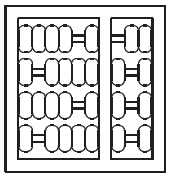

INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Quality of Service of Failure Detectors in the Presence of Loss Bursts

I. Sotoma *E. R. M. Madeira*

Technical Report - IC-03-019 - Relatório Técnico

October - 2003 - Outubro

Quality of Service of Failure Detectors in the Presence of Loss Bursts

Irineu Sotoma*† Edmundo Roberto Mauro Madeira*

Abstract

The work of Chen et al [3], *on Quality of Service (QoS) of failure detectors, did not explicitly model the possibility of loss bursts. They assumed only mean loss as the system parameter to message losses. This paper deals with the QoS of failure detectors when the probabilistic behavior of messages is extended with the probability distribution of loss burst lengths. The proposed Markov chain model, which treats loss bursts, is presented in this paper. Some simulation results are commented.*

1 Introduction

The Chen, Toueg and Aguilera's paper [3], hereafter refered as Chen et al, formalized the QoS of failure detectors and developed new failure detector (NFD) algorithms for synchronized clocks (NFD-S), and unsynchronized clocks (NFD-U and NFD-E). They assumed only *message mean loss probability* as the system parameter for message losses. However, there are networks, e.g. WANs, where the occurrence of loss bursts is very common [4, 7, 8]. So, it should be useful to extend their assumption to deal with loss bursts.

Yajnik et al [7], and Zhang [8] found out, from experiments performed on WANs, that Markov chains are adequate to model loss bursts. Sanneck [6] proposed an economic Markov chain model to loss bursts which needs only $m+1$ states, unlike traditional ones which need 2^m states. m is the order of the Markov chain, and it represents the last consecutive losses which are considered by the Markov chain. His approach uses the probability distribution of loss burst lengths to approximate both state and state transition probabilities. Therefore, we propose a Markov chain model, based on Sanneck one, to model the QoS of failure detectors in the presence of loss bursts. The simulation results show that the proposed model works better than, and similar to Chen et al work, respectively, when the system is bursty, and when it is not bursty.

This paper is organized as follows. Section 2 shows the used Sanneck model. The Sections 3 to 10, which shortly describe the results of Chen et al, are included only to make this report as self-contained as possible. Section 11 presents the proposed Markov chain model. Section 12 discusses the simulation results, and Section 13 offers some conclusions.

*Institute of Computing, University of Campinas, 13081-970 Campinas, SP.

†Research supported by FAPESP — Fundação de Amparo à Pesquisa do Estado de São Paulo, grant #00/05369-2.

2 Loss Run-Length Model

Sanneck [6] defined a model for loss run-length with a Markov chain with limited state space ($m+1$ states with limited m) (see Figure 1). The random variable X is defined as follows: $X = 0$ means *no* packet lost, $X = z$ $(0 < z < m)$ means *exactly* z packets lost, $X \geq z$ means *at least* z consecutive packets lost, and due to limited memory of the system, the last state $X = m$ is just defined as "m consecutive packets lost". A state transition occurs depending on transition probabilities p_{ij}, with $i < j$ (for loss burst lengths lower than or equal to m) or $i \geq j = 0$ (for a packet arrival), or $i = j = m$ (for loss bursts greater than m). The state probability of the system for $0 < z < m$ is $Pr(X \geq z)$, for $z = 0$ is $Pr(X = 0)$, and for $z = m$ is $Pr(X = m)$.

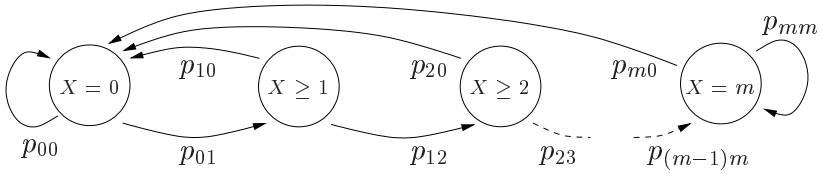

Figure 1. Sanneck model with limited state space.

3 The Failure Detector Model of Chen et al

The Chen et al model to the failure detector considers a system of two processes p and q, connected through a communication link. Process p may fail by crashing, and the link between p and q may delay and drop messages. There is a failure detector at q which monitors p, and that q does not crash. Henceforth, real time is continuous and ranges from 0 to ∞.

The output of the failure detector at q at time t is either S or T, which means that q suspects or trusts p at time t, respectively. A *transition* occurs when the output of the failure detector at q changes: an *S-transition* occurs when the output at q changes from T to S; and a *T-transition* occurs when the output at q changes from S to T. They assume that there is only a finite number of transitions during any finite time interval.

Since the behavior of the system is probabilistic, the precise definition of their model and their QoS metrics uses the theory of stochastic processes. They only consider failure detectors whose behavior eventually reaches the *steady state*. In the steady state, the probability law governing the behavior of the failure detector does not change over time. Their failure detector reaches the steady state soon after the first heartbeat message is sent (see Section 6).

4 QoS metrics for Failure Detectors of Chen et al

The QoS metrics, that Chen et al proposed, refer to the behavior of a failure detector after it reaches the steady state. They defined one metric to describe *speed* (how fast a failure detector detects crashes) and six ones to describe *accuracy* (how well the failure detector

avoids mistakes). A mistake occurs when the failure detector at q outputs T when p is still alive.

4.1 Primary Metrics

They proposed three primary metrics for the QoS specification of failure detectors. The first one measures the speed of a failure detector. It is defined in runs in which p crashes.

Detection time (T_D): this is a random variable representing the time that elapses from the time that p crashes to the time when the final *S-transition* (of the failure detector at q) occurs and there are no transitions afterward. If there is no such final *S-transition*, then $T_D = \infty$; if such an *S-transition* occurs before p crashes, then $T_D = 0$.

The following two metrics specify the accuracy of the failure detector. All accuracy metrics are defined with respect to *failure-free* runs, i.e., runs in which p does not crash.

However, the Chen thesis [2] notes that the output of any failure detector implementation at a time t should not depend on what happens after time t, i.e., the implementation does not predict the future. Therefore, the steady state behavior of a failure detector before a process p crashes is the same as its steady state behavior in runs in which p does not crash. Thus, all accuracy metrics also measure the accuracy of a failure detector in runs in which p eventually crashes (provided that this crash occurs after the failure detector has reached its steady state behavior).

There are two primary accuracy metrics:

Mistake recurrence time (T_{MR}): this is a random variable representing the time that elapses from an *S-transition* to the next one.

Mistake duration (T_M): this is a random variable representing the time that elapses from an *S-transition* to the next *T-transition*.

4.2 Derived Metrics

Besides these two accuracy metrics, they defined other four accuracy metrics which can be computed from T_{MR} and T_M:

Average mistake rate (λ_M): this measures the rate at which a failure detector makes mistakes.

Query accuracy probability (P_A): this is the probability that the failure detector's output is correct at a random time.

Good period duration (T_G): this is a random variable representing the time that elapses from a *T-transition* to the next *S-transition*.

Forward good period duration (T_{FG}): this is a random variable representing the time that elapses from a random time at which q trusts p to the time of the next *S-transition*.

4.3 How the Accuracy Metrics are Related

Theorem 1 of Chen et al explains how the six accuracy metrics are related. $Pr(A)$ denotes the probability of event A; $E(X)$, $E(X^k)$, and $V(X)$ denote the expected value (or mean), the kth moment, and the variance of random variable X, respectively.

Theorem 1. *For any ergodic failure detector, the following results hold: 1)* $T_G = T_{MR} - T_M$. *2) If* $0 < E(T_{MR}) < \infty$, *then* $\lambda_M = 1/E(T_{MR})$ *and* $P_A = E(T_G)/E(T_{MR})$. *3) If* $0 < E(T_{MR}) < \infty$ *and* $E(T_G) = 0$, *then* T_{FG} *is always 0. If* $0 < E(T_{MR}) < \infty$ *and* $E(T_G) \neq 0$, *then 3a) for all* $x \in [0, \infty)$, $Pr(T_{FG} \leq x) = \int_0^x Pr(T_G > y)dy/E(T_G)$, *3b)* $E(T_{FG}^k = E(T_G^{k+1})/[(k+1)E(T_G)]$. *In particular, 3c)* $E(T_{FG}) = [1 + V(T_G)/E(T_G)^2]E(T_G)/2$.

In failure-free runs, an ergodic failure detector is that which outputs histories which follow an ergodic probabilistic distribution. This means that, in failure-free runs, the failure detector slowly "*forgets*" its past history: from any given time on, its future behavior may depend only on its recent behavior.

5 The Probabilistic Network Model of Chen et al

Chen et al assume the following probabilistic network model:

1. Processes p (monitored process) and q (failure detector) are connected by a link that does not create or duplicate messages, but may delay or drop messages.

2. The message loss and the message delay through the link are probabilistic and are characterized by two parameters: i) *message loss probability* p_L, which is the probability that a message is dropped by the link, and ii) *message delay* D, which is a random variable with range $(0, \infty)$ representing the delay from the time a message is sent to the time it is received, under the condition that the message is not dropped by the link.

3. The expected value $E(D)$ and the variance $V(D)$ of D are finite.

4. Processes p and q have access to their own local clocks, and for these clocks there is no drift. They assert in practice, clock drift rate is usually very small.

5. The probabilistic behavior of the network does not change over time. In despite of this assumpion, they suggest ways to modify the algorithm so that it dynamically adapts to changes in the probabilistic behavior of the system.

6. The crashes can not be predicted.

7. The delay and loss behaviors of the messages that a process sends are independent of whether (and when) the process crashes.

8. Additionally, from Section 3.3 of Chen et al paper, they assume that the link from p to q satisfies the following *message independence* property: The behaviors of any two heartbeat messages sent by p are independent.

6 The NFD-S Algorithm of Chen et al

The NFD-S algorithm of Chen et al, in Figure 2, has two parameters: η and δ. The monitored process p sends periodically heartbeat messages $m_1, m_2, \ldots$ to the failure detector

process q every η time units. Every heartbeat message m_i is tagged with its sequence number i. Henceforth, σ_i denotes the sending time of message m_i.

q shifts the σ_is forward by δ to obtain the sequence of times $\tau_1 < \tau_2 < \ldots$, where $\tau_i = \sigma_i + \delta$, for $i \geq 1$. For $i = 0$, $\tau_0 = 0$. q uses the τ_is and the times on which it receives heartbeat messages to determine whether to trust or suspect p, by using every time period $[\tau_i, \tau_{i+1})$.

At time τ_i, q checks whether it has received some message m_j with $j \geq i$. If so, q trusts p during the entire period $[\tau_i, \tau_{i+1})$. If not, q starts suspecting p, and if, at some time t before τ_{i+1}, q receives some message m_j with $j \geq i$, then q starts trusting p from time t until τ_{i+1}. If q starts suspecting p at time τ_i, and by time τ_{i+1}, q has not received any message m_j with $j \geq i$, then q suspects p during the entire period $[\tau_i, \tau_{i+1})$.

Process p:

1 for all $i \geq 1$, at time $\sigma_i = i\eta$, send heartbeat m_i to q;

Process q:

2 Initialization: $output = S$; {suspect p initially}

3 for all $i \geq 1$, at time $\tau_i = \sigma_i + \delta$:

4 **if** did not receive m_j with $j \geq i$ **then** $output \leftarrow S$; {suspect p if no fresh message is received}

5 upon receive message m_j at time $t \in [\tau_i, \tau_{i+1})$:

6 **if** $j \geq i$ **then** $output \leftarrow T$; {trust p when some fresh message is received}

Figure 2. Failure detector algorithm NFD-S with parameters η and δ (clocks are synchronized).

From time τ_i to τ_{i+1}, only messages m_j with $j \geq i$ can affect the output of the failure detector. For this reason, τ_i is called a *freshness point*: from time τ_i to τ_{i+1}, messages m_j with $j \geq i$ are *still fresh*. So, NFD-S has the following property: q trusts p at time t if and only if q received a message that is still fresh at time t.

This property immediately implies that the failure detector reaches its steady state very quickly: It does so at time τ_1, i.e., δ time after the first heartbeat message is sent. This is because, after time τ_j, the state of process q only depends on what happens at or after time σ_j (the time when the jth message is sent).

7 The QoS Basic Model of Chen et al

The QoS Basic Model of Chen et al paper assumes the Lemma 2 and Proposition 13 which follow. **Lemma 2.** *For all $i \geq 0$ and all time $t \in [\tau_i, \tau_{i+1})$, q trusts p at time t if and only if q has received some message m_j with $j \geq i$ by time t.*

Proposition 13. *1) An S-transition can only occur at time τ_i for some $i \geq 2$ and it occurs at τ_i if and only if message m_{i-1} is received by q before time τ_i and no message m_j*

with $j \geq i$ is received by q by time τ_i; 2) Lemma 2 remains true if $j \geq i$ in the statement is replaced by $i \leq j \leq i+k$; 3) part 1) above remains true if $j \geq i$ in the statement is replaced by $i \leq j < i+k$.

The QoS Basic Model of Chen et al paper is described by the Definition 1, Proposition 3, and Theorem 5, which follow.

Definition 1.

1. *For any $i \geq 1$, let k be the smallest integer such that, for all $j \geq i+k$, m_j is sent at or after time τ_i.*
2. *For any $i \geq 1$, let $p_j(x)$ be the probability that q does not receive message m_{i+j} by time $\tau_i + x$, for every $j \geq 0$ and every $x \geq 0$; let $p_0 = p_0(0)$.*
3. *For any $i \geq 2$, let q_0 be the probability that q receives message m_{i-1} before time τ_i.*
4. *For any $i \geq 1$, let $u(x)$ be the probability that q suspects p at time $\tau_i + x$, for every $x \in [0, \eta)$.*
5. *For any $i \geq 2$, let p_s be the probability that an S-transition occurs at time τ_i.*

Proposition 3 below shows that Definition 1 can be expressed in a way independent of i. p_L is the probability of loss of heartbeats. $Pr(D > y)$ and $Pr(D < y)$ are respectively, the probability that a message delays more than y, and the probability that a message delays less than y.

Proposition 3.

1. $k = \lceil \delta/\eta \rceil$.
2. *For all $j \geq 0$ and for all $x \geq 0$,*
   $p_j(x) = p_L + (1 - p_L)Pr(D > \delta + x - j\eta)$.
3. $q_0 = (1 - p_L)Pr(D < \delta + \eta)$.
4. *For all $x \in [0, \eta)$,* $u(x) = \prod_{j=0}^{k} p_j(x)$.
5. $p_s = q_0 u(0)$.

Theorem 5. *Consider a system with synchronized clocks, where the probability of message losses is p_L and the distribution of message delays is $Pr(D \leq x)$. The failure detector NFD-S with parameters η and δ has the following properties:*

1. *The detection time is bounded as follows and the bound is tight:*
$$T_D \leq \delta + \eta. \tag{3.1}$$
2. *The average mistake recurrence time is:*
$$E(T_{MR}) = \frac{\eta}{p_s}. \tag{3.2}$$
3. *The average mistake duration is:*
$$E(T_M) = \frac{\int_0^\eta u(x)d(x)}{p_s}. \tag{3.3}$$

8 Configuring the Failure Detector to Satisfy QoS Requirements

This Section comes directly from Section 4 of Chen et al. They assume: 1) the local clocks of processes are synchronized and 2) one knows the probabilistic behavior of the messages, i.e., the message loss probability p_L and the distribution of message delays $Pr(D \leq x)$.

The goal is to find a configuration procedure, hereafter called configurator, which takes as input these assumptions and the QoS requirements (T_D^U, T_{MR}^L, T_M^U). T_D^U is an upper bound on the detection time, T_{MR}^L is a lower bound on the average mistake recurrence time, and T_M^U is an upper bound on the average mistake duration. In other words, the QoS requirements are that:

$$T_D \leq T_D^U, E(T_{MR}) \geq T_{MR}^L, E(T_M) \leq T_M^U. \tag{4.1}$$

Then, the configurator outputs *QoS cannot be achieved*, or the parameters η and δ satisfying the QoS requirements. To minimize the network bandwidth taken by the failure detector, the configurator intends to find the largest intersending interval η that satisfies these QoS requirements.

From Theorem 5, the goal can be restated as a mathematical programming problem:

$$\begin{aligned} &\textit{maximize } \eta \\ &\textit{subject to } \delta + \eta \leq T_D^U && (4.2) \\ &\frac{\eta}{p_s} \geq T_{MR}^L && (4.3) \\ &\frac{\int_0^\eta u(x)dx}{p_s} \leq T_M^U, && (4.4) \end{aligned}$$

where the values of $u(x)$ and p_s are given by Proposition 3. Chen et al replaced the problem (4.4) by a simpler and stronger constraint as follows.

Proposition 21. *If $p_0 > 0$ and $q_0 > 0$ (the nondegenerated case), then $E(T_M) \leq \eta/q_0$.*

So, the following configuration procedure to find η and δ results:

- *Step 1*: Compute $q_0' = (1 - p_L)Pr(D < T_D^U)$ and let $\eta_{max} = q_0' T_M^U$. If $\eta_{max} = 0$, then output *"QoS cannot be achieved"* and stop; else continue.

- *Step 2*: Let

$$f_\eta = \frac{\eta}{q_0' \prod_{j=1}^{\lceil T_D^U/\eta \rceil - 1} [p_L + (1 - p_L)Pr(D > T_D^U - j\eta)]} \tag{4.5}$$

  Find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$. Such an η always exists. To find such an η, we can use a simple numerical method, such as binary search (this works because, when η decreases, $f(\eta)$ increases exponentially fast).

- *Step 3*: Set $\delta = T_D^U - \eta$ and output η and δ.

Theorem 7. *Consider a system in which clocks are synchronized and the probabilistic behavior of messages is known. Suppose we are given a set of QoS requirements as in (4.1). The above procedure has two possible outcomes: 1) It outputs η and δ. In this case, with parameters η and δ, the failure detector NFD-S satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case,* no *failure detector can achieve the given QoS requirements.*

The above procedure may not find the optimal (largest) possible η that satisfies the QoS. The η found by the procedure is close to the optimal η depending on the distribution of message delay and the message loss. However, there is a conservative bound on the optimal η that always holds regardless of the distribution:

Proposition 8. *To satisfy the QoS constraint (4.1) with NFD-S, parameter η has to satisfy*

$$\eta \leq \eta_{max}/(p_L + (1 - p_L)Pr(D > T_D^U)),$$

where η_{max} is defined in Step 1 of the configuration procedure.

9 Dealing with Unknown Message Behavior

This Section comes directly from Section 5 of Chen et al. They assume: 1) the local clocks of processes are synchronized and 2) the probabilistic behavior of messages are unknown. In this case, it is still possible to compute η and δ by using only p_L, $E(D)$, and $V(D)$.

To do so, it is used the following *One-Sided Inequality* of the probability theory: for any random variable D with a finite expected value and a finite variance,

$$Pr(D > t) \leq \frac{V(D)}{V(D) + (t - E(D))^2}, \text{ for all } t > E(D). \tag{5.1}$$

With this, the following bounds on the QoS metrics of algorithm NFD-S can be derived:

Theorem 9. *Consider a system with synchronized clocks and assume $\delta > E(D)$. For algorithm NFD-S, we have $E(T_{MR}) \geq \eta/\beta$ and $E(T_M) \leq \eta/\gamma$, where*

$$\beta = \prod_{j=0}^{k_0} \frac{V(D) + p_L(\delta - E(D) - j\eta)^2}{V(D) + (\delta - E(D) - j\eta)^2},$$
$$k_0 = \lceil (\delta - E(D))/\eta \rceil - 1,$$

and

$$\gamma = \frac{(1 - p_L)(\delta - E(D) + \eta)^2}{V(D) + (\delta - E(D) + \eta)^2}.$$

The following configuration procedure assumes $T_D^U > E(D)$:

- *Step 1*: Compute $\gamma' = (1 - p_L)(T_D^U - E(D))^2/(V(D) + (T_D^U - E(D))^2)$ and let $\eta_{max} = min(\gamma' T_M^U, T_D^U - E(D))$. If $\eta_{max} = 0$, then output *"QoS cannot be achieved"* and stop; else continue.

- *Step 2*: Let

$$f(\eta) = \eta \cdot \prod_{j=1}^{\lceil (T_D^U - E(D))/\eta \rceil - 1} \frac{V(D) + (T_D^U - E(D) - j\eta)^2}{V(D) + p_L(T_D^U - E(D) - j\eta)^2}. \tag{5.2}$$

  Find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$. Such an η always exists.

- *Step 3*: Set $\delta = T_D^U - \eta$ and output η and δ.

Theorem 10. *Consider a system in which clocks are synchronized and the probabilistic behavior of messages is not known. Suppose we are given a set of QoS requirements as in (4.1). The above procedure has two possible outcomes: 1) It outputs η and δ. In this case, with parameters η and δ, the failure detector NFD-S satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case,* no *failure detector can achieve the given QoS requirements.*

Section 5.1 of Chen et al also shows how to estimate p_L, $E(D)$, and $V(D)$.

10 Dealing with Unknown Message Behavior and Unsynchronized Clocks

This Section comes directly from Section 6 of Chen et al. The preceding sections have assumed the clocks of p and q are synchronized. In the algorithm NFD-S, q sets the freshness points τ_is by shifting the sending times of heartbeats by a constant. When clocks are not synchronized, the local sending times of hearbeats at p cannot be used by q to set the τ_is and, thus, q needs to do it in a different way.

So, Chen et al developed a new failure detector algorithm, called NFD-U, for systems with unsynchronized clocks. The new algorithm is very similar to the NFD-S; the only difference is that q now sets the τ_is by shifting the *expected arrival times* of the heartbeats, rather than the *sending times* of heartbeats.

They assume that local clocks do not drift with respect to real time, i.e., they accurately measure time intervals. Let σ_i denote the sending time of m_i with respect to q's local clock. Then, the expected arrival time of m_i at q is $EA_i = \sigma_i + E(D)$, where $E(D)$ is the expected message delay.

They also assume that q knows the EA_is (but they show how to estimate them). To set the τ_is, q shifts the EA_is forward by α time units (i.e., $\tau_i = EA_i + \alpha$), where α is a new failure detector parameter that replaces δ.

The NFD-U algorithm of Chen et al, in Figure 3, has two parameters: η and α. NFD-U and NFD-S (see Section 6) differ only in the way they set the τ_is: in NFD-S, $\tau_i = \sigma_i + \delta$, while, in NFD-U, $\tau_i = EA_i + \alpha = \sigma_i + E(D) + \alpha$ (the last equality holds because $EA_i = \sigma_i + E(D)$). Thus, the QoS analysis of NFD-U is obtained by simply replacing δ by $E(D) + \alpha$ in the Proposition 3, Theorem 5, and Theorem 9.

Process p:{using p's local clock}

1 for all $i \geq 1$, at time $i\eta$, send heartbeat m_i to q;

Process q:{using q's local clock}

2 Initialization:

3 $\tau_0 = 0$;

4 $l = -1$; {l keeps the largest sequence number in all messages q received so far}

{if the current time reaches τ_{l+1}, then none of the messages received is still fresh}

5 upon τ_{l+1} = current time:

6 $output \leftarrow S$; {suspect p since no message received is still fresh at this time}

7 upon receive message m_j at time t:

8 **if** $j > l$ **then** {received a message with a higher sequence number}

9 $l \leftarrow j$;

10 $\tau_{l+1} \leftarrow EA_{l+1} + \alpha$; {set the next freshness point τ_{l+1} using the expected arrival time of m_{l+1}}

11 **if** $t < \tau_{l+1}$ **then** $output \leftarrow T$; {trust p since m_l is still fresh at time t}

Figure 3. Failure detector algorithm NFD-U with parameters η and α (clocks are not synchronized, but EA_is are known).

Replacing δ by $E(D) + \alpha$ in Theorem 9, the following bounds on accuracy metrics of NFD-U can be obtained:

Theorem 11. *Consider a system with drift-free clocks and assume $\alpha > 0$. For the algorithm NFD-U, we have $E(T_{MR}) \geq \eta/\beta$ and $E(T_M) \leq \eta/\gamma$, where*

$$\beta = \prod_{j=0}^{k_0} \frac{V(D) + p_L(\alpha - j\eta)^2}{V(D) + (\alpha - j\eta)^2}, \ k_0 = \lceil (\alpha/\eta \rceil - 1, \text{ and}$$

$$\gamma = \frac{(1 - p_L)(\alpha + \eta)^2}{V(D) + (\alpha + \eta)^2}.$$

Note that the bounds given in Theorem 11 use only p_L and $V(D)$, and $E(D)$ is not used.

Theorem 11 can be used to compute the parameters η and α of the failure detector NFD-U so that it satisfies the following QoS requirements:

$$T_D \leq T_D^u + E(D), E(T_{MR}) \geq T_{MR}^L, E(T_M) \leq T_M^U. \tag{6.1}$$

Note that the upper bound on the detection time T_D *is not* T_D^u, but T_D^u plus the *unknown* average message delay $E(D)$. So, the actual upper bound T_D^U on the detection time is $T_D^u + E(D)$.

So, the configuration procedure becomes as follows:

- *Step 1*: Compute $\gamma' = (1-p_L)(T_D^u)^2/(V(D)+(T_D^u)^2)$ and let $\eta_{max} = min(\gamma' T_M^U, T_D^u)$. If $\eta_{max} = 0$, then output *"QoS cannot be achieved"* and stop; else continue.
- *Step 2*: Let

$$f(\eta) = \eta \cdot \prod_{j=1}^{\lceil T_D^u/\eta \rceil - 1} \frac{V(D) + (T_D^u - j\eta)^2}{V(D) + p_L(T_D^u - j\eta)^2}. \tag{6.2}$$

  Find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$. Such an η always exists.
- *Step 3*: Set $\alpha = T_D^u - \eta$ and output η and α.

Theorem 12. *Consider a system with unsynchronized, drift-free clocks, where the probabilistic behavior of messages is not known. Suppose we are given a set of QoS requirements as in (6.1). The above procedure has two possible outcomes: 1) It outputs η and δ. In this case, with parameters η and δ, the failure detector NFD-U satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case,* no *failure detector can achieve the given QoS requirements.*

Section 6.2.2 of Chen et al also shows how to estimate p_L, and $V(D)$. Section 6.3 of Chen et al shows the NFD-E algorithm, which modifies the NFD-U algorithm only about the expected arrival times. While NFD-U assumes that q knows the exact value of all the EA_is, NFD-E estimates them.

So, each time q executes the line 10 of algorithm NFD-U, q considers the n most recent heartbeat messages, denoted by $m'_1, \ldots, m'_n$. Let $s_1, \ldots, s_n$ be the sequence numbers of such messages and $A'_1, \ldots, A'_n$ be their receipt times according to q's local clock. Then, EA_{l+1} is estimated by:

$$EA_{l+1} \approx \frac{1}{n}\left(\sum_{i=1}^{n} A'_i - \eta s_i\right) + (l+1)\eta.$$

They have shown, from simulations, that NFD-E and NFD-U are practically indistinguishable for values of n as low as 30.

11 A Model of Loss Bursts for Failure Detectors

This Section uses the Chen et al paper [3] as framework, and Chen thesis [2] to some proofs. Every Lemma, Definition, or Theorem with numeration greater than 23 is exclusive of our work; and those ones of Chen et al which were modified appear with a letter 'a' after the numeration.

11.1 Modified Probabilistic Network Model

The probabilistic network model considered in the proposed model is the same of the Chen et al one (see Section 5), except by the following changes:

1) Besides the message loss probability (p_L) and message delay (D), the link between p and q also has the additional probability distribution of loss burst lengths, given by all $p_{L,z}$'s, according to Table 1 of Section 11.2.

2) The *message independence* property is not required. There can be either the independent behavior of any two messages, or the dependent behavior of each message only with its predecessor one.

11.2 The Markov Model for Loss Bursts

The Markov model of Sanneck [6] (see Section 2) is the basis for the Definition 24.

Definition 24.

1. *Z_n is a sequence of random variables with values within the space $F = \{0, 1\}$. $Z_n = 0$ means a heartbeat message was received by q, and $Z_n = 1$ means a heartbeat message was lost.*

2. *h is the highest loss burst length which has been noted by q until the current time. We assume $h > 1$.*

3. *$S = [0, h']$, with $h' = h - 1$ and $S \subseteq \mathbf{N}$, is the set of possible states in the Markov chain.*

4. *$X_{n+1} = f(X_n, Z_{n+1})$ is the random variable which defines a Markov chain, with $X_n \in S$ and X_0 is the first observed state. If $X_n < h'$, then $X_{n+1} = Z_{n+1}X_n + Z_{n+1}$; else if $X_n = h'$, then $X_{n+1} = Z_{n+1}X_n$.*

5. *The definitions of state and state transition probabilities, and X_n values are the same of the random variable X in Section 2, by using h' in place of m, and the word "message" in place of "packet".*

The Definitions 25 and 26, at next, simplifly the notation for the state transition probabilities of the Markov chain from the Definition 24.

Definition 25. *The probability of forward state transitions, from a state b to a state $e \geq b$, is defined as $forw(b, e) = \prod_{n=b}^{e-1} p_{n(n+1)}$. When $n \geq h'$, it is used $p_{h'h'}$, according to* the Definition 24.

Definition 26. *The probability of a backward state transition, from a state i to the state 0 is defined as to$0(i) = p_{i0}$. When $i \geq h'$, it is used $p_{h'0}$, according to* the Definition 24.

The Table 1, based on Sanneck [6], shows how the probabilities used by the Markov chain of the Definition 24 could be approximated, by using only the loss probabilities, called $p_{L,z}$, for every loss burst of length z. If this information is not available in advance, some probability distribution (e.g. uniform) could be assumed before the model usage. Because a Markov chain with limited state space needs some $p_{L,n}$ with $n > h'$, which is $p_{L,h}$, to be possible the state transition probability $p_{h'h'}$, our approach is valid only when $h > 1$ (there are loss bursts). When $h = 1$, the Chen et al approach could be used. Hereafter in the formulae, $Pr(X_0 = i)$, for $i \in S$, is a short notation for $Pr(X = 0)$, $Pr(X \geq z)$, or $Pr(X = h')$ (see Table 1).

Table 1. The Markov chain probabilities calculation.

Markov model with $h'+1$ *states*	*a is the highest valid heartbeat message received*	$a \to \infty$
Burst loss $(0 < z \leq h)$	$p_{L,z} = \frac{o_z}{a}$ (o_z is the number of loss bursts of length z)	$Pr(X = z)$
Burst loss $(z = h')$ over window h'	$p'_{L,h'} = \sum_{n=h'}^{h} \frac{(n-h'+1)o_n}{a} = \frac{o_{h'}}{a} + \frac{2o_h}{a} = p_{L,h'} + 2p_{L,h}$	$Pr(X = h')$(state probability) $p_{L,cum}(h') = p'_{L,h'}$
Mean loss	$p_L = \sum_{z=1}^{h} \frac{zo_z}{a} = \sum_{z=1}^{h} zp_{L,z}$	$E[X]$
Cumulative loss $(0 < z < h')$	$p_{L,cum}(z) = \sum_{n=z}^{h} \frac{o_n}{a} = \sum_{n=z}^{h} p_{L,n}$	$Pr(X \geq z)$ (state probability)
Cumulative loss $(z = 0)$	$p_{L,cum}(0) = 1 - p_L$	$Pr(X = 0)$ (no loss case)
Conditional loss $(0 < z \leq h')$	$p_{L,cond}(z-1, z) = \frac{p_{L,cum}(z)}{p_{L,cum}(z-1)}$	$Pr(X \geq z \mid X \geq z-1)$ (state trans. prob. $p_{(z-1)z}$)
Conditional loss $(z = h')$	$p_{L,cond}(h', h') = \frac{\sum_{n=h'}^{h}(n-h')o_n}{\sum_{n=h'}^{h} no_n} = \frac{\frac{o_h}{a}}{h'(\frac{o_{h'}}{a}) + h(\frac{o_h}{a})} = \frac{p_{L,h}}{h'p_{L,h'} + hp_{L,h}}$	$Pr(X = h' \mid X = h')$ (state transition prob. $p_{h'h'}$)

The Definition 1a defines some probabilities which use the Markov chain of the Definition 24 for QoS of failure detectors in the presence of loss bursts. The Proposition 3a at next only mathematically describes the Definition 1a in a way independent of i.

Definition 1a.

1. *For any* $i \geq 1$, *let* k *be the smallest integer such that, for all* $j \geq i + k$, m_j *is sent at or after time* τ_i.

2. *For any* $i \geq 2$, *let* q_0 *be the probability that* q *receives the message* m_{i-1} *before time* τ_i. *In this case, the Markov chain goes to state 0.*

3. *For any* $i \geq 1$, *let* $u(x)$ *be the probability that* q *suspects* p, *by receiving no one of the messages* m_{i+j}, *for every* $0 \leq j \leq k-1$, *at time* $\tau_i + x$, *for all* $x \in [0, \eta)$. *From state 0, the Markov chain takes transitions.*

4. *For any* $i \geq 2$, *let* p_s *be the probability that an S-transition occurs at time* τ_i. *This characterizes the whole Markov chain.*

Proposition 3a.

1. $k = \lceil \delta/\eta \rceil$.

2. $q_0 = \sum_{n=0}^{h'} Pr(X_0 = n) to0(n) Pr(D < \delta + \eta)$.

3. *For all* $x \in [0, \eta)$, $u(x) = u_k(x)$. $u_w(x)$, *with* w *initially equal to* k, *is defined as*

follows:

$$
\begin{aligned}
u_1(x) &= forw(0,1) + to0(0)Pr(D > \delta + x - (k-1)\eta) \\
u_w(x) &= to0(0)Pr(D > \delta + x - (k-w)\eta)u_{w-1}(x) \\
&\quad + \sum_{a=1}^{w-2} forw(0,a)to0(a)Pr(D > \delta + x - (a+k-w)\eta)u_{w-(a+1)}(x) \\
&\quad + forw(0,w-1)to0(w-1)Pr(D > \delta + x - (k-1)\eta) \\
&\quad + forw(0,w)
\end{aligned}
$$

4. $p_s = q_0 u(0)$.

In the Proposition 3a.3: i) k messages could be received within $[\tau_{i-1}, \tau_i + x)$; ii) the total number of combination of losses (represented by bits 1) and delays (represented by bits 0) is 2^k; iii) the w index indicates how many messages are being considered, and from what one among $k-w$ to $k-1$. For example, $w = k$ considers the messages 0 to $k-1$, and $w = k-1$ considers the messages 1 to $k-1$. The following proof is based on the Proposition 4.2 proof in the Chen thesis.

Proof of Proposition 3a.

1) The proof of the Proposition 3a.1 is the same of Chen thesis: it is immediate from the fact that m_j is sent at time $\tau_i - \delta + (j-i)\eta$ for all $i \geq 1$.

2) The proof of the Proposition 3a.2 directly follows from the fact that q_0 is the probability of m_{i-1} is not lost and be received with delay less than $\delta + \eta$ time units, causing a state transition to state 0 ($to0(n)$).

3) The proof of $u(x)$ is built in parts:

a) $u_1(x) = forw(0,1) + to0(0)Pr(D > \delta + x - (k-1)\eta)$ represents the probability of the $(k-1)$-th message be lost, or be delayed with a delay greater than $\delta + x - (k-1)\eta$ time units to it be not received within $[\tau_{i-1}, \tau_i + x)$.

b) $forw(0,w-1)to0(w-1)Pr(D > \delta + x - (k-1)\eta)$ in $u_w(x)$ considers the probability of patterns with suffix $1^{w-1}0$. $forw(0,w-1)$ gives the probability of the sequence of $w-1$ losses of messages $k-w$ to $k-2$. So $to0(w-1)Pr(D > \delta + x - (k-1)\eta)$ gives the probability of the $(k-1)$-th message be received with delay greater than $\delta + x - (k-1)\eta$.

c) $forw(0,w)$ in $u_w(x)$ considers the probability of patterns with suffix 1^w, i.e., the probability of the sequence of w losses of messages $k-w$ to $k-1$.

d) $to0(0)Pr(D > \delta + x - (k-w)\eta)u_{w-1}(x)$ in $u_w(x)$ considers the probability of patterns with preffix 0. $to0(0)Pr(D > \delta + x - (k-w)\eta)$ means the probability of the delay of the $(k-w)$-th message be greater than $\delta + x - (k-w)\eta$ time units. $u_{w-1}(x)$ represents the recurrence which calculates the probabilities of the 2^{w-1} combinations of the following $w-1$ messages. This recurrence finishes when $w = 2$, by calling $u_1(x)$.

e) $\sum_{a=1}^{w-2} forw(0,a)to0(a)Pr(D > \delta + x - (a+k-w)\eta)u_{w-(a+1)}$ in $u_w(x)$ considers the probability of patterns which have prefix 1^a0, where a, $1 \leq a \leq w-2$, is the number of consecutive losses of messages. The probability resulting from losses are described by $forw(0,a)$. $to0(a)Pr(D > \delta + x - (a+k-w)\eta)$ represents the probability of the $(a+k-w)$-th message be delayed more than $\delta + x - (a+k-w)\eta$. $u_{w-(a+1)}(x)$ represents the probability of the $2^{w-(a+1)}$ combinations of the following $w-(a+1)$ messages. This recurrence finishes when $a = w-2$, by calling $u_1(x)$.

A strong induction proof follows to verify if the recurrence works. In the induction base, when $w = k = 1$, $u_1(x)$ works as explained in part 3a) above. When $w = k = 2$, $u_2(x)$ clearly uses the first, third, and fourth terms of $u_w(x)$ definition, and only $u_1(x)$ in the first term. In the induction hypothesis, we consider $u_w(x)$ works when $2 \leq w \leq k-1$. In the induction step, we verify if $u_w(x)$ works when $2 \leq w \leq k$. It is clear that when $w = k$, the first term of $u_w(x)$ uses $u_{w-1}(x)$, which by the induction hypothesis is calculated by $u_{k-1}(x)$. When $w = k$, the second term of $u_w(x)$ uses $u_{w-(a+1)}(x)$, where $w-(a+1)$ varies from $k-2$ to 1. Therefore, by the induction hypothesis, $u_{w-(a+1)}(x)$ is correctly calculated. The proofs of third and fourth terms directly follow from parts 3b) and 3c) above. So, $u_w(x)$ works when $2 \leq w \leq k$.

4) From the Proposition 3a.2, q outputs T, and from the Proposition 3a.3, q outputs S, leading to an *S-transition*. So, p_s really is the probability that an *S-transition* occurs at time τ_i. □

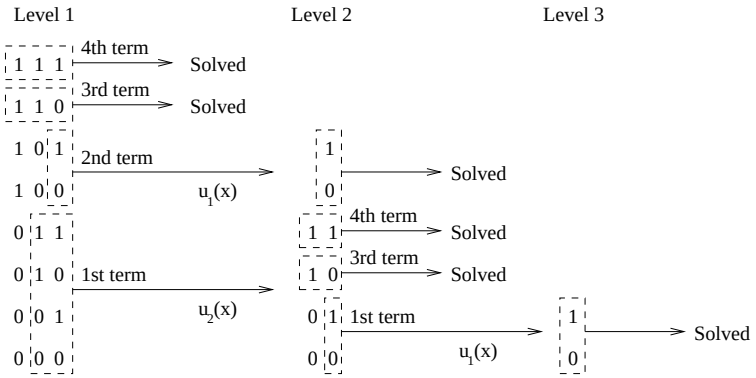

Figure 4. An example for the Proposition 3a.3 when $k = 3$.

Figure 4 presents an example which uses the terms of the Proposition 3a.3. $k = 3$, which leads to 3 levels of recursion and 2^3 bit combinations. This example shows the recursion in the building of the bit pattern:
In level 1: a) the first term treats patterns which begin with 0 (011, 010, 001, and 000), by calling $u_2(x)$; b) the second term deals with patterns which begin with sequences of 1's and are followed by a sequence of 0's (101 and 100), except those patterns which have sequences of 1's and only the last bit is 0; the patterns 1 and 0 are treated by calling $u_1(x)$; c) the third term treats patterns where all bits are 1, except the last one which is 0 (110); and d) the fourth term manages patterns where there is only 1's (111).
In level 2: The patterns (11, 10, 01, and 00) are treated by $u_2(x)$, whose: a) first term treats the patterns 01 and 00, and calls $u_1(x)$; b) second term solves the pattern 10; and c) fourth term solves the pattern 11. The patterns 1 and 0 are solved by $u_1(x)$.
In level 3: the patterns 1 and 0 are solved by $u_1(x)$.

The Proposition 27 describes the cases where the probability of message loss is different to 0 and 1, and the probability of a message to arrive within the expected time is greater than 0. It is used later on by the Proposition 21a.

Proposition 27. *The nondegenerated cases in the proposed model occur when* $0 <$

$Pr(X = 0) < 1$, $0 < Pr(X \geq 1) < 1$, *and* $q_0 > 0$.

Proof. $0 < Pr(X = 0) < 1$ means a heartbeat message is received with probability different to 0 and 1. $0 < Pr(X \geq 1) < 1$ means a heartbeat message is lost with probability different to 0 and 1. $q_0 > 0$ is assumed to guarantee p_s is different from 0, otherwise, no *S-transition* would occur. □

The Definition 28 defines the probability for q to suspect p, and the Proposition 29 mathematically describes this probability.

Definition 28. *For any $i \geq 1$, let $u'(x)$ be the probability that q suspects p at time $\tau_i + x$, for every $x \in [0, \eta)$. This suspicion occurs when no one of the messages m_{i+j} is received by time $\tau_i + x$, for every $0 \leq j \leq k-1$. Our $u'(x)$ assumes the Markov chain can be in any initial state $s \in S$.*

Proposition 29. $u'(x) = \sum_{s=0}^{h'} Pr(X_0 = s) u_{s,k}(x)$. *$u_{w-1}(x)$ and $u_{w-(a+1)}(x)$ use the $u(x)$ definition in the Proposition 3a and the Definition 28. $u_{s,w}(x)$, with w initially equal to k, is defined as follows:*

$$
\begin{aligned}
u_{s,1}(x) &= forw(s, s+1) + to0(s) Pr(D > \delta + x - (k-1)\eta) \\
u_{s,w}(x) &= to0(s) Pr(D > \delta + x - (k-w)\eta) u_{w-1}(x) \\
&\quad + \sum_{a=1}^{w-2} forw(s, s+a) to0(s+a) Pr(D > \delta + x - (a+k-w)\eta) u_{w-(a+1)}(x) \\
&\quad + forw(s, s+w-1) to0(s+w-1) Pr(D > \delta + x - (k-1)\eta) \\
&\quad + forw(s, s+w).
\end{aligned}
$$

Proof. The proof of $u'(x)$ is immediate from the fact that from any Markov chain state is possible to loss a message (forward transition) or to receive a message (transition to state 0). Additionally, the $u(x)$ definition in Proposition 3a can be used because the Markov chain is always in state 0 when $u_{w-1}(x)$ and $u_{w-(a+1)}(x)$ are called in $u_{s,w}(x)$ definition. □

The Proposition 30, at next, is used by the following Proposition 14a.

Proposition 30. *The Markov chain used in Proposition 3a has all state transition probabilities greater than 0.*

Proof. This is immediate from Definition 24 (X_{n+1} definition and $X_n \in [0, h']$). □

Proposition 14a. $u(0) > 0$, *in the nondegenerated cases, and for all $x \in [0, \eta)$, $u(0) \geq u(x)$.*

Proof. From $u(x)$ definition in Proposition 3a.4, $u(0) = u_k(0)$. In this case we also consider messages 0 to $k-1$. $u_w(0)$, with w initially equal to k, is then:

$$
\begin{aligned}
u_1(0) &= forw(0, 1) + to0(0) Pr(D > \delta - (k-1)\eta) \\
u_w(0) &= to0(0) Pr(D > \delta - (k-w)\eta) u_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-2} forw(0, a) to0(a) Pr(D > \delta - (a+k-w)\eta) u_{w-(a+1)}(0) \\
&\quad + forw(0, w-1) to0(w-1) Pr(D > \delta - (k-1)\eta) \\
&\quad + forw(0, w)
\end{aligned}
$$

When $k = 1$, by using Proposition 30, $u(0) = u_1 > 0$. When $k > 1$, it is enough to prove that one of the terms of $u_w(0)$ is greater than 0. Then, by using only the fourth term of $u_w(0)$ and Proposition 30: $u_w(0) \geq forw(0, w)$. So, $u(0) > 0$. To prove that for all $x \in [0, \eta)$, $u(0) \geq u(x)$, it is enough to note $u(x)$ is influenced by $Pr(D > \delta + x - y\eta)$, where y can be $k - w$, $a + k - w$, or $k - 1$. Since $0 \leq y \leq \delta$, $\delta + x - y\eta \geq 0$ for all $x \in [0, \eta)$. As greater the $\delta + x - y\eta$ value is, lower or equal is the $Pr(D > \delta + x - y\eta)$ value. Moreover, $Pr(D > \delta + x - y\eta)$ is always used in multiplications followed by sums of positive numbers. Therefore, if $x_1 = 0$ and $x_2 \in [0, \eta)$, $Pr(D > \delta + x_1 - y\eta) \geq Pr(D > \delta + x_2 - y\eta)$, which implies $u(x_1) \geq u(x_2)$, leading to $u(0) \geq u(x)$. □

The following Lemma 15a is used later on by Theorem 5a. Its proof is the same to that of Lemma 15 of Chen et al paper, except by the use of $u'(x)$ of the Proposition instead $u(x)$.

Lemma 15a. $P_A = 1 - \frac{1}{\eta} \int_0^\eta u'(x)dx$.

The following Lemma 17 of Chen et al paper is still valid in our model.

Lemma 17. $\{(T_{MR,n}, T_{M,n}), n = 1, 2, \ldots\}$ *is a delayed renewal reward process.*

The following Lemma 4 of Chen et al paper is still valid in our model because our model uses the same NFD-S algorithm.

Lemma 4. *NFD-S is an ergodic failure detector.*

The following Lemma 16a is still valid in our model, by using the p_s definition in the Proposition 3a. Its proof is the same of that of Lemma 16 of Chen et al paper, except by the use of $u(0) > 0$ of the Proposition 14a. For all $i \geq 2$, they let A_i be the event that an S-transition occurs at time τ_i. By the use of $u(0) > 0$ is possible to assert that $Pr(A_i) = p_s = q_0 u(0) > 0$ in nondegenerated cases.

Lemma 16a. $E(T_{MR}) = \eta / p_s$.

Because the same idea of algorithm NFD-S is used, the following Lemma 18 and its proof of Chen et al paper is still valid in our model.

Lemma 18. $T_D \leq \delta + \eta$ *and this bound is tight.*

The following theorem summarizes our QoS analysis of the NFD-S, by using the previous definitions and propositions which follow the Definition 24.

Theorem 5a. *Consider a system with synchronized clocks, where the probability of message loss p_L, the distribution of message delays $Pr(D \leq x)$, and the probability distribution of loss burst lengths are known. The failure detector NFD-S with parameters η and δ has the following properties:*

1. *The detection time is bounded as follows and the bound is tight:*

$$T_D \leq \delta + \eta. \tag{3.1}$$

2. *The average mistake recurrence time is:*

$$E(T_{MR}) = \frac{\eta}{p_s}. \tag{3.2a}$$

3. *The average mistake duration is:*

$$E(T_M) = \frac{\int_0^\eta u'(x)d(x)}{p_s}. \tag{3.3a}$$

Proof. The parts 1 and 2 of the theorem are direct from Lemmas 18 and 16a. Part 3 is derived from the relation between $E(T_M)$, P_A, and $E(T_{MR})$, as given in part 2 of the Theorem 1 and the results on P_A and $E(T_{MR})$ as given by Lemmas 15a and 16a. □

Our goal is to find a configuration procedure, hereafter called configurator, which takes as input the probabilistic behavior of heartbeats and the QoS requirements (T_D^U, T_{MR}^L, T_M^U), and outputs η and δ. T_D^U is an upper bound on the detection time, T_{MR}^L is a lower bound on the average mistake recurrence time, and T_M^U is an upper bound on the average mistake duration. In other words, the QoS requirements are that:

$$T_D \leq T_D^U, E(T_{MR}) \geq T_{MR}^L, E(T_M) \leq T_M^U. \tag{4.1}$$

From the Theorem 5a, the goal can be restated as a mathematical programming problem:

maximize η:

$$\text{subject to } \delta + \eta \leq T_D^U \tag{4.2}$$

$$\frac{\eta}{p_s} \geq T_{MR}^L \tag{4.3a}$$

$$\frac{\int_0^\eta u'(x)dx}{p_s} \leq T_M^U \tag{4.4a}$$

where the value of $u'(x)$ is given by the Proposition 29, and the value of p_s is given by the Proposition 3a. Similar to Chen et al, the problem (4.4a) was replaced by a simpler and stronger constraint as follows.

Proposition 21a. In the nondegenerated cases of the Proposition 14a, $E(T_M) \leq \frac{u'(0)\eta}{q_0 u(0)}$.

Proof. By Proposition 14a, $u(0) \geq u(x)$, for all $x \in [0, \eta)$, which is also valid to $u'(0) \geq u'(x)$. Thus, from equality (3.3a) and Proposition 3a:

$$E(T_M) = \frac{\int_0^\eta u'(x)dx}{p_s} \leq \frac{\int_0^\eta u'(0)dx}{q_0 u(0)} = \frac{u'(0)\eta}{q_0 u(0)}.$$

□

Hereafter, E and V are short notations for E(D) and V(D), respectively.

From the problem (4.2) and Propositions 3a and 21a, we obtain the following Proposition 31, which is used later on by the Theorem 5a configurator, which is defined after the Proposition 31.

Proposition 31. *Let be* $k = \lceil T_D^U/\eta \rceil$. *At next,* $u'(0)$ *and* $u''(0)$ *consider, like Chen et al, only the messages* 1 *to* $k-1$. $u'(0) = \sum_{s=0}^{h'} Pr(X_0 = s) u''_{s,k}(0)$. $u''_{s,w}(0)$, *which is based on Proposition 29, is defined as follows:*

$$\begin{aligned}
u''_{s,2}(0) &= forw(s, s+1) + to0(s) Pr(D > T_D^U - (k-1)\eta) \\
u''_{s,w}(0) &= to0(s) Pr(D > T_D^U - (k-w+1)\eta) u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-3} forw(s, s+a) to0(s+a) Pr(D > T_D^U - (a+k-w+1)\eta) u''_{w-(a+1)}(0) \\
&\quad + forw(s, s+w-2) to0(s+w-2) Pr(D > T_D^U - (k-1)\eta) \\
&\quad + forw(s, s+w-1).
\end{aligned}$$

The terms $u''_{w-1}(0)$ *and* $u''_{w-(a+1)}(0)$ *of* $u''_{s,w}(0)$ *use the following* $u''(0)$ *definition, which is*

based on Proposition 3a:

$$
\begin{aligned}
u_2''(0) &= forw(0,1) + to0(0)Pr(D > T_D^U - (k-1)\eta) \\
u_w''(0) &= to(0)Pr(D > T_D^U - (k-w+1)\eta)u_{w-1}''(0) \\
&\quad + \sum_{a=1}^{w-3} forw(0,a)to0(a)Pr(D > T_D^U - (a+k-w+1)\eta)u_{w-(a+1)}''(0) \\
&\quad + forw(0,w-2)to0(w-2)Pr(D > T_D^U - (k-1)\eta) \\
&\quad + forw(0,w-1).
\end{aligned}
$$

Proof. To prove $u''(0) = u(0)$, we follow a reasoning similar to Chen et al. By part 1 of Theorem 5a, we use $T_D \leq T_D^U = \eta + \delta$ in the following:
$u_1(0) = forw(0,1) + to0(0)Pr(D > \delta - (k-1)\eta)$, for $k = \lceil \delta/\eta \rceil$ and message $k-1$
$= forw(0,1) + to0(0)Pr(D > \delta + \eta - (k-1)\eta)$, for $k = \lceil (\delta+\eta)/\eta \rceil$ and message $k-1$
$= forw(0,1) + to0(0)Pr(D > T_D^U - (k-1)\eta) = u_2''(0)$, for $k = \lceil T_D^U/\eta \rceil$ and message $k-1$.
$u_w(0) = to(0)Pr(D > \delta - (k-w)\eta)u_{w-1}(0)$
$+ \sum_{a=1}^{w-2} forw(0,a)to0(a)Pr(D > \delta - (a+k-w)\eta)u_{w-(a+1)}(0)$
$+ forw(0,w-1)to0(w-1)Pr(D > \delta - (k-1)\eta)$
$+ forw(0,w)$, for $k = \lceil \delta/\eta \rceil$ and messages 0 to $k-1$
$u_w(0) = to(0)Pr(D > \delta + \eta - (k-w)\eta)u_{w-1}''(0)$
$+ \sum_{a=1}^{w-3} forw(0,a)to0(a)Pr(D > \delta + \eta - (a+k-w+1)\eta)u_{w-(a+1)}''(0)$
$+ forw(0,w-2)to0(w-2)Pr(D > \delta + \eta - (k-1)\eta)$
$+ forw(0,w-1)$, for $k = \lceil (\delta+\eta)/\eta \rceil$ and messages 1 to $k-1$
$u_w(0) = to(0)Pr(D > T_D^U - (k-w)\eta)u_{w-1}''(0)$
$+ \sum_{a=1}^{w-3} forw(0,a)to0(a)Pr(D > T_D^U - (a+k-w+1)\eta)u_{w-(a+1)}''(0)$
$+ forw(0,w-2)to0(w-2)Pr(D > T_D^U - (k-1)\eta)$
$+ forw(0,w-1) = u_w''(0)$, for $k = \lceil T_D^U/\eta \rceil$ and messages 1 to $k-1$. So, $u''(0) = u(0)$. The proof about $u'(0)$ of Proposition 29 be equal to $u'(0)$ of Proposition 31 follows directly from the proof above on $u''(0) = u(0)$. The only difference is the use of the state s on probabilities $to0$ and $forw$. □

From the problems (4.1), (4.2), (4.3a), (4.4a) and Propositions 21a and 31, we obtain the following configurator, called Theorem 5a configurator, to find η and δ:
Step 1: Compute $q_0' = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)Pr(D < T_D^U)$ and let $g(\eta) = u'(0)\eta/q_0'u''(0)$, where $u''(0) = u_k''(0)$. If $q_0'u''(0) = 0$, then output *"QoS cannot be achieved"* and stop; else continue. Find the largest $\eta_{max} \leq T_D^U$ such that $g(\eta_{max}) \leq T_M^U$.
Step 2: Let $f(\eta) = \eta/q_0'u''(0)$, find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$.
Step 3: Set $\delta = T_D^U - \eta$ and output η and δ.

Theorem 7a. *Consider a system in which clocks are synchronized and the probability of message loss p_L, the distribution of message delays $Pr(D \leq x)$, and the probability distribution of loss burst lengths are known. Suppose we are given a set of QoS requirements as in (4.1). The Theorem 5a configurator has two possible outcomes: 1) It outputs η and δ. In this case, with parameters η and δ, the failure detector NFD-S satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case,* no *failure detector can achieve the given QoS requirements.*

Proof. We prove the theorem in the following two parts:

1. Suppose that the configurator outputs *"QoS cannot be achieved"*. Then, the configurator stops at *Step 1* and, thus, $q_0' u''(0) = 0$. What follows about $q_0' = 0$ is from Chen et al. $q_0' = 0$ implies $Pr(D < T_D^U) = 0$. This means that, in such a system, no message is received within T_D^U time units after it is sent. Then, to satisfy $T_D \leq T_D^U$, we claim that, at any time $t > T_D^U$, any failure detector has to suspect p. In fact, since all messages q has received by time t are sent before time $t - T_D^U$, q does not obtain any information about whether p crashes at time $t - T_D^U$. Thus, to satisfy $T_D \leq T_D^U$, q has to suspect p at time t. Hence, for any failure detector, we have $E(T_M) = \infty$ and thus, it fails to satisfy $E(T_M) \leq T_M^U$. $u''(0) = 0$ implies $E(T_M) = g(\eta) = \infty$, which leads to the failure detector to fail to satisfy $E(T_M) \leq T_M^U$. Therefore, no failure detector can satisfy the given QoS in this case.

2. Suppose that the configurator outputs parameters η and δ. Then, by *Step 3*, we have $T_D^U = \eta + \delta$. By part 1 of Theorem 5a, $T_D \leq T_D^U$ is satisfied. By *Step 1* and Proposition 3a, $q_0' = \sum_{n=0}^{h'} Pr(X_0 = n) to0(n) Pr(D < \eta + \delta) = q_0$. Note that we have $q_0' u''(0) > 0$ since, otherwise, $g(\eta) = \infty$, and the configurator would output *"QoS cannot be achieved"* instead of η and δ. By Proposition 21a and *Step 1*, $E(T_M) \leq g(\eta) = \frac{u'(0)\eta}{q_0' u''(0)} = \frac{u'(0)\eta}{q_0 u(0)} \leq T_M^U$. So, $E(T_M) \leq T_M^U$ is satisfied. Thus, $f(\eta) = \eta / q_0' u''(0) = \eta / q_0 u(0) = \eta / p_s = E(T_{MR})$ by (3.2a). By *Step 2*, $f(\eta) \geq T_{MR}^L$ is satisfied. □

To compute η and δ when $Pr(D \leq x)$ is unknown, we can use, similarly to Chen et al, the following *One-Sided Inequality*: For any random variable D with a finite expected value and a finite variance, $Pr(D > t) \leq \frac{V}{V + (t-E)^2}$, for all $t > E$.

By applying that *One-Sided Inequality* on the Propositions 3a and 29, we obtain the Proposition 32 and the Theorem 9a, at next. Like Chen et al, the following Theorems 9a and 11a, and the Propositions 32 e 34, consider only the messages 0 to k_0, and their configurators consider only the messages 1 to k_0.

By applying the *One-Sided Inequality* on Propositions 3a and 29, we obtain the Proposition 32 at next.

Proposition 32. *Let be* $k_0 = \lceil (\delta - E)/\eta \rceil - 1$. *At next,* $u'(0)$ *and* $u''(0)$ *consider, like Chen et al, only the messages* 0 *to* k_0. $u'''(0) = \sum_{s=0}^{h'} Pr(X_0 = s) u''_{s,k}(0)$. $u''_{s,w}(0)$, *which is based on Proposition 29, is defined as follows:*

$$
\begin{aligned}
u''_{s,0}(0) &= forw(s, s+1) + to0(s)(V/(V + (\delta - E - k_0\eta)^2)) \\
u''_{s,w}(0) &= to0(s)(V/(V + (\delta - E - (k_0 - w)\eta)^2)) u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-1} forw(s, s+a) to0(s+a)(V/(V + (\delta - E - (a + k_0 - w)\eta)^2)) u''_{w-(a+1)}(0) \\
&\quad + forw(s, s+w) to0(s+w)(V/(V + (\delta - E - k_0\eta)^2)) \\
&\quad + forw(s, s+w+1).
\end{aligned}
$$

The terms $u''_{w-1}(0)$ *and* $u''_{w-(a+1)}(0)$ *of* $u''_{s,w}(0)$ *use the following* $u''(0)$ *definition, which is based on the Proposition 3a:*

$$
\begin{aligned}
u_0''(0) &= forw(0,1) + to0(0)(V/(V+(\delta - E - k_0\eta)^2)) \\
u_w''(0) &= to(0)(V/(V+(\delta - E - (k_0 - w)\eta)^2))u_{w-1}''(0) \\
&\quad + \sum_{a=1}^{w-1} forw(0,a)to0(a)(V/(V+(\delta - E - (a + k_0 - w)\eta)^2))u_{w-(a+1)}''(0) \\
&\quad + forw(0,w)to0(w)(V/(V+(\delta - E - k_0\eta)^2)) \\
&\quad + forw(0,w+1).
\end{aligned}
$$

Proof. Note that for all j such that $0 \le j \le k_0$, $\delta - j\eta > E(D) \Leftrightarrow j < (\delta - E(D))/\eta \Rightarrow max(j) = \lceil(\delta - E(D))/\eta\rceil - 1 = k_0$. $k_0 \le \lceil\delta/\eta\rceil - 1 = k - 1$. From Propositions 3a and 14a, with w initially equal to k, we have $u(x) \le u(0) = u_1(0)$, or $u(x) \le u(0) = u_w(0)$. For $w = k = 1$:
$u_1(0) = forw(0,1) + to0(0)Pr(D > \delta - (k-1)\eta)$. By the *One-Sided Inequality*, $u_1(0) \le forw(0,1) + to0(0)(V/(V+(\delta - (k-1)\eta - E)^2)$. By considering message $k_0 - 1$, we have $forw(0,1) + to0(0)(V/(V+(\delta - E - (k_0-1)\eta)^2)$, and by considering message k_0, we have $forw(0,1) + to0(0)(V/(V+(\delta - E - k_0\eta)^2) = u_0''(0)$. For $w = k > 1$:
$u_w(0) = to0(0)Pr(D > \delta - (k-w)\eta)u_{w-1}(0)$
$+\sum_{a=1}^{w-2} forw(0,a)to0(a)Pr(D > \delta - (a+k-w)\eta)u_{w-(a+1)}(0)$
$+forw(0,w-1)to0(w-1)Pr(D > \delta - (k-1)\eta)$
$+forw(0,w)$. By the *One-Sided Inequality*,
$u_w(0) \le to0(0)(V/(V+(\delta - (k-w)\eta - E)^2))u_{w-1}(0)$
$+\sum_{a=1}^{w-2} forw(0,a)to0(a)(V/(V+(\delta - (a+k-w)\eta - E)^2))u_{w-(a+1)}(0)$
$+forw(0,w-1)to0(w-1)(V/V+(\delta - (k-1)\eta - E)^2)$
$+forw(0,w) = temp1$. By considering messages 0 to $k_0 - 1$,
$temp1 \le to0(0)(V/(V+(\delta - (k_0-w)\eta - E)^2))u_{w-1}(0)$
$+\sum_{a=1}^{w-2} forw(0,a)to0(a)(V/(V+(\delta - (a+k_0-w)\eta - E)^2))u_{w-(a+1)}(0)$
$+forw(0,w-1)to0(w-1)(V/V+(\delta - (k_0-1)\eta - E)^2)$
$+forw(0,w) = temp2$. By considering messages 0 to k_0,
$temp2 \le to(0)(V/(V+(\delta - E - (k_0-w)\eta)^2))u_{w-1}''(0)$
$+\sum_{a=1}^{w-1} forw(0,a)to0(a)(V/(V+(\delta - E - (a+k_0-w)\eta)^2))u_{w-(a+1)}''(0)$
$+ forw(0,w)to0(w)(V/(V+(\delta - E - k_0\eta)^2))$
$+ forw(0,w+1) = u_w''(0)$.
By analogy, the proof of $u'''(0)$ follows directly from the proof of $u''(0)$. □

Theorem 9a. *By assuming the Proposition 32, consider a system with syncronized clocks and assume $\delta > E$. For the algorithm NFD-S, we have $E(T_{MR}) \ge \eta/\beta$ and $E(T_M) \le \frac{u'''(0)\eta}{\gamma' u''(0)}$, where*
$\gamma' = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)((\delta + \eta - E)^2/(V+(\delta + \eta - E)^2)$
$\beta = u''(0)$. For $E(T_M) \le \frac{u'''(0)\eta}{\gamma' u''(0)}$, we assume $\frac{u'(0)}{u(0)} \le \frac{u'''(0)}{u''(0)}$, where $u'(0)$ is from the Proposition 29 and $u(0)$ is from the Proposition 3a.

Proof. By the Proposition 3a and the *One-Sided Inequality*, we have
$q_0 = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)Pr(D < \delta + \eta)$

$\sum_{n=0}^{h'} Pr(X_0 = n)to0(n)(1 - Pr(D \geq \delta + \eta))$
$\geq \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)(1 - \frac{V}{V+(\delta+\eta-E)^2})$
$= \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)(\frac{(\delta+\eta-E)^2}{V+(\delta+\eta-E)^2}) = \gamma'.$

By applying the *One-Sided Inequality* on the Proposition 21a, $E(T_M) \leq \frac{u'(0)\eta}{q_0 u(0)}$, we have $u'(0) \leq u'''(0)$, $u(0) \leq u''(0)$, and $q_0 \geq \gamma'$. Therefore $E(T_M) \leq \frac{u'(0)\eta}{q_0 u(0)} \leq \frac{u'''(0)\eta}{\gamma' u''(0)}$ is valid only when $\frac{u'(0)}{u(0)} \leq \frac{u'''(0)}{u''(0)}$. In this case, if $\frac{u'''(0)\eta}{\gamma' u''(0)} \leq T_M^U$, then $E(T_M) \leq \frac{u'(0)\eta}{q_0 u(0)} \leq T_M^U$. To prove whether $E(T_{MR}) \geq \eta/\beta = \eta/u''(0)$, we use the fact that, by using the *One-Sided Inequality*, we have $u(0) \leq u''(0)$. So, by Propositions 3a and (3.2a), $E(T_{MR}) = \eta/p_s = \eta/q_0 u(0) \geq \eta/\beta = \eta/u''(0)$. Therefore, if $\eta/u''(0) \geq T_{MR}^L$, then $\eta/p_s \geq T_{MR}^L$. □

From the problem (4.2), and the Theorem 9a, we obtain the following Proposition 33, which is used later on by the Theorem 9a configurator, which is defined after the Proposition 32.

Proposition 33. *Let be* $k_0 = \lceil (T_D^U - E)/\eta \rceil - 1$. $u'''(0) = \sum_{s=0}^{h'} Pr(X_0 = s) u''_{s,k_0}(0)$. $u''_{s,w}(0)$ *based on the Proposition 32, is defined as follows:*

$$
\begin{aligned}
u''_{s,1}(0) &= forw(s, s+1) + to0(s)(V/(V + (T_D^U - k_0\eta - E)^2)) \\
u''_{s,w}(0) &= to0(s)(V/(V + (T_D^U - (k_0 - w + 1)\eta - E)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-2} forw(s, s+a)to0(s+a)(V/(V + (T_D^U - (a + k_0 - w + 1)\eta - E)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(s, s+w-1)to0(s+w-1)(V/(V + (T_D^U - k_0\eta - E)^2)) \\
&\quad + forw(s, s+w).
\end{aligned}
$$

The terms $u''_{w-1}(0)$ *and* $u''_{w-(a+1)}(0)$ *of* $u'''(0)$ *use the following* $u''(0)$ *definition, which is based on the Proposition 32:*

$$
\begin{aligned}
u''_1(0) &= forw(0, 1) + to0(0)(V/(V + (T_D^U - k_0\eta - E)^2)) \\
u''_w(0) &= to(0)(V/(V + ({T_D}^U - (k_0 - w + 1)\eta - E)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-2} forw(0, a)to0(a)(V/(V + (T_D^U - (a + k_0 - w + 1)\eta - E)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(0, w-1)to0(w-1)(V/(V + (T_D^U - k_0\eta - E)^2)) \\
&\quad + forw(0, w).
\end{aligned}
$$

From the problems (4.1), (4.2), (4.3a), (4.4a), Theorem 9a, and Propositions 21a and 32, we obtain the following configurator, called Theorem 9a configurator, to find η and δ:
Step 1: Compute $\gamma' = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)((T_D^U - E)^2/(V + (T_D^U - E)^2)$ and let $g'(\eta) = u'''(0)\eta/\gamma' u''(0)$. If $\gamma'_0 u''(0) = 0$, then output *"QoS cannot be achieved"* and stop; else continue. Find the largest $\eta_{max} \leq T_D^U - E$ such that $g'(\eta_{max}) \leq T_M^U$.
Step 2: Let $f(\eta) = \eta/\beta'$, where $\beta' = u''_w(0)$. Find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$.
Step 3: Set $\delta = T_D^U - \eta$ and output η and δ.

Theorem 10a. *Consider a system in which clocks are synchronized and the probability behavior of messages is not known. Suppose we are given a set of QoS requirements as in*

(4.1), and suppose $T_D^U > E(D)$. The Theorem 9a configurator has two possible outcomes: 1) It outputs η and δ. In this case, with parameters η and δ, the failure detector NFD-S satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case, no *failure detector can achieve the given QoS requirements.*

With drift-free clocks, we can replace δ by $E + \alpha$ in the Theorem 9a to obtain the Proposition 34 and Theorem 11a which follow. The QoS requirements are (T_D^u, T_{MR}^L, T_M^U), where $T_D^U = T_D^u + E$.

From the Proposition 32, by replacing δ with $E + \alpha$, we obtain the Proposition 34 at next.

Proposition 34. *Let be $k_0 = \lceil \alpha/\eta \rceil - 1$. At next, $u'(0)$ and $u''(0)$ consider, like Chen et al, only the messages 0 to k_0. $u'''(0) = \sum_{s=0}^{h'} Pr(X_0 = s) u''_{s,k}(0)$. $u''_{s,w}(0)$, which is based on Proposition 32, is defined as follows:*

$$
\begin{aligned}
u''_{s,0}(0) &= forw(s, s+1) + to0(s)(V/(V + (\alpha - k_0\eta)^2)) \\
u''_{s,w}(0) &= to0(s)(V/(V + (\alpha - (k_0 - w)\eta)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-1} forw(s, s+a)to0(s+a)(V/(V + (\alpha - (a + k_0 - w)\eta)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(s, s+w)to0(s+w)(V/(V + (\alpha - k_0\eta)^2)) \\
&\quad + forw(s, s+w+1).
\end{aligned}
$$

The terms $u''_{w-1}(0)$ and $u''_{w-(a+1)}(0)$ of $u''_{s,w}(0)$ use the following $u''(0)$ definition, which is based on the Proposition 32:

$$
\begin{aligned}
u''_0(0) &= forw(0, 1) + to0(0)(V/(V + (\alpha - k_0\eta)^2)) \\
u''_w(0) &= to(0)(V/(V + (\alpha - (k_0 - w)\eta)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-1} forw(0, a)to0(a)(V/(V + (\alpha - (a + k_0 - w)\eta)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(0, w)to0(w)(V/(V + (\alpha - k_0\eta)^2)) \\
&\quad + forw(0, w+1).
\end{aligned}
$$

Theorem 11a. *By assuming the Proposition 34, consider a system with drift-free clocks and assume $\alpha > 0$. For the algorithm NFD-U, we have $E(T_{MR}) \geq \eta/\beta$ and $E(T_M) \leq \frac{u'''(0)\eta}{\gamma' u''(0)}$, where*
$\gamma' = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)((\alpha + \eta)^2/(V + (\alpha + \eta)^2)$
$\beta = u''(0)$. *For $E(T_M) \leq \frac{u'''(0)\eta}{\gamma' u''(0)}$, we assume $\frac{u'(0)}{u(0)} \leq \frac{u'''(0)}{u''(0)}$, where $u'(0)$ is from the Proposition 29 and $u(0)$ is from the Proposition 3a.*

From the problem (4.2), the Theorem 11a, and the Proposition 34, and the fact that $T_D^U = T_D^u + E$, we obtain the following Proposition 35, which is used later on by the Theorem 11a configurator.

Proposition 35. *Let be $k_0 = \lceil T_D^u/\eta \rceil - 1$. $u'''(0) = \sum_{s=0}^{h'} Pr(X_0 = s) u''_{s,k}(0)$. $u''_{s,w}$, which is based on the Proposition 34, is defined as follows:*

$$
\begin{aligned}
u''_{s,1}(0) &= forw(s, s+1) + to0(s)(V/(V + (T_D^u - k_0\eta)^2)) \\
u''_{s,w}(0) &= to0(s)(V/(V + (T_D^u - (k_0 - w + 1)\eta)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-2} forw(s, s+a)to0(s+a)(V/(V + (T_D^u - (a + k_0 - w + 1)\eta)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(s, s+w-1)to0(s+w-1)(V/(V + (T_D^u - k_0\eta)^2)) \\
&\quad + forw(s, s+w).
\end{aligned}
$$

The terms $u''_{w-1}(0)$ *and* $u''_{w-(a+1)}(0)$ *use the following* $u''(0)$ *definition, which is based on the Proposition 34:*

$$
\begin{aligned}
u''_1(0) &= forw(0, 1) + to0(0)(V/(V + (T_D^u - k_0\eta)^2)) \\
u''_w(0) &= to(0)(V/(V + (T_D^u - (k_0 - w + 1)\eta)^2))u''_{w-1}(0) \\
&\quad + \sum_{a=1}^{w-2} forw(0, a)to0(a)(V/(V + (T_D^u - (a + k_0 - w + 1)\eta)^2))u''_{w-(a+1)}(0) \\
&\quad + forw(0, w-1)to0(w-1)(V/(V + (T_D^u - k_0\eta)^2)) \\
&\quad + forw(0, w).
\end{aligned}
$$

From the problems (4.1), (4.2a), (4.3a), (4.4a), Propositions 21a and 35, Theorem 11a, and due to $T_D^U = T_D^u + E$, we get the next configurator (called SM) to find η and α:
Step 1: Compute $\gamma' = \sum_{n=0}^{h'} Pr(X_0 = n)to0(n)((T_D^u)^2/(V + (T_D^u)^2)$ and let $g'(\eta) = u'''(0)\eta/\gamma' u''(0)$.
If $\gamma'_0 u''(0) = 0$, then output *"QoS cannot be achieved"* and stop; else continue. Find the largest $\eta_{max} \leq T_D^u$ such that $g'(\eta_{max}) \leq T_M^U$.
Step 2: Let $f(\eta) = \eta/\beta'$, where $\beta' = u''_w(0)$. Find the largest $\eta \leq \eta_{max}$ such that $f(\eta) \geq T_{MR}^L$.
Step 3: Set $\alpha = T_D^u - \eta$ and output η and α.

Theorem 12a. *Consider a system with unsynchronized, drift-free clocks, where the probabilistic behavior of messages is not known. Suppose we are given a set of QoS requirements as in (6.1). The configurator SM has two possible outcomes: 1) It outputs* η *and* δ. *In this case, with parameters* η *and* δ, *the failure detector NFD-U satisfies the given QoS requirements. 2) It outputs "QoS cannot be achieved". In this case,* no *failure detector can achieve the given QoS requirements.*

12 Simulation of the Proposed Model

In this Section, we have used the outputs (η and α) from Chen et al's configurator (Chen) and from our configurator SM (see the end of Section 11) to Chen et al's NFD-E algorithm (a variant of NFD-U) of the Section 10. We have performed two simulations for accuracy and two for detection time. These simulations have basic settings similar to Chen et al:

the message delay D follows the exponential distribution, $T_M^U = 1$, $T_{MR}^L = 10$ for $T_D^U = 1$, $T_{MR}^L = 100$ for $T_D^U \in \{1.5, 2.0\}$, and $T_{MR}^L = 10000$ for $T_D^U \in \{2.5, 3.0\}$. In the first simulation, $E(D) = 0.02$, $V(D) = 0.004$, and $p_L = 0.01$. In the second one, $E(D) = 0.1$, $V(D) = 0.01$, and $p_L = 0.03$. h is the maximum loss burst length used to generate the bursts (see Definition 24).

In the following figures, for each value of T_D^U, we plotted $E(T_{MR})$, or $E(T_M)$, by considering, respectively, the average of a run of mistake recurrences or mistake duration intervals. However, for limiting the simulation time, when the simulated time has reached $(x+10) * T_{MR}^L$, each simulation run stops. x is the number of intervals considered to obtain the $E(T_{MR})$ and $E(T_M)$. For $T_D^U \in \{1.0\}$, $x = 10000$. For $T_D^U \in \{1.5, 2.0\}$, $x = 1000$. For $T_D^U \in \{2.5, 3.0\}$, $x = 100$. This leads to the generation of at least $(x+10) * T_{MR}^L / \eta$ messages, when $\eta \leq 1$.

For $h \in \{3, 7, 12\}$, the bursty traffic was generated with an uniform distribution on $p_{L,z}$'s, by using the p_L formula in Table 1. The nonbursty traffic was generated only with p_L and $E(D)$ (without the $p_{L,z}$'s), similar to Chen et al paper. In this case, the configurator SM assumes $h = 2$, to get $h' = 1$ (see Definition 24).

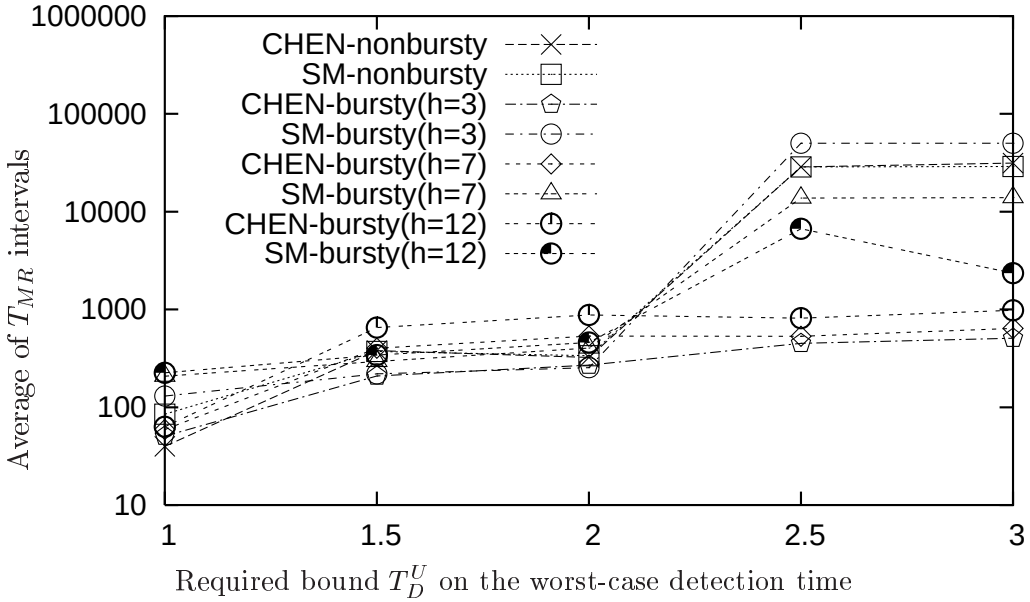

Figure 5. This simulation shows if $E(T_{MR}) \geq T_{MR}^L$ for $p_L = 0.01$ and $E(D) = 0.02$.

The figures 5 and 6 which follow show the simulation results for $p_L = 0.01$ and $E(D) = 0.02$. Figure 5 shows if $E(T_{MR})$ satisfies T_{MR}^L. On bursty traffic, the configurator SM satisfies T_{MR}^L in almost all cases, except in 2 ones when $h = 12$: $T_D^U \in \{2.5, 3\}$. The Chen's configurator satisfies T_{MR}^L in almost all cases, except in 6 ones: when $T_D^U \in \{2.5, 3\}$ and $h \in \{3, 7, 12\}$. Figure 6 shows if $E(T_M)$ satisfies T_M^U. The configurator SM satisfies T_M^U in all cases (bursty and nonbursty). On bursty traffic, the Chen's configurator does not satisfy $E(T_M) \leq 1$ in 9 cases: a) $h = 3$ and $T_D^U = 2$; and b) $h \in \{7, 12\}$ and $T_D^U \in \{1.5, 2, 2.5, 3\}$. On nonbursty traffic, the configurator SM behaves similarly to Chen's configurator, by

satisfying all cases for $E(T_{MR}) \geq T_{MR}^L$ and $E(T_M) \leq T_M^U$.

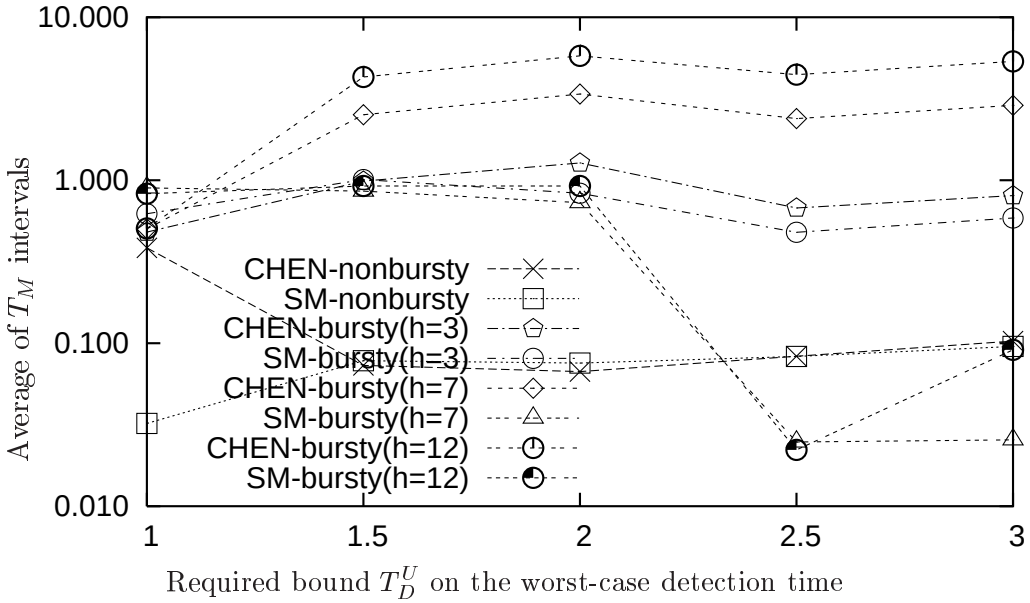

Figure 6. This simulation shows if $E(T_M) \leq T_M^U$ for $p_L = 0.01$ and $E(D) = 0.02$.

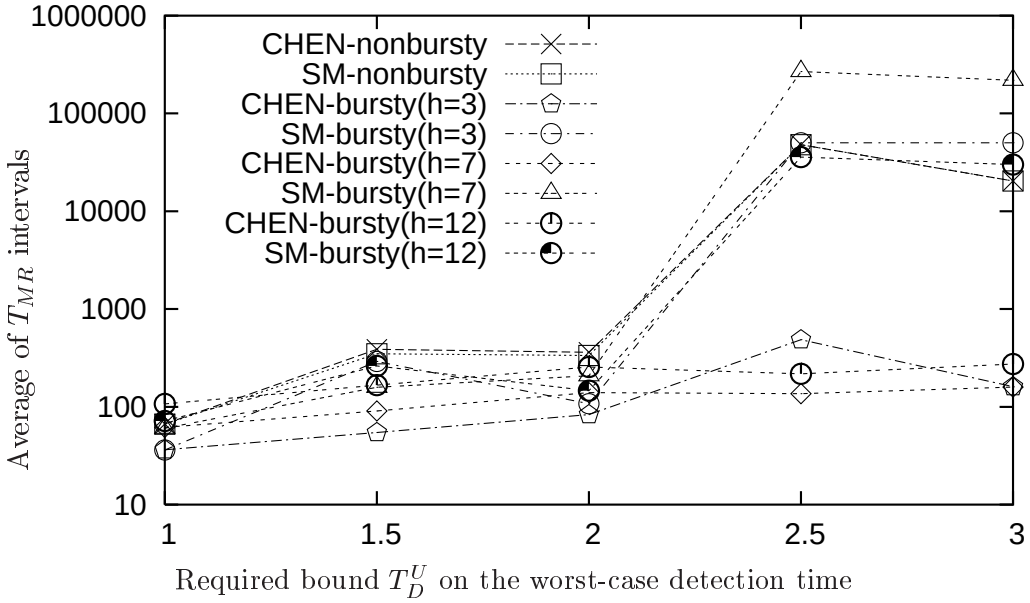

Figure 7. This simulation shows if $E(T_{MR}) \geq T_{MR}^L$ for $p_L = 0.03$ and $E(D) = 0.1$.

The figures 7 and 8 which follow show the simulation results for $p_L = 0.03$ and $E(D) = 0.1$. Figure 7 shows if $E(T_{MR})$ satisfies T_{MR}^L. On bursty traffic, the configurator SM satisfies T_{MR}^L in all cases. The Chen's configurator satisfies T_{MR}^L only in 6 cases, from 15 ones: when

$T_D^U = 1$ and $h \in \{3, 7, 12\}$, $T_D^U = 1.5$ and $h = 12$, and $T_D^U = 2$ and $h \in \{7, 12\}$. Figure 8 shows if $E(T_M)$ satisfies T_M^U. The configurator SM satisfies T_M^U in all cases (bursty and nonbursty), except when $T_D^U = 1$ and $h = 7$ in bursty traffic. On bursty traffic, the Chen's configurator does not satisfy $E(T_M) \leq 1$ in 11 cases: $h = 3$ and $T_D^U = 2$; and b) $h \in \{7, 12\}$ and $T_D^U \in \{1, 1.5, 2, 2.5, 3\}$. On nonbursty traffic, the configurator SM behaves similarly to Chen's configurator, by satisfying all cases for $E(T_{MR}) \geq T_{MR}^L$ and $E(T_M) \leq T_M^U$.

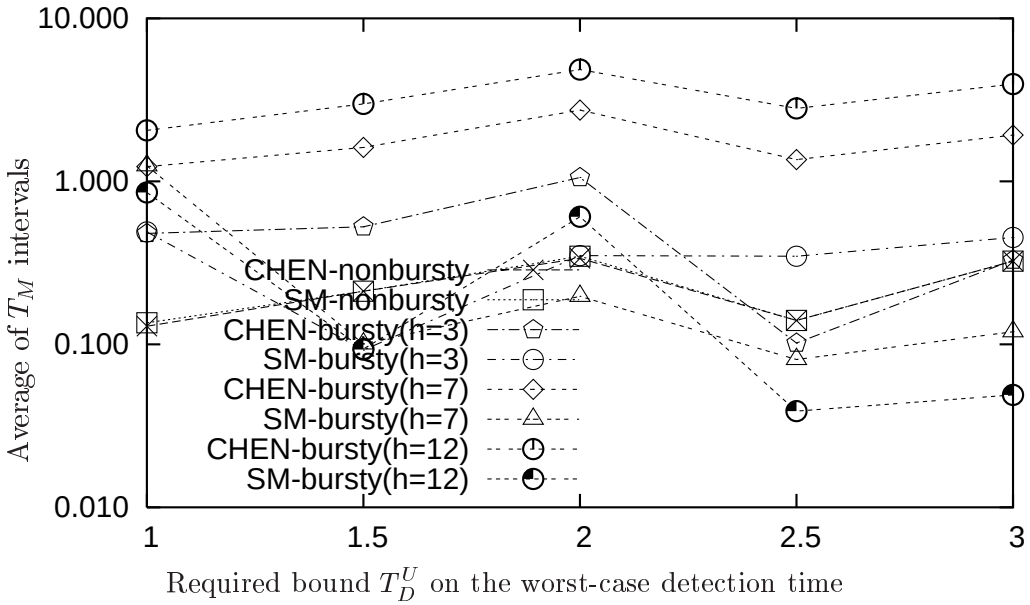

Figure 8. This simulation shows if $E(T_M) \leq T_M^U$ for $p_L = 0.03$ and $E(D) = 0.1$.

When $h = 3$ and $T_D^U \in \{2.5, 3\}$, the figures 5 and 7 show the configurator SM with $E(T_{MR}) \sim 50000$, which is much lower than real values. This was done to make the layout better because no T_{MR} interval has occurred in these cases.

On nonbursty traffic, both configurators behave similarly, by satisfying T_{MR}^L and T_M^U. The configurator SM meets better the QoS requirements because as the loss burst length increases, it generates lower η values and greater α values (see the end of Section 11). Unlike the Chen's configurator, which does not take into account the loss burst lengths.

Besides these simulations, we have verified if the required bound T_D^U is satisfied. For each value of T_D^U, in 50 runs with arbitrary crash times, both configurators always satisfy T_D^U.

It is important to highlight that the uniform distribution on loss bursts is rarely found in practice. The most common probability distributions on Internet are variations of geometric and exponential distributions [1, 5]. So, additional experiments should be made to evaluate both configurators. To guarantee the QoS, the geometric (exponential) distribution could be thought as the best case (all cases satisfied) to the configurators, and the uniform distribution could be thought as the worst one (some cases could be not satisfied).

13 Conclusions

This paper has extended the paper of Chen et al [3], by proposing a Markov model for QoS of failure detectors suitable to occurrence of loss bursts. The simulation results show that, on nonbursty traffic and on bursty traffic (when the original Chen et al work can fail), the new configurators guarantee the QoS requirements in all cases and in the greater number of cases, respectively.

14 Acknowledgements

The authors thank Jorge Stolfi by the useful comments about the Markov modeling and the impact of the *One-Sided Inequality* on the formulae.

References

[1] J. Andrén, M. Hilding, and D. Veitch. Understanding End-to-End Internet Traffic Dynamics. *Proceedings of IEEE 1998 Global Communications Conference (GLOBECOM 98)*. Sidney, Australia, November 1998.

[2] W. Chen. *On the Quality of Service of Failure Detectors*. PhD thesis, Cornell University, Cornell, May 2000.

[3] W. Chen, S. Toueg, and M. K. Aguilera. On the Quality of Service of Failure Detectors. *IEEE Transactions on Computers*, 51(5):561–580, May 2002.

[4] S. Floyd and V. Paxson. Difficulties in Simulating the Internet. *IEEE/ACM Transactions on Networking*, 9(4):392–403, August 2001.

[5] A. Konrad, B. Y. Zhao, A. D. Joseph, and R. Ludwig. A Markov-Based Channel Model Algorithm for Wireless Networks. *Proceedings of Fourth ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2001)*. Rome, Italy, July 2001.

[6] H. Sanneck. *Packet Loss Recovery and Control for Voice Transmission over the Internet*. PhD thesis, Technischen Universität Berlin, Berlin, October 2000.

[7] M. Yajnik, S. B. Moon, J. Kurose, and D. Towsley. Measurement and modeling of the temporal dependence in packet loss. In *Proceedings of INFOCOM'99*, New York, USA, March 1999.

[8] Y. Zhang. *Characterizing End-to-end Internet Performance*. PhD thesis, Cornell University, Cornell, August 2001.