Modeling the AIADD Paradigm in Networks with Variable Delays

G. Boggia, P. Camarda, A. D'Alconzo,
L. A. Grieco, and S. Mascolo*
DEE - Politecnico di Bari - Via Orabona, 4
70125 BARI, Italy
{g.boggia, camarda, a.dalconzo, a.grieco, mascolo}@poliba.it

E. Altman, C. Barakat
INRIA 2004, route des lucioles
06902 Sophia Antipolis
{eitan.altman, chadi.barakat}@sophia.inria.fr

ABSTRACT

Modeling TCP is fundamental for understanding Internet behavior. The reason is that TCP is responsible for carrying a huge quota of the Internet traffic. During last decade many analytical models have attempted to capture dynamics and steady-state behavior of standard TCP congestion control algorithms. In particular, models proposed in literature have been mainly focused on finding relationships among the throughput achieved by a TCP flow, the segment loss probability, and the round trip time (RTT) of the connection, which the flow goes through. Recently, Westwood+ TCP algorithm has been proposed to improve the performance of classic New Reno TCP, especially over paths characterized by high bandwidth-delay products. In this paper, we develop an analytic model for the throughput achieved by Westwood+ TCP congestion control algorithm when in the presence of paths with time-varying RTT. The proposed model has been validated by using the *ns*-2 simulator and Internet-like scenarios. Validation results have shown that this model provides relative prediction errors smaller than 10%. Moreover, it has been shown that a similar accuracy is achieved by analogous models proposed for New Reno TCP.

Keywords

TCP Congestion Control, TCP Modelling, Performance Evaluation

1. INTRODUCTION

TCP congestion control was introduced by Van Jacobson in the cornerstone paper [16] with the main objectives to obtain a high Internet utilization while avoiding network collapse due to congestion. Many improvements to the Jacobson's congestion control algorithm, also known as Tahoe TCP, have been proposed during last decade by the Internet Engineering Task Force (IETF) [21, 2, 1, 12]. All the proposed TCP enhancements are based on the well-known Additive Increase Multiplicative Decrease (AIMD) paradigm, which is made by a probing phase and a shrinking phase. During the probing phase, a TCP sender additively increases its own congestion window, which limits the number of outstanding segments, to probe network for new available bandwidth until a network congestion is revealed by the reception of 3 Duplicated Acknowledgments (DUPACKs) or the expiration of a retransmission timeout. During the shrinking phase, in order to recover the network from congestion, the congestion window is halved, when 3 DUPACKs are received, or reduced to one, if a retransmission timeout expires [10].

AIMD-based algorithms, such as New Reno TCP [12], exhibit poor performance in the presence of random losses and paths with high bandwidth-delay product [18]. So that, more sophisticated paradigms for finely tuning TCP congestion window would be required [19].

Recently, to overcome AIMD performance limitations, Westwood TCP and its Westwood+ variant [20, 14] have been proposed. Both of them are based on the new Additive Increase ADaptive Decrease (AIADD) paradigm, which leaves unchanged the probing phase of New Reno TCP and proposes an adaptive shrinking phase. In particular, the adaptive decrease used by Westwood TCP sets the congestion window according to the bandwidth left behind TCP at time of congestion, which is the end-to-end network available bandwidth as defined in [14]. The available bandwidth is estimated by properly processing the stream of returning ACKs. Westwood+ variant differs from the original one in that it employs an enhanced bandwidth estimation scheme, which is robust with respect to ACK compression [23].

Modeling TCP is fundamental for understanding Internet behavior. The reason is that TCP carries a very large quota of Internet traffic [8]. As a consequence, during the last decade, many analytical models have attempted to capture dynamic and steady-state behavior of standard TCP congestion control algorithms [25, 17, 22, 5]. In particular, those studies have mainly focused on the relationships between TCP throughput, the segment loss probability, and the round trip time (RTT) seen by a TCP sender.

*Corresponding author.

In [25] a simple analytical characterization of the steady-state throughput of Reno TCP [11] as a function of the loss rate and RTT for a bulk data transfer has been developed. Three models are proposed: the first one assumes only 3 DUPACKs reception as congestion indication; the second one considers also timeouts; the third one is a simplification of the second one. The models have been validated against traces collected from the real Internet; results have shown that even using the most accurate model, i.e., the second one, prediction errors can be up to 100% in some circumnstances.

In [17] an approach for dealing with issues concerning the stability of and the fairness of IP networks has been proposed. A part of the analytical arguments proposed in [17] has been later exploited in [15] to develop steady-state models for the throughput of Westwood+ TCP congestion control algorithm.

In [22] the attention has been focused on modeling the dynamics of TCP congestion control using a fluid approach. In particular, a stochastic dynamic model for TCP congestion window has been proposed in state-space domain. Then the proposed model has been used for designing Active Queue Management (AQM) algorithms. In [9] a similar approach has been followed for deriving a dynamic model for Westwood TCP.

Approaches cited so far do not consider the impact of RTT variability on TCP throughput. Recently the importance of this issue has been discussed in [5], where a closed-form expression for the throughput of a TCP connection going through a path with a time-varying delay has been derived. Anyway, models proposed in [5] consider classical TCP implementations deriving from the Reno algorithm and more recent TCP congestion control algorithms, such as Westwood+ TCP, have not been yet investigated. In order to bridge this gap, this paper proposes an analytical model for the throughput of a Westwood+ TCP connection going through a path with variable delays, using theoretical arguments from [5] and [15]. This study aims at capturing the behavior of the Westwood+ algorithm under realistic hypothesis, i.e., time varying delays, which hold in almost all packet switching networks and that are usually not taken into account.

Our findings have been validated by exploiting the *ns*-2 simulator with Internet-like scenarios. Validation results have shown that model we propose provides relative prediction errors smaller than 10%. Moreover, it has been shown that a similar accuracy is achieved by analogous models proposed for New Reno TCP in [5].

2. BACKGROUND ON RENO, NEW RENO, AND WESTWOOD+ TCP

A TCP connection is characterized by the following variables: (1) congestion window ($cwnd$); (2) slow start threshold ($ssthresh$); (3) round trip time of the connection (RTT); (4): minimum round trip time measured by the sender (RTT_{min}).

The pseudo-code of the Reno algorithm is reported below (see Algorithm 1).

```
if ACKs are successfully received then
    if cwnd < sstresh then
        cwnd = cwnd + 1;
    else
        cwnd = cwnd + 1/cwnd;
    end

else if 3 DUPACKs are received by the sender then
    ssthresh = cwnd/2;
    if ssthresh < 2 then ssthresh=2;
    cwnd = ssthresh;
 else if a coarse timeout expires then
    ssthresh = cwnd/2;
    if ssthresh < 2 then ssthresh = 2;
    cwnd=1;
end
```

Algorithm 1: Pseudo-code of Reno algorithm.

When 3 DUPACKs are received, the Reno TCP algorithm enters the *fast recovery phase* and retransmits the segment with the lowest unacknowledged sequence number. This phase is leaved when the retransmitted segment is successfully acknowledged. NewReno differs from Reno TCP in that it does not exit the *fast recovery* phase until all the segments within the current window are acknowledged. This feature, which is known as NewReno feature, improves the performance of Reno TCP when several segments within the same window get lost [12].

The key idea of TCP Westwood+ is to exploit the stream of returning acknowledgment packets to estimate the bandwidth $\widehat{B}$ that is available for the TCP connection. When a congestion episode happens at the end of the TCP probing phase, the bandwidth estimated from the stream of ACKs corresponds to the definition of best effort available bandwidth in a connectionless packet network. This bandwidth estimate is used to adaptively decrease the congestion window and the slow-start threshold after a timeout or three duplicate ACKs as it is described below (see Algorithm 2).

```
if ACKs are successfully received then
    if cwnd < sstresh then
        cwnd = cwnd + 1;
    else
        cwnd = cwnd + 1/cwnd;
    end

else if 3 DUPACKs are received by the sender then
    ssthresh = B̂ · RTT_min;
    if ssthresh <2 then ssthresh=2;
    cwnd = ssthresh;
 else if a coarse timeout expires then
    ssthresh = B̂ · RTT_min ;
    if ssthresh <2 then ssthresh=2;
    cwnd=1;
end
```

Algorithm 2: Pseudo-code of Westwood+ algorithm.

It is worth noting that the adaptive decrease mechanism employed by Westwood+ TCP improves the standard TCP multiplicative decrease algorithm. In fact, the adaptive win-

dow shrinking provides a congestion window that is sufficiently decreased in the presence of heavy congestion and not excessively so in the presence of light congestion or losses that are not due to congestion, such as in the case of unreliable radio links. Moreover, the adaptive setting of the congestion window increases the fair allocation of available bandwidth to different TCP flows. In fact, if for sake of simplicity we neglect the normalization factor *seg_size*, it could be noted that the setting $cwnd = (\widehat{B} \cdot RTT_{min})$ sustains a transmission rate $cwnd/RTT = (\widehat{B} \cdot RTT_{min})/RTT$ that is less than the bandwidth $\widehat{B}$ measured at the time of congestion; as a consequence, the TCP flow clears out its path backlog after the setting, thus leaving room in the buffers for coexisting and joining flows and, consequently, improving statistical multiplexing and fairness [14].

3. MODELING THE AIADD PARADIGM

3.1 Window Analysis

Mainly following the notation introduced in [5], let R_n be the sequence of round-trip times (RTTs), with $R_n = T_{n+1} - T_n$, and W_n be the window size at the beginning of the interval R_n (see Fig. 1).

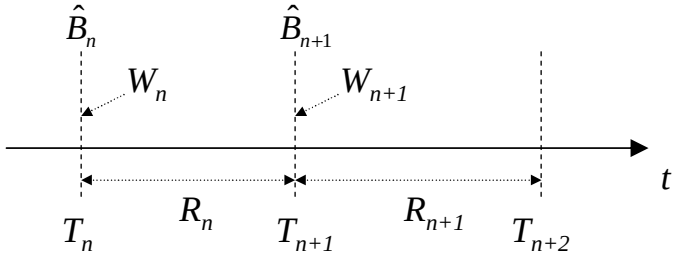

Figure 1: Reference time sequence.

We will study the dynamic of the window size in the presence of losses, following an approach similar to the one described in [5]. Let Z_n be the discrete random variable (r.v.) defined by relation

$$Z_n = \begin{cases} 0 & \text{there are no losses in the interval } R_n \\ 1 & \text{there is at least one loss in the interval } R_n \end{cases}. \quad (1)$$

As in [5], we suppose that losses[1] can be described by a Poisson process, with rate λ, independent of the window size and the RTT. This assumption makes the analysis tractable. Moreover, it is a good approximation when packets of a TCP connection are lost due to other factors that congestion induced by the connection itself, for example when the connection crosses many routers with exogenous traffic [4], or when it crosses wireless links with transmission errors.

Thus, probability that in a generic interval R_n there are no loss packets (i.e., $Z_n = 0$) is equal to

$$P\{\text{no loss event in } R_n\} = P(Z_n = 0|R_n) = e^{-\lambda R_n}. \quad (2)$$

[1]Model refers to a steady state TCP connection in Congestion Avoidance with losses due to 3DUPACK; timeouts are not considered.

Let $f_{R_n}(R)$ and $\mathcal{R}^*(s)$ be, respectively, the pdf of R_n and its Laplace-Stieltjes transform. We can write the probability of no losses as

$$\begin{aligned} P(Z_n = 0) &= \int_{R_n} P(Z_n = 0|R_n) f_{R_n}(R) dR \\ &= \int_{R_n} e^{-\lambda R_n} f_{R_n}(R) dR = \mathcal{R}^*(\lambda), \end{aligned} \quad (3)$$

where $\mathcal{R}^*(\lambda)$ is the Laplace-Stieltjes transform of R_n evaluated for $s = \lambda$.

Consequently, the probability of at least one loss in a RTT is given by

$$P(Z_n = 1) = 1 - P(Z_n = 0) = 1 - \mathcal{R}^*(\lambda). \quad (4)$$

Therefore, the mean value of Z_n is

$$E[Z_n] = 0 \cdot P(Z_n = 0) + 1 \cdot P(Z_n = 1) = 1 - \mathcal{R}^*(\lambda). \quad (5)$$

Considering TCP Westwood+ behavior [14], congestion window size W_{n+1} at the beginning of interval R_{n+1} changes according to the presence or not of packet loss during R_n. In particular, when there are no losses, W_{n+1} is equal to the window size at the previous step (i.e., W_n) increased by a factor β, equal to 1 as in TCP New-Reno if all packets are acknowledged by the receiver; if there is at least a loss, window is reduced only one time and its size is given by the product between the estimated bandwidth, $\widehat{B}_n$, at instant T_n, and the minimum round-trip time, RTT_{min}. Therefore,

$$W_{n+1} = \begin{cases} W_n + \beta, & \text{no losses in } R_n \Rightarrow Z_n = 0 \\ RTT_{min}\widehat{B}_n, & \text{at least one loss in } R_n \Rightarrow Z_n = 1 \end{cases} \quad (6)$$

where bandwidth $\widehat{B}_n$ is estimated by using the following low-pass filter [14]:

$$\widehat{B}_{n+1} = \alpha \widehat{B}_n + (1 - \alpha)\frac{W_n}{R_n}. \quad (7)$$

Exploiting matrix notation, eqs. (6) and (7) can be rewritten as

$$\begin{pmatrix} W_{n+1} \\ \widehat{B}_{n+1} \end{pmatrix} = \begin{pmatrix} 1 - Z_n & RTT_{min} \cdot Z_n \\ \dfrac{1-\alpha}{R_n} & \alpha \end{pmatrix} \begin{pmatrix} W_n \\ \widehat{B}_n \end{pmatrix} + \begin{pmatrix} (1 - Z_n)\beta \\ 0 \end{pmatrix}. \quad (8)$$

With a more compact notation, the previous system can be written as

$$\overline{Y}_{n+1} = \overline{\mathcal{A}}_n \overline{Y}_n + \overline{C}_n \ , \quad (9)$$

where $\overline{Y}_n = \left(W_n, \widehat{B}_n\right)^T$.

Eq. (9) is a stochastic vector recursive equation [7, 13, 3]. Applying results of [13] for which conditions can be easily checked (see Appendix), such an equation admits a unique solution as n grows to infinity, which is given by the ergodic

process $\overline{Y^*}_n$:

$$\overline{Y^*}_n = \sum_{j=0}^{\infty} \left(\prod_{l=n-j}^{n-1} \overline{\mathcal{A}}_l \right) \overline{C}_{n-j-1} \ . \tag{10}$$

For any initial value $\overline{Y}_0$, we have that

$$\lim_{n\to\infty} |\overline{Y}_n \ - \ \overline{Y^*}_n| \ \to \ 0;$$

and $\overline{Y}_n$ converges to $\overline{Y^*}_n$ almost sure.

To simplify discussion, in the sequel we will consider the system in steady state at time $t = 0$. The expression of $\overline{Y^*}_0$ will be evaluated when random variables $\{R_n\}$ are i.i.d. (independent and identically distributed) and in the case the process $\{R_n\}$ is Markov correlated.

3.2 Random variables R_n i.i.d.

Let R_n be i.i.d. random variables. Expectation of $\overline{Y^*}_0$ can be directly evaluated considering matrix equation (9). In fact, in this case mean values of coefficients in stationary regime are independent of the considered time instant and of $\overline{Y^*}_0$.

We have:

$$E[\overline{Y^*}_0] = E[\overline{\mathcal{A}}_0 \overline{Y^*}_0] + E[\overline{C}_0] \ . \tag{11}$$

Obviously:

$$E[\overline{Y^*}_0] = E\left[\begin{array}{c} W_0^* \\ \\ \widehat{B^*}_0 \end{array} \right] = \left(\begin{array}{c} E[W_0^*] \\ \\ E[\widehat{B^*}_0] \end{array} \right) \ . \tag{12}$$

Using eq. (5), we have::

$$E[\overline{C}_0] = E\left[\begin{array}{c} (1 - Z_0)\beta \\ \\ 0 \end{array} \right] = \left(\begin{array}{c} \beta \cdot \mathcal{R}^*(\lambda) \\ \\ 0 \end{array} \right) \tag{13}$$

The term $E[\overline{\mathcal{A}}_0 \overline{Y^*}_0]$ is given by

$$E[\overline{\mathcal{A}}_0 \overline{Y^*}_0] = E\left[\begin{array}{c} (1 - Z_0)\, W_0^* + RTT_{min} Z_0 \widehat{B^*}_0 \\ \\ (1-\alpha)\dfrac{W_0^*}{R_0} + \alpha \widehat{B^*}_0 \end{array} \right] \tag{14}$$

Z_0 is related to losses in the interval R_0 which starts at instant $t = 0$, whereas W_0^* and $\widehat{B^*}_0$ refer to the beginning of such an interval, therefore are independent on Z_0 and R_0. Thus, eq. (14) becomes

$$E[\overline{\mathcal{A}}_0 \overline{Y^*}_0] = \left(\begin{array}{c} E\left[1 - Z_0\right] E[W_0^*] + RTT_{min} E[Z_0] E[\widehat{B^*}_0] \\ \\ (1-\alpha) E\left[\dfrac{W_0^*}{R_0}\right] + \alpha E[\widehat{B^*}_0] \end{array} \right). \tag{15}$$

The r.v. Z_0 is independent on W_0^* due to Poisson hypothesis about process $\{R_n\}$ and the fact that the R_n are i.i.d.

Applying Jensen inequality [24],

$$E\left[\frac{W_0^*}{R_0}\right] \geq \frac{E[W_0^*]}{E[R_0]}. \tag{16}$$

Furthermore, approximating $E[1/R_0]$ by the first term of its Taylor expansion around the mean value $E[R_0]$, we can write

$$E\left[\frac{W_0^*}{R_0}\right] \approx \frac{E[W_0^*]}{E[R_0]}. \tag{17}$$

Thus, eq. (15) becomes:

$$E[\overline{\mathcal{A}}_0 \overline{Y^*}_0] \approx \left(\begin{array}{c} \mathcal{R}^*(\lambda) E[W_0^*] + RTT_{min}\left[1 - \mathcal{R}^*(\lambda)\right] E[\widehat{B^*}_0] \\ \\ (1-\alpha)\dfrac{E[W_0^*]}{E[R_0]} + \alpha E[\widehat{B^*}_0] \end{array} \right). \tag{18}$$

Now, considering system (11) and eqs. (12), (13), (18), after a bit of algebra we find that:

$$E[W_0^*] = \frac{\beta E[R_0]}{E[R_0] - RTT_{min}} \cdot \frac{\mathcal{R}^*(\lambda)}{1 - \mathcal{R}^*(\lambda)} \ . \tag{19}$$

3.3 Process R_n described by Markov model

Let R_n be described by a N-state Markov process, i.e., variables R_n are correlated. Let $\zeta(n)$ be the r.v. which denotes the state of the N-state Markov chain at instant T_n, P_{ij} be the state transition probability, and π_j be the steady state probabilities with $j = 1, \ldots, N$.

Assume the same hypotheses of [5], i.e., R_n is only a function of the Markov chain state at time T_n and not of the previous history. Hence, considering state of Markov chain at instants T_n and T_m with $m > n$, we have independence between matrices $\overline{\mathcal{A}}_m$ and $\overline{\mathcal{A}}_n$, and arrays $\overline{C}_m$ and $\overline{C}_n$ in eq. (9).

To evaluate the mean value $E[W_0^*]$, it is necessary to consider the mean value of W_0^* in each state of the Markov chain, following a procedure similar to the one described in [5]. We use the indicator function $1\{X\}$ defined as [24]:

$$1\{X\} = \left\{ \begin{array}{ll} 0 & \text{if event } X \text{ is false} \\ 1 & \text{if event } X \text{ is true} \end{array} \right. . \tag{20}$$

Therefore, if $p\{X\}$ is the probability of event X occurrence, probabilities that $1\{X\}$ takes values 0 and 1 are

$$p\{1\{X\} = 0\} = 1 - p\{X\}; \qquad p\{1\{X\} = 1\} = p\{X\}, \tag{21}$$

and mean value of $1\{X\}$ is given by

$$E[1\{X\}] = 0 \cdot p\{1\{X\} = 0\} + 1 \cdot p\{1\{X\} = 1\} = p\{X\}. \tag{22}$$

Moreover, the following properties is true for the indicator function:

$$1\{X\} = 1\{X\} \cdot 1\{X\}. \tag{23}$$

Now, we can evaluate $E[W_0^*]$. Let w_j be the mean value of W_0^* when in the state j of Markov chain at time $t = 0$ (i.e.,

T_0), that is, using indicator function,

$$w_j = E[W_0^* \cdot 1\{\zeta(0) = j\}]. \quad (24)$$

Similarly, let b_j be the mean value of estimated bandwidth $\widehat{B^*}_0$ when in the state j of Markov chain at time T_0:

$$b_j = E[\widehat{B^*}_0 \cdot 1\{\zeta(0) = j\}]. \quad (25)$$

Due to their definition and considering steady state conditions at instant T_1, we can also rewrite w_j and b_j as

$$\begin{aligned} w_j &= E[W_1^* \cdot 1\{\zeta(1) = j\}]; \\ b_j &= E[\widehat{B^*}_1 \cdot 1\{\zeta(1) = j\}]. \end{aligned} \quad (26)$$

Considering eqs. (6) and (7), we have the system

$$\begin{cases} W_1^* &= (1 - Z_0)W_0^* + RTT_{min} Z_0 \widehat{B^*}_0 + (1 - Z_0)\beta \\ \widehat{B^*}_1 &= (1 - \alpha)\dfrac{W_0^*}{R_0} + \alpha \widehat{B^*}_0 \end{cases} . \quad (27)$$

Hence,

$$\begin{aligned} w_j &= E\left[(1 - Z_0)W_0^* 1\{\zeta(1) = j\}\right] \\ &\quad + RTT_{min} E\left[Z_0 \widehat{B^*}_0 1\{\zeta(1) = j\}\right] \\ &\quad + \beta E\left[(1 - Z_0)1\{\zeta(1) = j\}\right]. \end{aligned} \quad (28)$$

Considering all the possible states at T_0,

$$\begin{aligned} w_j &= \sum_{i=1}^{N} E\left[(1 - Z_0)W_0^* 1\{\zeta(0) = i\}1\{\zeta(1) = j\}\right] \\ &\quad + RTT_{min} \sum_{i=1}^{N} E\left[Z_0 \widehat{B^*}_0 1\{\zeta(0) = i\}1\{\zeta(1) = j\}\right] \\ &\quad + \beta \sum_{i=1}^{N} E[(1 - Z_0)1\{\zeta(0) = i\}1\{\zeta(1) = j\}] \\ &= \sum_{i=1}^{N} E\left[(1 - Z_0)|\zeta(0) = i\right] E\left[W_0^* 1\{\zeta(0) = i\}\right] P_{ij} \\ &\quad + RTT_{min} \sum_{i=1}^{N} E[Z_0|\zeta(0) = i] E\left[\widehat{B^*}_0 1\{\zeta(0) = i\}\right] P_{ij} \\ &\quad + \beta \sum_{i=1}^{N} E[(1 - Z_0)|\zeta(0) = i]\pi_i P_{ij} \ . \end{aligned} \quad (29)$$

Now, $E[Z_0|\zeta(0) = i]$ is strictly related to R_n whose distribution depends on state of Markov chain. Therefore, $E[Z_0|\zeta(0) = i]$ can be written as

$$E[Z_0|\zeta(0) = i] = 1 - \mathcal{R}_i^*(\lambda) \quad (30)$$

that is, the mean value of Z_0 when in state i at time T_0, being function of the Laplace-Stieltjes transform $\mathcal{R}_i^*(s)$ of R_n in state i.

It follows that

$$E[(1 - Z_0)|\zeta(0) = i] = \mathcal{R}_i^*(\lambda). \quad (31)$$

Using expressions (30) and (31), eq. (28) becomes

$$\begin{aligned} w_j &= \sum_{i=1}^{N} \mathcal{R}_i^*(\lambda) w_i P_{ij} + RTT_{min} \sum_{i=1}^{N} \left[1 - \mathcal{R}_i^*(\lambda)\right) b_i P_{ij} \\ &\quad + \beta \sum_{i=1}^{N} \mathcal{R}_i^*(\lambda)\pi_i P_{ij}. \end{aligned} \quad (32)$$

Similarly, b_j can be evaluated as

$$\begin{aligned} b_j &= E[\widehat{B^*}_1 \cdot 1\{\zeta(1) = j\}] \\ &= E\left[\left(W_0^* \frac{1 - \alpha}{R_0} + \alpha \widehat{B^*}_0\right) 1\{\zeta(1) = j\}\right] \\ &= \sum_{i=1}^{N} E\left[\left.\frac{1 - \alpha}{R_0}\right| \zeta(0) = i\right] E\left[W_0^* 1\{\zeta(0) = i\}\right] P_{ij} \\ &\quad + \alpha \sum_{i=1}^{N} E\left[\widehat{B^*}_0 1\{\zeta(0) = i\}\right] P_{ij} = \\ &= \sum_{i=1}^{N} c_i w_i P_{ij} + \alpha \sum_{i=1}^{N} b_i P_{ij} \ . \end{aligned} \quad (33)$$

The parameters c_i are given by.

$$c_i = E\left[\left.\frac{1 - \alpha}{R_0}\right| \zeta(0) = i\right] \approx (1 - \alpha)\frac{1}{E[R_{0,i}]}. \quad (34)$$

where $E[R_{0,i}]$ is the mean value of R_n at state i and $= 1/E[R_{0,i}]$ is the approximation of $E[1/R_{0,i}]$, as in eq. (17).

When all w_j are computed, we have

$$E[W_0^*] = \sum_{j=1}^{N} w_j \ . \quad (35)$$

3.4 Throughput Analysis

The expression of $E[W_0^*]$ can be used for throughput evaluation. In fact, in general the throughput X of a long lived TCP connection is given by:

$$X = \frac{E[W_0^*]}{E[R_0]}. \quad (36)$$

When R_n are correlated, eq. (36) has to be evaluated numerically using the derivation in sec. 3.3, whereas when R_n are i.i.d. it is possible to find a closed form for throughput. In fact, from (19):

$$X = \frac{\beta}{E[R_0] - RTT_{min}} \cdot \frac{\mathcal{R}^*(\lambda)}{1 - \mathcal{R}^*(\lambda)} \ . \quad (37)$$

Such an expression can be rewritten as a function of probability p_{wr} that there is at least one congestion window reduction due to a loss event. It is worthwhile to note that p_{wr} is not the loss packet probability, given that in the same RTT time can be more packet losses, but only one window reduction.

Let λ' be the rate of window reduction. Obviously, $\lambda' \leq \lambda$,

given that λ is the packet loss rate, and

$$\lambda' = \frac{\text{mean number of events which reduce window}}{E[R_0]} = \frac{E[Z_0 = 1]}{E[R_0]} = \frac{1 - \mathcal{R}^*(\lambda)}{E[R_0]} . \quad (38)$$

Rate λ' is also given by product between probability of window reduction p_{wr} and throughput:

$$\lambda' = p_{wr} X. \quad (39)$$

Thus, from eqs. (38) and (39)

$$\mathcal{R}^*(\lambda) = 1 - E[R_0] p_{wr} X. \quad (40)$$

Substituting the last equation in (19), we can find throughput X as a function of p_{wr}. The term $\mathcal{R}^*(\lambda)$, which takes into account RTT variability, disappears, but impact of delay variation figures in p_{wr}.

For R_n i.i.d., using eq. (40) in (37):

$$X = \frac{\beta}{E[R_0] - RTT_{min}} \cdot \frac{1 - E[R_0] p_{wr} X}{E[R_0] p_{wr} X} . \quad (41)$$

It can be easily checked that this equation has a real and positive solution:

$$X = \frac{-p_{wr}\beta E[R_0] + \sqrt{(p_{wr}\beta E[R_0])^2 + 4\beta p_{wr} E[R_0] \Delta R}}{2 p_{wr} E[R_0] \Delta R}, \quad (42)$$

where $\Delta R = (E[R_0] - RTT_{min})$.

4. MODEL VALIDATION

In this section theoretical models represented by eqs. (19) and (35) will be validated by using the *ns*-2 simulator with Internet-like scenarios. Moreover, analogous models proposed in [5] for New Reno TCP will been validated.

We will consider the multihop topology depicted in Fig. 2, which is characterized by: (a) N hops; (b) one greedy TCP connection C_1 (which we take either Westwood+ or New Reno) going through all the N hops; (c) N cross traffic New Reno TCP connections $C_2 \div C_{N+1}$ transmitting data over each single hop. The simulation lasts 10000 s during which the C_1 connection always sends data. Each cross traffic source $C_2 \div C_{N+1}$ starts randomly in the interval [100 s, 3000s], and lasts for a random time uniformly distributed in the interval [1 s, 7000 s]. Links connecting nodes hosting FTP cross-traffic sources and sinks have a delay of 25 ms. Other links have a delay of 2.5 ms. The link capacity between routers is equal to 2 Mbps. The capacity of entry/exit links is 1000 Mbps. TCP segments are 1500 Bytes long. The overhead due to lower layer protocols is neglected. Buffers are 42 packets long, which is the bandwidth-delay product that would be obtained with a 2 Mbps link with RTT=250 ms.

The number of hops N has been varied from 2 to 20 and for each scenario 30 simulations have been executed, each one using a different seed for the ns-2 random number generator.

For each repetition, we have evaluated the average value of the congestion window of the C_1 connection over the time and we have compared the so obtained value with respect to the windows values predicted using eqs. (19) and (35) of the model, when the C_1 connection employs Westwood+ as congestion control algorithm, and using analogous models reported in [5] when New Reno is used.

In the following, when eq. (35) is used, we will consider a two state Markov model for describing R_n. In particular, for each of the 30 simulations, the set of collected R_n has been split into two subsets, one for each of the states of the Markov model. The first subset is made by R_n values smaller than a threshold value RTT_{th}, the second one contains the other R_n values. RTT_{th} has been varied by spanning all collected RTT values. For each selected threshold a different parameter set of the two-state R_n model has been obtained, to which corresponds a different prediction of the congestion window. Among all the obtained predictions, we will consider the one closer to the measured average congestion window. An analogous approach has been considered when we have used correlated R_n with the New Reno model proposed in [5].

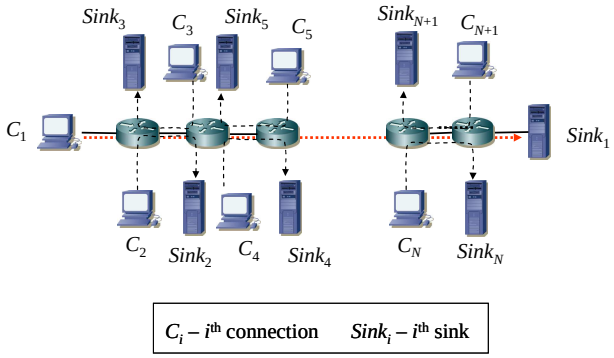

Figure 2: Multibottleneck topology.

Fig. 3 reports for each value of N both measured and estimated average *cwnd* obtained when the C_1 connection employs Westwood+ as congestion control algorithm. Also confidence intervals at 95% are reported for simulation results. In particular, Fig. 3.a reports the estimated average *cwnd* obtained assuming correlated R_n and using eq. (35), whereas Fig. 3.b reports analogous data obtained considering i.i.d. R_n and using eq. (19). By comparing Figs. 3.a and 3.b, it is straightforward to note that model using correlated R_n is more accurate than the model when i.i.d. R_n are assumed.

Note that in this scenario a lower throughput is achieved with Westwood+. In fact, the competing New Reno cross traffic, after congestion, does not deplete router buffers as Westwood+; thus, the C_1 connection with Westwood+ grabs an amount of bandwidth smaller than when it uses New Reno. However, this result does not affect model accuracy.

The same conclusions can be derived by looking at Fig. 4 where analogous data obtained for New Reno TCP have been reported.

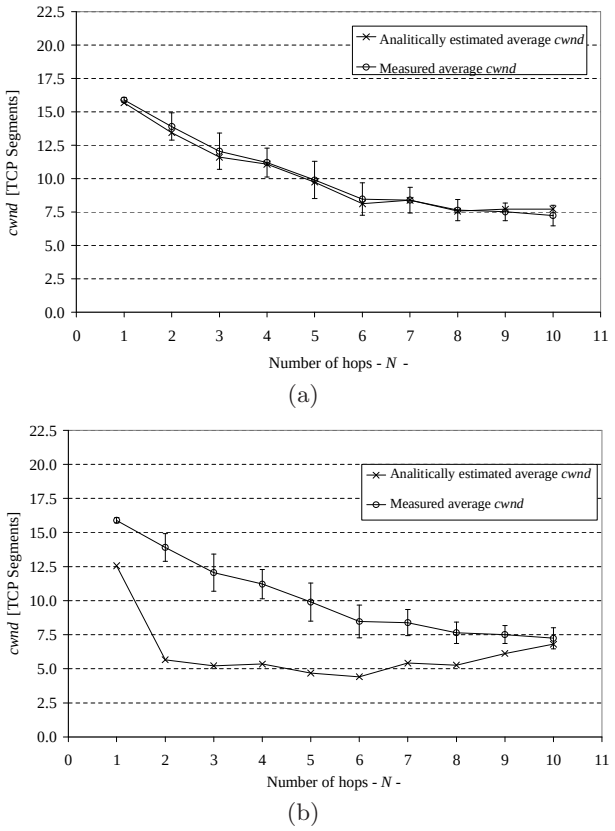

Figure 3: Analytically Estimated and Measured Average *cwnd* of Westwood+: (a) Correlated R_n; (b) i.i.d. R_n.

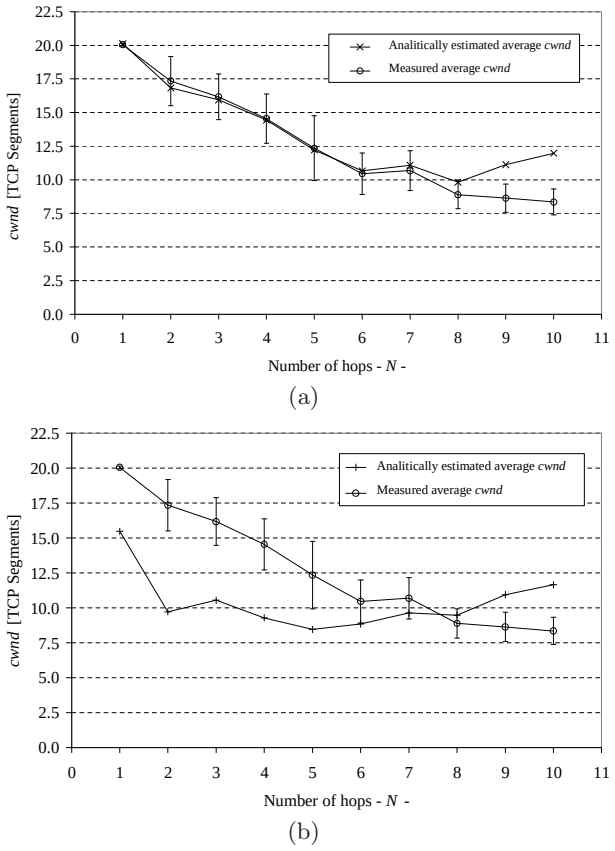

Figure 4: Analytically Estimated and Measured Average *cwnd* of New Reno: (a) Correlated R_n; (b) i.i.d. R_n.

In order to better quantify modeling improvements obtained assuming correlated R_n rather than i.i.d. R_n Fig. 5 reports the relative errors obtained for Westoowd+ and New Reno. It is straightforward to note that assuming correlated R_n leads to relative errors smaller than one order of magnitude with respect to the case of i.i.d. R_n.

By looking with more attention at Fig. 5, new insights into model behaviors can be found. In fact, when correlated R_n are assumed, in the case of Westwood+ (resp. New Reno) it can be noticed that relative errors start increasing monotonically when the number of hops traversed by the C_1 connections is larger than 7 (resp. 4). The reason is that when the number of hops increases it is more likely that *cwnd* dynamics are driven by timeouts expiration rather than 3 DUPACKs receptions. Thus, relative errors increases with the number of hops because slow start and *cwnd* backoff have not been modeled.

This can be confirmed by looking at Fig. 6 reporting the *cwnd* of the C_1 connection averaged over the 30 simulations. In particular, Fig. 6 clearly highlight that, when a scenario with 10 hops is considered, both New Reno and Westwood+ keep the *cwnd* at 1 for a very large quota of the simulation time. On the other hand, when a scenario with 4 hops is simulated, the measured average *cwnd* is larger than 1 for almost all the simulation time.

On the other hand, when the number of traversed hops is smaller than 7 (resp. 4), errors are smaller than 8% (4%) and exhibit statistical fluctuations that depend on the interactions between C_1 and the other $C_2 \div C_{N+1}$ connections, which are difficult to predict (see also [6]).

Thus it can be concluded that the proposed model is able to capture the behavior of Westwood+ TCP with a relative prediction error smaller than 10%. Moreover both the model we have proposed and the analogous one proposed for New Reno in [5] provide similar accuracies.

5. CONCLUSION

In this paper, an analytical model for the throughput of Westwood+ TCP in networks with variable delays has been developed. The model has been built starting from recent theoretical findings reported in literature [5] [15]. The proposed model, which does not take into account timeout events, has been validated by means of *ns*-2 simulations of Internet-like scenarios. Results have shown that analytical framework herein describes is able to capture Westwood+ TCP behavior with relative prediction errors smaller than 10%. Moreover, it has been shown that the same accuracy is provided by analogous models proposed in literature for New Reno TCP. The main finding of this paper is that a great modeling accuracy can be achieved by considering the

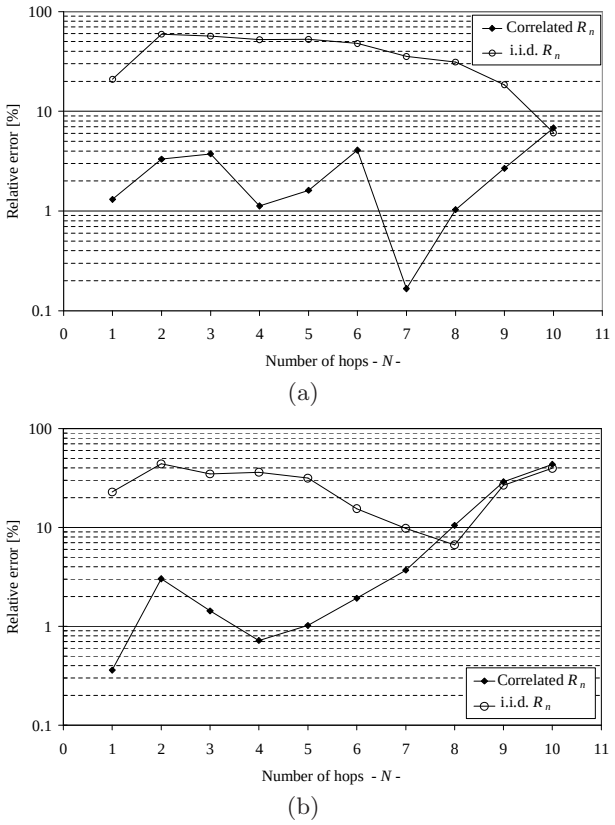

Figure 5: Model errors: (a) Westwood+; (b) New Reno.

correlation among RTT samples.

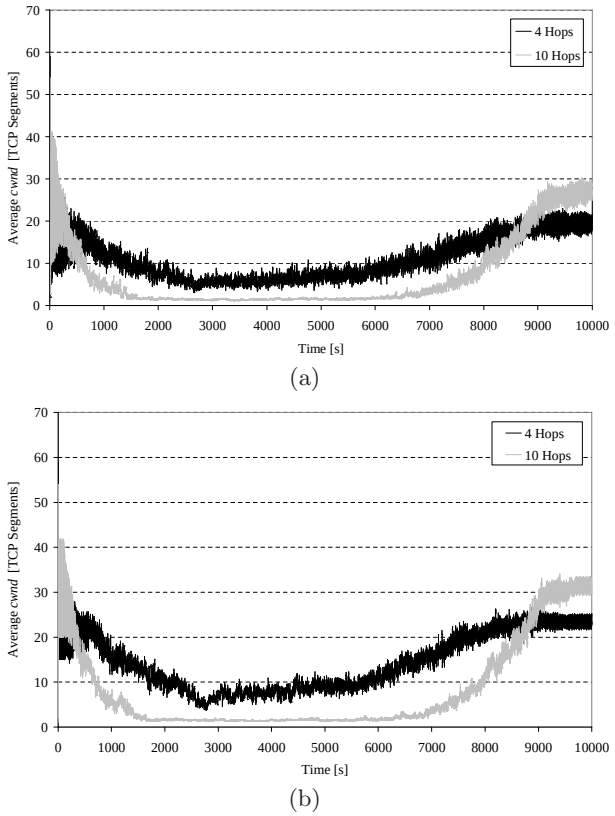

Figure 6: Average *cwnd*: (a) Westwood+; (b) New Reno.

6. REFERENCES

[1] M. Allman, S. Floyd, and C. Partridge. Increasing initial TCP's initial window. RFC 3390, October 2002.

[2] M. Allman, V. Paxson, and W. R. Stevens. TCP congestion control. IETF RFC 2581, April 1999.

[3] E. Altman. Stochastic recursive equations with applications to queues with dependent vacations. *Annals of Operations Research*, 112:43–61, 2002.

[4] E. Altman, K. Avratchenkov, and C. Barakat. A stochastic model of TCP/IP with stationary random losses. In *Proc. of ACM Sigcomm 2000*, Stockholm, Sweden, 2000.

[5] E. Altman, C. Barakat, and V. M. R. Ramos. Analysis of AIMD protocols over paths with variable delay. *Computer Networks*, 48(6):960–971, Aug. 2005.

[6] F. Baccelli and D. Hong. Interaction of TCP flows as billiards. In *Proc. of IEEE Infocom'03*, 2003.

[7] A. Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability*, 18:211–220, 1986.

[8] CAIDA. Cooperative association for internet data analysis. available at `http://www.caida.org/`, 2005.

[9] J. Chen, F. Paganini, R. Wang, M. Y. Sanadidi, and M. Gerla. Fluid-flow Analysis of TCP Westwood with RED. In *Proceedings of IEEE Globecom 2003*, San Francisco, CA, November 2003.

[10] D. Chiu and R. Jain. Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. *Computer Networks and ISDN Systems*, 17(1):1–14, June 1989.

[11] K. Fall and S. Floyd. Simulation-based Comparisons of Tahoe, Reno, and SACK TCP. *Computer Communication Review*, 26(3), July 1996.

[12] S. Floyd, T. Henderson, and A. Gurtov. NewReno modification to TCP's fast recovery. IETF RFC 3782, April 2004.

[13] P. Glasserman and D. D. Yao. Stochastic vector difference equations with stationary coefficients. *Journal of Applied Probability*, 32:851–866, 1995.

[14] L. A. Grieco and S. Mascolo. Performance evaluation and comparison of Westwood+, New Reno and Vegas TCP congestion control. *ACM Computer Communication Review*, 34(2):25–38, April 2004.

[15] L. A. Grieco and S. Mascolo. Mathematical Analysis of Westwood+ TCP Congestion Control. *IEE Proceedings Control Theory and Applications*, 152(1):35–42, January 2005.

[16] V. Jacobson. Congestion avoidance and control. In *ACM Sigcomm '88*, pages 314–329, Stanford, CA, USA, August 1988.

[17] F. Kelly. Mathematical modeling of the internet. In *Fourth International Congress on Industrial and Applied Mathematics*, Edinburgh, July 1999.

[18] T. V. Lakshman and U. Madhow. The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss. *IEEE Transactions on Networking*, 5(3):336–350, June 1997.

[19] S. H. Low, F. Paganini, and J. C. Doyle. Internet congestion control. *IEEE Control Systems Magazine*, 22(1):28–43, February 2002.

[20] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, and R. Wang. TCP Westwood: End-to-end bandwidth estimation for efficient transport over wired and wireless networks. In *ACM Mobicom 2001*, Rome, Italy, July 2001.

[21] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow. TCP Selective Acknowledgment Options. IETF RFC 2018, October 1996.

[22] V. Misra, W.-B. Gong, and D. F. Towsley. Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED. In *SIGCOMM*, pages 151–160, 2000.

[23] J. C. Mogul. Observing TCP dynamics in real networks. In *ACM Sigcomm '92*, pages 305–317, Baltimore, MD, USA, August 1992.

[24] R. Nelson. *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag, 1995.

[25] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: A simple model and its empirical validation. In *ACM Sigcomm '98*, pages 303–314, Vancouver BC, Canada, 1998.

APPENDIX

A. CHECKING OF CONDITION FOR STOCHASTIC EQUATION

Herein, we will check conditions reported in [13] to establish that the stochastic vector recursive equation (9) admits a unique solution given by eq. (10) as n grows to infinity. The conditions to be verified are [13]:

$$-\infty \leq E[\log \left\|\overline{\mathcal{A}}_0\right\|] < 0; \qquad E[\log \left\|\overline{C}_0\right\|] < \infty, \tag{43}$$

where $\|\cdot\|$ is a matrix (vector) norm. Note that index 0 refers to initial condition for $\overline{\mathcal{A}}_n$ and $\overline{C}_n$, not to the steady state.

It is well known that a matrix norm [13] has to verify the following properties:

(i) $\left\|\overline{\mathcal{S}}\right\| \geq 0$ and $\left\|\overline{\mathcal{S}}\right\| = 0$ if and only if $\overline{\mathcal{S}} = \overline{0}$;

(ii) $\forall a \in \mathbb{R},\ \left\|a\overline{\mathcal{S}}\right\| \leq |a| \left\|\overline{\mathcal{S}}\right\|$;

(iii) $\left\|\overline{\mathcal{S}} + \overline{\mathcal{T}}\right\| \leq \left\|\overline{\mathcal{S}}\right\| + \left\|\overline{\mathcal{T}}\right\|$;

(iv) $\left\|\overline{\mathcal{S}} \cdot \overline{\mathcal{T}}\right\| \leq \left\|\overline{\mathcal{S}}\right\| \cdot \left\|\overline{\mathcal{T}}\right\|$;

while a vector norm has to verify only properties (i), (ii) e (iii).

It can be easily checked that the following function is a norm:

$$\left\|\overline{\mathcal{S}}\right\| = \max_{i,j}\{S_{ij}\} \ . \tag{44}$$

Now, using norm (44), we can show that conditions (43) are verified.

For what concerns vector $\overline{C}_0$ defined in eq. (9) we have:

$$\left\|\overline{C}_0\right\| = |(1 - Z_0)\beta| = (1 - Z_0)\beta \tag{45}$$

and, evaluating its mean value:

$$E[\left\|\overline{C}_0\right\|] = E[(1 - Z_0)\beta] = \mathcal{R}^*(\lambda)\beta \ . \tag{46}$$

Now, by Jensen inequality [24] the following result holds:

$$E(\log \left\|\overline{C}_0\right\|) \leq \log E[\left\|\overline{C}_0\right\|] < \infty \tag{47}$$

that is, the second condition in (43).

For what concerns matrix $\overline{\mathcal{A}}_0$, note that W_n and $\widehat{B}_n$ in eq. (9) are dimensional variables. An appropriate selection of units of measurement can allow us to verify the first condition in (43).

In particular, if W_n is expressed in TCP segments and each R_n is expressed as multiple of the minimum value RTT_{min}, we have:

$$RTT_{min} = 1. \tag{48}$$

Moreover,

$$\begin{aligned} R_n^s \geq RTT_{min} \ \ \forall n \quad &\Rightarrow R_n^s \geq 1 \\ \Rightarrow \frac{1}{R_n^s} \leq 1 \quad &\Rightarrow E\left[\frac{1}{R_n^s}\right] \leq 1, \end{aligned} \tag{49}$$

where R_n^s represents the R_n value scaled by RTT_{min}.

Now, by definition $Z_n \leq 1$ and $E[Z_n] = 1 - \mathcal{R}^*(\lambda) < 1$, so that each term of matrix $\overline{\mathcal{A}}_n$ in eq. (9) with $n = 0$ (i.e., the initial condition) is less or equal than 1. Therefore, considering the expectation of each term, we have:

$$\begin{aligned} E\left[1 - Z_0\right] < 1; \quad E[RTT_{min} Z_0] < 1; \\ E\left[\frac{1-\alpha}{R_0^s}\right] < 1; \quad \alpha < 1 \ ; \end{aligned} \tag{50}$$

that is

$$E[\left\|\overline{\mathcal{A}}_0\right\|] < 1 \tag{51}$$

Fianlly, applying Jensen inequality [24], also the first condition in (43) is verified:

$$-\infty \leq E[\log \left\|\overline{\mathcal{A}}_0\right\|] \leq \log E[\left\|\overline{\mathcal{A}}_0\right\|] < 0 \ . \tag{52}$$