

Non-Asymptotic Throughput and Delay Distributions in Multi-Hop Wireless Networks

Florin Ciucu, Oliver Hohlfeld and Pan Hui

Deutsche Telekom Laboratories / Technical University Berlin

{florin, oliver, ben}@net.t-labs.tu-berlin.de

Abstract—The class of Gupta-Kumar results give the asymptotic throughput in multi-hop wireless networks but cannot predict the throughput behavior in networks of typical size. This paper addresses the non-asymptotic analysis of the multi-hop wireless communication problem and provides, for the first time, closed-form results on multi-hop throughput and delay distributions. The results are non-asymptotic in that they hold for any number of nodes and also fully account for transient regimes, i.e., finite time scales, delays, as well as bursty arrivals. Their accuracy is supported by the recovery of classical single-hop results, and also by simulations from empirical data sets with realistic mobility settings. Moreover, for a specific network scenario and a fixed pair of nodes, the results confirm Gupta-Kumar's $\Omega(1/\sqrt{n \log n})$ asymptotic scaling law.

I. INTRODUCTION

The problem of multi-access communication is about the fundamental performance limits of communication in wireless networks. Metrics of interest include network capacity, referring to the maximal data rates which can be reliably sustained by the network, and network delay.

The classical information theory has played a fundamental role to the field of wireless communications. However, its approach to the network capacity problem still faces two major drawbacks: it leads to solutions only for few topologies (e.g., Shannon's two-node model), and it ignores the fundamental aspects of data burstiness and delay characteristic to packet switching networks [15]. A more recent approach which ignores multi-user coding schemes, as in the work of Gupta and Kumar [18], leads to solutions in asymptotic form. Although such results explain how network capacity scales in the number of nodes, they are unable to predict the exact capacity in networks of typical size, and are thus being questioned on their practicality [3]. As far as the network delay problem is concerned, this is in itself a fundamentally difficult problem even in wired networks where solutions are limited to product-form networks [23].

This paper proposes a stochastic network calculus approach to analyze the performance of multi-hop wireless networks from a non-asymptotic point of view. Concretely, the paper considers networks of any size, accounts for the burstiness and delay aspects of data networks, and also encompasses the notion of time scales for transient analysis; these aspects were recently posed as a challenge for the development of a functional network information theory [4]. The contribution of this paper consists in the derivation of explicit and non-asymptotic bounds on throughput and delay distributions over a multi-hop path in a wireless network. For a specific network

scenario the results confirm Gupta-Kumar's $\Omega(1/\sqrt{n \log n})$ asymptotic scaling law for a fixed pair of nodes. Other consequences include the recovery of classical results on single-hop throughput averages and a sufficient part of a stability condition conjectured in [33]. Simulations from an empirical data set with realistic mobility settings further indicate the accuracy of the distribution results.

A. Solution's Generality and Overview

The network model considers both static and random topologies by representing the number of nodes contending for the same channel by a distribution function. This node density model accounts for mobility and random node placement, and is supported by empirical data with realistic mobility settings. The paper considers the total delay comprising both the queueing delay due to arrival burstiness and the access delay due to multi-access [13]; also, the paper applies to the broad class of bursty arrivals with finite moment generating functions (MGFs). The network model assumes the existence of a multi-hop path at the network layer. Although the paper is restricted to the slotted-Aloha MAC protocol [1] and does not explicitly consider SINR and network scheduling (there is no queueing at the nodes due to scheduling), it shows how to extend the analysis to account for such aspects and other MAC protocols. The results on throughput and delays are mathematically rigorous in that they do not involve simplifying assumptions, commonly made in the literature for technical reasons (e.g., statistical independence), but they fully address the difficult problem of spatial-time correlations arising in multi-hop communication.

The analysis in this paper follows the approach of the stochastic network calculus [9], [21], which is a probabilistic extension of the deterministic network calculus conceived by Cruz [12]. The main features of the calculus relevant to the analysis in this paper are 1) service abstraction (i.e., the characterization of the exclusive service received by network nodes due to contention at the MAC layer) and 2) multi-node analysis as a relatively straightforward extension of single-node analysis by using (min, +) algebra operations.

B. Related Work

Results on throughput and delays in multi-access networks are mostly restricted to averages, specific arrivals, or single-hop scenarios, whereas multi-hop results rely on simplifying technical assumptions. Throughput averages were derived in both static [25], [2], [5] and random [24] networks. Queueing

delay averages were derived in a scenario with two identical nodes [29], [31]; an approximation analysis for more than two nodes was considered in [13]. Access delay distributions, restricted to Bernoulli or Poisson processes, were derived in terms of z-transforms [32] and in explicit form [35]. Multi-hop throughput averages were obtained by decomposing the network into single-hop links and deriving the local throughputs using a set of fixed-point equations [16]. The total delay in a single line multi-hop network was derived using simplifying independence assumptions to deal with the spatial-time correlations [34].

As far as network capacity results are concerned, the pure information theory approach produced partial solutions in network size [15]. In an attempt to simplify the problem, Gupta and Kumar considered certain ideal assumptions on power-control, routing, or scheduling, and also dispensed with multi-user coding schemes [18], and obtained the network capacity's order of growth in the number of nodes n : the per-node capacity scales as $\Theta(1/\sqrt{n})$ in arbitrary networks and as $\Theta(1/\sqrt{n \log n})$ in random networks.

Necessary and sufficient stability conditions in Aloha multi-access networks exist in explicit form only for two nodes [33]. In scenarios with more than two nodes solutions are exact and recursive [30], or approximative and asymptotically exact [6].

The rest of the paper is structured as follows. Section II describes the single-hop wireless network model and the main analytical tools. Section III applies these tools to derive single-hop per-node throughput and delay distributions, and also stability conditions. Further extensions to the multi-hop case are considered in Section IV. Brief conclusions are presented in Section VI.

II. NETWORK MODEL AND ANALYTICAL TOOLS

In this section we describe the wireless network model and the main analytical tools used for throughput and delay analysis. One such tool is the interfering process model for successful transmissions of a node running a MAC protocol. From this interfering process we then construct a node's service curve which provides a lower bound on the amount of the node's successful transmissions as a function of time. Finally we describe two density models capturing the randomness of the number of nodes within a contention region.

For the single-hop analysis we consider that the number of nodes is a random process $N(t)$. All the nodes hear each other, share a single communication channel, and all run the slotted-Aloha MAC protocol (the multi-hop scenario will be considered in Section IV). We are interested in the throughput and delay analysis for a fixed node A transmitting to some arbitrary node D , whenever $N(t) > 1$. We assume that all the nodes except for A are saturated (always backlogged), e.g., having infinite arrivals. In turn, the node A can be either saturated or not; the latter situation corresponds to the case of bursty arrivals. The node A has an infinite sized buffer such that in the case of bursty arrivals the model accounts for total (queueing plus access) delay.

We use an underlying discrete time model with increment Δt representing the transmission time of one data unit (packet). For each slot $(t, t + \Delta t]$, each node transmits with probability $p_{N(t)}$ depending on $N(t)$; when $N(t) = 1$ we let $p_1 = 0$, i.e., no transmissions. The communication channel offers a maximal service rate of one packet per one time unit, such that at most Δt aggregate throughput can be achieved by all the nodes during $(t, t + \Delta t]$. We also use continuous time approximations—in order to simplify computations with differential calculus—when the time scale of interest is large, e.g., when computing the asymptotic (long-term) throughput.

For this network model we are interested in two performance metrics. The first is A 's throughput for both transient (i.e., time dependent) and asymptotic regimes. Formally, in the case when A is saturated, the transient throughput refers to the maximal rates λ_t such that for all $t \geq 0$

$$\Pr(D(t) \leq \lambda_t t) \geq \varepsilon, \quad (1)$$

where $D(t)$ denotes the cumulative departure process at node D (more exactly the number of packets received exclusively from A by time t) and ε is some fixed violation probability, e.g., $\varepsilon = 10^{-3}$. In other words, λ_t represents a probabilistic lower bound on A 's achievable throughput by time t .

The other metric is A 's delay process which we consider only in the case of bursty arrivals, denoted by the cumulative process $A(t)$. The delay process is denoted by $W(t) = \inf\{d : A(t-d) \leq D(t)\}$, i.e., $W(t)$ represents the delay of the last packet successfully transmitted by time t . In particular we look for probabilistic delay bounds d satisfying

$$\Pr(W(t) > d) \leq \varepsilon, \quad (2)$$

for some violation probability ε . Note that A 's delay process is meaningless when A is saturated.

A. Interfering Process and Service Curve Representation

The next definition introduces the key analytical tool used in this paper, i.e., the interfering process of a node.

Definition 1: The virtual interfering process $V(t)$ of a node A is defined by its Bernoulli increment process $V(t, t + \Delta t) := V(t + \Delta t) - V(t)$ as

$$V(t, t + \Delta t) = \begin{cases} 0 & \text{if } A \text{ virtually and successfully} \\ & \text{transmits during } (t, t + \Delta t] \\ \Delta t & \text{otherwise} \end{cases} \quad (3)$$

The probabilities are for $q_{N(t)} := 1 - p_{N(t)}(1 - p_{N(t)})^{N(t)-1}$

$$\begin{cases} \Pr(V(t, t + \Delta t) = 0) = 1 - q_{N(t)}, \text{ and} \\ \Pr(V(t, t + \Delta t) = \Delta t) = q_{N(t)}. \end{cases} \quad (4)$$

By “ A virtually and successfully transmits” it is meant that $V(t, t + \Delta t)$ is defined as if A was always saturated, whence the tag “virtual” for $V(t)$.

The cumulative interfering process $V(t)$ can be written as $V(t) = \sum_{s=0}^{\lceil t/\Delta t \rceil - 1} V(s, s + \Delta t)$ and accounts for the amount of time when A 's (virtual) attempts are unsuccessful, i.e.,

either when at least one node other than A transmits or when the channel is idle. Consequently, $t - V(t)$ represents A 's (virtual) throughput during $[0, t]$. Note that in the single-hop scenario $V(t)$ depends on all the nodes in the transmission range of A . In multi-hop scenarios, $V(t)$ would additionally depend on the hidden nodes of A relative to D .

From this virtual interfering process we next construct a service curve for node A . A service curve is a key network calculus concept modelling a lower bound on how much A can transmit as a function of time. If A 's arrival process is also known then the calculus yields bounds on throughput and delay distributions. The benefits of relying on service curves will become more obvious when analyzing multi-hop scenarios: the fundamentally hard problem of deriving end-to-end results is elegantly solved by service curves convolution.

Formally, a bivariate random process $S(s, t)$ is a (probabilistic) service curve for node A if for all arrival and departure processes $A(t)$ and $D(t)$, respectively, and all times $t \geq 0$ [9]

$$D(t) \geq A * S(t) . \quad (5)$$

Here, $*$ is the (min, +) convolution operator defined as $A * S(t) = \inf_{0 \leq s \leq t} \{A(s) + S(s, t)\}$.

In order to construct a service curve $S(s, t)$ for node A , in the considered single-hop network scenario, we interpret A 's service in the following way. A server of capacity one serves two arrival flows: one is $A(t)$ and the other is the (virtual) arrival process $V(t)$ from Definition 1. The critical observation is that the arrivals of $V(t)$ receive higher priority. In other words, we interpret A 's service as if its arrivals were served at a static priority (SP) scheduler which gives higher priority to the interfering process. Using this interpretation, we use a result from network calculus [14] which gives that A has the service curve

$$S(s, t) = t - s - V(s, t) , \quad (6)$$

which is referred to as a leftover service curve, since it expresses the capacity left unused by $V(t)$.

To clearly see that Eq. (5) holds with $S(s, t)$ from Eq. (6) for *all* the arrival processes $A(t)$, let us fix $t \geq 0$. Taking s^* as the beginning of the last busy period of A before t implies that $A(s^*) = D(s^*)$ and A has always data to serve during $(s^*, t]$. Therefore

$$\begin{aligned} D(t) &= D(s^*) + D(s^*, t) \\ &= A(s^*) + t - s^* - V(s^*, t) \\ &\geq A * S(t) . \end{aligned}$$

For understanding the correctness of the follow-up results, we make the critical observation that $S(s, t)$ is *statistically independent* of *any* arrival process $A(t)$ which may induce unsaturated situations, as long as $A(t)$ is independent of the Aloha protocol. Indeed, note that the only source of randomness in Eqs. (3) and (4) stems from the MAC protocol. Also, the process $V(s, t)$, which appears in $S(s, t)$, is *decoupled* by definition from any actual arriving process $A(t)$ by defining it under virtual saturation of A .

For a brief check-up of the accuracy of this service curve construction, let us consider a fixed number of nodes $N(t) = n$ and probability $p_{N(t)} = p$. If A is saturated then $S(s, t)$ is a strict service curve [7] for $A(s, t)$, i.e., $D(s, t) \geq S(s, t)$ for all $0 \leq s \leq t$. Taking expectations yields

$$E[D(s, t)] \geq t - s - E[V(s, t)] = p(1 - p)^{n-1}(t - s) .$$

Optimizing $p = \frac{1}{n}$ we get that the rate of $D(s, t)$ approaches $\frac{1}{ne}$ when n is sufficiently large. This is exactly the normalized maximal throughput in a slotted-Aloha network [24].

B. Density Models

Our network calculus analysis needs the MGF of the interfering process $V(t)$ from Definition 1, i.e.,

$$M_t(\theta) = E \left[e^{\theta V(t)} \right] ,$$

for some $\theta > 0$. We use the shorthand M_t for $M_t(\theta)$. Note that $V(t)$ is modulated by the process $N(t)$ which represents the number of nodes interfering with node A .

In this subsection we derive M_t for two density models of $N(t)$, depending whether the process has memory or not. The first model assumes that $N(t)$ is a Markov chain and is suitable in mobile scenarios. The second model assumes that $N(t)$ is memoryless, i.e., $N(t)$ is independent of $N(s)$ for all $s < t$. This model is suitable in static scenarios with random node placement, e.g., if nodes are uniformly placed within an area, as in [18], then $N(t)$ has the binomial distribution or its Poisson approximation. The memoryless model can be derived from the conflict graph model for capturing interference from [20], by letting $N(t)$ represent the node-degree of the graph. Also, given mobility models used in the literature (see [26] for a survey), the distribution of $N(t)$ can be in principle derived from the initial state and the mobility description.

1) Markov-Modulated Interfering Process: Here we consider that $V(t)$ is a Markov-modulated interfering process, i.e., $N(t)$ is a continuous Markov process. The process $V(t)$ interferes with rate one and probabilities depending on the current state (see Eqs. (3) and (4)). We restrict the Markov chain to n states, i.e., the node A can interfere with at most $n - 1$ other nodes. Also, we make the simplification that transitions exist only between consecutive states; the transitions between states i and $i + 1$ are denoted by α_i and β_i , respectively, for $i = 1, \dots, n - 1$. By convention we set $\alpha_n = 0$ and $\beta_0 = 0$. The time spent in state n is exponentially distributed with mean $1/(\alpha_n + \beta_{n-1})$. The chain guarantees the existence of the stationary steady-state distribution $\pi = (\pi_1, \dots, \pi_n)$, with $\pi_i = \lim_{t \rightarrow \infty} \Pr(N(t) = i)$.

The value of M_t can be computed (not shown here) using standard techniques based on conditioning (see [11]). We obtained the continuous time approximation

$$E \left[e^{\theta V(t)} \right] \leq e^{\lambda_n t} , \quad (7)$$

where, by convention, $\lambda_n = \max_i \lambda_i$ is the spectral radius of a tridiagonal matrix depending on the Markov chain parameters and transmission probabilities.

2) *Memoryless-Modulated Interfering Process*: We now consider that $N(t)$ is memoryless with the stationary distribution $\pi = (\pi_1, \dots, \pi_n)$, with $\pi_i = \Pr(N(t) = i)$. This simplified model has the advantage of being amenable to an easier mathematical analysis. Using conditioning, one can derive the MGF

$$E[e^{\theta V(t)}] = (1 + q(e^\theta - 1))^t, \quad (8)$$

where $q = \sum_i \pi_i q_i$ and $q_i = 1 - p_i(1 - p_i)^{i-1}$. Also, one may obtain the continuous time approximation

$$E[e^{\theta V(t)}] = e^{\theta q t}. \quad (9)$$

We note that the continuous time approximations of the MGF of $V(t)$ from Eqs. (7) and (9), which have the advantage of having simpler expressions than in discrete time, are to be used in asymptotic regimes where the continuous time approximation has negligible effects.

III. THROUGHPUT AND DELAY ANALYSIS. THE SINGLE-HOP CASE

In the next two subsections we compute bounds on the distribution of A 's throughput and delay.

A. Saturated Case. Throughput Analysis

When A is saturated we determine lower bounds on the achievable throughput λ_t defined in Eq. (1).

Theorem 1: (TRANSIENT AND ASYMPTOTIC THROUGHPUT) Let the memoryless interfering process from Subsection II-B2. Denote $q_i = 1 - p_i(1 - p_i)^{i-1}$ and $q = \sum_i \pi_i q_i$. A lower bound on the transient throughput is

$$\lambda_t = \sup_{\mathbf{p}, \theta} \left\{ 1 - \frac{\log b}{\theta} + \frac{\log \varepsilon}{t\theta} \right\}, \quad (10)$$

where $b = 1 + q(e^\theta - 1)$, $\mathbf{p} = (p_1, \dots, p_n)$ is the vector of transmission probabilities, and ε is some fixed violation probability. A lower bound on the asymptotic throughput is

$$\lambda := \lim_{t \rightarrow \infty} \lambda_t = \sup_{\mathbf{p}} \{1 - q\}. \quad (11)$$

Similar bounds can be derived for the Markov-modulated interfering process from Subsection II-B1. For instance, a bound on the asymptotic throughput is as in Eq. (11) with q replaced by $\frac{\lambda_n}{\theta}$, where λ_n is the spectral-radius from Eq. (7).

PROOF. We can assume without loss of generality that $A(s, t) \geq t - s$ for all $s \leq t$, such that the node A is saturated. Using the definition of the service curve from Eq. (5) and the Chernoff bound for some $\theta > 0$, we can write for all $t \geq 0$

$$\begin{aligned} \Pr(D(t) \leq t\lambda_t) &\leq \Pr\left(\inf_s \{A(s) + S(s, t)\} \leq t\lambda_t\right) \\ &\leq \Pr(V(t) \geq t(1 - \lambda_t)) \leq E[e^{\theta V(t)}] e^{-\theta(1 - \lambda_t)t}. \end{aligned} \quad (12)$$

To derive the throughput λ_t for finite time scales we use the MGF of $V(t)$ from Eq. (8) and get

$$\Pr(D(t) \leq t\lambda_t) \leq \left(be^{-\theta(1 - \lambda_t)} \right)^t.$$

Equating the right-hand side to some $\varepsilon > 0$, we obtain the expression from Eq. (10).

In turn, the direct way to derive the asymptotic throughput λ is to take the limit in t in Eq. (10) and use that $f(\theta) = \frac{\log(1 + q(e^\theta - 1))}{\theta}$ is non-increasing for $\theta \geq 0$ and $q \leq 1$, whereas $\lim_{\theta \rightarrow 0} f(\theta) = q$. An alternative way is to use the continuous time approximation of the MGF of $V(t)$ from Eq. (9) and continue Eq. (12) as $\Pr(D(t) \leq t\lambda_t) \leq e^{-\theta(1 - q - \lambda_t)t}$. From here we obtain the bound $\lambda_t = 1 - q - x$ for all $x > 0$, since the probability converges to 0 as $t \rightarrow \infty$. Letting $x \rightarrow 0$ yields Eq. (11). \square

B. Unsaturated Case. Stability and Delay Analysis

When A is not saturated we determine the maximal arrival rate r of $A(t)$ such that the probability

$$\Pr(D(t) \leq A(t) - \sigma)$$

decays to zero in σ . This value provides a bound on A 's rate to ensure stability. We also compute probabilistic upper bounds on the distribution of the delay process $W(t)$ as in Eq. (2).

Theorem 2: (STABILITY CONDITION. DELAY BOUNDS) Let the memoryless interfering process from Subsection II-B2, and assume that the arrival process $A(t)$ is Poisson with mean rt . Denote $q_i = 1 - p_i(1 - p_i)^{i-1}$, $q = \sum_i \pi_i q_i$, and $b = 1 + q(e^\theta - 1)$. A bound on the maximal arrival rate of node A preserving stability is

$$r \leq \sup_{\mathbf{p}, \theta > 0} \left\{ \frac{1}{e^\theta - 1} \log \frac{e^\theta}{b} \right\}, \quad (13)$$

where $\mathbf{p} = (p_1, \dots, p_n)$ is the vector of transmission probabilities. For r and θ satisfying this bound we have the delay bound for all $d \geq 0$

$$\Pr(W(t) > d) \leq M e^{-\theta(1 - \frac{\log b}{\theta})d}, \quad (14)$$

where $M = e^{-\theta(1 - \frac{\log b}{\theta} - r \frac{e^\theta - 1}{\theta})}$.

Note that the throughput can be computed as in Theorem 1. *PROOF.* Using Eq. (5) we have for $t, \sigma \geq 0$

$$\begin{aligned} \Pr(D(t) \leq A(t) - \sigma) \\ \leq \Pr\left(\sup_{0 \leq s \leq t} \{A(s, t) - t + s + V(s, t)\} \geq \sigma\right). \end{aligned} \quad (15)$$

To bound the sample-path probability we follow [10]. First note that the processes $A(s, t)$ and $V(s, t)$ are statistically independent, since the definition of $V(t)$ is independent of the actual arrival process $A(t)$. Let $r, \theta > 0$ such that

$$E[e^{\theta(A(1) + V(1) - 1)}] = be^{\theta(r \frac{e^\theta - 1}{\theta} - 1)} \leq 1,$$

i.e., the condition from Eq. (13). Then, using the independent increments property of the processes $A(s, t)$ and $V(s, t)$, it can be shown that the process $T(s) = e^{\theta(A(t-s, t) + V(t-s, t) - s)}$ is a supermartingale relative to the filtration of σ -algebras $\mathcal{F}_s = \sigma\{A(t-s, t), V(t-s, t)\}$. Formally it means that $E[T(s+1) | \mathcal{F}_s] \leq T(s)$, i.e., the mean value at time $s+1$ conditioning on the past is smaller than the value at time

s [17]. We can continue Eq. (15) using Doob's maximal inequality [17] yielding

$$\Pr(D(t) \leq A(t) - \sigma) \leq \Pr\left(\sup_s T_s \geq \sigma\right) \leq e^{-\theta\sigma},$$

which proves the claim on the throughput. The proof for the delay bound follows from similar network calculus arguments. \square

In the theorem, Eq. (13) provides a sufficient condition for stability from A 's perspective, as it is obtained by assuming that the other nodes are saturated. In the case of a static scenario with a fixed number of nodes $N(t) = n$, and where each node transmits according to slotted-Aloha with fixed probability p_i , the condition becomes

$$r \leq \sup_{\theta > 0} \left\{ \frac{1}{e^\theta - 1} \log \frac{e^\theta}{1 + \left(1 - p_1 \prod_{j \neq 1} (1 - p_j)\right) (e^\theta - 1)} \right\} = p_1 \prod_{j \neq 1} (1 - p_j), \quad (16)$$

assuming that node A is the first node. This stability condition recovers the sufficiency part of the classical conjecture on Aloha's stability from [33], which can be thus extended to networks with random topologies.

The derived stability condition can also be extended to arrival processes with finite MGFs. To see this, assume for instance that there exists a rate r , depending on a parameter $\theta > 0$, such that $E[e^{\theta A(t)}] \leq e^{\theta r t}$. Following the proof and using Eq. (9) and $e^{\theta E[A(t)]} \leq E[e^{\theta A(t)}]$ (Jensen's inequality), the stability condition from Eq. (13) becomes $E[A(t)] \leq 1 - q$; this recovers Eq. (16) in the case of a static scenario. Similarly, the delay bounds can be extended as well to arrival processes with finite MGFs. The Poisson assumption from the theorem is not critical but only enables the derivation of sample-path bounds using the supermartingale argument which simplifies notation and sharpens the bounds; in the case of more general arrival processes, such sample-path bounds involve the computation of a sum [9], [14].

IV. THROUGHPUT AND DELAY ANALYSIS. THE MULTI-HOP CASE

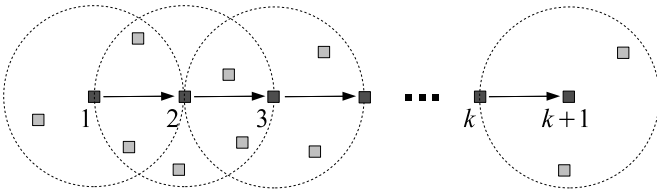


Fig. 1. A network of $n = m(1 + (1 - \phi)k)$ fixed nodes; each circle has m nodes and the intersection of any two circles has ϕm nodes. We are interested in the throughput and delay analysis of node 1 transmitting to node $k + 1$ using the nodes $2, 3, \dots, k$ as relays.

To keep the notation simple we consider the static network topology from Figure 1. Node 1 transmits to node $k + 1$ using nodes $2, 3, \dots, k$ as relays; moreover, nodes $2, 3, \dots, k$ receive

data only from their predecessors, so that a single flow transits each. The $k + 1$ circles from the figure represent the interfering ranges (for simplicity identical to the transmission ranges) of the $k + 1$ nodes. We assume that each circle contains m nodes, and the intersection of any two circles contains ϕm nodes, where $0 < \phi \leq .5$; we omit the details when ϕm is not integer. The whole network has $n = m(1 + (1 - \phi)k)$ nodes.

As in the single-hop network model from Section II, we assume that the wireless communication channel offers a maximal service rate of one packet per one time unit. All the nodes run the slotted-Aloha protocol with time unit $\Delta t = 1$. The transmission probability of each node is p . All the nodes except for $1, 2, \dots, k + 1$ are saturated; when computing the end-to-end throughput the node 1 is also saturated. For convenience we denote a direct transmission from node i to node j by $[i \rightarrow j]$.

The fundamental difficulty of an end-to-end throughput and delay analysis, in contrast to the single-hop analysis, is represented by the correlated transmissions involving the relaying nodes. For instance, the transmissions $[1 \rightarrow 2]$ and $[2 \rightarrow 3]$ are correlated through the behavior of the m nodes located within the second circle. Additionally, the analysis must consider that the transmissions $[i \rightarrow i + 1]$ cannot take place unless there is positive backlog at node i , for $i = 2, 3, \dots, k$. In the rest of this section we will show that the leftover service curves constructed in Eq. (6) lead to a clean solution for the end-to-end analysis, without relying of simplifying technical assumptions.

Let us first recall that the interfering process $V_1(t)$ of node 1 must account not only for the nodes within the circle of node 1, but also for the hidden nodes relative to node 2, i.e., $(1 - \phi)m$ additional nodes. Also, let the following equivalent representation of the interfering process from Eqs. (3) and (4)

$$V(t, t + 1) = 1 - X_1(t) \prod_{i=2}^{N(t)} (1 - X_i(t)), \quad (17)$$

where $X_i(t)$ are i.i.d. Bernoulli random variables $B(p)$, i.e., with probability p take value Δt , and with probability $1 - p$ take value 0. Then the interfering process $V_1(s, t)$, written here in bivariate form $V_1(s, t) := V_1(t) - V_1(s)$, is for all $0 \leq s \leq t$

$$V_1(s, t) = t - s - \sum_{u=s+1}^t X_1(u) \prod_{i=2}^{m(2-\phi)} X_i(u), \quad (18)$$

where $X_i(u)$ are i.i.d. Bernoulli random variables $B(p)$. The interfering processes $V_j(t)$ of nodes $2, 3, \dots, k$ have similar expressions. For instance, in the case of node 2,

$$V_2(s, t) = t - s - \sum_{u=s+1}^t Y_1(u) \prod_{i=2}^{m(2-\phi)} Y_i(u), \quad (19)$$

where $Y_i(u)$ are i.i.d. $B(p)$ random variables. As remarked before the sets $\{X_1(u), X_2(u), \dots, X_{(2-\phi)m}(u)\}$ and $\{Y_1(u), Y_2(u), \dots, Y_{(2-\phi)m}(u)\}$ have in common m random variables which correlate the transmissions $[1 \rightarrow 2]$ and $[2 \rightarrow 3]$.

Having the interfering processes, the leftover service curves $S_i(s, t)$ modelling lower bounds on the transmissions $[i \rightarrow i+1]$ follow directly from Eq. (6) for all $i = 1, 2, \dots, k$. For instance, in the case of node 1,

$$S_1(s, t) = \sum_{u=s+1}^t X_1(u) \prod_{i=2}^{m(2-\phi)} X_i(u). \quad (20)$$

Furthermore, by applying the concatenation result from network calculus [9], it follows that the bivariate process

$$\begin{aligned} S(s, t) &= S_1 * S_2 * \dots * S_k(s, t) \\ &= \inf_{s \leq u_1 \leq \dots \leq u_{k-1} \leq t} \{S_1(s, u_1) + \dots + S_k(u_{k-1}, t)\} \end{aligned} \quad (21)$$

is a network service curve for the end-to-end transmission $[1 \rightarrow 2 \rightarrow \dots \rightarrow k+1]$. This process essentially models a lower bound on the available end-to-end service by capturing the interferences at all the relay nodes.

A critical observation for our further analysis is that for fixed u_1, u_2, \dots, u_{k-1} , the (random) processes $S_1(s, u_1), S_2(u_1, u_2), \dots, S_k(u_{k-1}, t)$, and implicitly $V_1(s, u_1), V_2(u_1, u_2), \dots, V_k(u_{k-1}, t)$ are statistically independent. In particular, note that $V_1(s, u_1)$ and $V_2(u_1, u_2)$ from Eqs. (18) and (19) are independent because their expressions do not share any random variables: the intervals $(s, u_1]$ and $(u_1, u_2]$ are disjoint, whereas the randomization process in slotted-Aloha is memoryless (e.g., $X_i(u+1)$ is independent of $X_i(u)$ in Eq. (18)).

We are now ready to provide the results on the end-to-end throughput and delay distributions, and stability.

Theorem 3: (TRANSIENT AND ASYMPTOTIC THROUGHPUT; STABILITY CONDITION; DELAY BOUNDS) Consider the network scenario from Figure 1. Denote $q = 1 - p(1 - p)^{m(2-\phi)-1}$. If the node 1 is saturated then a lower bound on the transient multi-hop throughput, as defined in Eq. (1), is for all $t \geq 0$

$$\lambda_t = \sup_{p, \theta} \left\{ 1 - \frac{\log b}{\theta} + \frac{\log \varepsilon}{t\theta} - \frac{\log \binom{t+k-2}{k-1}}{t\theta} \right\}, \quad (22)$$

where $b = 1 + q(e^\theta - 1)$ and ε is some fixed violation probability. A lower bound on the asymptotic throughput $\lambda := \lim_{t \rightarrow \infty} \lambda_t$ is

$$\lambda = 1 - q. \quad (23)$$

In turn, if the arrival process $A(t)$ at node 1 satisfies $E[e^{\theta A(t)}] \leq e^{\theta r t}$, for some $r > 0$, depending on θ , a bound on the maximal arrival rate of A preserving stability is

$$r \leq \sup_{\theta > 0} \left\{ 1 - \frac{\log b}{\theta} \right\}. \quad (24)$$

For r and θ satisfying this bound we have the delay bound for all $d \geq 0$

$$Pr(W(t) > d) \leq \left(\frac{1}{\theta \left(1 - \frac{\log b}{\theta} - r \right)} \right)^k e^{-\theta \left(1 - \frac{\log b}{\theta} \right) d}. \quad (25)$$

The formulas in the theorem are very concise because of the perfect symmetry in the $k+1$ circles from Figure 1 (e.g., m nodes in all circles, same transmission probabilities for all nodes); in asymmetric scenarios the corresponding formulas would increase in notation.

With respect to throughput, Eq. (22) differs from the single-hop parallel result from Theorem 1 (restricted to slotted-Aloha for a static network with same transmission probabilities for all nodes) by the last term involving the binomial factor; this factor depends both on time and the number of hops. Moreover, optimizing Eq. (23) with $p = \frac{1}{m(2-\phi)}$ we obtain the lower bound on the end-to-end asymptotic throughput

$$\lambda = \frac{1}{m(2-\phi)e}, \quad (26)$$

which is independent of the number of hops and represents the normalized asymptotic throughput in a slotted-Aloha network with $m(2-\phi)$ nodes.

With respect to the stability condition, Eq. (24) is slightly weaker than the single-hop condition from Eq. (13) in that it is obtained using a summation argument, rather than the stronger supermartingale argument used in Theorem 1. A simple inspection into Eq. (24) reveals the natural condition that $r \leq \lambda$, with λ from Eq. (26). Stronger conditions can be further derived by considering different transmission probabilities at the nodes, as in Theorem 2.

On the other hand, an inspection into the delay bound from Eq. (25) reveals that the end-to-end delay quantiles $d(\varepsilon)$ scale as $\Theta(k)$ in the number of hops, $\mathcal{O}(m)$ in the number of nodes m within a circle, and also $\mathcal{O}(1-\phi)$ in the fraction of nodes ϕ within an intersection.

Before giving the proof we also mention that the results in the theorem can be easily extended to a random scenario with $N(t)$ nodes within each circle. Under symmetric conditions, as pointed out earlier, the formulas in the theorem would still hold with q replaced by the weighted average $\sum_i \pi_i q_i$, as in Theorem 2.

PROOF. Denote $A(t)$ the arrivals at node 1 and $D(t)$ the departures at node $k+1$. Following the proof of Theorem 1 we assume that $A(s, t) \geq t - s$ for all $s \leq t$. Following the derivation of Eq. (12), but with the network service curve $S(s, t)$ from Eq. (21), we can write for all $t \geq 0$

$$\begin{aligned} Pr(D(t) \leq t\lambda_t) &\leq Pr\left(\inf_{0 \leq s \leq t} \{A(s) + S(s, t)\} \leq t\lambda_t\right) \\ &\leq Pr\left(\sup_{1 \leq u_1 \leq \dots \leq u_{k-1} \leq t} \sum_{l=1}^k V_l(u_{l-1}, u_l) \geq t(1 - \lambda_t)\right) \\ &\leq \sum_{1 \leq u_1 \leq \dots \leq u_{k-1} \leq t} E\left[\theta \sum_{l=1}^k V_l(u_{l-1}, u_l)\right] e^{-\theta t \lambda_t} \\ &\leq \sum_{1 \leq u_1 \leq \dots \leq u_{k-1} \leq t} L = \binom{t+k-2}{k-1} L, \end{aligned}$$

where $L = (be^{-\theta(1-\lambda_t)})^t$, $u_0 = 0$ and $u_k = t$. In the third line we applied the Chernoff bound for some $\theta > 0$. Then we used the independence of the processes $V_l(u_{l-1}, u_l)$; recall the

explanation after Eq. (21). Finally, the binomial term $\binom{t+k-2}{k-1}$ is the number of combinations with repetition. In turn, to show the asymptotic rate, we can apply the Stirling's approximation for the factorial, and obtain that $\binom{t+k-2}{k-1}/(t\theta) \rightarrow 0$ as $t \rightarrow \infty$. The rest of the proof follows as in the proof of Theorem 1 and using network calculus arguments (see for instance [14]). \square

A. Further Extensions

Here we make some remarks on how to further extend the methodology from this paper to take into account other MAC scheduling algorithms, the noise at the physical layer, scheduling at the network layer, and also the ability of Theorem 3 to confirm Gupta-Kumar's scaling law (specifically the lower bound) for the specific network considered in Figure 1.

For other MAC scheduling algorithms, such as 802.11 DCF, what is needed is the redefinition of the interfering process from Definition 1. In the single-hop case one may use the success transmission probabilities from [5], which are based on simplifying independence assumptions, whereas the multi-hop extension is not so straightforward. In the case of the slotted-Aloha network with the capture effect [28], i.e., a transmission can succeed in the presence of overlapping transmissions if the SINR is sufficiently large. To account for this situation we can change the expression of $q_{N(t)}$ from Definition 1, using conditioning, to

$$q = 1 - p \left[(1 - p)^{n-1} + \left(1 - (1 - p)^{n-1} \right) \Pr(SINR \geq \zeta) \right],$$

where ζ is the threshold, and we let $N(t) = n$. Also, $1 - q$ can be set to available success probabilities for packet reception (see [28] for slotted-Aloha and [27] for CSMA/CD).

Let us now consider the aspect of network scheduling, i.e., there is queueing at nodes $1, 2, \dots, k$ from Figure 1 due to serving multiple flows (so far we implicitly assumed that only the flow $[1 \rightarrow k + 1]$ is being served, while the rest of the nodes have only the role of interfering). Denote by $A_i(t)$ the aggregate of other flows, but $[1 \rightarrow k + 1]$, at each node i . Then, the service curves from Eq. (20), for the flow traversing the network, become $S_1(s, t) = \sum_{u=s+1}^t X_1(u) \prod_{i=2}^{m(2-\phi)} X_i(u) - A_1(s, t)$ using the leftover theorem in network calculus [14]. In other words, $A_1(s, t)$ is an additional high priority process in the service interpretation from Section II-A.

Lastly, to fit Gupta-Kumar's network model with n nodes [18], in the model from Figure 1, we let $m = \log n$ and $k = \sqrt{n}$. We must consider now that the nodes $1, 2, \dots, k$ serve other flows, in addition to the flow shown in Figure 1. Because of the uniform distribution of the n nodes in Gupta-Kumar's model, it follows that each node $1, 2, \dots, k$ is transited on average by $\sqrt{n/\log n}$ flows; this is the same as the average number of hops between two independently chosen nodes. Combining this with the result from Eq. (26), it follows that the throughput of the multi-hop flow from Figure 1 scales according to $\Omega(1/\sqrt{n \log n})$. We point out that the recovery of this result was possible by fixing the pair of nodes, given the simplifying network scenario we treated in this section.

In principle, one may extend the analysis from this section to asymmetric scenarios which would result in less concise expressions on end-to-end throughputs and delays.

V. NUMERICAL RESULTS

We first validated the two density models from Subsections II-B1 and II-B2 by using two experimental datasets gathered by the Huggle Project, referred to as *Infocom05* and *Infocom06* [19]; these are the most densely connected empirical datasets publicly available to the research community. In *Infocom05*, 41 iMotes devices were distributed to approximately 40 students attending the Infocom 2005 student workshop and in *Infocom06* the number of participants increased to 98. The number of internal contacts for the two datasets are 22459 and 191336, respectively, whereas the average number of contacts/pair/day are 4.6 and 6.7, respectively. Previous works have used these traces to extract important characteristics of human inter-contact times [8], [22]. For several randomly selected nodes we plotted the distribution of uninterrupted time spent by $N(t)$ (the number of neighbors at time t) in several states. For both datasets we observed a geometric distribution which justifies the Markov model from Subsection II-B1. Also, we observed a geometric distribution in time for $N(t)$, which suggests that the binomial distribution, arising for the commonly used uniform node placement distribution may not be suitable for modelling networks with realistic mobility settings (note: the plots are not shown here for lack of space).

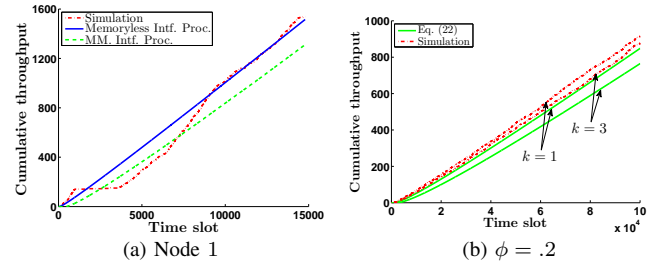


Fig. 2. In (a): cumulative throughput of Node 1 from *Infocom05* in an emulated mobile scenario, and with fitted Markov-modulated and memoryless interfering processes. In (b): cumulative throughput for node 1 in Figure 1 with $m = 10$ (violation probability $\epsilon = 10^{-3}$)

Figure 2.(a) illustrates the accuracy of the non-asymptotic throughput bounds from Theorem 1 by using simulation results from the *Infocom05* dataset. In this dataset, 41 nodes collected contact data among themselves for four consecutive days with time granularity 120 s. We interpolated the neighboring information with time slot 10 s, such that the difference between the number of neighbors in two consecutive slots is at most one. Then we fitted the Markov-modulated interfering process from Subsection II-B1 and computed its spectral radius; also, from the stationary distribution of the number of neighbors we fitted the memoryless interfering process from Subsection II-B2. The figure shows the cumulative throughput quantiles ($\epsilon = 10^{-3}$) of node 1 for the entire trace duration by simulating slotted-Aloha for the emulated mobility trace, and

also the bound from Eq. (10) for the two fitted interfering processes. We observe that the bound obtained from the memoryless process is reasonably accurate; the other bound, obtained from the Markov-modulated process, underestimates the throughput because of the spectral radius upper bound approximation from Eq. (7).

Figure 2.(b) illustrates the behavior of the multi-hop throughput from Eq. (22) for an overlap fractions $\phi = .2$ and two numbers of hops k . We plot the cumulative throughput as a function of time using Eq. (22), and by simulating slotted-Aloha, for $\varepsilon = 10^{-3}$. We observe that the time scale used (10^5 slots) is insufficient to capture the independence of the multi-hop throughput of the number of hops, as asymptotically predicted in Eq. (26); this shows the limitations of asymptotic results for throughput predictions over finite time scales. We also observed (plots not shown here) that higher ϕ yields larger throughput; thus, the number of nodes contending for a channel has a slightly higher impact on the throughput than the number of nodes correlating adjacent transmissions.

VI. CONCLUSIONS

In this paper we have opened a new perspective on the performance analysis of multi-hop wireless networks. Concretely, we have provided explicit bounds on throughput and delay distributions, which hold not only asymptotically, as existing results from the literature, but also account for finite network sizes, time scales, random topologies, and broad classes of arrivals. The obtained results can be immediately applied to further understand the capacity and delay behavior in multi-hop wireless networks, e.g., network capacity is fundamentally influenced by randomness in network topology. In the future we plan to use our methodology to evaluate new communication protocols, which, for instance, use different backoff distributions than the exponential at the MAC layer and various scheduling algorithms at the network layer. By making such intrinsically complex and difficult problems analytically tractable, we conclude that this paper brings a significant contribution to connecting fundamental research and practice in the field of multi-hop wireless networks.

REFERENCES

- [1] N. Abramson. The Aloha system: another alternative for computer communications. In *Proceedings of AFIPS Joint Computer Conferences*, pages 281–285, 1970.
- [2] N. Abramson. The throughput of packet broadcasting channels. *IEEE Transactions on Communications*, 25(1):117–128, 1977.
- [3] I. Akyildiz and X. Wang. *Wireless Mesh Networks*. Wiley, 2009.
- [4] J. G. Andrews, N. Jindal, M. Haenggi, R. Berry, S. Jafar, D. Guo, S. Shakkottai, R. Heath, M. Neely, S. Weber, and A. Yener. Rethinking Information Theory for Mobile Ad Hoc Networks. *IEEE Communications Magazine*, 46(12):94–101, 2008.
- [5] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.
- [6] C. Bordenave, D. McDonald, and A. Proutiere. Performance of random medium access control, an asymptotic approach. In *ACM Sigmetrics*, pages 1–12, 2008.
- [7] J. Y. L. Boudec and P. Thiran. *Network Calculus*. Springer Verlag, Lecture Notes in Computer Science, LNCS 2050, 2001.
- [8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606–620, 2007.
- [9] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer Verlag, 2000.
- [10] F. Ciucu. Network calculus delay bounds in queueing networks with exact solutions. In *International Teletraffic Congress (ITC)*, pages 495–506, 2007.
- [11] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. *Journal of Applied Probability*, 33(3):886–903, Sept. 1996.
- [12] R. Cruz. A calculus for network delay, parts I and II. *IEEE Transactions on Information Theory*, 37(1):114–141, Jan. 1991.
- [13] A. Ephremides and R.-Z. Zhu. Delay analysis of interacting queues with an approximate model. *IEEE Transactions on Communications*, 35(2):194–201, 1987.
- [14] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *IEEE International Workshop on Quality of Service (IWQoS)*, pages 261–270, 2006.
- [15] R. G. Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, 31(2):124–142, Mar. 1985.
- [16] Y. Gao, D.-M. Chiu, and J. C. S. Lui. Determining the end-to-end throughput capacity in multi-hop networks: methodology and applications. In *ACM Sigmetrics/Performance*, pages 39–50, June 2006.
- [17] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- [18] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, Mar. 2000.
- [19] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *ACM Mobihoc*, pages 241–250, 2008.
- [20] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. In *ACM Mobicom*, pages 66–80, 2003.
- [21] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.
- [22] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *ACM Mobicom*, pages 183–194, 2007.
- [23] L. Kleinrock. *Queueing Systems*, volume 1. John Wiley and Sons, 1975.
- [24] L. Kleinrock and J. Silvester. Optimum transmission radii for packet radio networks or why six is a magic number. In *Proceedings of IEEE National Telecommunication Conference*, pages 4.3.1–4.3.5, 1978.
- [25] L. Kleinrock and F. A. Tobagi. Packet switching in radio channels: Part I—carrier sense multiple-access modes and their throughput-delay characteristics. *IEEE Transactions on Communications*, 23(12):1400–1416, 1975.
- [26] M. Musolesi and C. Mascolo. Mobility models for systems evaluation. A survey. In B. Garbinato, H. Miranda, and L. Rodrigues, editors, *Middleware for Network Eccentric and Mobile Applications*, pages 43–62. Springer, 2009.
- [27] C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Measurement-based models of delivery and interference in static wireless networks. In *ACM Sigcomm*, pages 51–62, 2006.
- [28] L. G. Roberts. Aloha packet system with and without slots and capture. *SIGCOMM Computer Communication Review*, 5(2):28–42, 1975.
- [29] M. Sidi and A. Segall. Two interfering queues in packet-radio networks. *IEEE Transactions on Communications*, 31(1):123–129, 1983.
- [30] W. Szpankowski. Stability conditions for some distributed systems: Buffered random access systems. *Advanced in Applied Probability*, 26(2):498–515, 1994.
- [31] H. Takagi and L. Kleinrock. Mean packet queueing delay in a buffered two-user CSMA/CD system. *IEEE Transactions on Communications*, 33(10):1136–1139, 1985.
- [32] F. A. Tobagi. Distributions of packet delay and interdeparture time in slotted Aloha and carrier sense multiple access. *Journal of the ACM*, 29(4):907–927, 1982.
- [33] B. Tsybakov and V. Mikhailov. Ergodicity of slotted Aloha system. *Problemy Peredachi Informatsii*, 4(15):73–87, 1979.
- [34] M. Xie and M. Haenggi. Towards an end-to-end delay analysis of wireless multihop networks. *Ad Hoc Networks*, 7(5):849–861, 2009.
- [35] Y. Yang and T.-S. P. Yum. Delay distributions of slotted Aloha and CSMA. *IEEE Transactions on Communications*, 51(11):1846–1857, 2003.