Transient Analysis for Multi-Hop Wireless Networks Under Static Routing

Jaya Prakash Champati, *Member, IEEE,* Hussein Al-Zubaidy, *Senior Member, IEEE,* and James Gross, *Senior Member, IEEE*

***Abstract*—In this article, we investigate the transient behavior of a sequence of packets/bits traversing a multi-hop wireless network under static routing. Our work is motivated by novel applications from the domain of process automation, Machine-Type Communication (MTC) and cyber-physical systems, where short messages are communicated and statistical guarantees need to be provided on a per-message level. In order to optimize such a network, apart from understanding the stationary system dynamics, an understanding of the short-term dynamics (i.e. transient behavior) is also required. To this end, we derive novel Wireless Transient Bounds (WTB) for end-to-end delay and backlog in a multi-hop wireless network using stochastic network calculus approach. We start by analyzing a single end-to-end path, i.e. a line topology, and then we show how the obtained results can be applied to a mesh network with static routing using a concept called '*leftover service*'. WTB depends on the initial backlog at each node as well as the instantaneous channel states. We numerically compare WTB with Kernel-Based-Transient Bound (KBTB), which can be obtained by adapting existing stationary bound, as well as simulated end-to-end delay of the investigated network. While KBTB and stationary bounds are not able to capture the short-term system dynamics well, WTB provides relatively tight upper bound and has a decay rate that closely matches the simulation. This is achieved by WTB only with a slight increase in the computational complexity, by a factor of $O(T+N)$, where T is the duration of the arriving sequence and N is the number of hops in the network. We believe that the presented analysis and the bounds are necessary tools for future work on transient network optimization for many important emerging applications, e.g., massive MTC, critical MTC, edge computing and autonomous vehicle.**

***Index Terms*—Transient analysis; machine type communication; stochastic network calculus; time-critical networks; wireless**

I. Introduction

Since the introduction of wireless communication networks, there has always been a significant interest in characterizing their real-time and delay performance. Over the last decade this interest has been revived by novel applications from the domain of automation and control, and cyber-physical systems. The introduction of Ultra-Reliable Low-Latency Communication (URLLC) as one major innovation of fifth generation (5G) cellular networks [1], [2] is a good example of the contemporary industrial interest in such applications. Another example is the evolution of Time-Sensitive Networking (TSN)[1] – which has originally been targeting multimedia applications run over Ethernet – into a new automation-centric networking standard with extensions to wireless networks.

These practical developments are accompanied by renewed research efforts in communication theory as well as queuing analysis for wireless networks. On the communication-theoretic level, for instance, there has been an intense interest in so-called finite-blocklength link performance models [3]. In queuing analysis, on the one hand, latency analysis has received significant attention, while on the other hand, over the last five years, there is an increasing interest in new metrics like the Age-of-Information [4]. Much of this work is essentially motivated by the desire to understand system trade-offs with respect to latency, freshness of packets, etc. which is essential for automation and control applications.

In automation and control, the operation mode generally can be distinguished into time-triggered versus event-triggered systems. Time-triggered systems are characterized by periodic state transmissions, resembling more or less packetized constant-rate traffic. In contrast, event-triggered systems generate state transmissions only in case that a significant state change has happened. Thus, depending on the system dynamics and the impact of noise, event-triggered systems can expose an underlying communication system to sudden and bursty communication requests, for which nevertheless stringent latency constraints need to be met per packet.

This event-triggered nature and its implications for per-packet latency guarantee motivates us to turn to corresponding performance models. In analyzing queuing systems, we distinguish two perspectives, *transient* and *steady-state* analysis. It is well known that under certain conditions of the random arrival and service process a queuing system has a steady state, which is characterized by stationarity of performance measures such as delay/sojourn time and queue length. If the arrival and service processes do not meet those conditions, the system only has transient states, i.e. the distribution of the delay for instance changes continuously over time. From the discussion above, it is immediately clear that steady-state performance models are not the best choice for event-triggered control systems. Instead, transient analysis appears more suitable as only the current network state and its consequences for the immediate, near future are relevant.

However, most of the research dealing with general network performance analysis using queuing theory consider systems in steady state. This is also true for alternative approaches to (wireless) network performance analysis like effective capacity [5] and stochastic network calculus [6]–[8]. In contrast,

J.P. Champati, H. Al-Zubaidy, and J. Gross are with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {jpra,hzubaidy,jamesgr}@kth.se). This research was supported in part by the Swedish Research Council (VR) under grant 2016-04404.

[1] https://1.ieee802.org/tsn/

the literature on transient analysis is generally sparse. This is due to the fact that queuing analysis in transient state quickly becomes intractable. For example, while for simple M/M/1 or more general Markovian queuing systems the steady state is governed by the (conceptually simple) flow balance equations, in case of transient analysis the involved differential equations lead to intractable expressions even for M/M/1 systems [9]. Subsequently, either approximations or numerical methods for the transient analysis have been proposed [10]–[12].

Furthermore, despite the potential relevance of transient analysis for communication networks, it has received little attention when analyzing practically relevant networking effects. In [13] the selection of the TCP congestion window is studied by applying transient analysis for flows of short lengths. Through a simple recursive formula for the average completion time of the flow transmission, the authors showed a significant impact of different window settings. Nevertheless, their model does not account for queuing effects along the route, among other issues. A second application example is the analysis of ATM networks [14], where a transient analysis based on an extension of Petri nets is presented. While demonstrating a very strong aspect of transient analysis in general - for example, the ability to characterize practically relevant overload situations (which cannot be dealt with using steady-state analysis) - the presented approach nevertheless rests heavily on numerical methods that limit the analytical insight. With respect to stochastic network calculus, one exception that considers transient behavior of wireless networks is [15]. In this work the effects of transient phases on delay and backlog for sleep scheduling in wireless networks are investigated by proposing non-stationary service curves. However, they numerically evaluate the probabilistic bounds for specific service processes, and do not consider the effect of initial backlog in the system.

In this work, we strive to establish novel tools for transient network performance analysis, in particular, for wireless networking. To this end, we propose an analytic model based on stochastic network calculus, and present novel bounding techniques that allow an accurate characterization of the transient probabilistic delay of a multi-hop wireless network. We consider a sequence of packets/bits of interest that arrive abruptly at a source node which is to be delivered to a destination node via a dedicated route consisting of N consecutive wireless links on a first-come-first-serve (FCFS) basis. This multi-hop route is modeled as a tandem of queues with time varying random service process. Importantly, our model considers the initial queue occupancy at the time of injection of the packet/bit sequence, since it has a considerable effect on the sequence's transient delay performance under FCFS. As we show in Section IV-E, the obtained results can also be applied to a more general mesh network topology wherein a static source-to-destination route is assumed to be known a priori, for example, pre-computed by a routing algorithm.

The main contribution of this work lies in deriving new performance bounds for single- and multi-hop transmission scenarios that can be used to characterize the probabilistic end-to-end delay of instantaneously arriving packets given a certain backlog in the system. The proposed bounds do not rely on any restriction on the arrival process and are applicable to i.i.d. wireless channels with general service distribution. They are also shown (numerically) to be significantly tighter compared with the available stationary bounds in the literature. Furthermore, we derive probabilistic bounds for the backlog in the system, and present a method to extend the bounds for the case where initial backlog information is delayed. An important takeaway message of this work is that the currently available approaches - which study network performance while assuming steady-state network operation - *can significantly misrepresent* the transient performance of individual packets. We believe that our results can contribute to efficient control algorithms for future networks, where individual delay performance per packet has significant application-layer impact. In the context of automation applications, this applies for instance to messages belonging to safety applications.

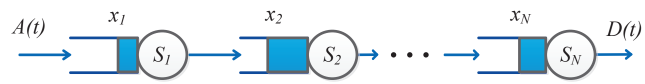


Fig. 1: Model of the considered multi-hop wireless network.

The rest of the paper is organized as follows. In Section II, we describe the network model and state basic assumptions. In Section III, we provide some background for the problem and the used methodology. In Section IV, we present our main contribution: the derivation of novel performance bounds for transient network operation. In Section V, we present numerical results to evaluate the proposed bounds and compare them to simulation results for multi-hop wireless networks. We conclude in Section VI.

II. Network Model

We investigate a wireless network composed of multiple store-and-forward relay nodes in tandem with one arrival process, $A(t)$, and one departure process, $D(t)$, as shown in Figure 1. Arriving delay-critical packets are to be delivered across the multi-hop network without violating a given probabilistic end-to-end delay constraint. Our goal is to develop a model that facilitates end-to-end delay analysis of a time-critical packet sequence on a per-message basis. Later, in Section IV-E, we show how the devised model and the obtained results can be applied to the analysis of a mesh topology with static routing. We next describe the queueing model in Figure 1 and provide a precise problem statement.

A. Queuing Model

We consider a fluid-flow, discrete-time queuing model for the N-hop wireless network shown in Figure 1. We are interested in studying the delay performance of the network with respect to a message sequence that abruptly arrives at time t_0 and lasts for T time slots, where $t_0 \geq 0$ is any arbitrary time. At time t, we denote the buffer state of the network by $\mathbf{x}(t) \in \mathbb{Z}^N_{\geq 0}$, where $x_n(t)$ is the backlog of wireless link n at time t, $n \in \{1, 2, \ldots, N\}$ and $\mathbb{Z}_{\geq 0}$ is the set of non-negative integers. To simplify notation, we designate the initial backlog at time t_0 as $\mathbf{x}(t_0) \equiv \mathbf{x}$. Given $\mathbf{x}$, we ignore the arrivals before time t_0 and simply consider that the system started at time t_0 with initial backlog $\mathbf{x}$. Later, in Section IV-C we also consider the case where $\mathbf{x}$ is not known instantaneously, but at time t_0 delayed backlog information is available, e.g. through later feedback from other links, from time $t_0 - d$ for $0 \le d \le t_0$.

The arrivals of interest are described by the cumulative arrival process $A(t_0, t)$, $0 \le t_0 \le t$ where $A(u,t) = \sum_{i=u}^{t-1} a_i$, for all $t_0 \le u \le t$, a_i is the arrival increment during time slot i and $a_i = 0$, for all $i \notin [t_0, t_0 + T)$. After being served by the system, the arrivals result in a departure process $D(t_0, t)$. Given t_0, we define $A(t) = A(t_0, t)$ and $D(t) = D(t_0, t)$. In this work, we do not impose any restriction on $A(t)$, i.e. arrival increments can take independent and arbitrary values, in deriving the bounds. We also present expressions for the case where $A(t)$ obeys the (σ, ρ)-bounded traffic characterization introduced by Cruz [16] – a typical assumption for deriving bounds in the network calculus literature.

Definition 1. *The cumulative arrival process $A(t)$ is (σ, ρ)-bounded if,*

$$A(t) - A(u) \le \rho(t-u) + \sigma, \; \forall \, t_0 \le u \le t, \tag{1}$$

for some $\sigma \ge 0$ and $\rho \ge 0$.

The cumulative service provided by the n^{th} wireless link is given by

$$S_n(u,t) = \sum_{i=u}^{t-1} \mu_{n,i}, \tag{2}$$

where $\mu_{n,i}$ is the capacity of the n^{th} wireless link in time slot i. We assume that the service processes are i.i.d. across both time slots and links (i.e., no cross link interference). We consider general distribution for the service processes and derive the results, but for the purposes of numerical evaluation and simulation we use point-to-point Rayleigh-block fading channel model. We assume first-come-first-serve discipline for the designated arrival sequence at each store-and-forward node along the path.

The total backlog $B(t)$ and the end-to-end virtual delay $W(t)$ of the queuing network described above are then given by

$$B(t) = A(t) + \sum_{n=1}^{N} x_n - D(t), \tag{3}$$

and

$$W(t) = \inf\left\{ w \ge 0 : A(t) + \sum_{n=1}^{N} x_n \le D(t+w) \right\}. \tag{4}$$

B. Problem Statement

The arriving message of interest $A(t)$ is constrained by end-to-end delay deadline w, i.e. data arrived at link 1 by time t must must be transmitted before $t + w + 1$. We strive for an analysis of the *delay violation probability* of the virtual delay $W(t)$ for $t \in [t_0, t_0 + T)$, i.e. an analysis of $\mathbb{P}(W(t) > w)$ for the time-critical data as it traverses the multi-hop wireless network described above under the crucial assumption of the initial backlog $\mathbf{x}$ along the route. As discussed, the usual perspective in analyzing queuing systems is steady state analysis,

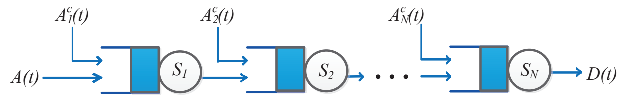


Fig. 2: Equivalent model.

where the metrics of interest are stationary [17]. In contrast, the case where the stochastic properties of the metrics are not stationary is commonly referred to as transient state. In this spirit, we refer to our analysis goal as *transient analysis*, as the stochastic properties of the virtual delay of interest are clearly non-stationary.

For convenience, we define $\tau = t + w$ and $x_{\max} = \max_{n \in \{1,\ldots,N\}} x_n$. We summarize the symbols used throughout the paper in Table I.

TABLE I: List of Symbols

n	Index of a wireless link
t_0	Start time of observation
w	Delay deadline
τ	$t + w$
$A(t)$	Cumulative arrivals from t_0 to $t-1$
$D(t)$	Cumulative departures from t_0 to $t-1$
$S_n(t)$	Cumulative service from t_0 to $t-1$
x_n	Backlog at link n at time t_0
$\mathbf{x}$	Backlog vector at time t_0
$B(t)$	Total backlog at time t
$W(t)$	Virtual delay at time t
T	Duration of the message of interest

III. Methodology and Existing Results

We resort to stochastic network calculus theory [6], [7] as principle starting point for our analysis. The key benefit from using stochastic network calculus is the ability to extend single hop results to multi-hop settings with reasonable efforts. Also, the analysis is applicable to arrival and service-time processes with general distribution. Nevertheless, this theory in general provides bounds rather than exact results which is a necessary compromise to achieve tractability. In the literature, there are several stochastic network calculus approaches that are suitable for the performance analysis of wireless networks. These approaches include computing the effective capacity of the channels [5], computing the Moment Generating Function (MGF) of a Markov process abstraction of the wireless fading channel [18], and ON-OFF service characterization of slotted Aloha access over shared wireless channel [19]. A more recent approach, namely $(\min, \times)$ network calculus – that provides end-to-end performance characterization of wireless networks in terms of the fading channel physical parameters, i.e. fading distribution and average signal-to-noise ratio (SNR) – is developed in [8] and is based on $(\min, \times)$ dioid algebra. In this paper, we pursue transient analysis by utilizing the $(\min, \times)$ network calculus, while we note that in principle this could also be pursued for example by MGF-based network calculus.

A. Equivalent Model

In order to analyze the network in Figure 1 using stochastic network calculus approach, we must first overcome the following incompatibility: in the network calculus framework it is assumed that initially all buffers are empty and there are no arrivals (from the considered traffic stream) before the start time t_0, i.e. $\sum_{n=1}^{N} x_n = 0$, and $A(0,t_0) = 0$. Furthermore, it is assumed that no service is rendered by time t_0, i.e. $S(0,t_0) = 0$. To this end, we define an alternative, yet equivalent, queuing model for our system shown in Figure 2. We treat the initial backlog x_n at link n as cross-traffic, $A_n^c(t)$, given by

$$A_n^c(t) = \min(\kappa(t-t_0), x_n), \tag{5}$$

where $\kappa(t)$ is a burst function with $\kappa(t) = 0$, for $t = 0$, and $\kappa(t) = \infty$, otherwise. The devised model satisfies the requirements for a network-calculus-based analysis. Given the equivalent model, without loss of generality, we let $t_0 = 0$ in the rest of the paper. In this case, we are interested in the network performance during the period $0 \le t \le T$.

For the ease of exposition we use $A_n(t)$ and $D_n(t)$ to denote the cumulative arrivals and cumulative departures, respectively, at link n. Note that $A_1(t) = A(t) + A_1^c(t)$,

$$A_n(t) = D_{n-1}(t) + A_n^c(t),\ \forall n \in \{2,\ldots,N\}, \tag{6}$$

and $D(t) = D_N(t)$.

B. $(\min,\times)$ Network Calculus for Wireless Network Analysis

The main objective of $(\min,\times)$ network calculus is to obtain probabilistic performance bounds for multi-hop wireless networks in terms of the underlying fading channel parameters. A key concept of this approach is the transformation of the system model into an alternative analysis domain, known as *SNR domain* [8], using the exponential function. In this domain, the random service rendered by a wireless fading channel is characterized in terms of the variability of the instantaneous SNR, that is the SNR service process at wireless link n is given by

$$\mathcal{S}_n(u,t) = e^{S_n(u,t)} = \prod_{i=u}^{t-1} e^{\mu_{n,i}}\,, \tag{7}$$

where we use the calligraphic font to represent corresponding processes in the SNR domain. Similarly, the cumulative arrivals and departures in the SNR domain are given by

$$\mathcal{A}_n(u,t) = e^{A_n(u,t)} = \mathcal{D}_{n-1}(u,t)\cdot\mathcal{A}_n^c(u,t),$$

where

$$\mathcal{A}_n^c(u,t) = \min(e^{\kappa(t-u)}, e^{x_n}),\ \text{and}\ \mathcal{D}_n(u,t) = e^{D_n(u,t)}\,.$$

Then using (3), the SNR backlog process is described by

$$\mathcal{B}(t) = e^{B(t)} = \frac{\mathcal{A}(t)}{\mathcal{D}(t)}\cdot\prod_{n=1}^{N} e^{x_n}. \tag{8}$$

However, the transformation does not affect time. Therefore the delay in the SNR domain is given by

$$\mathcal{W}(t) = W(t) = \inf\left\{w \ge 0 : \mathcal{A}(t)\cdot\prod_{n=1}^{N} e^{x_n} \le \mathcal{D}(t+w)\right\}.$$

At link n, the input/output relationship in $(\min,\times)$-algebra is given by $\mathcal{D}_n(0,t) \ge \mathcal{A}_n \otimes \mathcal{S}_n(0,t)$, where $\otimes$ is the $(\min,\times)$-convolution operator defined as

$$\mathcal{X}\otimes\mathcal{Y}(u,t) = \inf_{u\le v\le t}\{\mathcal{X}(u,v)\cdot\mathcal{Y}(v,t)\}\ .$$

Performance analysis of communication networks often focuses on a stochastic characterization of virtual delay. As shown in [8] an upper bound for the delay violation probability can be derived in terms of an integral transform, namely, the Mellin transform of the cumulative arrival and service processes in the SNR domain and by using the moment bound which are defined next.

Definition 2. *The Mellin transform of a non-negative random variable $\mathcal{X}$ is defined as*

$$\mathcal{M}_{\mathcal{X}}(s) = \mathbb{E}\left[\mathcal{X}^{s-1}\right], \tag{9}$$

for any $s \in \mathbb{R}$, whenever the expectation $\mathbb{E}[\cdot]$ exists.

For a cumulative process $\mathcal{X}(u,t)$, we denote its Mellin transform by $\mathcal{M}_{\mathcal{X}}(s,u,t) = \mathcal{M}_{\mathcal{X}(u,t)}(s)$.

Definition 3. *The moment bound for a random variable X is given by*

$$\mathbb{P}\{X \le x\} \le \min_{s>0} x^s\mathbb{E}[X^{-s}] = \min_{s>0} x^s\mathcal{M}_X(1-s).$$

Note that the Mellin transform uniquely identifies the distribution of a random variable. However, an inverse Mellin transform may not be as easy to obtain. On the other hand, the moment bound provides a useful mechanism to recover a bound on the distribution of that random variable in terms of its Mellin transform.

The following theorem is restated from [8] for convenience and due to its relevance to this work.

Theorem 1 (Theorem 1, [8])**.** *A probabilistic bound for the virtual delay at time t is given by $\mathbb{P}(\mathcal{W}(t) > w^\varepsilon) \le \varepsilon$, where w^ε is the smallest $w \ge 0$ that satisfies*

$$\inf_{s>0}\{\mathcal{K}(s,\tau,t)\} \le \varepsilon\,, \tag{10}$$

where $\tau = t + w$ and

$$\mathcal{K}(s,\tau,t) = \sum_{u=0}^{t}\mathcal{M}_{\mathcal{A}}(1+s,u,t)\cdot\mathcal{M}_{\mathcal{S}}(1-s,u,\tau)\,. \tag{11}$$

Furthermore, a probabilistic bound for the stationary virtual delay $\mathbb{P}(\mathcal{W} > w^\varepsilon) \le \varepsilon$ of the system is obtained likewise by considering the limit of the function $\mathcal{K}$ as $t \to \infty$.

In what follows, we refer to the function $\mathcal{K}(s,\tau,t)$ above as the *kernel*. The bound given by Theorem 1 is applicable to any type of network as long as the Mellin transforms exist and are obtainable. Its usefulness is surmount when applied to the analysis of wireless fading channels as the Mellin transform

$\mathcal{M}_{\mathcal{S}}(1-s,u,\tau)$ is already derived for many fading channel models in the literature, e.g., [20]–[23], which makes this approach attractive for wireless networks analysis.

For the convenience of exposition of the results in this paper we define the function $V(s)$ as follows:

$$V(s) \triangleq [\mathcal{M}_{\mathcal{S}}(1-s,u,\tau)]^{\frac{1}{\tau-u}}. \quad (12)$$

Rayleigh block-fading channel – For Rayleigh block-fading wireless channel that provides a service process with achievable Shannon capacity per slot, we have for channel n

$$S_n(u,t) = W \sum_{i=u}^{t-1} \log_2(1+\gamma_n(i)), \quad (13)$$

where W is the bandwidth and $\gamma_n(i)$ is the instantaneous SNR for channel n during time slot i. Since the channels are i.i.d. both across links and time slots, we write $\gamma_n(i) = \bar{\gamma}Y$, where $\bar{\gamma}$ is the average SNR and the channel gain Y is an exponentially distributed random variable with unit mean. In this case, the Mellin transform of the service increment is given by:

$$\begin{aligned}\mathcal{M}_{\mathcal{S}}(1-s,u,\tau) &= \mathbb{E}[\mathcal{S}^{-s}(u,\tau)] = \mathbb{E}\left[\prod_{i=u}^{\tau-1}(1+\gamma_n(i))^{\frac{-sW}{\log 2}}\right] \\ &= \prod_{i=u}^{\tau-1}\int_0^\infty (1+\bar{\gamma}y_i)^{\frac{-sW}{\log 2}} e^{-y_i} dy_i \\ &= \left[e^{\frac{1}{\bar{\gamma}}}\bar{\gamma}^{\frac{-sW}{\log 2}}\Gamma\left(1-\frac{sW}{\log 2},\bar{\gamma}^{-1}\right)\right]^{\tau-u} = [V(s)]^{\tau-u},\end{aligned}$$

where $\Gamma(x,a)$ is the lower incomplete Gamma function.

C. Stationary Bound for Transient Analysis

For transient analysis, one may consider the stationary bound of Theorem 1 and apply it to the equivalent model in Figure 2. However, this approach is not entirely accurate as we assume the arrival increments $\{a_i\}$ to be zero for $i > T$. Nevertheless, it is worthwhile to establish some form of stationary bound as reference for the transient analysis presented later. One straightforward way to establish such a reference is by assuming the deterministic instantaneous arrivals $\{a_i\}$ to occur over the infinite time horizon and invoke Theorem 1. In this subsection, we follow this approach and discuss a motivating numerical evaluation.

In order to obtain a bound for the stationary virtual delay, one essentially has to determine the limit of the kernel, given in (11), as t goes to infinity. Using this approach, a bound with a closed-form expression was derived in [8] which is restated here for convenience.

Theorem 2 (Section V-C, [8]). *A probabilistic stationary end-to-end delay bound for a cascade of N i.i.d. wireless channels with homogeneous average SNRs, (σ,ρ)-bounded arrival traffic, and (σ_c,ρ_c)-bounded cross-traffic is given by*

$$\mathbb{P}(\mathcal{W} > w) \le \inf_{s>0}\left\{\frac{e^{s(-\rho w+\sigma+N\sigma_c)}}{(1-V_0(s))^N}\cdot \min\{1,(V_0(s))^w(w+1)^{N-1}\}\right\},$$

where

$$V_0(s) = e^{s(\rho+\rho_c)}V(s).$$

In order to apply this result to the model described above, we dimension the cross-traffic at each link using Eq. (5) by setting $\sigma_c = x_{\max}$ and $\rho_c = 0$. In the rest of the paper, we refer to this reference bound given in Theorem 2 as *stationary bound*.

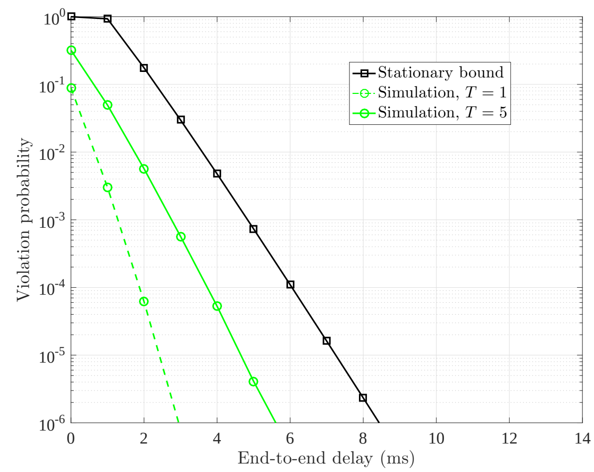


Fig. 3: Delay violation probability vs. end-to-end delay for a single link with SNR = 5 dB, $x_1 = 0$, $\rho = 20$, $\sigma = 0$.

To evaluate the viability of this approach, we compare in Figure 3 the transient violation probability obtained using simulations with the stationary bound for a single wireless link. The considered scenario is parameterized by a time-slotted system where per slot a time-critical packet of size $\rho = 20$ bits arrives, while we set the backlog to $\mathbf{x} = 0$ (i.e. $\rho_c = 0$ and $\sigma_c = 0$). Accordingly, for determining the stationary bound we set the instantaneous arrival to a constant arrival process with rate $\rho = 20$ bits per time slot. In Figure 3, we plot the resulting delay violation probability of the end-to-end delay from simulation and compare it to the stationary bound for two arrival processes with $T = 1$ and $T = 5$. The plot reveals the challenge with respect to transient analysis, as there is a considerable gap between the simulations and the stationary bound. For smaller T, the gap becomes arbitrarily large, and in particular the decay rate of the bound does not match the decay rate of the simulations. This slackness is due to the extra terms added to the summation when letting $t \to \infty$ in (11) in order to obtain the stationary bound since $\lim_{t\to\infty}\mathbb{P}(\mathcal{W}(t) > w) \to \mathbb{P}(\mathcal{W} > w)$. Although the stationary delay bound has its validity with respect to steady state analysis, Figure 3 shows that it is worthwhile to pursue better bounds for transient scenarios.

IV. Transient Analysis

In this section, we present our main contribution: new transient upper bounds for the delay violation probability for the N-hop wireless network described in Sections II and III. To this end, we present two different approaches and later (in Section V) compare the probabilistic delay bounds resulting from both approaches. A first, straightforward approach is to start from the definition of the kernel as presented in Section III-B and derive a transient bound. We refer to this as *Kernel-Based-Transient Bound (KBTB)*. However, as we will show later, this bound - while improving over the stationary bound - is still loose. This motivates us to present a new analysis for the transient behaviour of the system which results in a new *Wireless Transient Bound* (WTB). Furthermore, we analyse the computational complexity of WTB, and show how it can be used when the initial backlog information is delayed. Finally, we show how the analysis can be extended for mesh networks.

A. Kernel-Based-Transient Bound (KBTB)

We derive a transient upper bound for the violation probability based on the results from [8]. We first note that the bound on the violation probability of the virtual delay at time t in Theorem 1 corresponds to the transient analysis problem at hand. We hence focus on the components of the kernel. Since the sequence of arrivals is deterministic, we have

$$\mathcal{M}_{\mathcal{A}}(s+1,u,t) = \mathbb{E}[(\mathcal{A}(u,t))^s] = [\mathcal{A}(t)/\mathcal{A}(u)]^s. \quad (14)$$

Let $\mathcal{S}_0$ denote the dynamic SNR server [17] that describes the network service offered by the multi-hop route to the through traffic. The following lemma evaluates the Mellin transform of $\mathcal{S}_0$.

Lemma 1 (Lemma 6, [8]). *Given $\sigma_c = x_{max}$ and $\rho_c = 0$, the Mellin transform of $\mathcal{S}_0(u,\tau)$ satisfies for $s<1$ that*

$$\mathcal{M}_{\mathcal{S}_0}(1-s,u,\tau) \le e^{sNx_{max}} \binom{N-1+\tau-u}{\tau-u} V(s)^{\tau-u}. \quad (15)$$

Substituting (14) and (15) into the definition of the kernel (11), we obtain:

$$\mathcal{K}(s,\tau,t) \le e^{sNx_{\max}} \left[\sum_{u=0}^{t} \left(\frac{\mathcal{A}(t)}{\mathcal{A}(u)}\right)^s \binom{N-1+\tau-u}{\tau-u} V^{\tau-u}(s)\right]. \quad (16)$$

The KBTB bound is then computed by minimizing the RHS of (16) over s. For the case of a single link, a closed form expression can be obtained for the KBTB bound under (σ,ρ)-bounded arrivals which is given by the following corollary.

Corollary 1. *Assuming $\mathcal{A}(t)$ follows the (σ,ρ)-bounded traffic characterization, defined in* (1)*, for a single wireless link, an upper bound for $\mathbb{P}(\mathcal{W}(t) > w)$ is given by*

$$\min_{s>0} \left\{ e^{s(x_1-\rho w)} (V_0(s))^w \left[e^{s\sigma} \cdot \frac{V_0(s) - (V_0(s))^{t+1}}{1 - V_0(s)} + 1 \right] \right\}.$$

Proof. The proof is given in Appendix A. □

Later, in Section V, we will show that for multi-hop scenarios the KBTB bound can become very loose in particular in cases with non-zero initial backlog. Nevertheless, it still captures the exponential decay rate of the delay tail distribution. The slackness in the KBTB bound is mainly due to the fact that it is based on results that are initially derived for stationary settings. Although we believe that asymptotically we may not be able to improve on this bound, there is plenty of room for improvement for short sequences (i.e., small T) and short delay target. To this end, we next derive the proposed WTB.

B. Wireless Transient Bound (WTB)

Our derivation of WTB is inspired by the bounding techniques used in [8]. However, we conduct independent analysis starting with the basic principles of network calculus and tailor the result, from the beginning, to our system with initial backlog $\mathbf{x}$, which makes the analysis more involved. We start by presenting our analysis for a single-hop case $N=1$; then, we generalize the obtained result for the multi-hop case.

In the following theorem, we state the proposed WTB bound for the single-hop case.

Theorem 3. *Given an arrival sequence $\mathcal{A}$ traversing a wireless channel and given an initial backlog x_1, an upper bound for the delay violation probability, $\mathbb{P}\{\mathcal{W}(t) > w\}$, is given by*

$$\min_{s>0} \left[\mathcal{A}^s(t) e^{sx_1} V^{\tau}(s) + \sum_{u=1}^{t-1} [\mathcal{A}(t)/\mathcal{A}(u)]^s V^{\tau-u}(s) \right],$$

where $\tau = t + w$.

Proof. Let $\mathcal{S}(t)$ be the SNR service process for wireless channel given by (7). From the definition of dynamic server [8], we have for all $\tau \ge 0$

$$\mathcal{D}(\tau) \ge \min_{0\le u\le\tau} \{\mathcal{S}(\tau-u)\cdot\mathcal{A}(u)\cdot\mathcal{A}_1^c(u)\}.$$

Recall that $\mathcal{A}_1^c(0) = 1$, and $\mathcal{A}_1^c(u) = e^{x_1}$, for all $u > 0$. Now, the event $\{\mathcal{W}(t) > w\}$ is equivalent to the event that the cumulative departures at time τ is strictly less than the cumulative arrivals at time t plus the initial backlog x_1. Therefore,

$$\begin{aligned}
&\mathbb{P}\{\mathcal{W}(t) > w\} = \mathbb{P}\{\mathcal{D}(\tau) < \mathcal{A}(t)e^{x_1}\} \\
&= \mathbb{P}\left\{ \min_{0\le u\le\tau} [\mathcal{S}(\tau-u)\cdot\mathcal{A}(u)\cdot\mathcal{A}_1^c(u)] < \mathcal{A}(t)e^{x_1} \right\} \\
&= \mathbb{P}\left\{ \{\mathcal{S}(\tau) < \mathcal{A}(t)e^{x_1}\} \cup \left(\bigcup_{u=1}^{\tau} \{\mathcal{S}(\tau-u)\cdot\mathcal{A}(u) < \mathcal{A}(t)\} \right) \right\} \\
&= \mathbb{P}\left\{ \{\mathcal{S}(\tau) < \mathcal{A}(t)e^{x_1}\} \cup \left(\bigcup_{u=1}^{t-1} \{\mathcal{S}(\tau-u)\cdot\mathcal{A}(u) < \mathcal{A}(t)\} \right) \right\} \\
&\le \mathbb{P}\{\mathcal{S}(\tau) < \mathcal{A}(t)e^{x_1}\} + \sum_{u=1}^{t-1} \mathbb{P}\{\mathcal{S}(\tau-u)\cdot\mathcal{A}(u) < \mathcal{A}(t)\} \\
&\le \min_{s>0} \left[[\mathcal{A}(t)]^s e^{sx_1} \mathbb{E}[\mathcal{S}^{-s}(\tau)] + \sum_{u=1}^{t-1} \left[\frac{\mathcal{A}(t)}{\mathcal{A}(u)}\right]^s \mathbb{E}[\mathcal{S}^{-s}(\tau-u)] \right] \\
&= \min_{s>0} \left[[\mathcal{A}(t)]^s e^{sx_1} V^{\tau}(s) + \sum_{u=1}^{t-1} \left[\frac{\mathcal{A}(t)}{\mathcal{A}(u)}\right]^s V^{\tau-u}(s) \right].
\end{aligned}$$

In the fourth step above we have used the fact that $\mathbb{P}\{\mathcal{S}(\tau-u)\cdot\mathcal{A}(u) < \mathcal{A}(t)\} = 0$, for $u \ge t$. The fifth step utilizes the union bound and the sixth step follows from using moment bound for each of the terms in step four. □

It is worth noting that the first term in the bound in Theorem 3 is contributed by the departures of the initial backlog and the second term is due to the departures of the arrivals of interest. Next, we extend the bound in Theorem 3 to the homogeneous N-hop case. Even though the analysis for this case is more involved, it essentially uses the same bounding techniques from the proof of Theorem 3. However, one key aspect to consider for deriving the bound for the N-hop case is whether to unfold the $(\min, \times)$-convolution iteratively starting with input/output relation at link 1 or link N. While in the previous work [8], the authors considered link 1, in this paper we start with link N and then iteratively bound the departure processes $\{\mathcal{D}_n\}$ in the decreasing order of n. We note that if one starts with link 1, then the cumulative service process obtained using $(\min, \times)$-convolution does not carefully accommodate for the initial backlogs at each node. On the other hand, starting with link N allowed us to systematically account for the initial backlog at each node and consequently the bound obtained is tighter than that of the former approach.

Theorem 4. *When the sequence $\mathcal{A}$ traverses the N-hop wireless network in Figure 2 and given the initial backlog vector* $\mathbf{x}$*, we compute $\mathbb{P}\{\mathcal{W}(t) > w\} \leq \min_{s>0} \ \Phi(s)$, where $\Phi(s)$ is given by*

$$\Phi(s) = \sum_{i=0}^{N-1} \binom{i+\tau-1}{\tau-1} [\mathcal{A}(t)]^s e^{s\sum_{n=1}^{N-i} x_n} V^{\tau}(s) + \binom{N+\tau-2}{\tau-1} \sum_{u=1}^{t-1} [\mathcal{A}(t)/\mathcal{A}(u)]^s V^{\tau-u}(s). \quad (17)$$

Proof. Since the event $\{\mathcal{W}(t) > w\}$ is equivalent to the event that the cumulative departures at node N at time τ is strictly less than the cumulative arrivals by time t plus the total initial backlog $\sum_{n=1}^{N} x_n$, we have

$$\mathbb{P}\{\mathcal{W}(t) > w\} = \mathbb{P}\{\mathcal{D}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\}. \quad (18)$$

Recall that $\mathcal{D}_N(t) \geq \mathcal{A}_N \otimes \mathcal{S}_N(t)$ and $\mathcal{A}_N(t) = \mathcal{D}_{N-1}(t) \cdot \mathcal{A}_N^c(t)$. Therefore,

$$\begin{aligned}
&\mathbb{P}\{\mathcal{D}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} \\
&\leq \mathbb{P}\{\mathcal{A}_N \otimes \mathcal{S}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} \\
&= \mathbb{P}\{(\mathcal{D}_{N-1} \cdot \mathcal{A}_N^c) \otimes \mathcal{S}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} \\
&= \mathbb{P}\Big\{\min_{0 \leq u \leq \tau} [\mathcal{D}_{N-1}(u) \cdot \mathcal{A}_N^c(u) \cdot \mathcal{S}_N(\tau-u)] < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\Big\} \\
&= \mathbb{P}\Big\{\{\mathcal{S}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} \\
&\qquad \cup \big(\bigcup_{u=1}^{\tau} \{\mathcal{D}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\}\big)\Big\} \\
&\leq \mathbb{P}\{\mathcal{S}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} \\
&\qquad + \sum_{u=1}^{\tau} \mathbb{P}\{\mathcal{D}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\}\}.
\end{aligned} \quad (19)$$

In the following we find an upper bound for the probabilities in the summation term of the last step in (19).

$$\begin{aligned}
&\mathbb{P}\{\mathcal{D}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\} \\
&\leq \mathbb{P}\{(\mathcal{D}_{N-2} \cdot \mathcal{A}_{N-1}^c) \otimes \mathcal{S}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\} \\
&= \mathbb{P}\Big\{\min_{0 \leq u_1 \leq u} [\mathcal{D}_{N-2}(u_1) \cdot \mathcal{A}_{N-1}^c(u_1) \cdot \\
&\qquad \mathcal{S}_{N-1}(u-u_1) \cdot \mathcal{S}_N(\tau-u)] < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\Big\} \\
&\leq \mathbb{P}\{\mathcal{S}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\} + \\
&\sum_{u_1=1}^{u} \mathbb{P}\{\mathcal{D}_{N-2}(u_1) \cdot \mathcal{S}_{N-1}(u-u_1) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-2} x_n}\}.
\end{aligned} \quad (20)$$

Substituting (20) in (19), we obtain

$$\begin{aligned}
&\mathbb{P}\{\mathcal{W}(t) > w\} \\
&\leq \mathbb{P}\{\mathcal{S}_N(\tau) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N} x_n}\} + \\
&\quad \mathbb{P}\{\mathcal{S}_{N-1}(u) \cdot \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-1} x_n}\} + \\
&\quad \sum_{u=1}^{\tau} \sum_{u_1=1}^{u} \mathbb{P}\Big\{\mathcal{D}_{N-2}(u_1) \mathcal{S}_{N-1}(u-u_1) \mathcal{S}_N(\tau-u) < \mathcal{A}(t) \cdot e^{\sum_{n=1}^{N-2} x_n}\Big\}.
\end{aligned}$$

One can again use similar manipulation as in (20) to bound the probabilities in the double summation of the RHS of the above inequality. Repeating this step iteratively, and using the convention $u_0 = u$ and $u_{-1} = \tau$, we arrive at (21) (top of the next page). The first and second terms in the RHS of (21) are upper bounded as shown in (22) and (23), respectively. In the first inequality of (22) we have used the moment bound, and in the second inequality we have used the fact

$$\sum_{u=1}^{\tau} \sum_{u_1=1}^{u} \cdots \sum_{u_{i-1}=1}^{u_{i-2}} 1 = \binom{i+\tau-1}{\tau-1}.$$

In the first inequality of (23), we have used the fact that the probability terms are zero for $u_{N-1} \geq t$.

Finally, substituting (22) and (23) in (21), we obtain the result. □

We again note that the first term in (17) is contributed by the total initial backlog while the second term is contributed by the arrivals of interest. Also, note that the expression in (17) is valid for any service process for which the Mellin transform exists and any sequence of arrivals in the interval $[0, T]$. One may further simplify this expression by assuming that the cumulative arrival process obeys (σ, ρ)-bounded traffic characterization. This results in a simpler upper bound which is given by the following corollary.

Corollary 2. *Assuming $\mathcal{A}(t)$ obeys (σ, ρ)-bounded traffic characterization, defined in* (1)*, the proposed transient bound in Theorem 4 is reduced to the following:*

$$\mathbb{P}\{\mathcal{W}(t) > w\} \leq \min_{s>0} \ V^{\tau}(s) \Big[\sum_{i=0}^{N-1} \binom{i+\tau-1}{\tau-1} e^{s\sum_{n=1}^{N-i} x_n} + \binom{N+\tau-2}{\tau-1} \cdot e^{s(\sigma+\rho t)} \cdot \frac{1-V_0^{-t}(s)}{V_0(s)-1}\Big],$$

where $V_0(s) = e^{s\rho} V(s)$.

Proof. The proof is given in Appendix B. □

Transient Backlog Bound – Given the initial backlog vector $\mathbf{x}$, using Theorem 4 we derive an upper bound for the

$$\mathbb{P}\{\mathcal{W}(t)>w\} \leq \sum_{i=0}^{N-1}\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{i-1}=1}^{u_{i-2}}\mathbb{P}\Big\{\mathcal{S}_{N-i}(u_{i-1})\cdot\Pi_{n=1}^{i-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\cdot e^{\sum_{n=1}^{N-i}x_n}\Big\}$$
$$+\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{N-1}=1}^{u_{N-2}}\mathbb{P}\Big\{\mathcal{A}(u_{N-1})\cdot\Pi_{n=1}^{N-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\Big\}. \tag{21}$$

$$\sum_{i=0}^{N-1}\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{i-1}=1}^{u_{i-2}}\mathbb{P}\{\mathcal{S}_{N-i}(u_{i-1})\cdot\Pi_{n=1}^{i-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\cdot e^{\sum_{n=1}^{N-i}x_n}\}$$
$$\leq\sum_{i=0}^{N-1}\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{i-1}=1}^{u_{i-2}}\min_{s>0}\,[\mathcal{A}(t)]^s\cdot e^{s(\sum_{n=1}^{N-i}x_n)}\mathbb{E}[\{\mathcal{S}_{N-i}(u_{i-1})\cdot\Pi_{n=1}^{i-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)\}^{-s}]$$
$$\leq\min_{s>0}\sum_{i=0}^{N-1}[\mathcal{A}(t)]^s\cdot e^{s\sum_{n=1}^{N-i}x_n}V^{\tau}(s)\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{i-1}=1}^{u_{i-2}}1=\min_{s>0}\sum_{i=0}^{N-1}\binom{i+\tau-1}{\tau-1}[\mathcal{A}(t)]^s\cdot e^{s\sum_{n=1}^{N-i}x_n}V^{\tau}(s). \tag{22}$$

$$\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{N-1}=1}^{u_{N-2}}\mathbb{P}\{\mathcal{A}(u_{N-1})\cdot\Pi_{n=1}^{N-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\}$$
$$\leq\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{N-1}=1}^{t-1}\mathbb{P}\{\mathcal{A}(u_{N-1})\cdot\Pi_{n=1}^{N-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\}$$
$$=\sum_{u_{N-1}=1}^{t-1}\mathbb{P}\{\mathcal{A}(u_{N-1})\cdot\Pi_{n=1}^{N-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u)<\mathcal{A}(t)\}\sum_{u=1}^{\tau}\sum_{u_1=1}^{u}\dots\sum_{u_{N-2}=1}^{u_{N-3}}1$$
$$\leq\binom{N+\tau-2}{\tau-1}\sum_{u_{N-1}=1}^{t-1}\min_{s>0}\,[\mathcal{A}(t)/\mathcal{A}(u_{N-1})]^s\mathbb{E}[\{\Pi_{n=1}^{N-1}\mathcal{S}_{N-n}(u_{n-1}-u_n)\cdot\mathcal{S}_N(\tau-u))\}^{-s}]$$
$$\leq\binom{N+\tau-2}{\tau-1}\cdot\min_{s>0}\sum_{u=1}^{t-1}[\mathcal{A}(t)/\mathcal{A}(u)]^sV^{\tau-u}(s). \tag{23}$$

---

total backlog $B(t)$ in the system at time t which we state in the following corollary.

Corollary 3. *Given the initial backlog vector* $\mathbf{x}$*, an upper bound for the violation probability for the total backlog in the system at time* t *is given by*

$$\mathbb{P}\{\mathcal{B}(t)>e^x\}\leq\min_{s>0}\quad e^{-xs}\Phi(s).$$

Proof. Using (8), we obtain

$$\mathbb{P}\{\mathcal{B}(t)>e^x\}=\mathbb{P}\{\mathcal{A}(t)\cdot e^{\sum_{n=1}^{N}x_n}/\mathcal{D}(t)>e^x\}$$
$$=\mathbb{P}\{\mathcal{D}(\tau)<\mathcal{A}(t)\cdot e^{\sum_{n=1}^{N}x_n-x}\} \tag{24}$$

Note that the expression in (24) is same as (18), except for the additional factor e^{-x}. Therefore, we repeat the same steps of the proof of Theorem 4 and arrive at the desired result. □

C. Complexity Analysis

From (17), we infer that the computational complexity for computing $\Phi(s)$ is $O(t+N)$ for $0\leq t\leq T$. Further, to obtain the WTB we need to solve the optimization problem of minimizing $\Phi(s)$ over $s>0$. Thankfully, this is a convex optimization problem as $\Phi(s)$ is convex which we state in the following lemma.

Lemma 2. *Assuming* $V(s)$*, defined in* (12)*, is differentiable, for* $s>0$*, the function* $\Phi(s)$ *is convex.*

Proof. Since sum of convex functions is convex, it is sufficient to prove that the terms $[\mathcal{A}(t)/\mathcal{A}(u)]^sV^{\tau-u}(s)$ and $V^{\tau}e^{s\sum_{n=1}^{N-i}x_n}$ are convex. We will show that $[\mathcal{A}(t)/\mathcal{A}(u)]^sV^{\tau-u}(s)$ is convex and proof for the latter term is similar. From (12), we have $V^{\tau-u}(s)=\mathbb{E}[\mathcal{S}^{-s}(u,\tau)]$. Therefore, for $\theta\in[0,1]$, we obtain

$$[\mathcal{A}(t)/\mathcal{A}(u)]^{\theta s_1+(1-\theta)s_2}V^{\tau-u}(\theta s_1+(1-\theta)s_2)$$
$$=\mathbb{E}\big[[\mathcal{A}(t)/(\mathcal{A}(u)\mathcal{S}(u,\tau))]^{(\theta s_1+(1-\theta)s_2)}\big]$$
$$\leq\mathbb{E}\big[\theta[\mathcal{A}(t)/(\mathcal{A}(u)\mathcal{S}(u,\tau))]^{s_1}+(1-\theta)[\mathcal{A}(t)/(\mathcal{A}(u)\mathcal{S}(u,\tau))]^{s_2}\big]$$
$$=\theta[\mathcal{A}(t)/\mathcal{A}(u)]^{s_1}\mathbb{E}[\mathcal{S}^{-s_1}(u,\tau)]+(1-\theta)[\mathcal{A}(t)/\mathcal{A}(u)]^{s_2}\mathbb{E}[\mathcal{S}^{-s_2}(u,\tau)]$$
$$=\theta[\mathcal{A}(t)/\mathcal{A}(u)]^{s_1}V^{\tau-u}(s_1)+(1-\theta)[\mathcal{A}(t)/\mathcal{A}(u)]^{s_2}V^{\tau-u}(s_2).$$

In the third step above, we have used the fact that $[\frac{\mathcal{A}(t)}{\mathcal{A}(u)\mathcal{S}(u,\tau)}]^s$ is convex in s, since $\frac{\mathcal{A}(t)}{\mathcal{A}(u)\mathcal{S}(u,\tau)}>0$. Therefore, $[\mathcal{A}(t)/\mathcal{A}(u)]^sV^{\tau-u}(s)$ is convex and the result is proven. □

From Lemma 2 we infer that, for any given service process, WTB can be computed by solving a convex optimization problem in a single variable s. Further, we note that stationary bound given in Theorem 2 is also known to be a convex optimization problem [24] and the objective function is in closed form. Since WTB doesn't have a closed-form expression, it has a factor of $O(T+N)$ higher computational complexity compared with the stationary bound. This increase in computational complexity can be attributed to the fact that WTB does not restrict the sequence of arrivals by choosing (σ,ρ)-bounded arrivals, and it carefully incorporates the initial backlog of each node. Nevertheless, the increase in the computational complexity for WTB can be justified for the relative tightness it provides compared with the stationary bound (as we show in Section V). Finally, a similar analysis shows that the KBTB bound has a factor of $O(T)$ higher computational complexity compared with the stationary bound.

D. Delayed Backlog Information

To compute WTB, node 1 requires to know the backlogs of all nodes at time t_0. However, in practical systems, the current backlog information at node $n>1$ may not be known to node 1 instantaneously; instead, it may be received with some delay due to the time it takes to relay this information back to node 1. For example, if a node relays its current backlog and the backlogs of the successor nodes to its predecessor node with a delay of one time slot, then the backlog at node $n>1$ at time t_0 will be received by node 1 at time t_0+n-1. In this case, at time t_0, node 1 will have the backlog information of all the N nodes at time t_0-d, where $d=N-1$. To incorporate this delayed backlog information in our transient analysis, we extend the delay bound for the case where at time t_0 node 1 is only aware of the initial backlog vector at time t_0-d, which we refer to by $\mathbf{x}(t_0-d)$, for some $d>0$.

To find the delay bound, we reuse the model in Figure 1 and consequently apply Theorem 4. In the following, we refer to the cumulative arrivals in the interval $[t_0-d,t_0-1]$ as the overhead traffic $\mathcal{A}^{\mathrm{o}}(t_0-d,t)$, where $t\in[t_0-d,t_0-1]$. We define a cumulative arrival process $\mathcal{A}'(t)$ as follows.

$$\mathcal{A}'(t)=\begin{cases}\mathcal{A}^{\mathrm{o}}(t_0-d,t) & t_0-d\le t\le t_0,\\ \mathcal{A}^{\mathrm{o}}(t_0-d,t_0)\cdot\mathcal{A}(t_0,\min(t,t_0+T)) & t>t_0.\end{cases}$$

Note that $\mathcal{A}'(t)$ only accounts for arrivals until the time at which the time-critical sequence ends. Now, given $\mathbf{x}(t_0-d)$ the arrivals that occurred before t_0-d can be ignored and we may consider that the system started at time t_0-d with initial backlog $\mathbf{x}(t_0-d)$. This is equivalent to the model in Figure 1, except that the starting time is t_0-d and the arrival sequence of our interest starts at t_0 instead of t_0-d. Note that our analysis that leads to the derivation of WTB is independent of the starting time, but depends on the cumulative arrivals since the starting time and the initial backlogs at the nodes. Thus, for the system that starts at time t_0-d, it is easy to see that Theorem 4 can be applied using the cumulative arrival process $\mathcal{A}'(t)$ and $\mathbf{x}(t_0-d)$. This result is stated in the following theorem without proof.

Theorem 5. *When the sequence $\mathcal{A}$ traverses the N-hop wireless network in Figure 2, and given the delayed initial backlog vector $\mathbf{x}(t_0-d)$ at time t_0, we compute $\mathbb{P}\{\mathcal{W}(t)>w\}\le\min_{s>0}\ \Phi_d(s)$, where $\Phi_d(s)$ is given by*

$$\Phi_d(s)=\sum_{i=0}^{N-1}\binom{i+\tau-1}{\tau-1}[\mathcal{A}'(t)]^s e^{s\sum_{n=1}^{N-i}x_n(t_0-d)}V^\tau(s)+\binom{N+\tau-2}{\tau-1}\sum_{u=1}^{t-1}[\mathcal{A}'(t)/\mathcal{A}'(u)]^s V^{\tau-u}(s).$$

Note that the delay bound in Theorem 5 is loose compared to the delay bound where the initial backlogs at time t_0 are known, i.e., $\mathbf{x}(t_0)$ is known to node 1. The difference between these bounds depends on delay d in obtaining the backlog measurements and on the overhead traffic $\mathcal{A}^{\mathrm{o}}$ intensity.

E. Extension to Wireless Mesh Network

The bounds obtained so far are based on the assumption of a single through flow traversing a pre-selected route, modeled as a tandem of queues, with no interfering traffic from other flows in the network. In this section we extend our analysis to a wireless mesh network; see for example Figure 4. Each pair of nodes in the network are connected via point-to-point wireless link with data rate equivalent to the fading channel capacity. For illustration purposes, we assume that cross-traffic, i.e., traffic from flows other than the through flow, arrives at each intermediate relay node in the through flow path and departs afterwards with no further crossing of the path of the through flow. Analysis of cases where the two flows share more than one link is possible but more involved and can only be solved on a per-case basis. For all flows, we assume static routing. Let $A_n^e(t)$ denote the cumulative cross-traffic entering node n. Given the initial backlog x_n, we are interested in deriving WTB for the static route of interest in the network, shown in Figure 5.

In the presence of cross-traffic, let $S'_n(t)$ denote the cumulative *left-over service* received by the *flow of interest* at node n, where the flow of interest comprises of the total initial backlog and the arrivals of interest $A(t)$. If $S'_n(t)$ for all n is given, then one can compute the WTB. However, computing $S'_n(t)$ exactly is hard as the number of packets that belong to the flow of interest that are served in each time slot at node n depends on the random service received by the flow of interest and the cross-traffic at the preceding nodes and also on the order of service, which depends on the selected scheduling algorithm. However, in the absence of information regarding the scheduling algorithm, a lower bound for $S'_n(t)$ can still be obtained by assuming the worst-case scenario of static priority given to cross traffic and can be expressed by the *leftover service* as follows [17].

$$S'_n(t)\ge\max(0,S_n(t)-E^e(t)),\ \forall n,$$

where $E^e(t)$ is the envelope process for $A_n^e(t)$ for any n. We note that the envelope process $E^e(t)$ can be computed by measuring the characteristics of the cross-traffic at the nodes. For the worst-case scenario, we consider $S'_n(t)=\max(0,S_n(t)-E^e(t))$, for all n, and compute

$$V'(s)\triangleq[\mathcal{M}_{\mathcal{S}'}(1-s,u,\tau)]^{\frac{1}{\tau-u}}. \qquad (25)$$

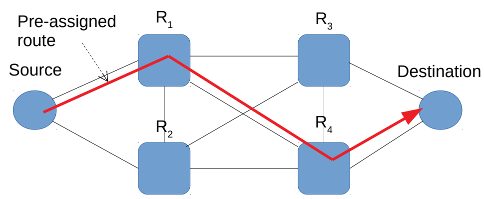


Fig. 4: Example of wireless mesh topology with a static route.

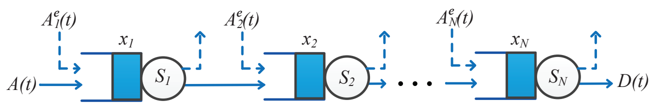


Fig. 5: Multi-hop wireless network queueing model.

Since $S_n'(t) \leq S_n(t)$, we have $V(s) \leq V'(s)$. Therefore, we substitute (25) in (17) and obtain the WTB. Note that in the case of static priority given to the through flow, the cross-traffic will have no effect on the performance of the through flow and the WTB computed in the previous section will hold again. Results for other scheduling algorithms, e.g., FCFS, earliest-deadline-first (EDF), are also obtainable in the stochastic network calculus framework. Nevertheless, they are outside the scope of this article.

V. Numerical Analysis

In this section, we present a numerical evaluation of the proposed bounds and compare them to simulation results. More specifically, we first present a comparison of the proposed WTB bound with the stationary bound and the KBTB bound. We then evaluate the tightness of WTB in comparison with the violation probability obtained using simulation. Finally, we present a routing example illustrating the utility of WTB and in turn the merit of transient analysis.

Throughout the section, we assume Rayleigh block-fading channel model for the links and use the corresponding service process given in (13). We also use a base set of parameters as follows: time slotted-system with slot duration of 1 ms, $\rho = 25, \sigma = 25$, bandwidth $W = 20$ kHz, average SNR $= 5$ dB, and total initial backlog of 100 bits, which is equally distributed along a multi-hop route (whenever a multi-hop route is considered). Note that with an average SNR of 5 dB, the average service rate (assuming Shannon capacity) amounts to about 34 bits per time slot. Thus, by setting $\rho = 25$, the system basically operates in a transient regime where asymptotically it becomes stable. We consider two types of arrival processes: (1) a burst arrival with $T = 1, \sigma = 25, \rho = 0$, modeling a single packet, with 25 bits, passed to the network at t_0; (2) an arrival process over multiple time slots with $T = 5, \sigma = 0, \rho = 25$, i.e. a train of packets. The numerical bounds are computed using MATLAB and the Discrete Event Simulation is done using C.

A. Comparison of Upper Bounds

Recall that the stationary bound cannot be directly applied to the problem at hand. We use it as a reference by assuming the arrival process occurs over infinite time horizon and

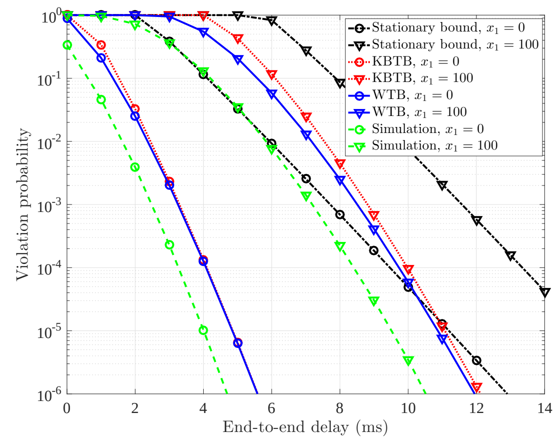


Fig. 6: Delay violation probability vs. end-to-end delay for a single link with burst arrival ($T = 1$), SNR $= 5$ dB, $\rho = 0$, and $\sigma = 25$.

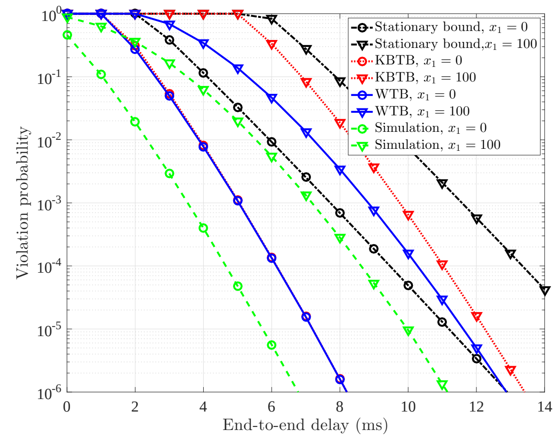


Fig. 7: Delay violation probability vs. end-to-end delay for a single link with the packet train arrival process ($T = 5$), SNR $= 5$ dB, $x_1 = 100, \rho = 25$ and $\sigma = 0$.

accordingly set the corresponding parameters in the bound. We start our numerical investigation by considering the burst arrival model ($T = 1$) and a single wireless link with and without initial backlog. The corresponding bounds and the simulation are presented in Figure 6, where we plot the delay violation probability versus an increasing delay target w. We observe that both WTB and KBTB are significantly lower than the stationary bound. Note that the proposed WTB is not significantly lower than the KBTB bound for such simple case of burst arrival. We also note that both WTB and KBTB capture the tail decay rate of the delay distribution while the stationary bound is drastically off. In Figure 7, we consider the packet train arrival model with $T = 5$ for the same single link system with an initial backlog of 100, considering again the violation probability over an increasing delay target. The figure reveals that in this case WTB outperforms the KBTB bound over the entire range of delay target values by about one order of magnitude. In comparison to the simulated system performance, the proposed WTB is still about one order of magnitude higher, with a larger gap for longer delay targets.

Furthermore, in Figure 8, we extend the scenario to a 2-

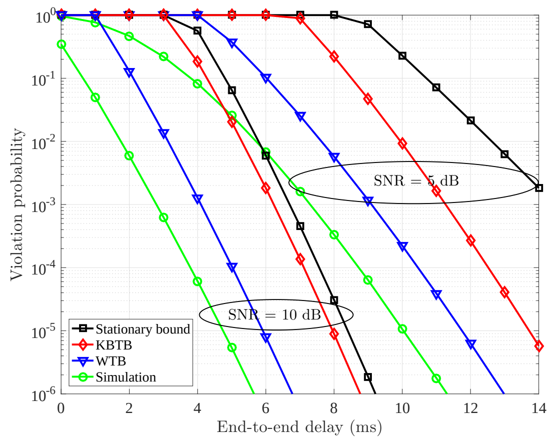


Fig. 8: Delay violation probability vs. end-to-end delay for a 2-hop network with the packet train arrival process ($T = 5$), $x_n = 100, \rho = 25$ and $\sigma = 0$.

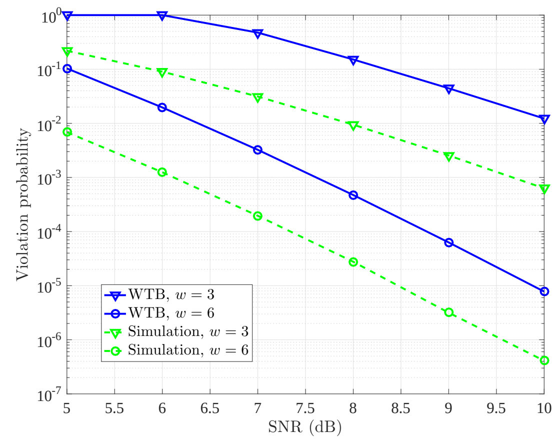


Fig. 10: Delay violation probability vs. average SNR for a 2-hop network for different target delays w with packet train arrival process ($T = 5$), $x_n = 50, \rho = 25$, and $\sigma = 0$.

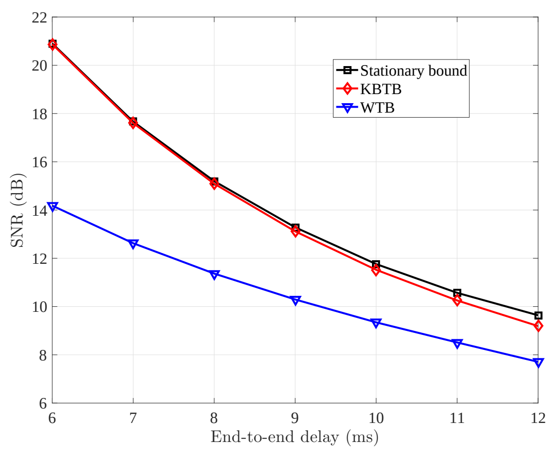


Fig. 9: SNR vs. end-to-end delay for a delay violation probability requirement of 10^{-9} for a 2-hop network with packet train arrival process ($T = 5$), $x_n = 100, \rho = 25$ and $\sigma = 0$.

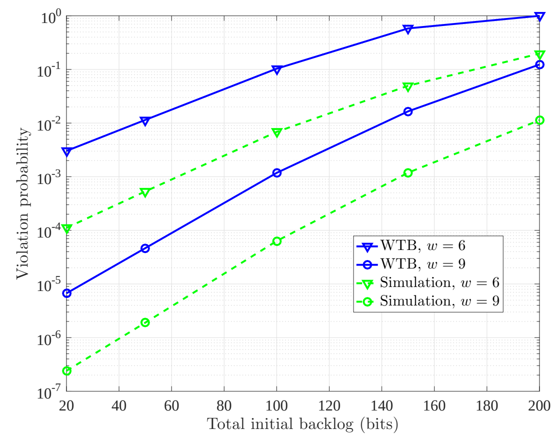


Fig. 11: Backlog violation probability vs. total initial backlog for a 2-hop network for different delays w with packet train arrival process ($T = 5$), $\rho = 25$, $\sigma = 0$, and SNR $= 5$ dB.

hop wireless system while considering the packet train arrival with $T = 5$. The plot shows a comparison for two cases of average SNR: 5 dB, and 10 dB. We observe that for the 2-hop network the KBTB bound performs worse than that for a one hop case. In particular for an average SNR of 10 dB, i.e., at lower utilization (43%), the proposed WTB bound is tighter by two orders of magnitude compared with the KBTB bound. The relevance of this improvement is illustrated in Figure 9. Here we plot the predicted SNR requirement as computed using different bounds, for a QoS requirement of violation probability smaller than 10^{-9}, and for varying delay target w in a 2-hop network. We observe that for all delay target values, the proposed WTB bound results in a significantly lower average SNR requirement per link than the two comparison schemes. In other words, the proposed WTB bound provides a much better starting point, for instance, for accurate channel adaptation for mission-critical data transmissions in the short-term regime. In summary, the results above demonstrate that the proposed WTB bound significantly outperforms the KBTB and stationary bounds under general settings involving multiple hops, large initial backlogs and different utilizations. Therefore, we conclude that the KBTB bound that is derived directly using the results from stationary analysis is inadequate for transient analysis.

B. *Evaluation of WTB*

In the previous subsection, we have seen that WTB is typically within one order of magnitude of the simulated violation probability. In this subsection, we further investigate its performance for different parameter settings, and concentrate explicitly on comparing it with the simulated system behavior. In Figures 10 and 11, we present performance results by varying the average SNR and the total initial backlog, respectively, in a 2-hop network with the packet train arrival process. These results confirm that for a 2-hop network, the gap between the proposed WTB bound and the simulated system performance remains around one order of magnitude despite significant variations in the average link SNRs or the initial backlog of the system.

In Figure 12, we present the comparison for a 3-hop network with burst arrival and train arrival processes. In this case, we

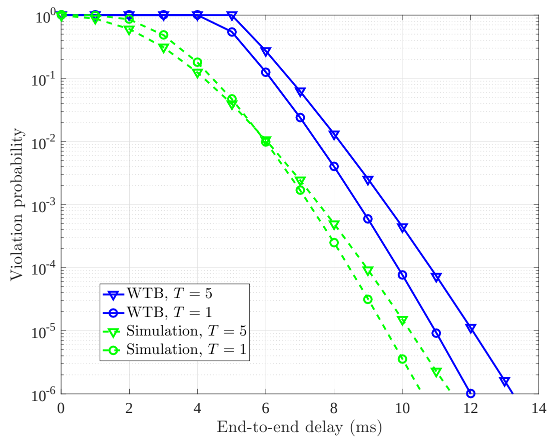


Fig. 12: Delay violation probability vs. end-to-end delay for a 3-hop network for different T, $x_n = 33, \rho = 25, \sigma = 0$, and SNR $= 5$ dB.

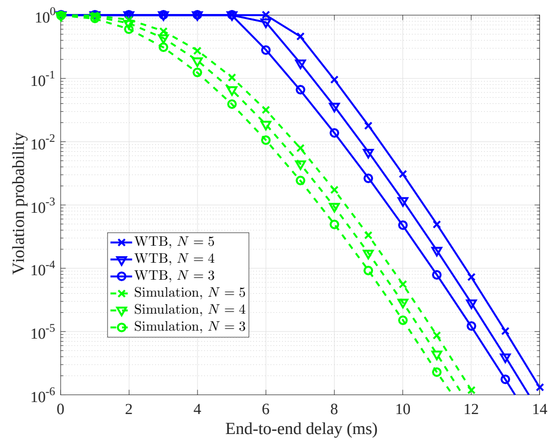


Fig. 14: Delay violation probability vs. end-to-end delay for different N with packet train arrival process ($T = 5$), $x_n = 100/N, \rho = 25, \sigma = 0$, and SNR $= 5$ dB.

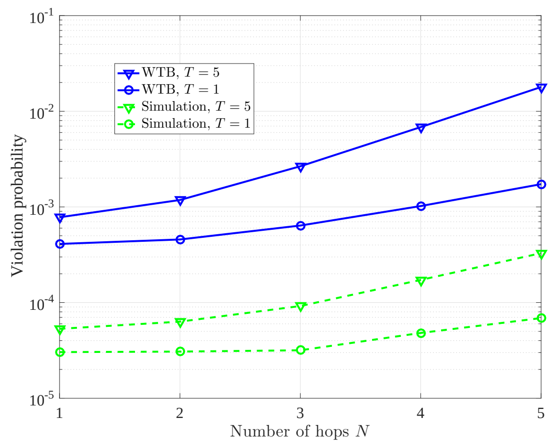


Fig. 13: Delay violation probability vs. N for different T, $w = 9$, $x_n = 100/N, \rho = 25, \sigma = 0$, and SNR $= 5$ dB.

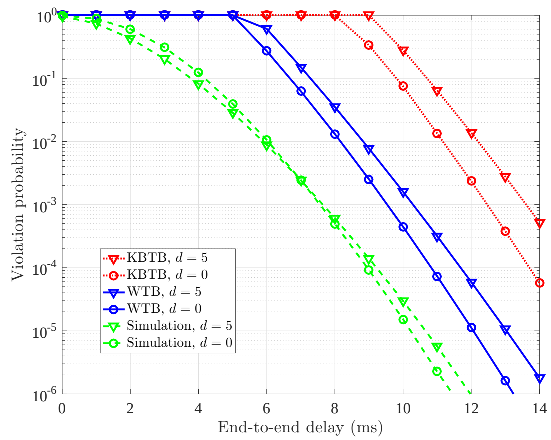


Fig. 15: Delay violation probability vs. end-to-end delay for a 3-hop network for different d with packet train arrival process $(T = 5), x_n = 33, \rho = 25, \sigma = 0$, and SNR $= 5$ dB.

observe that the gap increases beyond one order of magnitude for the considered arrival process types. Further, in Figure 13 we study the performance with increasing number of hops. We observe that for $T = 1$, i.e., burst arrival process, the gap between WTB and simulation is almost insensitive to the number of hops, while there is a steady increase in the gap for $T = 5$. We expect this trend to continue as the number of hops increases. Nevertheless, from Figure 14 we observe that WTB has a decay rate that matches closely with the decay rate of the simulated violation probability with increasing number of hops. The significance of this property is that, an optimization of the proposed WTB bound can yield good heuristic solutions for the optimization of the end-to-end delay in the network operating in transient state.

Delayed Backlog Information: Finally, in Figures 15 and 16 we study the impact of delayed backlog information on the predicted system behavior as captured by the bound presented in Section IV-C. In these figures we use the packet train arrival model while varying on the x-axis the target delay w as well as backlog information delay d. From Figure 15, we observe similar trends as before for varying w, but recognize instantly a cost of the outdated backlog information of between a factor of 2 up to 5. In Figure 16 we observe that - in correspondence to the previous figure - the WTB bound becomes loose as d increases by the same ratios as observed previously. This is expected and is a consequence of using union bound, in which the number of terms increase as d increases.

Remark 1: For the case of delayed backlog information, ideally, in the simulation one would want to simulate the system starting from t_0 for a given initial backlog $\mathbf{x}(t_0)$. However, the question then arises what is the backlog $\mathbf{x}(t_0-d)$ at time $t_0 - d$, since different $\mathbf{x}(t_0 - d)$ values may lead to the observed $\mathbf{x}(t_0)$. Thus, we are not able to find a meaningful comparison with the bounds in this case. Instead, we simulate the system starting at time t_0-d for a given $\mathbf{x}(t_0-d)$, compute the bounds using $\mathbf{x}(t_0 - d)$ and present the comparison. As a consequence the simulated violation probability in Figure 16 varies with d.

Discussion: From the above observations we conclude that WTB can be potentially used for end-to-end delay optimization problems and outperforms any existing steady-state bound. It also performs consistently with increasing number

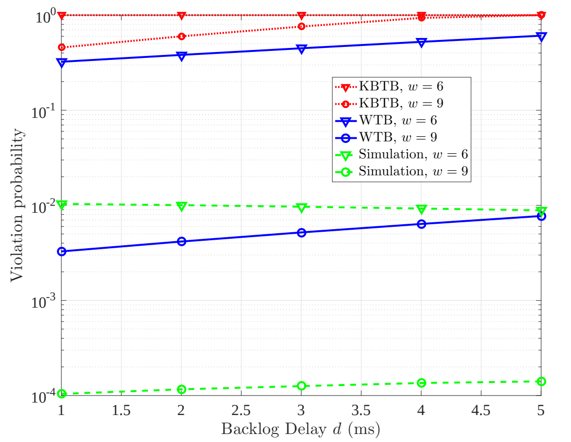


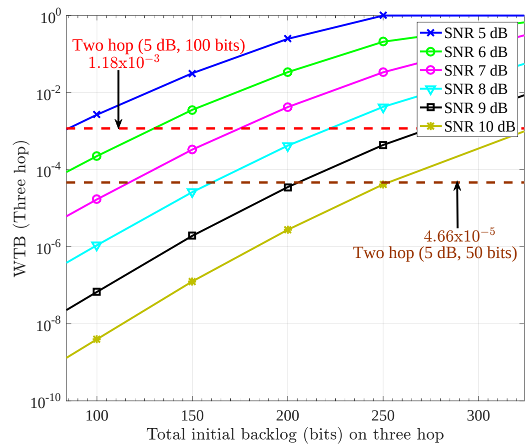
Fig. 16: Delay violation probability vs. backlog information delay d for a 3-hop network for different w with packet train arrival process $(T = 5), x_n = 33, \rho = 25, \sigma = 0$, and SNR $= 5$ dB.

of hops for the burst arrival process. However, WTB becomes loose with increasing number of hops. This looseness can be attributed to the use of the union bound repeatedly. The bound can be tightened for the cases where probability distribution of the cumulative service can be computed exactly, as in this case the use of the union bound can be avoided.

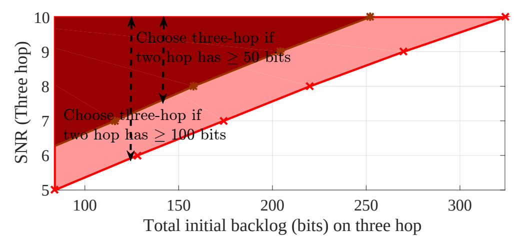
C. Routing Example

In this section, we use WTB for best route selection (in terms of end-to-end delay statistics) for the arriving time-critical traffic. It is known that most available routing algorithms use number of hops, and possibly SNR in a wireless setting, to decide the best route. Although this is a good strategy for stationary flows where long term statistics are prevailing, on the short term, it may result (temporarily) in large buffers which may disperse after a short while. If a time-critical message arrives during this brief period of bloating buffers, an alternative route that is not normally used can be a better choice. We present next an example scenario that is engineered to shed light on such a case and show the utility of WTB in this regard.

Consider a time-critical message arrives at a wireless network node and can take one of two possible routes: a 2-hop route, and a 3-hop route. Assume (for the sake of simpler exposition) that the routes are homogeneous, i.e., the average SNRs are the same for each node along a route. We set $T = 5$, $\rho = 25$, and end-to-end delay $w = 9$ ms. For the 2-hop route, setting SNR $= 5$ dB at each link, for total initial backlogs of 50 and 100 bits, we compute WTB values of 4.66×10^{-5} and 1.18×10^{-3}, respectively. Given these WTB values for the 2-hop route, in Figure 17(a), we present WTB for the 3-hop route for varying total initial backlog and for different SNR values. The crossing points with the dotted lines provide thresholds for total initial backlog and SNR values where it is preferable to choose the 3-hop (with lower WTB) over the 2-hop route. Based on this, decision regions can be identified for choosing the 3-hop route as shown in Figure 17(b). We



(a)



(b)

Fig. 17: (a) WTB for the 3-hop route (solid lines) versus the total initial backlog for different SNRs and $w = 9$ ms; WTB for the 2-hop route (dotted lines) with SNR $= 5$ dB, and total inital backlogs 50 bits (bottom) and 100 bits (top). (b) Decision region for choosing 3-hop route over 2-hop route with different total inital backlogs.

conclude that, instead of using only the hop count, using the initial backlogs on a route and transient analysis would enable a routing algorithm to choose a route that results in better QoS for a time-critial message.

It is important to note that the proposed routing algorithm is based on comparison of upper bounds which may not always yield the desired outcome since the actual processes may or may not have the same order as their upper bounds. Nevertheless, WTB has a decay rate that closely matches the decay rate of the vilation probability, and it becomes loose as the number of hops increases; see Figure 13. Therefore, the proposed policy will most likely perform better or at least the same as policies that consider only hop count to decide the best route. To avoid false positives, a routing algorithm using WTB may choose a different policy in the decision region close to the threshold. For example, when in that region, the algorithm may randomly choose one of the two routes, or it may always choose the 2-hop route.

VI. Conclusions

We have studied the problem of characterizing the end-to-end delay of a sequence of time-critical messages traversing through a multi-hop wireless network with non-zero initial backlog at each hop. As this requires the network to be analysed in the transient state, we attempt to find upper bounds for the end-to-end delay using stochastic network calculus. We have studied the state-of-the-art upper bounds and have demonstrated their poor performance for the problem at hand. We have derived WTB by using the first principles of network calculus and the state-of-the-art bounding techniques. A key aspect of WTB is that it carefully incorporates the known initial backlog in the network. We also extended WTB for the case where the initial backlog information is delayed. Through extensive simulations we have showed that WTB is significantly better than the alternatives. Also, we have observed that its decay rate closely matches the decay rate of the simulated violation probability. Further, using a routing example, we have shown that WTB could potentially be used in routing decisions to improve QoS for time-critical messages. We believe that these key features of WTB makes it a useful metric in the design and optimization of the networks for safety-critical machine-type applications.

REFERENCES

[1] O. N. C. Yilmaz, Y. P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *Proc. IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015, pp. 1190–1195.

[2] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Seln, and J. Skld, "5G wireless access: requirements and realization," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 42–47, Dec. 2014.

[3] Y. Polyanskiy, H. V. Poor, and S. Verd?u, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[4] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. IEEE SECON*, 2011.

[5] W. Dapeng and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Transactions Wireless Communications*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

[6] Y. Jiang, "A basic stochastic network calculus," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 123–134, Aug. 2006.

[7] M. Fidler, "An end-to-end probabilistic network calculus with moment generating functions," in *Proc. IEEE IWQoS*, Jun. 2006, pp. 261–270.

[8] H. Al-Zubaidy, J. Liebeherr, and A. Burchard, "Network-layer performance analysis of multihop fading channels," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 204–217, Feb. 2016.

[9] P. Morse, *Queues, Inventories and Maintenance: The Analysis of Operational Systems with Variable Demand and Supply*. Wiley, 1958.

[10] J. Zhang and E. J. Coyle, "The transient solution of time-dependent M/M/1 queues," *IEEE Transactions on Information Theory*, vol. 37, no. 6, pp. 1690–1696, Nov. 1991.

[11] T. Matis and R. Feldman, "Transient Analysis of State-Dependent Queuing Networks via Cumulative Functions," *Journal of Applied Probability*, vol. 38, 2001.

[12] T. Czachrski, *Queueing Models for Performance Evaluation of Computer Networks - Transient State Analysis*. Springer International Publishing Switzerland, 2015.

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

[14] C. Wang, D. Logothetis, K. Trivedi, and I. Viniotis, "Transient behavior of ATM networks under overloads," in *Proc. IEEE INFOCOM*, Mar. 1996.

[15] N. Becker and M. Fidler, "A non-stationary service curve model for performance analysis of transient phases," in *Proc. International Teletraffic Congress (ITC)*, Sep. 2015, pp. 116–124.

[16] R. L. Cruz, "A calculus for network delay, part I: Network elements in isolation," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–131, Jan 1991.

[17] C.-S. Chang, *Performance Guarantees in Communication Networks*. Springer-Verlag, 2000.

[18] M. Fidler, "A network calculus approach to probabilistic quality of service analysis of fading channels," in *Proc. IEEE GLOBECOM*, Nov. 2006, pp. 1–6.

[19] F. Ciucu, "Non-asymptotic capacity and delay analysis of mobile wireless networks," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 39, no. 1, pp. 359–360, Jun. 2011.

[20] S. Schiessl, J. Gross, and H. Al-Zubaidy, "Delay analysis for wireless fading channels with finite blocklength channel coding," in *Proc. ACM MSWiM*, Nov. 2015, pp. 13–22.

[21] H. Forssell, R. Thobaben, H. Al-Zubaidy, and J. Gross, "On the impact of feature-based physical layer authentication on network delay performance," in *Proc. IEEE GLOBECOM*, Dec. 2017, pp. 1–6.

[22] H. Al-Zubaidy, V. Fodor, G. Dn, and M. Flierl, "Reliable video streaming with strict playout deadline in multihop wireless networks," *IEEE Transactions on Multimedia*, vol. 19, no. 10, pp. 2238–2251, Oct. 2017.

[23] F. Naghibi, S. Schiessl, H. Al-Zubaidy, and J. Gross, "Performance of wiretap rayleigh fading channels under statistical delay constraints," in *Proc. IEEE ICC*, May 2017, pp. 1–7.

[24] N. Petreska, H. Al-Zubaidy, R. Knorr, and J. Gross, "Power-minimization under statistical delay constraints for multi-hop wireless industrial networks," *CoRR*, vol. abs/1608.02191, 2016.

APPENDIX

A. Proof of Corollary 1

Using (1), we obtain

$$[\mathcal{A}(t)/\mathcal{A}(u)]^s \leq e^{s(\sigma+\rho(t-u))}. \tag{26}$$

Now, using Theorem 1 with $N = 1$ and substituting (26) in (16), we obtain

$$\begin{aligned}
&\mathbb{P}(\mathcal{W}(t) > w) \\
&\leq \min_{s>0} \left\{ e^{s x_{\max}} \Big[\sum_{u=0}^{t-1} e^{s(\sigma+\rho(t-u))} V^{\tau-u}(s) + (V(s))^{\tau-t} \Big] \right\} \\
&= \min_{s>0} \left\{ e^{s(x_{\max}-\rho w)} \Big[\sum_{u=0}^{t-1} e^{s\sigma} V_0^{\tau-u}(s) + (V_0(s))^{w} \Big] \right\} \\
&= \min_{s>0} \left\{ e^{s(x_1-\rho w)} \Big[e^{s\sigma} V_0^{\tau}(s) \frac{(V_0(s))^{-t}-1}{(V_0(s))^{-1}-1} + (V_0(s))^{w} \Big] \right\} \\
&= \min_{s>0} \left\{ e^{s(x_1-\rho w)} (V_0(s))^{w} \Big[e^{s\sigma} \frac{V_0(s)-(V_0(s))^{t+1}}{1-V_0(s)} + 1 \Big] \right\}.
\end{aligned}$$

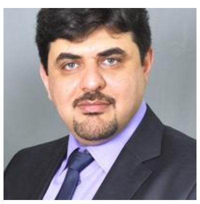
B. Proof of Corollary 2

Using (26) in the summation part of the first term of $\Phi(s)$, given in (17), we obtain

$$\begin{aligned}
\sum_{u=1}^{t-1} [\mathcal{A}(t)/\mathcal{A}(u)]^s V^{-u}(s) &\leq \sum_{u=1}^{t-1} e^{s(\sigma+\rho(t-u))} V(s)^{-u} \\
&= e^{s(\sigma+\rho t)} \cdot \sum_{u=1}^{t-1} [e^{s\rho} V(s)]^{-u} \\
&= e^{s(\sigma+\rho t)} \cdot \frac{1-V_0^{-t}(s)}{V_0(s)-1}.
\end{aligned}$$

The corollary follows by substituting the above inequality in (17).

Jaya Prakash Champati is a post-doctoral researcher from the division of Information Science and Engineering, EECS, KTH Royal Institute of Technology, Sweden. He finished his PhD in Electrical and Computer Engineering, University of Toronto, Canada in 2017. He obtained his master of technology degree from Indian Institute of Technology (IIT) Bombay, India, and bachelor of technology degree from National Institute of Technology Warangal, India. His general research interest is in the design and analysis of algorithms for scheduling problems that arise in networking and information systems. Currently, his focus is on freshness and delay analysis for time-critical control applications in Cyber-Physical Systems (CPS) and the Internet of Things (IoT). In the past, he worked on task scheduling and job assignment problems with motivations from computational offloading in edge computing systems. Prior to joining PhD he worked at Broadcom Communications, where he was involved in developing the LTE MAC layer. He was a recipient of the best paper award at IEEE National Conference on Communications, India, 2011.



Hussein Al-zubaidy (S07M11SM16) received the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2010. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, from 2011 to 2013. In the Fall of 2013, he joined the School of Electrical Engineering (EES) at the Royal Institute of Technology (KTH), Stockholm, Sweden, as a Post-Doctoral Fellow. Since Fall 2015, he has been a Senior Researcher with EES at the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Al-Zubaidy is the recipient of many honors and awards, including the Ontario Graduate Scholarship (OGS), NSERC Visiting Fellowship, NSERC Summer Program in Taiwan, OGSST, and NSERC Post-Doctoral Fellowship.

James Gross received his Ph.D. degree from TU Berlin in 2006. From 2008-2012 he was Assistant Professor and head of the Mobile Network Performance Group at RWTH Aachen University, as well as a member of the DFG-funded UMIC Research Centre of RWTH. Since November 2012, he has been with the Electrical Engineering and Computer Science School, KTH Royal Institute of Technology, Stockholm, where he is professor for machine-to-machine communications. He served as Director for the ACCESS Linnaeus Centre from 2016 to 2019, while he is currently a member of the board of KTHs Innovative Centre for Embedded Systems. His research interests are in the area of mobile systems and networks, with a focus on critical machine-to-machine communications, cellular networks, resource allocation, as well as performance evaluation methods. He has authored about 150 (peer-reviewed) papers in international journals and conferences. His work has been awarded multiple times, including the Best Paper Award at ACM MSWiM 2015, the Best Demo Paper Award at IEEE WoWMoM 2015, the Best Paper Award at IEEE WoWMoM 2009, and the Best Paper Award at European Wireless 2009. In 2007, he was the recipient of the ITG/KuVS dissertation award for his Ph.D. thesis. He is also co-founder of R3 Communications GmbH, a Berlin-based start-up in the area of ultrareliable low-latency wireless networking for industrial automation.