] that will enable a faster and more efficient loss recovery).

This model lays the emphasis on understanding the bursts and the way they impact the individual performance. The model recursively describes the different scenarios that may happen depending on the mean RTT and loss probability.

Whereas the upper bound value that the initial burst can reached is still a sensitive topic [3] [2], this model enables to weight the pros and cons of transmitting a large initial burst.

Furthermore, the proposed model verifies our observations on the bursts impact and validates the extensive simulations we carried out to introduce Initial Spreading, a fast start-up TCP mechanism.

The model confirms that Initial Spreading enables the IW to be safely enlarged from 3 to 10 segments, without suffering the detrimental effects of the bursts. The use of Initial Spreading with an IW of 10 segments generates important gains when compared with any other IW used without Initial Spreading, and also enables a better stability and predictability.

References

[1] N. Dukkipati, T. Refice, Y. Cheng, J. Chu, T. Herbert, A. Agarwal, A. Jain, and N. Sutin, "An Argument for Increasing TCP's Initial Congestion Window," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 3, pp. 26–33, Jun. 2010.

[2] J. Chu, N. Dukkipati, Y. Cheng, and M. Mathis, "Increasing tcp's initial window," RFC 6928, IETF, Experimental, Jan. 2013.

[3] A. Allman and S. Floyd, "Increasing tcp's initial window," RFC 3390, IETF, Proposed Standard, 2002.

[4] R. Sallantin, C. Baudoin, E. Chaput, E. Dubois, F. Arnal, and A. Beylot, "Initial spreading: a fast start-up tcp mechanism," *LCN*, 2013.

[5] R. Sallantin, C. Baudoin, E. Chaput, F. Arnal, E. Dubois, and A. Beylot, "Safe increase of the tcp's initial window using initial spreading," *Internet draft*, 2014.

[6] N. Cardwell, S. Savage, and T. Anderson, "Modeling tcp latency," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, 2000, pp. 1742–1751 vol.3.

[7] B. Sikdar, S. Kalyanaraman, and K. Vastola, "Analytic models for the latency and steady-state throughput of tcp tahoe, reno, and sack," *Networking, IEEE/ACM Transactions on*, vol. 11, no. 6, pp. 959–971, 2003.

[8] M. Mellia and H. Zhang, "Tcp model for short lived flows," *Communications Letters, IEEE*, vol. 6, no. 2, pp. 85–87, 2002.

[9] K. Zhou, K. Yeung, and V.-K. Li, "On bursty packet loss model for tcp performance analysis," in *High Performance Switching and Routing, 2005. HPSR. 2005 Workshop on*, 2005, pp. 292–296.

[10] P. Dirnopoulos, P. Zeephongsekul, and Z. Tari, "Modeling the burstiness of tcp," in *Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, 2004. (MASCOTS 2004). Proceedings. The IEEE Computer Society's 12th Annual International Symposium on*, 2004, pp. 175–183.

[11] A. Aggarwal, S. Savage, and T. Anderson, "Understanding the performance of TCP pacing," in *INFOCOM*, vol. 3, mar 2000, pp. 1157 –1165.

[12] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling tcp reno performance: a simple model and its empirical validation," *Networking, IEEE/ACM Transactions on*, vol. 8, no. 2, pp. 133–145, 2000.

[13] K. Zhou, K. Yeung, and V.-K. Li, "Throughput modeling of tcp with slow-start and fast recovery," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, vol. 1, 2005, pp. 5–10.

[14] A. Paxson, V. Allman and M. Chu, J. Sargent, "Computing tcp's retransmission timer," RFC 6298, IETF, Proposed Standard, 2011.

[15] N. Dukkipati, N. Cardwell, Y. Cheng, and M. Mathis, "Tail Loss Probe (TLP): An Algorithm for Fast Recovery of Tail Losses," 2013.

[16] M. Allman, K. Avrachenkov, U. Ayesta, J. Blanton, and P. Hurtig, "Early Retransmit for TCP and Stream Control Transmission Protocol (SCTP)," no. 5827, 2010.