This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

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

Jingjing Luo†, Li Yu†, Jinbei Zhang‡, Xinbing Wang‡
†School of Electronic Information and Communications, Huazhong University of Science and Technology, China
‡Dept. of Electronic Engineering, Shanghai Jiao Tong University, China
Email: †{luojingjing, hustlyu}@hust.edu.cn, ‡{abelchina, xwang8}@sjtu.edu.cn

***Abstract*—Previous works on multicast capacity mainly focus on deriving asymptotic order results in large-scale wireless networks. While they can explore the general scaling laws of throughput capacity, it is also of great interest for practical concern to predict the exact achievable throughput in networks with an arbitrary finite number of nodes. In this paper, we investigate the non-asymptotic throughput and delay of multihop wireless networks for multicast applications wherein for each source node, k nodes are randomly selected as receivers. It is challenging for the exact performance analysis since multicast transmission has a dynamic nature, due to the following factors: 1) random distribution of nodes, 2) bursty traffic arrivals, and 3) different time scales for transient analysis. To tackle the problem, we propose an explicit analytical model and develop a multicast routing scheme, which accounts for the above aspects. With our proposed model, we derive lower and upper bounds on non-asymptotic multicast throughput and delay using stochastic network calculus. We show that the performance results hold for all time scales and network sizes, and are strongly correlated to data burstiness and the number of receivers. While we investigate from an non-asymptotic point of view, our results can also cover the asymptotic scaling laws. Simulations are conducted to further verify the accuracy of the analytical bounds.**

I. Introduction

There has been a growing interest to understand the fundamental capacity limits of wireless networks, since the seminal work of Gupta and Kumar [1]. In [1], they studied the asymptotic unicast capacity of wireless ad hoc networks consisting of n nodes and showed that the scalability of wireless multihop routing is limited. Specifically, the per-node throughput capacity decays as $\Omega\left(1/\sqrt{n\log n}\right)$[1] in random networks while $O\left(1/\sqrt{n}\right)$ in arbitrary networks. Since then, a large variety of studies [2]–[6] have been conducted trying to improve the network capacity. These works mainly focus on unicast transmissions in ad hoc networks. Another line of research is devoted to multicast traffic pattern since it is widely adopted in many practical applications. In multicast routing, packets originating from a source have to be delivered to a set of destination nodes. One advantage of multicast routing is that it reduces the total bandwidth required to communicate with all destinations, since some links can be shared among different destinations.

Under various settings, the throughput capacity with regard to multicast flows is analyzed in [7]–[13]. In [7], Jacquet and Rodolakis investigated the scaling properties of multicast capacity for random wireless networks. They showed that a node can transmit packets to its k receivers with a maximum rate of $O\left(\frac{1}{\sqrt{kn\log n}}\right)$. In [8], Shakkottai *et al*. revealed that by employing a novel routing architecture to transfer multicast data, named the multicast comb, it is possible to achieve a per-flow throughput of $O\left(\frac{1}{\sqrt{n^{\in}\log n}}\right)$ when the number of multicast sources is $n^{\in}$. In [9], Li provided a more detailed study on the multicast scaling laws for large-scale networks. In particular, it concentrated on the asymptotic multicast capacity of a static random wireless network and assumed that packets are sent from node to node in a multi-hop manner until they reach all their destinations. Li showed that the per-flow throughput of n multicast sessions is $\Theta\left(\sqrt{\frac{1}{n\log n}}\cdot\frac{1}{\sqrt{k}}\right)$ when $k=O\left(\frac{n}{\log n}\right)$, and is $\Theta\left(\frac{1}{n}\right)$ when $k=\Omega\left(\frac{n}{\log n}\right)$. Following this line, it is later proved that a per-flow throughput of $O\left(\frac{1}{\sqrt{kn\log n}}\right)$ can also be achieved under other scenarios, like in [12], [13].

It is notable that all the works mentioned above mainly concentrate on dealing with the order results of multicast throughput in large-scale networks. These results are obtained under a common assumption that both the time and the network size are infinite. Although such results provide a good understanding of the general scaling laws and thus the growth rate of the throughput with respect to network size n and the number of receivers k, they cannot completely capture the features of small or medium sized networks. This motivates us to investigate the non-asymptotic multicast throughput, which can hold for any finite network size and time scale. Besides, in these works, they all assume that the transmissions are in steady state and the buffer is large enough so that packets will not be dropped by any intermediate node, which is not feasible especially in scenarios with bursty arrivals. Moreover, delay is roughly inversely proportional to per-node throughput in previous scaling results of static networks.

[1]Given two functions $f(n)\geq 0$ and $g(n)\geq 0$: $f(n)=O(g(n))$ means that $\limsup_{n\to\infty}\frac{f(n)}{g(n)}=c<\infty$; $f(n)=\Omega(g(n))$ is equivalent to $g(n)=O(f(n))$; $f(n)=\Theta(g(n))$ means that $f(n)=O(g(n))$ and $g(n)=O(f(n))$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

The time spent in each hop is assumed to be constant and therefore out of specific consideration in asymptotic analysis. However, for exact analysis, delay is a random process which is correlated to many factors, such as data burstiness and routing strategy. These factors make the analysis non-trivial and the investigation into the impact of these factors critically important. Therefore, non-asymptotic performance is receiving growing interest in the literature, e.g., [15]–[19].

In [15]–[17], the authors investigated the non-asymptotic performance guarantees for single-hop scenarios with regard to specific arrivals. Later on, Ciucu *et al*. [18], [19] explored the non-asymptotic throughput of multi-hop wireless networks for unicast transmissions. Specifically, they considered a fixed multihop path in a line network, where data burstiness may come into the source node and incur some randomness for the transmission process. Explicit upper and lower bounds are obtained on the end-to-end throughput, and it is shown that the behavior of these bounds is strongly related to the degree of data burstiness and the number of hops. The main tool that enables the non-asymptotic analysis in these works is stochastic network calculus [20]–[22]. In the framework of stochastic network calculus, traffics are described by envelope functions, which provide upper bounds on traffics over time intervals, and service is characterized by service curves, which give lower bounds on the service available to a flow. These bounds are permitted to be violated with a small probability. The main attractive features of network calculus enhancing the performance analysis are as follows: 1) a stochastic service curve can model the amount of data that a node can send over a single hop, independent of the arrival process; 2) the end-to-end service for a multi-hop path can be derived with the convolution of each node's service curve.

Observed that multicast has a great impact over multiple unicast scenarios for scaling results [7]–[9], we are thus motivated to investigate its non-asymptotic performance in this paper. We consider a two-dimensional network, where the source has multiple receivers. In such a scenario, there may be many options for routing paths' scheduling and the routing scheme should be carefully designed. Moreover, the link may be shared by different receivers, making traffic processes coupled. Therefore, it is difficult to directly employ previous approaches (such as [18], [19], [27]) for our analysis. Our previous work [14] provides one first step towards this problem. In particular, we construct a novel multicast tree from CDS (connected dominating set) for multicast routing, and then model the interference of nodes, which depends on node degree in the tree. By using stochastic network calculus, we derive the exact achievable multicast throughput. In this paper, we further study the non-asymptotic multicast delay based on the proposed explicit analytical model. Practical factors make the exact delay analysis difficult even in unicast scenarios. When it comes to multicast scenarios, the route is dynamic due to randomly distributed source and multiple receiver nodes, which makes the analysis much more challenging. We define multicast delay as the time required to deliver packets from S to all receivers, which is equivalent to the maximal time spent by one S-D pair among all k S-D pairs. While the longest S-D path in a multicast tree incurs the maximum expected delay, we calculate the delay by analyzing its longest S-D path, accounting for data burstiness at the source. To model the data burstiness at the source node, we consider the class of EBB (exponentially bounded burstiness) processes [23]. The effects of data burstiness on throughput and delay are investigated via both theoretical analysis and simulations in this paper. Besides the non-asymptotic performance analysis, we also extend our analytical model to asymptotic regimes. In summary, our main contributions are as follows:

- We first investigate the non-asymptotic performance of multi-hop wireless networks for multicast route, which is dynamic due to the randomly distributed source and receivers, rather than a fixed multi-hop path in a line network. We develop an analytical model for multicast routing in which stochastic network calculus can be employed for the analysis. Furthermore, instead of assuming infinite buffer, we present a sufficient condition in terms of available service rate, to ensure transmission stability.
- The exact end-to-end delay for multicast transmissions is first derived. Different with asymptotic analysis, arrival burstiness on the queueing behavior at each relay node comes into play. To deal with it, each relay node is modeled as a queue which is fed by EBB arrivals and its service is described by a statistical service curve. With the closed-form results, the impact of the degree of burstiness on the end-to-end delay can be characterized. Moreover, our results can also cover the scaling laws in asymptotic regimes.
- While interference of each transmitting node is assumed to be identical in the analysis of unicast transmissions [18], we describe the interference among simultaneous transmissions according to the degree of the transmitting node in the constructed multicast tree. This is because the interference is affected by overlapping transmissions and the number of receivers of the transmitting node. Moreover, compared with the well-known protocol model in [1], our interference model can measure each node's interference more precisely in the proposed scheme.

The rest of this paper is organized as follows. In section II, we provide definitions and network model along with the multicast scheme and interference model. In section III, we introduce the main analytical tool, interfering process, for performance analysis. By using such a tool, we derive the upper and lower bounds on non-asymptotic multicast throughput and delay, based on stochastic network calculus. Section IV presents brief simulation results to validate it. Finally we conclude the paper in Section V.

II. Network Model and Definitions

A. Cell-Partitioned Network Model

We consider a static random network as illustrated in Figure 1. n nodes (indexed as $\{w_1, w_2, ..., w_n\}$) are randomly and uniformly distributed on a unit square. The transmission range

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

of each node is r and omnidirectional antennas are adopted. For any two nodes w_i and w_j, they can communicate directly if $\|w_i - w_j\| \leq r$ and there is no other interference. The unit square is divided into nonoverlapping square cells, each with side length $\frac{\sqrt{2}r}{2}$.

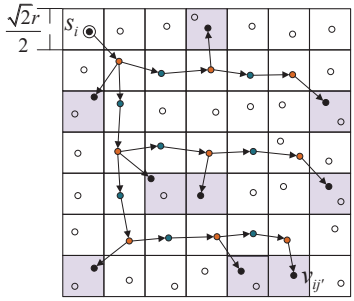


Fig. 1: Cell-partitioned network model. The black, orange and green solid nodes represent receiver nodes, dominating relay nodes and assistant relay nodes, respectively. Shaded squarelets are cells with at least one receiver node.

n nodes form a set $\mathcal{W} = \{w_1, w_2, ..., w_n\}$. Assume that a subset $\mathcal{S} \subseteq \mathcal{W}$ of $n_{\mathcal{S}} = |\mathcal{S}|$ nodes will serve as the source nodes of $n_{\mathcal{S}}$ multicast sessions. Given a source node s_i, we randomly select a set $\mathcal{V}_i = \{v_{i1}, v_{i2}, ..., v_{ik}\}$ ($\mathcal{V} \subseteq \mathcal{W}$) of k nodes as its destinations. We define such a relationship as a multicast session $\mathcal{M}_i = \{s_i\} \cup \mathcal{V}_i$. To make a successful transmission, a source node has to deliver packets to all its k destinations. A source node transmits packets to a destination node by multiple intermediate relay nodes, which are divided into two types: dominating relay nodes and assistant relay nodes. A set of dominating relay nodes is a subset $\mathcal{W}' \subset \mathcal{W}$ such that each node in $\mathcal{W}/\mathcal{W}'$ is adjacent to some node in $\mathcal{W}'$. We say that a transmitting node can *cover* a node within its transmission range. Therefore, the dominating relay node set $\mathcal{W}'$ can cover all receiver nodes. A connected dominating set (CDS) is a dominating set which induces a connected graph [24]. For any two neighboring dominating relay nodes, there will be at least one node between them and randomly choose one such node (named assistant relay node) to connect them. For simplicity, we assume that each receiver node is covered by a dominating relay node, as shown in Figure 1. Each relay node receives data either from the source node or its previous relay node. In this paper, we analyze the multicast performance based on the multicast tree spanning s_i and $\mathcal{V}_i$ as depicted in Figure 1, in the presence of interference from other nodes' transmissions. The details of multicast tree construction are presented in the next section.

B. Multicast Tree Construction

We propose a multicast scheme for the network model, as shown in Figure 3. Specifically, it constructs a multicast tree from CDS, spanning the source and all receivers in a multicast session.

Consider a multicast session from Figure 1, which includes a randomly selected source node s_i and a set of receiver nodes $\mathcal{V}_i = \{v_{i1}, v_{i2}, ..., v_{ik}\}$. The method to construct a multicast tree spanning s_i and $\mathcal{V}_i$ is illustrated in Figure 2.

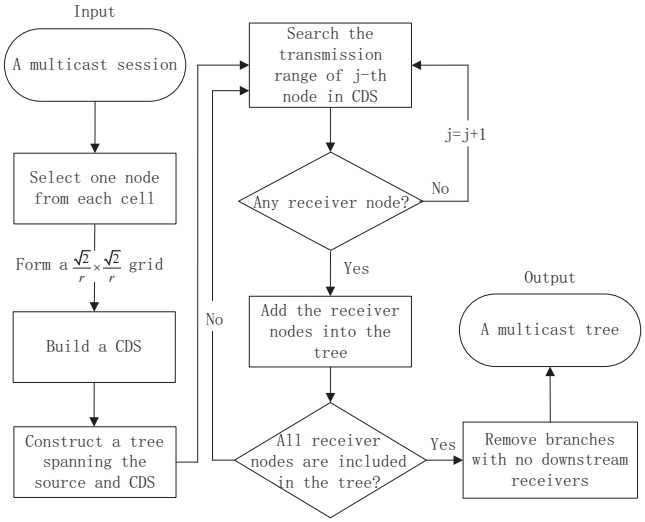


Fig. 2: A flow chart of the multicast tree construction.

First, for each cell with at least one node, we select one node according to the following rules: If there exists a source node, select it. If there exists a receiver node[2], select it. Otherwise, select one node at random. The set of selected nodes will form a $\frac{\sqrt{2}}{r} \times \frac{\sqrt{2}}{r}$ grid approximatively, as shown in Figure 3(a). Here we use (p_i, p_j) to denote the position of node in the p_i-th row and p_j-th column, where $1 \leq p_i, p_j \leq \frac{\sqrt{2}}{r}$.

Second, in the grid, we build a CDS to search the receiver nodes $\mathcal{V}_i = \{v_{i1}, v_{i2}, ..., v_{ik}\}$, using the following method: choose all the nodes located at $(2p'_i, 2p'_j)$, where $p'_i, p'_j > 0$ are integers and $1 \leq 2p'_i, 2p'_j \leq \frac{\sqrt{2}}{r}$. The set of selected nodes $\mathcal{U}_i = \left\{u_{i1}, u_{i2}, ..., u_{i\frac{1}{2r^2}}\right\}$ can cover all the other nodes in the network and form a CDS, which are represented by solid square nodes as depicted in Figure 3(b).

Third, we use breadth-first-search algorithm to construct a tree $T^*(\mathcal{U}_i)$, which is rooted at the source node s_i and connects nodes in $\mathcal{U}_i$. Obviously, for each link $u_{ij'}u_{ij''}$ composed of two points in $\mathcal{U}_i$, the distance between these two points is exactly two hops. Then connect the two points via an intermediate node[3] which is crossed by line $u_{ij'}u_{ij''}$, as illustrated in Figure 3(c). The resulted structure will serve as the routing guideline for the multicast session.

Fourth, according to the index of nodes in CDS $\mathcal{U}_i = \left\{u_{i1}, u_{i2}, ..., u_{i\frac{1}{2r^2}}\right\}$, we search the receiver nodes in the transmission range of each node in CDS[4] and add these receiver nodes into tree $T^*(\mathcal{U}_i)$ until all receiver nodes are included in the tree, as shown in Figure 3(d). Then remove branches with no downstream receivers from $T^*(\mathcal{U}_i)$. The final

[2]Note that there may exist several receiver nodes within a cell. In this case, regard these receiver nodes as a virtual node.

[3]If the intermediate node happens to be a receiver node, it also acts as the relay node for multicast route.

[4]The receiver nodes included in CDS can also act as routing guideline for searching other receivers.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

multicast tree is denoted as $T(\mathcal{V}_i)$.

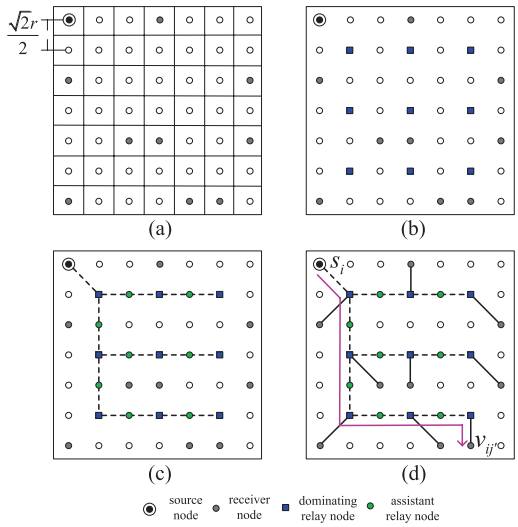


Fig. 3: A multicast tree construction with grid layout.

C. Interference Model

There are several interference models employed to describe the interference in wireless networks. In this paper, we model the interference among simultaneous transmissions, by calculating the number of interfering nodes for each link [25]. Two types of interfering nodes are considered, namely, neighboring interfering nodes and hidden interfering nodes. For example, in Figure 4, node 3 and node 5 are the neighboring interfering nodes for node 1. Node 7, node 8 and node 9 are hidden interfering nodes of node 1.

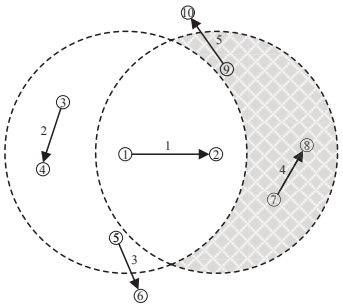


Fig. 4: Interfering node problems.

When it comes to multicast scenario, the interference of each link largely depends on the multicast tree structure, which makes the analysis more difficult. Consider the network scenario in Figure 1. The multicast flow is from the source node s_i to the destination nodes $\mathcal{V}_i = \{v_{i1}, v_{i2}, ..., v_{ik}\}$. A transmitting node is said to be with degree l if it has l receiving nodes within its transmission range. It has been proved that when CDS is constructed, the degree of each node in CDS is bounded by a constant [24]. Different methods of constructing CDS will lead to different degree bounds. It is straightforward to find that each node in CDS has a degree l bounded by 3 if the method presented in Figure 3 is used. In particular, we build a CDS on top of a grid containing the source, all receivers and other nodes. When the multicast tree is constructed, each node in CDS are connected with at most 3 child nodes. For a flow going through a transmitter to its receivers, the interfering nodes include not only the neighboring interfering nodes of the transmitter, but also the hidden interfering nodes relative to all the receivers. The number of interfering nodes U_{lx} is thus determined by the product of interfering area and nodes' density. The interfering area varies as transmitting node's degree l and the relative positions of transmitting node and receiving nodes, which is illustrated in Figure 5 and classified as follows.

i) when $l = 1$, there are two types as in Figure 5(a)-(b);
ii) when $l = 2$, there are four types as in Figure 5(c)-(f);
iii) when $l = 3$, there are four types as in Figure 5(g)-(j).

The dotted circles in Figure 5 represent the sensing range of nodes. Suppose that each node has a constant sensing range r_s and it can sense the transmissions of any node falling within that range. For simplicity, the sensing range is assumed to be identical to the transmission range. We suppose that each circle contains m nodes and the shaded area 1, 2, 3, 4 in Figure 5 contain $\psi_1 m$, $\psi_2 m$, $\psi_3 m$, $\psi_4 m$ $(0 < \psi_1, \psi_2, \psi_3, \psi_4 < 1)$ nodes, respectively. Then the number of interfering nodes U_{lx} for a transmitting node with degree l can be summarized as follows:

Case 1: $l = 1$

$$U_{11} = m + (1 - \psi_1)\, m = (2 - \psi_1)\, m,$$
$$U_{12} = m + (1 - \psi_2)\, m = (2 - \psi_2)\, m.$$

Case 2: $l = 2$

$$U_{21} = m + 2\,(1 - \psi_2)\, m - \psi_3 m = (3 - 2\psi_2 - \psi_3)\, m,$$
$$U_{22} = m + 2\,(1 - \psi_2)\, m = (3 - 2\psi_2)\, m,$$
$$U_{23} = m + (1 - \psi_1)\, m + (1 - \psi_2)\, m - \psi_4 m = (3 - \psi_1 - \psi_2 - \psi_4)\, m,$$
$$U_{24} = m + (1 - \psi_1)\, m + (1 - \psi_2)\, m = (3 - \psi_1 - \psi_2)\, m.$$

Case 3: $l = 3$

$$U_{31} = m + 3\,(1 - \psi_2)\, m - 2\psi_3 m = (4 - 3\psi_2 - 2\psi_3)\, m,$$
$$U_{32} = m + (1 - \psi_1)\, m + 2\,(1 - \psi_2)\, m - 2\psi_4 m = (4 - \psi_1 - 2\psi_2 - 2\psi_4)\, m,$$
$$U_{33} = m + (1 - \psi_1)\, m + 2\,(1 - \psi_2)\, m - \psi_3 m = (4 - \psi_1 - 2\psi_2 - \psi_3)\, m,$$
$$U_{34} = m + (1 - \psi_1)\, m + 2\,(1 - \psi_2)\, m - \psi_3 m - \psi_4 m = (4 - \psi_1 - 2\psi_2 - \psi_3 - \psi_4)\, m.$$

D. Traffic Pattern

We employ a continuous time model to describe the arrival and departure processes. In particular, the cumulative arrival and departure processes in a time interval $[0, t]$ are denoted by left-continuous processes $A(t)$ and $D(t)$, respectively. By convention, $A(0) = D(0) = 0$. Then the corresponding bivariate processes of arrivals and departures in the time interval $(s, t]$ are represented by $A(s, t) := A(t) - A(s)$ and $D(s, t) := D(t) - D(s)$.

We consider bursty data sources. It is widely known that the class of EBB processes can well characterize the traffic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

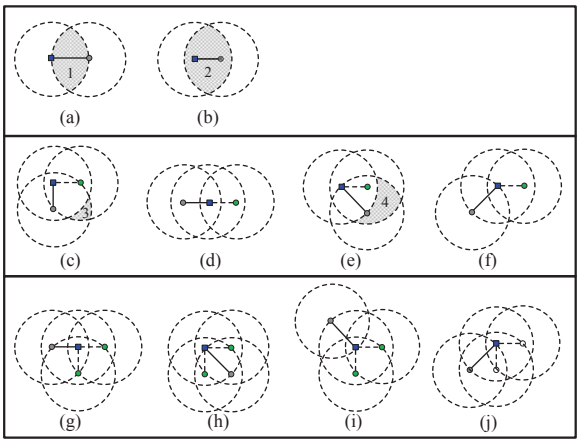


Fig. 5: Interference model.

burstiness. The EBB class includes a rich category of traffic patterns, e.g., Markov-modulated On-Off processes, Poisson process and multiplexed regulated traffic.

EBB is defined and modeled in [23]. According to the definition, an arrival process $A(t)$ has an exponentially bounded burstiness (EBB) if it satisfies for all $\sigma \geq 0$ and $0 \leq s \leq t$

$$\mathrm{P}\left(A(t) - A(s) > r(t-s) + \sigma\right) \leq Me^{-\theta\sigma} \tag{1}$$

where $r > 0$ is a bound on the arrival rate, the term $\sigma > 0$ allows for some burstiness, $\theta > 0$ is the decay rate of the violation probability function, and $M > 0$ is a constant. It implies that the probability of violating the arrival bound during a given time interval decays exponentially.

Based on the pioneering stochastic $(\sigma(\theta), r(\theta))$-calculus [15], [26], we establish the envelopes of moment generating functions (MGF) of EBB arrival processes. In particular, for any $\theta > 0$, there exists an upper rate r^U that for all $s \leq t$, the moment generating function of the EBB processes is bounded by

$$E\left[e^{\theta A(s,t)}\right] \leq e^{\theta r^U (t-s)}. \tag{2}$$

We also assume that there exists a lower rate r^L that for all $s \leq t$, the Laplace transform of the EBB processes satisfies

$$E\left[e^{-\theta A(s,t)}\right] \leq e^{-\theta r^L (t-s)}. \tag{3}$$

By using Jensen's inequality, we can further bound the average rate of the arrival process $A(t)$ that for any time scale $t \geq 0$,

$$r^L \leq \frac{E\left[A(t)\right]}{t} \leq r^U. \tag{4}$$

E. Definitions

Let us first define the convolution operator $*$ for two doubly indexed processes f and g as:

$$(f*g)(t) = \begin{cases} \inf\limits_{0\leq s\leq t}\{f(t-s)+g(s)\}, & t>0 \\ 0, & t\leq 0 \end{cases}. \tag{5}$$

Statistical Service Curve: A nonnegative, doubly indexed random process $S(s,t)$ is a statistical service curve for an arrival process $A(t)$ if for all $t \geq 0$ and $\sigma \geq 0$

$$\mathrm{P}\left(D(t) < A * [S-\sigma]_+(t)\right) \leq \omega(\sigma) \tag{6}$$

where $[x]_+ = \max(x, 0)$ and $\omega(\sigma)$ is a non-increasing violation probability function. The function satisfies the integrability condition:

$$\int_0^\infty \omega(x)dx < \infty. \tag{7}$$

Non-Asymptotic Per-Flow Multicast Throughput: Given a multicast session $\mathcal{M}_i = \{s_i\} \cup \mathcal{V}_i$, let $\Lambda_i = \left(\lambda_t^{i1}, \lambda_t^{i2}, ..., \lambda_t^{ik}\right)$ denote a rate vector of the throughput rates of all receiver nodes at time t. For a rate vector, we define the per-flow throughput of session $\mathcal{M}_i$ as $\min_{v_{ij}\in\mathcal{V}_i}\lambda_t^{ij}$ $(j = 1, 2, ..., k)$, which is the minimum departure throughput rate at which multicast packets can be received by all k destination nodes successfully. Denote $D_i^{\min}(t)$ as the departure process of the receiver node with minimum throughput rate. For $D_i^{\min}(t)$, there exists corresponding violation probabilities γ_C^L and γ_C^U such that

$$\begin{aligned} &\mathrm{P}\left(D_i^{\min}(t) \leq \lambda_t^{iL} t\right) \leq \gamma_C^L, \\ &\mathrm{P}\left(D_i^{\min}(t) \geq \lambda_t^{iU} t\right) \leq \gamma_C^U. \end{aligned} \tag{8}$$

Here λ_t^{iL} and λ_t^{iU} are the lower and upper bounds of non-asymptotic per-flow throughput for multicast session $\mathcal{M}_i$. For briefness, they are written by λ_t^L and λ_t^U thereinafter, respectively.

Therefore, the non-asymptotic per-session multicast throughput is defined as the aggregated feasible per-flow multicast throughput $\sum_{v_{ij}\in\mathcal{V}_i}\lambda_t^{ij}$. Its lower bound equals $k\lambda_t^L$, and similarly, its upper bound is $k\lambda_t^U$.

Non-Asymptotic Multicast Delay: For a multicast session $\mathcal{M}_i$, we describe the delay process as $W_{ij}(t) = \inf\{d_{ij} \geq 0 : A_i(t-d_{ij}) \leq D_{ij}(t)\}$ $(j = 1, 2, ..., k)$. $W_{ij}(t)$ represents the delay process at node v_{ij} at time t. Let $\mathcal{D}_i = \{d_{i1}, d_{i2}, ..., d_{ik}\}$ denote the corresponding delay vector, then the per-flow delay of session $\mathcal{M}_i$ is defined as $\max_{v_{ij}\in\mathcal{V}_i} d_{ij}$, which is the time required to deliver all packets to k receiver nodes by time t successfully. For delay process $W_i^{\max}(t)$, there exists corresponding violation probabilities γ_D^L and γ_D^U such that

$$\begin{aligned} &\mathrm{P}\left(W_i^{\max}(t) \leq d_L\right) \leq \gamma_D^L, \\ &\mathrm{P}\left(W_i^{\max}(t) \geq d_U\right) \leq \gamma_D^U. \end{aligned} \tag{9}$$

Here d_L and d_U are the lower and upper bounds of non-asymptotic multicast delay for session $\mathcal{M}_i$, respectively.

III. Non-Asymptotic Multicast Throughput of Multi-Hop Networks

A. Interfering Process

In this section, we define the interfering process for multicast transmissions. Interfering process was previously introduced in [27] for unicast transmissions. By taking into account the impact of multicast traffic, we derive node's service curve

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

which provides a lower bound on the service available to a flow.

Assume that all the nodes run the slotted-Aloha protocol with time unit τ_0. A transmitting node can be regarded as a node with degree l if it has l receiver nodes.

Definition of Interfering Process: For a transmitting node Q with degree l, denote $I_l(t)$ as its interfering process. It can be represented by its Bernoulli increment process $I_l(t, t+\tau_0) := I_l(t+\tau_0) - I_l(t)$,

$$I_l\left(t, t+\tau_0\right) = \begin{cases} 0, \text{ if Q successfully transmits to} \\ \quad l \text{ receiver nodes during } (t, t+\tau_0] \\ \tau_0, \text{ otherwise} \end{cases}. \tag{10}$$

The probabilities are:

$$\begin{cases} \mathrm{P}\left(I_l(t, t+\tau_0) = 0\right) = 1 - P_{U_l} \\ \mathrm{P}\left(I_l(t, t+\tau_0) = \tau_0\right) = P_{U_l} \end{cases} \tag{11}$$

where $P_{U_l} := 1 - p(1-p)^{U_l - 1}$, which implies that during time interval $(t, t+\tau_0]$, the probability that the sender cannot successfully transmit to all l receiver nodes is P_{U_l}. p is the probability that the sender gets a transmission opportunity and U_l is the average number of interfering nodes for a transmitting node with degree l.

Obviously, the interfering process $I_l(t)$ is modulated by U_l depending on l. From the previous analysis, we find that each node of the multicast session in Figure 1 has a degree l bounded by 3. Based on the proposed interference model, we can obtain U_l by averaging U_{lx} for any given l ($l = 1, 2, 3$),

$$\begin{aligned} U_1 &= \tfrac{1}{2}\sum_{x=1}^{2} U_{1x} = (2 - \tfrac{\psi_1+\psi_2}{2})m, \\ U_2 &= \tfrac{1}{4}\sum_{x=1}^{4} U_{2x} = (3 - \tfrac{2\psi_1+6\psi_2+\psi_3+\psi_4}{4})m, \\ U_3 &= \tfrac{1}{6}(2U_{31} + U_{32} + U_{33} + 2U_{34}) \\ &= (4 - \tfrac{4\psi_1+14\psi_2+7\psi_3+10\psi_4}{6})m. \end{aligned} \tag{12}$$

The cumulative interfering process $I_l(t)$ can be written as $I_l(t) = \sum_{s=0}^{\lceil t/\tau_0 \rceil - 1} I_l(s, s+\tau_0)$, and accounts for the amount of time when the sender cannot successfully transmit. From the expression of $I_l(t)$, we can observe that the interfering process depends on the degree of the transmitting node and the corresponding independent Bernoulli random variables.

For $\theta \neq 0$, the MGF of $I_l(t)$ is given by,

$$E[e^{\theta I_l(t)}] = e^{\theta \frac{\log q_l}{\theta} t} \tag{13}$$

where $q_l = 1 + P_{U_l}\left(e^{\theta} - 1\right)$. The term $\frac{\log q_l}{\theta}$ is the relative rate of $I_l(t)$.

B. Single-Hop Case

We next present the exact throughput bounds of a single-hop multicast session as shown in Figure 6. Node Q transmits packets to k receiver nodes (i.e., R_1 to R_k) simultaneously for each time slot $(t, t+\tau_0]$. We suppose that only the multicast flow $[\mathrm{Q} \to \mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R}_k]$ is served while the rest of nodes play only the role of interfering. Node Q is fed by EBB arrivals and its service is described by a statistical service curve. Compared with the single-hop unicast analysis, the difficulty of multicast performance analysis is mainly caused by the interference of multiple receivers rather than a single receiver and the coupled traffic. In the following, we will show that the leftover service curve can lead to a clean solution for the multicast analysis, combined with our interference model.

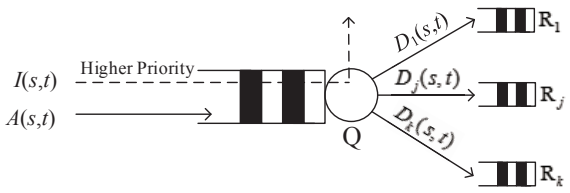


Fig. 6: Illustration of Q's service.

Leftover service curve can model the service available to a tagged flow at a node with a constant service rate, where there exists interfering flows. In order to construct a leftover service curve $S(s,t)$ [21] for node Q, we describe Q's service in the following way (see Figure 6 for illustration). Node Q serves two arrival traffics: one is through traffic $A(t)$ and the other is cross traffic $A_c(t)$ which has a higher priority. Here the cross traffic is referred to as the interfering process, i.e., $A_c(t) = I_k(t)$ when there are k receiver nodes.

For node Q with k receivers and 1 packet/timeslot service rate, its leftover service curve is

$$S(s,t) = t - s - I_k(s,t) \tag{14}$$

which expresses the service available to the through traffic $A(t)$ at this link.

Therefore, for each transmission $[\mathrm{Q} \to \mathrm{R}_j]$ $(j = 1, ..., k)$, its cumulative arrivals and departures in $(s,t]$ are represented by $A(s,t)$ and $D_j(s,t)$, respectively. The service for arrival process $A(s,t)$ is given with $S(s,t)$.

Theorem 1. *(Lower and Upper Bounds of the Non-Asymptotic Multicast Throughput) For a particular node Q transmitting to k nodes, let the arrival process at node Q be $A(t)$ with relative upper and lower rates r^U and r^L, depending on $\theta > 0$, as in (2) and (3). For the departure process $D_j(t)$ of receiver node R_j, we have the following stochastic lower bound on the non-asymptotic throughput for all $\lambda_t^L, \varepsilon \geq 0$ and $t \geq 0$*

$$\mathrm{P}\left(D_j(t) \leq \lambda_t^L t\right) \leq \inf_{\theta>0, \sigma>0} \left\{ C_k e^{-\theta\left(r^L - \lambda_t^L - \varepsilon\right)t} e^{\theta\sigma} + \omega(\sigma) \right\} \tag{15}$$

where $C_k = \frac{1}{\theta\left(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon\right)}$, $q_k = 1 + P_{U_k}\left(e^{\theta} - 1\right)$, $P_{U_k} = 1 - p(1-p)^{U_k - 1}$, $p = \frac{1}{U_k}$, U_k is the number of interfering nodes, and $\omega(\sigma)$ is a violation probability of the service guarantee in (14) satisfying (6). Note that the infimum should also satisfy $1 - \frac{\log q_k}{\theta} > r^L - \varepsilon$, which is the stability condition for the process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

The corresponding upper bound for all λ_t^U and $t \geq 0$ is

$$\mathrm{P}\left(D_j(t) \geq \lambda_t^U t\right) \leq \inf_{\theta>0,\sigma>0} \left\{ e^{-\theta t\left(\lambda_t^U - r^U\right)} e^{-\theta\sigma} + \omega\left(\sigma\right) \right\}. \tag{16}$$

In the stability condition $1 - \frac{\log q_k}{\theta} > r^L - \varepsilon$, the left-hand side term represents the available service rate for a single-hop transmission. By introducing a parameter $\varepsilon \geq 0$, we can relax r^L in the exponential and make an optimization space for θ in the stability condition. Note that r^U and r^L are non-asymptotic both in time and network size.

Proof: Fix σ and some discrete time t. Assume that for receiver node R_j in a single-hop multicast session, the following inequality holds

$$D_j(t) \geq A * \left[S - \sigma\right]_+ (t) \tag{17}$$

such that we can write

$$\begin{aligned}
&\mathrm{P}(D_j(t) \leq \lambda_t^L t) \\
&\quad = \mathrm{P}(D_j(t) \leq \lambda_t^L t \,|(16) \text{ satisfies})\mathrm{P}((16) \text{ satisfies}) \\
&\qquad + \mathrm{P}(D_j(t) \leq \lambda_t^L t \,|(16) \text{ fails})\mathrm{P}((16) \text{ fails}) \\
&\quad \leq \mathrm{P}(A * [S-\sigma]_+(t) \leq \lambda_t^L t) + \omega(\sigma) \\
&\quad \leq \mathrm{P}(\inf_{0\leq s\leq t}\{A(s) + S(s,t) - \sigma\} \leq \lambda_t^L t) + \omega(\sigma) \\
&\quad \leq \sum_{s=0}^{t} \mathrm{P}(A(s) + S(s,t) - \sigma \leq \lambda_t^L t) + \omega(\sigma) \\
&\quad \leq \sum_{s=0}^{t} \mathrm{P}(A(s) + t - s - I_k(s,t) - \sigma \leq \lambda_t^L t) + \omega(\sigma).
\end{aligned} \tag{18}$$

Using Boole's inequality and Chernoff bound, we can bound the sum above by

$$\begin{aligned}
&\sum_{s=0}^{t} \mathrm{P}(e^{\theta(\lambda_t^L t + \sigma + I_k(s,t) - A(s) - (t-s))} \geq 1) \\
&\quad \leq \sum_{s=0}^{t} E[e^{\theta(\lambda_t^L t + \sigma + I_k(s,t) - A(s) - (t-s))}] \\
&\quad \leq \sum_{s=0}^{t} e^{-\theta t(r^L - \lambda_t^L)} e^{-\theta(1 - r^L - \frac{\log q_k}{\theta})(t-s)} e^{\theta\sigma}.
\end{aligned} \tag{19}$$

Since the sum is non-increasing in r^L, we can relax this rate by some $\varepsilon \geq 0$ satisfying the stability condition $1 - \frac{\log q_k}{\theta} > r^L - \varepsilon$. Then the sum can be further bounded by

$$\sum_{s=0}^{t} e^{-\theta(r^L - \lambda_t^L - \varepsilon)t} e^{-\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon)(t-s)} e^{\theta\sigma}. \tag{20}$$

Assume $s^* = t - s$ ($s^* \geq 0$), since $e^{-\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon)} > 0$, $\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon) < 1$. Thus for all $\theta > 0$ and $\varepsilon \geq 0$, we can bound the last sum by

$$\begin{aligned}
&\sum_{s^* \geq 0} e^{-\theta(r^L - \lambda_t^L - \varepsilon)t} e^{-\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon)s^*} e^{\theta\sigma} \\
&\qquad = \frac{e^{-\theta(r^L - \lambda_t^L - \varepsilon)t} e^{\theta\sigma}}{1 - e^{-\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon)}} \\
&\qquad \leq \frac{e^{-\theta(r^L - \lambda_t^L - \varepsilon)t} e^{\theta\sigma}}{\theta(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon)}.
\end{aligned} \tag{21}$$

In (21) we applied the inequality $\sum_{s^* \geq 1} e^{-cs^*} \leq \frac{1}{c}$, for $c > 0$. The derivation of the lower bound of throughput is completed by minimizing over θ, ε, and σ.

In turn, for the upper bound, we can write

$$\begin{aligned}
&\mathrm{P}(D_j(t) \geq \lambda_t^U t) \\
&\quad \leq \mathrm{P}(A * [S-\sigma]_+ \geq \lambda_t^U t) + \omega(\sigma) \\
&\quad \leq \mathrm{P}(A(s) + t - s - I_k(s,t) - \sigma \geq \lambda_t^U t) + \omega(\sigma) \\
&\quad \leq e^{\theta(1 - r^U - \frac{\log q_k}{\theta})(t-s)} e^{-\theta t(\lambda_t^U - r^U)} e^{-\theta\sigma}.
\end{aligned} \tag{22}$$

The infimum is obtained at $s = t$, then we can bound the last sum by minimizing over θ and σ.

■

Set the right-hand side term of (15) (the throughput bound) equal to some fixed violation probability γ_C^L, we get

$$\begin{aligned}
\lambda_t^L &= \frac{\log \frac{\gamma_C^L - \omega(\sigma_1)}{C_k^1} - \theta_1\sigma_1}{\theta_1 t} + r^L - \varepsilon_1 \\
&= r^L - \frac{\sigma_1}{t} + \frac{\log \theta_1 \left(\gamma_C^L - \omega(\sigma_1)\right)\left(1 - r^L - \frac{\log q_k}{\theta_1} + \varepsilon_1\right)}{\theta_1 t} - \varepsilon_1
\end{aligned} \tag{23}$$

where σ_1, θ_1 ε_1 are parameters that minimize the violation probability in (15) under the constraint $1 - \frac{\log q_k}{\theta_1} > r^L - \varepsilon_1$.

In turn, for the upper bound λ_t^U, we have

$$\lambda_t^U = r^U - \frac{\sigma_2}{t} - \frac{\log\left(\gamma_C^U - \omega\left(\sigma_2\right)\right)}{\theta_2 t}. \tag{24}$$

Theorem 2. *(Lower and Upper Bounds of the Non-Asymptotic Multicast Delay) Consider the same assumption for the network as Theorem 1. For the delay process $W_j(t)$ of the multicast transmission, we have the following stochastic lower bound on the non-asymptotic delay for all d_L and $t \geq 0$*

$$\mathrm{P}\left(W_j\left(t\right) \leq d_L\right) \leq \inf_{\theta>0,\sigma>0} \left\{ e^{\theta d_L \left(1 - \frac{\log q_k}{\theta}\right)} e^{-\theta\sigma} + \omega\left(\sigma\right) \right\}. \tag{25}$$

The corresponding upper bound for all d_U and $t \geq 0$ is

$$\mathrm{P}\left(W_j\left(t\right) \geq d_U\right) \leq \inf_{\substack{\theta>0,\varepsilon\geq 0,\sigma>0, \\ 1 - \frac{\log q_k}{\theta} > r^L - \varepsilon}} \left\{ \begin{array}{l} B_k e^{-\theta d_U \left(r^L - \varepsilon\right)} e^{\theta\sigma} \\ +\omega\left(\sigma\right) \end{array} \right\} \tag{26}$$

where $B_k = \frac{1}{\theta\left(1 - r^L - \frac{\log q_k}{\theta} + \varepsilon\right)}$.

Proof: Fix σ, time t and d_L. Since the inequality in (17) still holds, the corresponding delay process for transmission Q-R_j is denoted by $W_j(t) = \inf\{d : A(t-d) \leq D_j(t)\}$

$$W_j(t) \geq d_U \Rightarrow A(t - d_U) \geq D_j(t). \tag{27}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

It follows that

$$\begin{aligned}
&\mathrm{P}(W_j(t) \geq d_U) \\
&\quad \leq \mathrm{P}(A(t-d_U) \geq D_j(t) \,|(16) \text{ satisfies })\mathrm{P}((16) \text{ satisfies}) \\
&\qquad + \mathrm{P}(A(t-d_U) \geq D_j(t) \,|(16) \text{ fails })\mathrm{P}((16) \text{ fails}) \\
&\quad \leq \mathrm{P}(A(t-d_U) \geq A * [S-\sigma]_+(t)) \ + \omega(\sigma) \\
&\quad \leq \mathrm{P}(A(t-d_U) \geq \inf_{0\leq s\leq t}\{A(s)+S(s,t)-\sigma\}) + \omega(\sigma) \\
&\quad \leq \mathrm{P}\left(\sup_{s\leq t-d_U}\{A(s,t-d_U)-S(s,t)+\sigma\} \geq 0\right) + \omega(\sigma) \\
&\quad \leq \sum_{s=0}^{t-d_U} \mathrm{P}\left(A(s,t-d_U)+I(s,t)-(t-s)+\sigma \geq 0\right) + \omega(\sigma)
\end{aligned} \tag{28}$$

Using Boole's inequality and Chernoff bound for some $\theta > 0$, we have

$$\begin{aligned}
&\sum_{s=0}^{t-d_U} \mathrm{P}\left(e^{\theta(A(s,t-d_U)+I(s,t)-(t-s)+\sigma)} \geq 1\right) \\
&\quad \leq \sum_{s=0}^{t-d_U} E\left[e^{\theta(A(s,t-d_U)+I(s,t)-(t-s)+\sigma)}\right] \\
&\quad \leq \sum_{s=0}^{t-d_U} e^{\theta r^L(t-s-d_U)} e^{-\theta\varepsilon(t-s-d_U)} e^{\theta\frac{\log q_k}{\theta}(t-s)} e^{-\theta(t-s)} e^{\theta\sigma} \\
&\quad \leq \sum_{s=0}^{t-d_U} e^{-\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right)(t-s)} e^{-\theta d_U\left(r^L-\varepsilon\right)} e^{\theta\sigma}.
\end{aligned} \tag{29}$$

Assume $s^* = t-s$ ($s^* \geq 0$), since $e^{-\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right)} > 0$, $\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right) < 1$. Thus for all $\theta > 0$, we can further bound the sum by

$$\begin{aligned}
&\sum\nolimits_{s^*\geq 0} e^{-\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right)s^*} e^{-\theta d_U\left(r^L-\varepsilon\right)} e^{\theta\sigma} \\
&\qquad = \frac{e^{-\theta d_U\left(r^L-\varepsilon\right)} e^{\theta\sigma}}{1-e^{-\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right)}} \\
&\qquad \leq \frac{e^{-\theta d_U\left(r^L-\varepsilon\right)} e^{\theta\sigma}}{\theta\left(1-r^L-\frac{\log q_k}{\theta}+\varepsilon\right)}.
\end{aligned} \tag{30}$$

Then we can bound the last sum by minimizing over θ and σ.

We now derive the lower delay bound d_L for the same scenario. Starting from the definition of delay process $W_j(t)$, we have

$$\begin{aligned}
&\mathrm{P}\left(W_j\left(t\right) \leq d_L\right) \\
&\quad \leq \mathrm{P}\left(A\left(t-d_L\right) \leq A * \left[S-\sigma\right]_+\left(t\right)\right) + \omega\left(\sigma\right) \\
&\quad \leq \mathrm{P}(A\left(t-d_L\right) \leq \inf_{0\leq s\leq t}\{A(s)+S(s,t)-\sigma\}) + \omega\left(\sigma\right) \\
&\quad \leq \mathrm{P}(\inf_{s\leq t-d_L}\{S(s,t)-A(s,t-d_L)-\sigma\} \geq 0) + \omega\left(\sigma\right) \\
&\quad \leq \mathrm{P}\left(t-s-I(s,t)-A(s,t-d_L)-\sigma \geq 0\right) + \omega\left(\sigma\right).
\end{aligned} \tag{31}$$

We remark that the points $s = t-d_L+1, ..., t$ do not contribute to the infimum in (31) (due to the positivity constraint). Then in a similar way

$$\begin{aligned}
&\mathrm{P}\left(e^{\theta(t-s-I(s,t)-A(s,t-d_L)-\sigma)} \geq 1\right) \\
&\quad \leq E\left[e^{\theta(t-s-I(s,t)-A(s,t-d_L)-\sigma)}\right] \\
&\quad \leq e^{\theta(t-s)} e^{-\theta\frac{\log q_k}{\theta}(t-s)} e^{-\theta r^U(t-s-d_L)} e^{-\theta\sigma} \\
&\quad \leq e^{\theta d_L\left(1-\frac{\log q_k}{\theta}\right)} e^{-\theta\sigma}.
\end{aligned} \tag{32}$$

The infimum is attained at $s = t-d_L$. We can bound the last sum by minimizing over θ and σ.

■

Solving for the delay bound in (25) with the right-hand side term equal to γ_D^L, we get

$$d_L = \frac{\sigma_3}{1-\frac{\log q_k}{\theta_3}} + \frac{\log\left(\gamma_D^L - \omega(\sigma_3)\right)}{\theta_3\left(1-\frac{\log q_k}{\theta_3}\right)}. \tag{33}$$

Similarly, for the upper bound d_U, we have

$$\begin{aligned}
d_U &= \frac{\theta_4\sigma_4 - \log\frac{\gamma_D^U-\omega(\sigma_4)}{B_k^4}}{\theta_4\left(r^L-\varepsilon_4\right)} \\
&= \frac{\sigma_4}{r^L-\varepsilon_4} - \frac{\log\theta_4\left(1-r^L-\frac{\log q_k}{\theta_4}+\varepsilon_4\right)\left(\gamma_D^U-\omega\left(\sigma_4\right)\right)}{\theta_4\left(r^L-\varepsilon_4\right)}.
\end{aligned} \tag{34}$$

C. Multi-Hop Case

Unlike the interfering processes in single-hop case, $I(t)$ in multihop case would additionally depend on the correlated transmissions of neighboring links.

Under the multicast scenario in Figure 1, the longest multihop path is from the source s_i to the destination node $v_{ij'}$. Let $\mathrm{P_T}\left(s_i, v_{ij'}\right)$ denote such a path with H nodes. A flow will generally be served by some dominating relay nodes whose degree are bounded by 3 and some assistant relay nodes during the path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$. The number of dominating relay nodes from CDS with degree l are denoted as $E(l)$ $(l = 1, 2, 3)$. Then in path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$, we have

$$E\left(1\right) = 1, \ E\left(2\right) = E\left(3\right) = \sqrt{k} - 1. \tag{35}$$

The number of assistant relay nodes in path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$ is represented by $E_a(1)$ (because their degree is 1), and is given by

$$E_a(1) = E(2) + E(3) = 2(\sqrt{k} - 1). \tag{36}$$

Based on the analysis above, path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$ can be transformed into the path as shown in Figure 7. Thus, the total number of hops of path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$ is $H = 4\sqrt{k} - 2$.

From the leftover service curve in (14), for any choice of

Fig. 7: Transformation of path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$ with different cross-traffics in each nodes. Nodes of larger degree have more cross-traffics, which are represented by dotted lines.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

$\delta > 0$, the function

$$S_l^h(t) = ((1 - \frac{\log q_l}{\theta}) - \delta)t \tag{37}$$

is the service available to the through flows at node h ($h = 1, 2, ..., H$) with degree l. The parameter δ is a relaxation of the service, which implies that the guaranteed service can be reduced by a small rate.

The resulting error function is

$$\omega^{s,h}(\sigma) = \frac{e^{\theta\tau_0}}{\theta\delta\tau_0} e^{-\theta\sigma} \tag{38}$$

where θ is the decay rate of the error function and τ_0 is a discretization parameter indicating a time step. We choose specific values of σ so that the error value does not exceed some violation probability.

Then we derive performance bounds of path $\mathrm{P_T}(s_i, v_{ij'})$ by statistical network service curve. Here we apply Theorem 1 in [28] to derive it. By convolving the service curves of all the nodes along path $\mathrm{P_T}(s_i, v_{ij'})$, we can obtain the statistical network service curve $S^{\mathrm{e}}(t)$ of the path. For any choice of $0 < \delta \leq \frac{1 - r^U - \frac{\log q_3}{\theta}}{H+1}$, we have

$$\begin{aligned} S^{\mathrm{e}}(t) = {} & S_1(t) * S_1^1(t) * S_2^2(t) * \cdots * S_2^{\sqrt{k}}(t) * \\ & S_3^{\sqrt{k}+1}(t) * \cdots * S_3^{2\sqrt{k}-1}(t) * \\ & S_{a1}^{2\sqrt{k}}(t) * \cdots * S_{a1}^{4\sqrt{k}-3}(t) \\ = {} & ((1 - \tfrac{\log q_3}{\theta}) - (4\sqrt{k} - 2)\delta)t \end{aligned} \tag{39}$$

with an error function

$$\omega^{\mathrm{e}}(\sigma) = C_0 \frac{H e^{\theta\tau_0}}{(\theta\delta\tau_0)^{\frac{2H-1}{H}}} e^{-\frac{\theta}{H}\sigma} \tag{40}$$

where $C_0 \geq 1$ is a constant and $H = 4\sqrt{k} - 2$. $\omega^{\mathrm{e}}(\sigma)$ is calculated based on Lemma 3 in [28], which is omitted here.

The derivation of the statistical network service curve in (39) and (40) helps convert the multihop analysis into single-hop analysis. It is independent of the arrival process at the source node, and holds for all deterministic or stochastic traffic. However, in order to compute the performance bounds for a flow, both the network service curve and the traffic characterization are required. We next state the main results of this paper: non-asymptotic performance guarantees on end-to-end multicast throughput and delay, where arrivals are bounded by stochastic envelopes satisfying (2) and (3), and service is expressed in the form of statistical service curves satisfying (39) and (40).

Theorem 3. *(Lower and Upper Bounds on Non-Asymptotic End-to-End Multicast Throughput) Consider the scenario from Figure 1. Let the arrival process $A_i(t)$ at node s_i with relative upper and lower rates r^U and r^L, depending on $\theta > 0$, as in (2) and (3). For departure process $D_{\mathrm{e}}^{\min}(t)$ of one receiver node with minimum throughput rate, we have the following stochastic lower bound on the end-to-end per-flow multicast throughput for all $\lambda_t^{L_{\mathrm{e}}}$, $\varepsilon \geq 0$ and $t \geq 0$*

$$\begin{aligned} & \mathrm{P}\left(D_{\mathrm{e}}^{\min}(t) \leq \lambda_t^{L_{\mathrm{e}}} t\right) \\ & \quad \leq \inf_{\theta>0,\sigma>0} \left\{ C_{\mathrm{e}} e^{-\theta\left(r^L - \lambda_t^{L_{\mathrm{e}}} - \varepsilon\right)t} e^{\theta\sigma} + \omega^{\mathrm{e}}(\sigma) \right\} \end{aligned} \tag{41}$$

where $C_{\mathrm{e}} = \frac{1}{\theta\left(1 - r^L - \frac{\log q_3}{\theta} - \left(4\sqrt{k}-2\right)\delta + \varepsilon\right)}$ and the infimum should also satisfy the stability condition $1 - \frac{\log q_3}{\theta} > r^L - \varepsilon$. The corresponding upper bound for all $\lambda_t^{U_{\mathrm{e}}}$ and $t \geq 0$ is

$$\begin{aligned} & \mathrm{P}\left(D_{\mathrm{e}}^{\min}(t) \geq \lambda_t^{U_{\mathrm{e}}} t\right) \\ & \quad \leq \inf_{\theta>0,\sigma>0} \left\{ e^{-\theta\left(\lambda_t^{U_{\mathrm{e}}} - r^U\right)t} e^{-\theta\sigma} + \omega^{\mathrm{e}}(\sigma) \right\}. \end{aligned} \tag{42}$$

The proof of Theorem 3 relies on the analysis of the single-hop case. The path $\mathrm{P_T}(s_i, v_{ij'})$ can be regarded as a single-hop path with a service curve $S \sim_{sc} \langle S^{\mathrm{e}}(t), \omega^{\mathrm{e}}(\sigma) \rangle$, where $S^{\mathrm{e}}(t)$ and $\omega^{\mathrm{e}}(\sigma)$ are given by (39) and (40), respectively. The results can be derived by replacing $S(t)$ and $\omega(\sigma)$ in (18) and (22) with $S^{\mathrm{e}}(t)$ and $\omega^{\mathrm{e}}(\sigma)$, respectively, which is omitted here.

Note that in our settings we optimize over the values of δ, τ_0 and σ. An exact optimization for these parameters may be difficult if the error function is not exponential. However, the throughput and delay bounds are not very sensitive to the choice of these parameters in proximity to the optimum. Thus, an explicit optimization can be replaced by an iterative procedure.

Set the right-hand side term of (41), i.e., the probability that the end-to-end throughput violates the lower bound $\lambda_t^{L_{\mathrm{e}}} t$, equal to $\gamma_C^{L_{\mathrm{e}}}$, we get

$$\begin{aligned} \lambda_t^{L_{\mathrm{e}}} &= \frac{\log \frac{\gamma_C^{L_{\mathrm{e}}} - \omega^{\mathrm{e}}(\sigma_5)}{C_{\mathrm{e}}} - \theta_5\sigma_5}{\theta_5 t} + r^L - \varepsilon_5 \\ &= r^L - \frac{\sigma_5}{t} - \frac{\log C_{\mathrm{e}}}{\theta_5 t} + \\ &\quad \frac{\log\left(\gamma_C^{L_{\mathrm{e}}} - \omega^{\mathrm{e}}(\sigma_5)\right)}{\theta_5 t} - \varepsilon_5. \end{aligned} \tag{43}$$

In turn, for the upper bound $\lambda_t^{U_{\mathrm{e}}}$, we have

$$\lambda_t^{U_{\mathrm{e}}} = r^U - \frac{\sigma_6}{t} - \frac{\log\left(\gamma_C^{U_{\mathrm{e}}} - \omega^{\mathrm{e}}(\sigma_6)\right)}{\theta_6 t}. \tag{44}$$

Then consider that all the nodes in the single multicast session from Figure 1 also serve flows from other multicast sessions. For each cell, the total number of multicast sessions (or flows) passing through it is not more than $\Theta(\sqrt{kn\log n})$ [9]. Thus the probability that each node in the given multicast session cannot successfully transmit to all its l receivers is $P_{U_l} = 1 - \frac{p(1-p)^{U_l - 1}}{\sqrt{kn\log n}}$. Since $q_l = 1 + P_{U_l}(e^\theta - 1)$, we have

$$\lim_{\theta\to 0} \frac{\log q_l}{\theta} = \lim_{\theta\to 0} \frac{1 + P_{U_l}(e^\theta - 1)}{\theta} = P_{U_l}. \tag{45}$$

Combined with the stability condition in Theorem 3, we can obtain that $r^L = \Theta\left(\frac{1}{\sqrt{kn\log n}}\right)$. For some fixed violation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

probability and time scale t, the non-asymptotic lower bound from (43) and upper bound from (44) scale as

$$\lambda_t^{L_{\rm e}} = \lambda_t^{U_{\rm e}} = \Theta\left(\frac{1}{\sqrt{kn\log n}}\right) \tag{46}$$

which captures the results in [7]–[11] from an asymptotic point of view.

Theorem 4. *(Lower and Upper Bounds on Non-Asymptotic End-to-End Multicast Delay) Consider the scenario from Figure 1 and the same assumption for the network as Theorem 3. For the per-session delay process $W_{\rm e}^{\max}(t)$, we have the following stochastic lower bound on the end-to-end multicast delay for all $d_{L_{\rm e}}$ and $t \geq 0$*

$$\mathrm{P}\left(W_{\rm e}^{\max}\left(t\right) \leq d_{L_{\rm e}}\right) \leq \inf_{\theta>0,\sigma>0} \left\{ \begin{array}{l} ke^{\theta d_{L_{\rm e}}\left(1-\frac{\log q_3}{\theta}\right)}e^{-\theta\sigma}+ \\ k\omega^{\rm e}\left(\sigma\right) \end{array} \right\}. \tag{47}$$

The corresponding upper bound for all $d_{U_{\rm e}}$ and $t \geq 0$ is

$$\begin{aligned} &\mathrm{P}\left(W_{\rm e}^{\max}\left(t\right) \geq d_{U_{\rm e}}\right) \\ &\leq \inf_{\substack{\theta>0,\varepsilon\geq 0,\sigma>0,\\ 1-\frac{\log q_3}{\theta}>r^L-\varepsilon}} \left\{B_{\rm e}ke^{-\theta d_{U_{\rm e}}\left(r^L-\varepsilon\right)}e^{\theta\sigma}+k\omega^{\rm e}\left(\sigma\right)\right\} \end{aligned} \tag{48}$$

where $B_{\rm e} = \frac{1}{\theta\left(1-r^L-\frac{\log q_3}{\theta}-\left(4\sqrt{k}-2\right)\delta+\varepsilon\right)}$.

Proof: Here we provide the key idea behind the proof. Denote $W_{\rm e}(t)$ as the average delay of path $\mathrm{P_T}\left(s_i, v_{ij'}\right)$. With its statistical network service curve $S \sim_{sc} \langle S^{\rm e}\left(t\right), \omega^{\rm e}\left(\sigma\right)\rangle$ in (39) and (40), we can obtain the delay bounds by replacing $S\left(t\right)$ and $\omega\left(\sigma\right)$ in (28) and (31) with $S^{\rm e}\left(t\right)$ and $\omega^{\rm e}\left(\sigma\right)$, respectively. Then for the per-session delay process $W_{\rm e}^{\max}(t)$ (i.e., the maximal time spent by one S-D pair among all S-D pairs), we have

$$\begin{aligned} \mathrm{P}(W_{\rm e}^{\max}(t) \leq d_{L_{\rm e}}) &\leq k\mathrm{P}(W_{\rm e}(t) \leq d_{L_{\rm e}}) \\ &\leq \inf_{\theta>0,\sigma>0} \left\{ \begin{array}{l} ke^{\theta d_{L_{\rm e}}\left(1-\frac{\log q_3}{\theta}\right)}e^{-\theta\sigma}+ \\ k\omega^{\rm e}\left(\sigma\right) \end{array} \right\}. \end{aligned} \tag{49}$$

Similarly, we can also derive $\mathrm{P}(W_{\rm e}^{\max}(t) \geq d_{U_{\rm e}})$. ■

Solving for the delay bound in (47) with the right-hand side term equal to $\gamma_D^{L_{\rm e}}$, we get

$$d_{L_{\rm e}} = \frac{\sigma_7}{1-\frac{\log q_3}{\theta_7}} - \frac{\log k}{\theta_7\left(1-\frac{\log q_3}{\theta_7}\right)} + \frac{\log\left(\gamma_D^{L_{\rm e}} - k\omega^{\rm e}(\sigma_7)\right)}{\theta_7\left(1-\frac{\log q_3}{\theta_7}\right)}. \tag{50}$$

Similarly, for the upper bound $d_{U_{\rm e}}$, we have

$$\begin{aligned} d_{U_{\rm e}} &= \frac{\theta_8\sigma_8 - \log\frac{\gamma_D^{U_{\rm e}}-\omega^{\rm e}(\sigma_8)}{B_{\rm e}}}{\theta_8(r^L-\varepsilon_8)} \\ &= \frac{\sigma_8}{r^L-\varepsilon_8} - \frac{\log k}{\theta_8(r^L-\varepsilon_8)} + \frac{\log B_{\rm e}}{\theta_8(r^L-\varepsilon_8)} \\ &\quad - \frac{\log\left(\gamma_D^{U_{\rm e}} - k\omega^{\rm e}(\sigma_8)\right)}{\theta_8(r^L-\varepsilon_8)}. \end{aligned} \tag{51}$$

Then consider that all the nodes in the single multicast session from Figure 1 also serve flows from other multicast sessions. Following the scaling analysis of end-to-end multicast throughput, we have $P_{U_3} = 1 - \frac{p(1-p)^{U_3-1}}{\sqrt{kn\log n}}$ and $r^L = \Theta\left(\frac{1}{\sqrt{kn\log n}}\right)$. Plug P_{U_3} and r^L into (50) and (51), respectively. For some fixed violation probability, the non-asymptotic lower bound from (50) and upper bound from (51) scale as

$$d_{L_{\rm e}} = d_{U_{\rm e}} = \Theta(\sqrt{kn\log n}\log k) \tag{52}$$

which implies that the per-session multicast delay increases as $\sqrt{kn\log n}\log k$.

IV. Numerical Results

In this section, we analyze the impacts of the degree of burstiness and multicast (i.e., the number of receivers k) on the performance bounds of single-hop case and multi-hop case, respectively. Moreover, we validate theoretical results from Theorem 3 and 4 by simulating multi-hop case (i.e., the scenario in Figure 1). The source traffics are modeled as an aggregate of independent Markov Modulated On-Off (MMOO) processes. This kind of process falls into the category of the EBB traffic model. The MMOO arrival process is a continuous time process supported by a homogeneous two-state Markov chain $X(t)$, which is described in terms of the generator matrix

$$G = \left(\begin{array}{cc} -\mu & \mu \\ \lambda & -\lambda \end{array} \right). \tag{53}$$

Here, μ denotes the transition rate from "On" state to "Off" state and λ denotes the transition rate from "Off" state to "On" state. In the "On" state, the arrival process transmits at the peak rate r^U, and no arrivals occur in the "Off" state. The MGF of a MMOO process (arrival flow) is bounded by (2) and (3), and the effective arrival rate is given by $r = \frac{r^U\theta-\mu-\lambda+\left(\left(r^U\theta-\mu+\lambda\right)^2+4\lambda\mu\right)^{\frac{1}{2}}}{2\theta}$. r satisfies $r^L \leq r \leq r^U$, where $r^L = \frac{\lambda}{\lambda+\mu}r^U$.

We introduce a parameter $T = \frac{1}{\mu} + \frac{1}{\lambda}$ to describe the burstiness of a flow. T is the expected time for the Markov chain to change states twice. For a flow with given r^U and r^L, a larger value of T indicates a higher degree of burstiness.

In order to validate the correctness and tightness of our performance bounds, we conduct simulations of multihop case and compare analytical results with simulation results. Consider the multicast scenario in Figure 1, which is transformed into the following topology: 49 nodes are placed in a 7×7 grid as shown in Figure 3(a). The distance between two neighboring nodes is set to be 70m. The simulation is constructed in two settings: 1) the packets are sent by multicast transmissions; 2) the packets are sent by multiple unicast transmissions to all the receivers. We will compare the throughput and delay bounds obtained by these two settings to show the benefits of multicast transmissions.

The simulation parameters of multihop case are as follows: The capacity of each node is $C = 1$Mbps, the source node

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

s_i generates MMOO flows with peak rate $r^U = 600$Kbps. Both the transmission range and sensing range are set to be 100m. The throughput rate of destination node $v_{ij'}$ is tracked every 200 seconds, and the delay is recorded after all packets are sent to k receiver nodes successfully. k is set to be 9 for throughput simulations while k varies from 1 to 16 for delay simulations.

For the numerical settings of theoretical analysis, we choose a specific value of σ so that the error probability $\omega^{\mathrm{e}}(\sigma)$ in (40) does not exceed a desired value, e.g., 10^{-3}. The number of nodes in transmission range is set to $m = 8$.

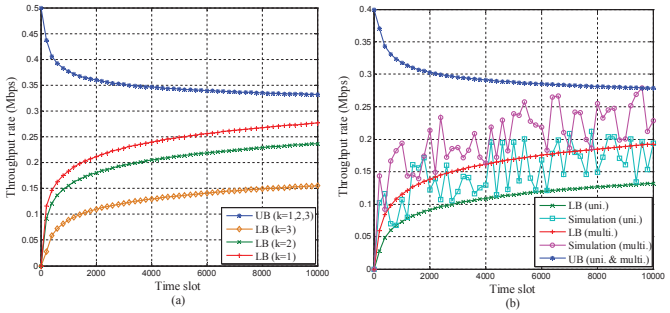


Fig. 8: (a) Non-asymptotic throughput rates for different number of receiver nodes k in single-hop case ($T = 1$), (b) non-asymptotic analytical throughput rates vs. simulation throughput rates in multi-hop case ($T = 1$).

Figure 8(a) shows the behavior of upper bound (UB) and lower bound (LB) on the exact throughput rate of a single-hop multicast session with respect to the number of receivers k ($k = 1, 2, 3$). Here the input burstiness is set to $T = 1$. As the figure illustrated, LB decreases with k while UB is independent of k. The intuitive explanation is as follows. The output flow will have the minimal throughput rate λ_t^L when the input flow with minimal rate r^L is served with minimal service rate, i.e., all the potential interfering nodes also want to transmit packets. As k increases, the number of interfering nodes U_k will also increases and so does the rate of the interfering flow. Thus the leftover service rate for through flows decreases, resulting in a loss of the lower bound of throughput λ_t^L. However, the output flow has maximal throughput rate λ_t^U when a node happens to get a transmission opportunity to serve a through flow with maximal rate r^U. In this case, no other nodes interfere the transmission, which implies that λ_t^U is independent of k. Figure 8(b) illustrates the comparison of analytical and simulation results in two transmission situations of multihop case, with the same setting of input burstiness. We can observe that LB and UB tightly bound their corresponding simulation results in multicast setting, which validates the correctness of the analytical bounds. However, the UB of multiple unicast setting is relatively loose and its simulation throughput rates (λ_1) are generally lower than that of multicast setting (λ_2), since there are no shared links between different S-D pairs in multiple unicast transfers, and multicast transfers have a more effective use of the bandwidth of service nodes. Moreover, λ_1 exceeds λ_2 with very small probability, due to the flow burstiness.

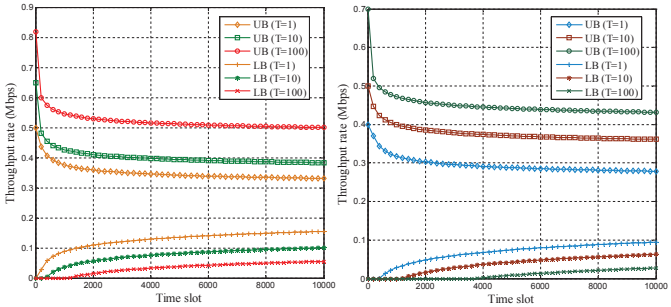


Fig. 9: Upper and lower bounds of the non-asymptotic throughput rates for different burstiness parameters T: (a) single-hop case ($k = 3$), (b) multi-hop case.

Figure 9 illustrates the LB and UB of the exact throughput rate under different burstiness parameters ($T = 1, 10, 100$). From this figure, we can see that both LB and UB greatly depend on the burstiness of flows. Specifically, UB increases with T while LB decreases with T, which implies that the higher the degree of burstiness, the bigger the gap between UB and LB. We present an intuitive insight for this phenomenon. When T increases, it means that the time of state "on" and "off" will both increase. Based on the description of the EBB process, we can observe that the "on" time has a greater impact on the upper rate while the "off" time has a greater impact on the lower rate. Therefore, as the time of these two states increases, the upper rate will increase and lower rate will decrease. In addition, the throughput rates are more sensitive to burstiness in multi-hop case than in single-hop case, as shown in Figure 9(a) and (b). This is because our analytical multihop bounds fully account for the burstiness at the relay nodes.

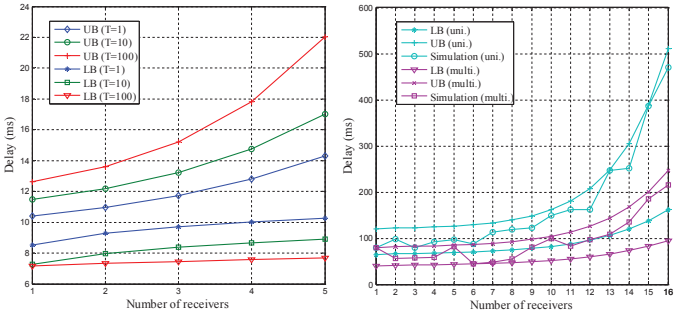


Fig. 10: (a) Upper and lower bounds of the non-asymptotic delay for different burstiness parameters T, in single-hop case, where k varies from 1 to 5, (b) non-asymptotic analytical delay vs. simulation delay in multi-hop case ($T = 1$).

Figure 10(a) presents UB and LB of the exact delay for different T and k in the single-hop case. As the figure shows, UB increases with T while LB decreases with T. Because given the number of receivers k, delay varies with the upper rate and lower rate of the input flow. The impacts of T on the upper rate and lower rate are provided in the previous analysis. Moreover, both UB and LB increase with k. This is in accordance with the fact that the number of transmission opportunities will decrease as k increases. Figure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology

8(b) illustrates the comparison of analytical and simulation results in two transmission situations of multihop case, with the same setting of input burstiness. We can observe that LB and UB tightly bound its corresponding simulation results in multicast setting. However, for multiple unicast setting, since there are very few situations experiencing low delays, its corresponding LB is relatively loose. It indicates that our analytical bounds provide more reliable estimates. Besides, simulation delay results of multiple unicast setting (W_1) are higher than those of multicast setting (W_2), for the same reason discussed in Figure 8(b). Compared with the single-hop case, both LB and UB increase with k, at a faster growth rate. This is due to the additional effects from the number of hops H, which depends on k. As k increases, it will need more hops to deliver packets to the furthest nodes, incurring higher delay. On the other hand, the randomness will also expand and more time is required to cover the worst case.

V. Conclusion

In this paper, we derived non-asymptotic multicast throughput and delay of multihop wireless networks with a certain number of nodes. An explicit analytical model is proposed for multicast routing, in which stochastic network calculus is employed for the exact analysis. The upper and lower bounds provide insights on how the multicast throughput and delay are affected by practical factors, such as the network size, the degree of data burstiness, the time scales and the number of receivers. Furthermore, our results also hold in asymptotic regimes. Simulation results from realistic network settings can be tightly bounded by the analytical performance bounds, which validates the accuracy of our proposed method.

For future work, we want to investigate the impact of mobility models on non-asymptotic performance. In mobile networks, nodes can store and carry packets until they meet the destinations. Therefore, it is expected that compared with static networks, mobile networks will can achieve a larger throughput, at the cost of larger delay, as shown in [2] for asymptotic analysis. However, for non-asymptotic analysis, it is still not clear how to model the arrival and service processes and how to apply the network calculus in mobile scenarios, which therefore is quite challenging and of great interest.

Acknowledgment

This work was partially supported by NSF China (No. 61325012, 61271219, 61221001, 61428205, 61231010); China Ministry of Education Doctor Program (No.20130073110025); Shanghai Basic Research Key Project (12JC1405200, 11JC1405100); Shanghai International Cooperation Project: (No. 13510711300); Research Fund for the Doctoral Program of MOE (No. 20120142110015).

References

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory,* vol. 46, no. 2, pp. 388-404, Mar. 2000.

[2] M. Grossglauser and D. N. C. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.

[3] P. Li, Y. Fang, J. Li, and X. Huang, "Smooth trade-offs between throughput and delay in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 427-438, Mar. 2012.

[4] D. M. Shila, Y. Cheng, and T. Anjali, "Throughput and delay analysis of hybrid wireless networks with multi-hop uplinks," in *Proceedings of IEEE INFOCOM*, Apr. 2011.

[5] J. Abouei, A. Bayesteh, and A. K. Khandani, "On the delay-throughput tradeoff in distributed wireless networks," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2159-2174, Apr. 2012.

[6] P. Li and Y. Fang, "On the throughput capacity of heterogeneous wireless networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 2073-2086, Dec. 2012.

[7] P. Jacquet and G. Rodolakis, "Multicast scaling properties in massively dense ad hoc networks," in *Proceedings of ICPADS-Workshops*, 2005.

[8] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of large multihop wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1691-1700, Dec. 2010.

[9] X. Li, "Multicast capacity of wireless ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 950-961, Jun. 2009.

[10] X. Li, Y. Liu, S. Li, and S. Tang, "Multicast capacity of wireless ad hoc networks under gaussian channel model," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1145-1157, Aug. 2010.

[11] J. Zhang, X. Wang, X. Tian, Y. Wang, X. Chu, and Y. Cheng, "Optimal multicast capacity and delay tradeoffs in MANETs," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1104-1117, May. 2014.

[12] X. Li, S. Tang, O. Frieder, "Multicast capacity for large scale wireless ad hoc networks," in *Proceedings of ACM MobiCom*, Sept. 2007.

[13] S. Li, Y. Liu, and X. Li, "Capacity of large scale wireless networks under gaussian channel model," in *Proceedings of ACM MobiCom*, 2008.

[14] J. Luo, J. Zhang, L. Yu and X. Wang, "Non-asymptotic multicast throughput capacity in multi-hop wireless networks," in *Proceedings of IEEE Globecom*, Dec. 2013.

[15] C. S. Chang, "Performance guarantees in communication networks," Springer Verlag, 2000.

[16] Y. Jiang, "A basic stochastic network calculus," in *Proceedings of ACM SIGCOMM*, Oct. 2006.

[17] Y. Jiang, "Stochastic service curve and delay bound analysis: A single node case," in *Proceedings of IEEE International Teletraffic Congress (ITC)*, Sept. 2013.

[18] F. Ciucu, "On the scaling of non-asymptotic capacity in multi-access networks with bursty traffic," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Aug. 2011.

[19] F. Ciucu and J. Schmitt, "On the catalyzing effect of randomness on the per-flow throughput in wireless networks," in *Proceedings of IEEE INFOCOM*, Apr. 2014.

[20] Y. Jiang and Y. Liu, "*Stochastic Network Calculus*," Springer, 2008.

[21] J. Liebeherr, A. Burchard, and F. Ciucu, "Delay bounds in communication networks with heavy-tailed and self-similar traffic," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1010-1024, Feb. 2012.

[22] F. Ciucu and J. Schmitt, "Perspectives on network calculus: no free lunch, but still good value," in *Proceedings of ACM SIGCOMM*, 2012.

[23] O. Yaron and M. Sidi, "Performance and stability of communication networks via robust exponential bounds," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 372-385, Jun. 1993.

[24] P. J. Wan, K. Alzoubi, and O. Frieder, "Distributed construction of connected dominating set in wireless ad hoc networks," in *Proceedings of IEEE INFOCOM*, Jun. 2002.

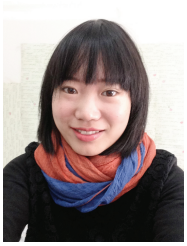
[25] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proceedings of ACM Mobicom*, 2003.

[26] M. Fidler, "An end-to-end probabilistic network calculus with moment generating functions," in *IEEE International Workshop on Quality of Service (IWQoS)*, 2006.

[27] F. Ciucu, O. Hohlfeld, and P. Hui, "Non-asymptotic throughput and delay distributions in multi-hop wireless networks," in *Proceedings of Allerton Conference on Communications, Control and Computating*, Sept. 2010.

[28] F. Ciucu, A. Burchard, and J. Liebeherr, "Scaling properties of statistical end-to-end bounds in the network calculus," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2300-2312, Jun. 2006.

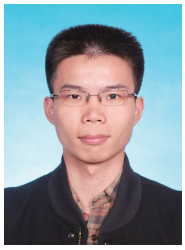
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2015.2465963, IEEE Transactions on Vehicular Technology



Jingjing Luo received the B.S. degree from the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2010, and is currently pursuing the Ph.D. degree in the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China.

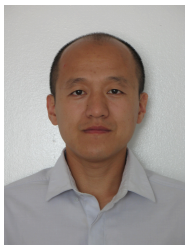
Her current research interests include capacity scaling laws and mobility models in wireless networks.

Li Yu received the B.S. degree and Ph.D. degree both from the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 1992, and 1999, respectively. Currently, she is a professor and and serves as the director of Multimedia and Communication Network Center, in the School of Electronic Information and Communications, Huazhong University of Science and Technology, China.



Jinbei Zhang received the B.S. degree in Electronic Engineering from Xidian University, Xi'an, China, in 2010, and is currently pursuing the Ph.D. degree in electronic engineering at Shanghai Jiao Tong University, Shanghai, China.

His current research interests include network security, capacity scaling law and mobility models in wireless networks.



Xinbing Wang received the B.S. degree (with hons.) from the Department of Automation, Shanghai Jiaotong University, Shanghai, China, in 1998, and the M.S. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001. He received the Ph.D. degree, major in the Department of electrical and Computer Engineering, minor in the Department of Mathematics, North Carolina State University, Raleigh, in 2006. Currently, he is a professor in the Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai, China. Dr. Wang has been an associate editor for IEEE/ACM Transactions on Networking and IEEE Transactions on Mobile Computing, and the member of the Technical Program Committees of several conferences including ACM MobiCom 2012, ACM MobiHoc 2012, 2013, IEEE INFOCOM 2009-2014.