



Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Towards an end-to-end delay analysis of wireless multihop networks

Min Xie^{a,*}, Martin Haenggi^{b,1}^a University College London Adastral Park Campus, Martlesham Heath, Suffolk IP5 3RE, United Kingdom^b Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, United States

ARTICLE INFO

Article history:

Received 5 June 2007

Accepted 21 April 2008

Available online 22 August 2008

Keywords:

Multihop

MAC

TDMA

ALOHA

Delay

Correlation

ABSTRACT

In wireless multihop networks, end-to-end (e2e) delay is a critical parameter for quality of service (QoS) guarantees. We employ discrete-time queueing theory to analyze the end-to-end (e2e) delay of wireless multihop networks for two MAC schemes, m -phase TDMA and slotted ALOHA. In one-dimensional (1-D) networks, due to the lack of sufficient multiplexing and splitting, a space–time correlation structure exists, the nodes are spatially correlated with each other, and the e2e performance cannot be analyzed as in general two-dimensional networks by assuming all nodes independent of each other. This paper studies an 1-D network fed with a single flow, an extreme scenario in which there is no multiplexing and splitting. A decomposition approach is used to decouple the whole network into isolated nodes. Each node is modeled as a GI/Geo/1 queueing system. First, we derive the complete per-node delay distribution and departure characterization, accounting for both the queueing delay and access delay. Second, based on the departure process approximation, we define a parameter to measure the spatial correlation and its influence on the e2e delay variance. Our study shows that traffic burstiness of the source flow and MAC together determines the sign of the correlation.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

With the growing demand for real-time applications over wireless networks, increasing attention is paid to the delay analysis of transmissions over error-prone channels. In multihop networks, like ad hoc, mesh, and multihop cellular networks, the analysis is more challenging than in single-hop networks due to the delay accumulation at each hop. Many factors affect the end-to-end (e2e) delay, including the routing algorithm, the MAC and packet scheduling algorithm and error-prone wireless channels. The analysis is unlikely to be tractable if all these factors are considered together. However, if there is only a single active path in the network, the two-dimensional (2-D) topology (Fig. 1a) can be reduced to one-dimension (1-D) (Fig. 1b) and routing could be ignored.

In the 1-D network, referred to as line network, there is no inter-flow interference, and the corresponding performance is an upper bound for general 2-D networks. On the other hand, it is easier to approximate 2-D networks than 1-D networks because in 2-D networks the delays are closer to be independent while in single flow 1-D networks the correlation cannot be ignored. The e2e delay is determined by the joint distribution of the successive delays of a packet traversing multiple nodes. In 2-D networks, with network-wide traffic integration, all nodes may be assumed to be independent and analyzed in isolation such that the joint distribution can be approximated in a product-form [1]. Generally speaking, the conditions to permit the “independence” assumption are: (i) the peak rate of each source does not exceed 5% of the total link capacity; and (ii) no more than 10% of the departing sources go to the same immediate downstream link [2], *i.e.*, large-scale multiplexing and splitting. However, in networks with convergecasting (*i.e.*, information gathering towards a central node Fig. 1a), the above conditions do not hold since there exists a space–time correlation structure. Then it is

* Corresponding author. Tel.: +44 1473 663701; fax: +44 1473 635199.

E-mail addresses: m.xie@adastral.ucl.ac.uk (M. Xie), mhaenggi@nd.edu (M. Haenggi).¹ Tel.: +1 574 631 6103; fax: +1 574 631 4393.

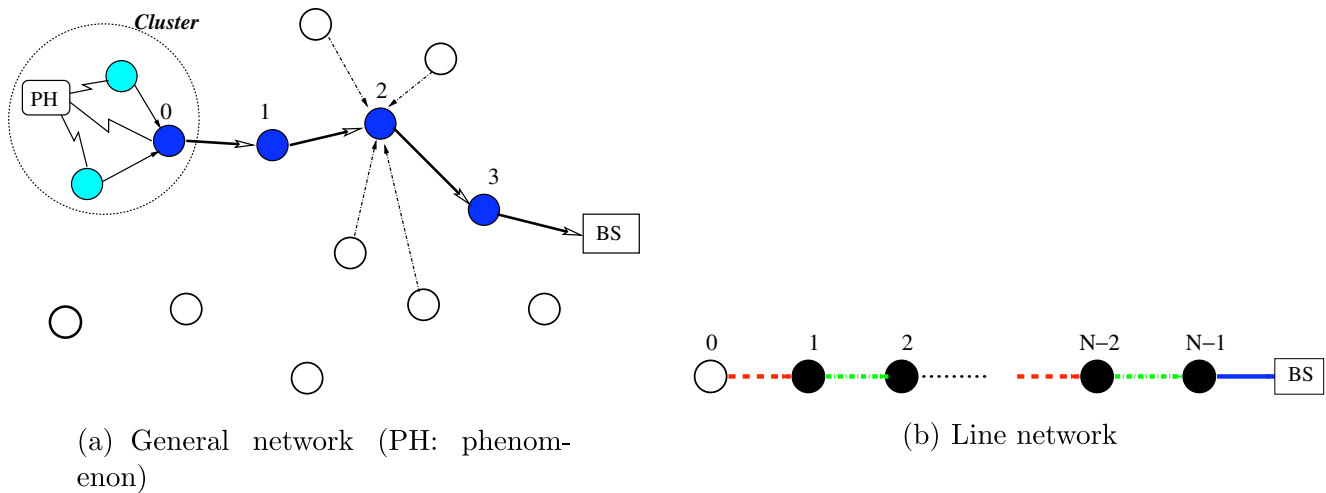


Fig. 1. Wireless multihop networks.

difficult to derive the closed-form joint distribution. Specially, in an extreme case where all intermediate nodes are pure relays (Fig. 1b, which is a representative of the area closer to the base station in random networks with convergecasting), the departure process of node i is exactly the arrival to node $i + 1$ and so forth and thus the space-time correlation is too strong to be ignored. Such correlations substantially complicate the e2e analysis.

The spatial correlation is mainly caused by the temporal correlation of the traffic flow, to which several factors contribute as well as the original traffic statistics. Here the *temporal correlation* is referred to as the correlation in two consecutive packet arrivals while the *spatial correlation* is the dependence between the activities of two nodes. Channel errors cause distortions to the traffic flow, which, in turn, change the temporal correlation. Such distortions may be further accumulated with multihop transmission [3]. The other factor is multiple access control (MAC) that schedules the node transmission order and may incur access delays, which certainly change the packet arrival pattern. Therefore, the study of the spatial correlation should take into account both the traffic statistics and the distortions caused by wireless channel errors and MAC.

1.1. Previous work

The throughput and single-hop delay of many MAC schemes have been comprehensively studied in the literature [4,5]. However, little work has been carried out on their multihop delay. Moreover, previous MAC studies usually assume that traffic is generated in a way that incurs no queueing delay, e.g., a new node is generated to represent the newly generated packet; or new packets are generated only when the buffer is empty [5–7]. These models are simplified and unrealistic. In practice, new packets may be generated when the buffer is non-empty and thus experience a queueing delay. On the other hand, the study of queueing networks is concerned with the queueing delay only, ignoring the access delay [8–11].

Due to the presence of the queueing delay, queueing models are needed. If we assume independent wireless

channel errors, the service time is geometrically distributed, and a single node can be modeled as a GI/Geo/1 system. In the literature, the queue length distribution of general GI/Geo/1 queues has been well studied [12]. However, to analyze multihop networks, the requirement for a departure process characterization arises. In the literature, only a few papers address the departure process when the arrival process has correlation in time, e.g., [13]. Moreover, for non-Bernoulli and non-Poisson arrivals, it is known that the departure process is correlated with the queue length and arrival process [14], which results in cumbersome expressions [13,15] that prohibit a scalable e2e analysis. Closed-form solutions for the delay of wireless regular line networks with a single source (like Fig. 1b) are available only if the arrival is Bernoulli [9] or the channels are error-free [11]. For other cases, approximations are needed. [16] analyzed discrete-time tandem queueing networks with bursty and correlated input traffic by ignoring the correlation between nodes. An IEEE 802.11 wireless ad hoc network is modeled as a series of independent M/G/1 systems to obtain a delay distribution in product-form [17]. Similarly, in [18], the e2e delay variance of a two-node tandem network is derived by assuming that the two nodes are independent. The “independence” assumption usually holds for general network topologies with flow multiplexing. For line networks without multiplexing, such an assumption may lead to a very pessimistic or overly optimistic performance expression, especially in terms of delay variance.

1.2. Our contributions

This paper studies the e2e delay of a wireless line network (Fig. 1b) fed with a single source. We consider both the access delay and queueing delay. The e2e performance is investigated under two simple but typical MAC schemes, m -phase spatial TDMA [4] and slotted ALOHA. In TDMA, a node is scheduled to transmit once in m time slots, and nodes m hops apart may transmit simultaneously. In ALOHA, every node independently transmits with probability p_m whenever it has packets. TDMA (with nodes fully cooperative) and ALOHA (with nodes completely indepen-

dent) represent the two extremes in terms of the level of the node coordination and will provide upper and lower performance bounds for other meaningful MAC schemes, respectively. The arrival processes to every node are all relayed versions of the original traffic flow generated at the source node. To explore the influence of original traffic statistics on the e2e delay, three traffic models with different traffic burstiness and memory properties are studied, CBR (for voice data [19] and periodic traffic in sensor networks), correlated on-off, and Bernoulli (for bursty data).

Our contributions are twofold. First, we use discrete-time queueing theory to analyze the MAC-controlled nodes, deriving a complete delay and departure process characterization. Particularly, the analysis for TDMA involves a GI/Geo/1 queueing model with *non-integer interarrival times*. Second, based on the single-node analysis, we not only confirm the convergence phenomenon of the relayed flows, which has been proved in previous work via entropy theory, but also reveal that the direction of convergence depends on both the original traffic flow and the MAC scheme. Moreover, based on the departure characterization, we define a parameter to measure the spatial correlation, which determines the e2e delay variance. Simulation results are provided to verify our analysis in both the single node and the e2e delay variance.

The rest of the paper is organized as follows. The system model is introduced in Section 2. In Section 3, we present two models to derive the delay and departure process characterization of GI/Geo/1 systems for CBR and on-off traffic. Then we establish and analyze GI/Geo/1 systems for single nodes in the TDMA and ALOHA networks in Sections 4 and 5, respectively. Section 6 compares the single-node delays of TDMA and ALOHA and studies the direction of convergence of the departure processes. Section 7 extends the analysis to the e2e delay and derive a parameter to measure the sign of the spatial correlation. Section 8 concludes the paper.

2. System model

The line network under consideration (Fig. 1b) is composed of N transmitting nodes and a sink or base station (BS). Denote node i by n_i ($i = 0, 1, 2, \dots, N-1$) and the delay experienced at n_i by D_i with mean \bar{D}_i and variance σ_i^2 . The probability mass function (pmf) of D_i is denoted by $\{d_j^{(i)}\}$. The e2e delay is given by $D = \sum_{i=0}^{N-1} D_i$ with mean \bar{D} and variance σ^2 . A FIFO discipline is used at n_i . A flow of fixed-length packets is generated at the source n_0 at rate λ , and all remaining nodes are pure relays. The time is slotted to the duration of one packet transmission. So the network is modeled as a discrete-time tandem queueing network. For non-Bernoulli and non-Poisson arrivals, the departure process of a node is correlated with the queue length and its arrival. Therefore, the D_i 's are correlated, which leads to $\sigma^2 \neq \sum_{i=0}^{N-1} \sigma_i^2 \triangleq \beta$. If D_i 's are positively (negatively) correlated, then $\sigma^2 > \beta$ ($\sigma^2 < \beta$). Note that previous work, by assuming “independent” nodes and ignoring the spatial correlation, usually assumed $\sigma^2 = \beta$.

The channels are subject to independent errors (e.g., AWGN or block fading channels) and characterized by a

“capture” model [20] with a capture probability $\mu \triangleq \Pr(\text{SNIR} \geq \Theta)$, i.e., a transmission is successful with probability μ . To guarantee 100% reliability, the failed packets will be retransmitted at each hop until received successfully. The number of transmission attempts to successfully send a packet is geometrically distributed with parameter μ , denoted by \mathcal{G}_μ . Note that in practice, TDMA and ALOHA result in different capture probabilities [21]. So, we denote the capture probability of TDMA and ALOHA by μ_T and μ_A , respectively.

The traffic flow to n_i is characterized by the interarrival time A_i , whose probability mass function (pmf) is $a_k(i) = \Pr\{A_i = k\}$ and probability generating function (pgf) is $A_i(z) = \sum_{k=0}^{\infty} a_k(i)z^k$. The departure processes of n_i ($i > 0$) is characterized by interdeparture time T_i . We consider three typical traffic models, (i) smooth CBR, where the packet interarrival time is an integer constant $r = 1/\lambda$; (ii) memoryless Bernoulli, where a packet is generated with probability λ in each time slot; (iii) bursty and correlated On-off, where the arrival process is modulated by a two-state Markov chain that alternates between ON (1) and OFF (0) states with transition probabilities a_{01} and a_{10} . The pmf is

$$a_k = \begin{cases} 1 - a_{10} & k = 1 \\ a_{10}(1 - a_{01})^{k-2}a_{01} & k > 1 \end{cases} \quad (1)$$

The on-off source generates a stream of *correlated* bursty and silent periods both of which are geometrically distributed in length. The mean burst size is $B = 1/a_{10}$. The average rate is $\lambda = a_{01}/(a_{10} + a_{01})$. Bernoulli is a special on-off process with $a_{01} + a_{10} = 1$ so that the burst and silent periods are *independent*.

The delay D_i consists of two parts, the queueing delay and access delay, as shown in Fig. 2. In TDMA, define m time slots as a frame. The transmission is successful with probability μ_T . So the service time is $S \sim \mathcal{G}_{\mu_T}$ and a TDMA node can be modeled as a GI/Geo/1 system at the frame level though the interarrival time is not an integer as usual. In this model, the access delay is hidden in the frame. In ALOHA, a packet is successfully transmitted if and only if the node attempts to transmit *and* the transmission is successful, with probability $\mu_s \triangleq \mu_A p_m$ (given that the arrival and the channel state are independent²). Both the access delay and the failed transmission attempts can be regarded as unsuccessful transmission attempts. Since the channel errors are independent and the transmit probability p_m is fixed, the service time is $S \sim \mathcal{G}_{\mu_s}$ at the slot level. So, an ALOHA node can also be modeled as GI/Geo/1 although the arrival process characterization is different from TDMA.

We use a decomposition approach to decompose the tandem queueing network into single nodes in isolation [22,23]. The e2e analysis is based on the single-node analysis, including both the node delay distribution and the departure characterization. Since both TDMA and ALOHA nodes can be modeled as GI/Geo/1, we start with the analysis of GI/Geo/1 systems.

² To account for the half-duplex restriction, here μ_A is the conditional capture probability given that the receiver is listening.

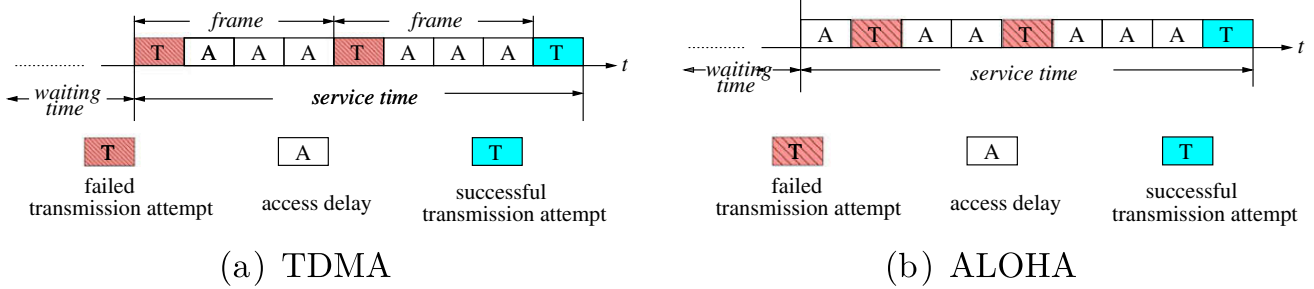


Fig. 2. Packet transmission procedure in TDMA and ALOHA.

3. GI/Geo/1 queueing systems

We use two models to derive the delay distribution of GI/Geo/1 systems. The first one is a conventional queueing model that provides the queue length distribution [9,24] and will be used for CBR traffic with non-integer interarrival times. The second one is a delay model [25] that directly tracks the evolution of the delay of the head-of-line (HOL) packet in the queue and could be used for CBR traffic in TDMA and on-off traffic.

3.1. Conventional queueing model for CBR traffic

In a conventional queueing model, the system state is denoted by the queue length [9,24]. For a GI/Geo/1 system, the pmf of the queue length is [12,26]

$$\pi_k = \begin{cases} 1 - \rho & k = 0 \\ \rho(1 - \gamma)\gamma^{k-1} & k > 0 \end{cases} \quad (2)$$

where ρ is the traffic intensity, defined as the ratio of the average arrival rate to the service rate. γ is the unique solution of $z = A(1 - \mu + \mu z)$ that lies in the region $(0, 1)$. From (2), the pmf of the queue length viewed by an arrival is in a geometric form $q_k = (1 - \gamma)\gamma^k$ ($k \geq 0$) [27].

Eq. (2) is derived based on the condition that the interarrival times are integers. However, in TDMA systems, the system is analyzed at the frame level and packet arrivals occur at the slot level. Then more than one packet may arrive during one frame and thus the interarrival times are no longer integer. For CBR arrivals with frame length of m and interarrival time r slots ($r > m$), even though the system is reduced to D/Geo/1 with $A(z) = z^{r/m}$, Theorem 1 shows that the pmf of the queue length is more complex than (2) if r/m is irreducible.

Theorem 1. Consider a discrete-time D/Geo/1 system with a geometric server \mathcal{G}_μ and constant interarrival time r/m ($r > m$, $r, m \in \mathbb{N}$ and r/m is irreducible). The pmf of the queue length distribution is

$$\pi_k = \begin{cases} 1 - \rho & k = 0 \\ \sum_{j=1}^m C_j \gamma_j^{k-1} & k > 0 \end{cases} \quad \rho = m/r\mu, \quad (3)$$

where C_j is a normalizing constant and $\{\gamma_j | j = 1, 2, \dots, m\}$ are the m roots of $z^m = (1 - \mu + \mu z)^r$ that lie inside the unit circle.

Proof. Denote the system states at the beginning of frame t by a two-dimensional Markov chain $\{Q(t), Y(t)\}$, where $Q(t) \geq 0$ is the queue length and $Y(t) = 1, 2, \dots, r$ is the number of slots to the next packet arrival. Divide the set $\{1, 2, \dots, r\}$ into two parts $\mathbb{Y}_0 \triangleq \{1, \dots, \Delta\}$ and $\mathbb{Y}_1 \triangleq \{\Delta + 1, \dots, r\}$ ($\Delta \triangleq r - m$), where the subscript represents the number of packets arriving during one frame. Denote the steady-state system probability by $Q(k, y) := \lim_{t \rightarrow \infty} \Pr\{Q(t) = k, Y(t) = y\}$. The balance equations are

$$\begin{aligned} y \in \mathbb{Y}_0 : Q(k, y) &= \begin{cases} (1 - \mu)Q(k, y + m) + \mu Q(k + 1, y + m) & k > 0, \\ Q(0, y + m) + \mu Q(1, y + m) & k = 0, \end{cases} \\ y \in \mathbb{Y}_1 : Q(k, y) &= \begin{cases} (1 - \mu)Q(k - 1, y - \Delta) + \mu Q(k, y - \Delta) & k > 1, \\ Q(0, y - \Delta) + \mu Q(1, y - \Delta) & k = 1. \end{cases} \end{aligned} \quad (4)$$

Define the row vector $\vec{v}_k := \{Q(k, 1), \dots, Q(k, r)\}$ ($k \geq 0$). For $n \geq 1$, (4) can be rewritten in a matrix form $\vec{v}_k \mathbf{M}_0 + \vec{v}_{k+1} \mathbf{M}_1 + \vec{v}_{k+2} \mathbf{M}_2 = 0$, where

$$\begin{aligned} \mathbf{M}_0 &= \begin{bmatrix} \mathbf{0} & (1 - \mu)\mathbf{I}_m \\ \mathbf{0}_\Delta & \mathbf{0} \end{bmatrix}, \quad \mathbf{M}_1 = \begin{bmatrix} \mathbf{0} & \mu\mathbf{I}_m \\ (1 - \mu)\mathbf{I}_\Delta & \mathbf{0} \end{bmatrix} - \mathbf{I}, \\ \mathbf{M}_2 &= \begin{bmatrix} \mathbf{0} & \mathbf{0}_m \\ \mu\mathbf{I}_\Delta & \mathbf{0} \end{bmatrix}. \end{aligned}$$

This is a homogeneous vector difference equation with constant coefficients. Its characteristic matrix polynomial is $\mathbf{Q}(z) = \mathbf{M}_0 + \mathbf{M}_1 z + \mathbf{M}_2 z^2$. Using the eigenvalue method [28], \vec{v}_k is solved as $\vec{v}_k = \mathbf{CZ}^k \Phi$, where the diagonal matrix $\mathbf{Z} = \text{diag}(z_j)$ and the matrix $\Phi = [\vec{\phi}_j]^T$ are composed of the eigenvalues $\{z_j\}$ and eigenvectors $\{\vec{\phi}_j\}$ of $\mathbf{Q}(z)$ in the form of $\vec{\phi} \mathbf{Q}(z) = 0$ with $\vec{\phi}_j = \{\phi_j(1), \phi_j(2), \dots, \phi_j(r)\}$. The eigenvalues are solved from $\det|\mathbf{Q}(z)| = 0$, which leads to $z^m = (1 - \mu + \mu z)^r$. Then, $Q(k, y)$ is

$$Q(k, y) = \sum_{j=1}^m \frac{C_j (1 - \xi_j) \gamma_j^k \xi_j^{-y}}{1 - \gamma_j}, \quad (5)$$

which gives rise to the queue length probability $\pi_k = \sum_{y=1}^r Q(k, y)$ as in (3). \square

From (5), we derive the pmf of the queue length viewed by an arrival and then calculate the delay distribution in Theorem 2.

Theorem 2. Consider a discrete-time D/Geo/1 system with a geometric server \mathcal{G}_μ and constant interarrival time r/m ($r > m$, $r, m \in \mathbb{N}$ and r/m is irreducible). The pmf of the delay is

$$d_k = \frac{1}{\rho} \sum_{j=1}^m \frac{C_j}{1 - \gamma_j} \cdot (1 - \xi_j) \xi_j^{k-1}, \quad k \geq 1, \quad (6)$$

where $\{\gamma_j | j = 1, 2, \dots, m\}$ are the m roots of $z^m = (1 - \mu + \mu z)^r$ inside the unit circle and $\xi_j = \gamma_j^{1/r}$ is the root of $\mu x^r - x^m + 1 - \mu = 0$.

Proof. The packet delay D_0 is composed of three independent parts, the access delay $D_A \in \{0, \dots, m - 1\}$, the waiting time D_W , and the service time D_S . For a packet arriving during frame t , its access delay is $D_A(t) = m - Y(t)$ ($Y(t) < m$) and $Y(t)$ evolves as $Y(t + 1) = Y(t) + \Delta = r - D_A(t) \in \mathbb{Y}_1$. The probability that the packet sees $k - 1$ packets in the buffer is

$$Q(k | D_A) \triangleq \frac{Q(k, r - D_A)}{\sum_{k=1}^{\infty} \sum_{y \in \mathbb{Y}_1} Q(k, y)} = \frac{r}{m} \sum_{j=1}^m \frac{C_j (1 - \xi_j) \gamma_j^{k-1} \xi_j^{D_A}}{1 - \gamma_j}. \quad (7)$$

The waiting time D_W is the sum of service times $D_{S_0} \sim \mathcal{G}_\mu$ (at the frame level) of the $k - 1$ buffered packets. At the slot level, the pgf is $G_{D_{S_0}}(z) = \frac{\mu z^m}{1 - (1 - \mu)z^m}$ and the pgf of D_W is $G_{D_W}(z) = (G_{D_{S_0}}(z))^{k-1}$. The service time of the observed packet has a pmf $\Pr\{D_S = km + 1\} = \mu(1 - \mu)^k$ ($k \geq 0$) and pgf $G_{D_S}(z) = \frac{\mu z}{1 - (1 - \mu)z^m}$. The pgf of the delay $D_0 = D_A + D_W + D_S$ is

$$G_{D_0}(z) = \sum_{n=1}^{\infty} \sum_{D_A=0}^{m-1} G_{D_W}(z) G_{D_S}(z) \pi(k | D_A) z^{D_A} \\ = \frac{1}{\rho} \sum_{j=1}^m \frac{C_j (1 - \xi_j) z}{\xi_j (1 - \gamma_j) (1 - \xi_j z)}.$$

Inverse z-transform yields (6). \square

Theorem 1 and 2 can be regarded as the analysis of generalized D/Geo/1 systems that involve with non-integer interarrival times. The previous result (2) is a special case of Theorem 1 with $m = 1$.

The delay distribution of a D/Geo/1 system involves m roots $\{\gamma_j | j = 1, 2, \dots, m\}$ inside the unit circle. If $m > 1$ (non-integer interarrival times), some roots are complex and negative real and only numerical results are available for large m and r . To obtain a closed-form solution, for $r < 2m$, we simplify the delay distribution by ignoring the complex and negative roots and considering only the real positive root $\xi_1 \in (0, 1)$. This way, (6) is reduced to $d_k \approx (1 - \xi_1) \xi_1^{k-1}$ ($k \geq 1$), i.e., $D_0 \sim \mathcal{G}_{1-\xi_1}$. However, since ξ_1 is the root of a high degree polynomial $\mu x^r - x^m + 1 - \mu$, it is still difficult to calculate D_0 's distribution. Lemma 3 approximately expresses ξ_1 with (m, r, μ) .

Lemma 3. Given a polynomial $\mu x^r - x^m + 1 - \mu = 0$ with $0 < \mu < 1$, $0 < m/(r\mu) < 1$ and $r < 2m$, the real positive root ξ_1 in the region $(0, 1)$ can be approximated by

$$\xi_1 \approx 1 - \frac{2(1 - \rho)}{\Delta \rho}, \quad \text{where } \rho = \frac{m}{r\mu} < 1, \quad \Delta = r - m < m. \quad (8)$$

Proof. Based on Descartes' Sign Rule, for the polynomial $f(x) = \mu x^r - x^m + 1 - \mu$, there are exactly two real positive roots, 1 and $\xi_1 \in (0, 1)$. The derivative of $f(x)$ is a continuous function $f'(x) = r\mu x^{r-1} - mx^{m-1}$. Given $f(\xi_1) = 0$ and $f(1) = 0$, Rolle's Theorem states that there must be at least one point $x_{\min} \in (\xi_1, 1)$ with $f'(x_{\min}) = 0$. Since $f'(x) = 0$ yields two solutions, $x = 0$ with $f(0) = 1 - \mu > 0$ and $x = x_{\min} = (m/r\mu)^{1/(r-m)}$ with $f(x_{\min}) < 0$, we say x_{\min} is the local minimum between ξ_1 and 1. Using two inequalities, $-\frac{1-\rho}{\rho} < \ln \rho \leq \Delta(\rho^{\frac{1}{\Delta}} - 1)$ [29], x_{\min} is lower bounded by $x_{\min} \gtrsim 1 - (1 - \rho)/(\Delta \rho)$. Numerical results verify that ξ_1 is very close to 1 (Fig. 3). Then it is reasonable to assume an equal distance from x_{\min} to 1 and ξ_1 , i.e., $\xi_1 \approx 2x_{\min} - 1$, which leads to (8). \square

The approximation (8) is tight when Δ is large and ρ is close to 1, both of which also guarantee $\xi_1 \lesssim 1$. Now that $D_0 \sim \mathcal{G}_{1-\xi_1}$, the corresponding delay mean and variance are approximately

$$\bar{D}_0 \approx \frac{1}{1 - \xi_1} \approx \frac{\Delta \rho}{2(1 - \rho)}, \quad \sigma_0^2 \approx \frac{\xi_1}{(1 - \xi_1)^2} = \bar{D}_0(\bar{D}_0 - 1). \quad (9)$$

3.2. Delay models

The delay model can be used to analyze GI/Geo/1 systems with on-off traffic and characterize the departure process. In the delay model, the system state is denoted by the current delay of the HOL packet [25]. For a GI/Geo/1 system with on-off arrivals (a_{01}, a_{10}) (1), letting $D = \infty$ and extending the result [25], Eq. (9) the delay is $D_0 \sim \mathcal{G}_{1-\alpha}$, where

$$\alpha = \frac{1 - \mu}{\mu a_{10} + (1 - \mu)(1 - a_{01})}. \quad (10)$$

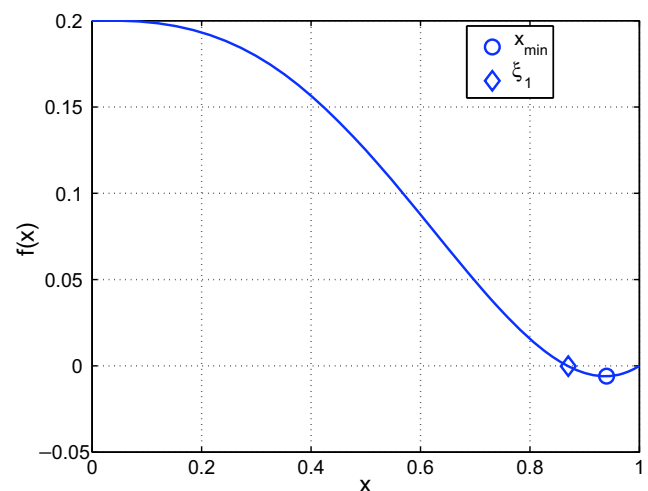


Fig. 3. Approximation of real positive root ξ_1 for the polynomial $f(x) = \mu x^r - x^m + 1 - \mu$ with $m = 3, r = 4, \mu = 0.8$.

The advantage of the delay model is its capability of characterizing the departure process, as shown in Lemma 4.

Lemma 4. Consider a discrete-time GI/Geo/1 queueing system with service rate μ and on-off arrival with transition probabilities (a_{01}, a_{10}) . The interdeparture time T has the pgf

$$G_T(z) = \tilde{\pi}_B S(z) + (1 - \tilde{\pi}_B) \frac{a_{01}z}{1 - (1 - a_{01})z} S(z), \quad (11)$$

where $\tilde{\pi}_B = 1 - \frac{a_{01}(1-\rho)}{\lambda}$ is the system busy probability viewed by a departure and $S(z) = \frac{\mu z}{1 - (1-\mu)z}$

Proof. Let the system state be the delay of the HOL packet at the moment of a packet departure. The transition probabilities are

$$P_{jk} = \mu(1 - \mu)^{l-1} a_h, \quad \begin{cases} k = j + l - h, & j \geq 0 \\ k = l - h, & j < 0 \end{cases} \quad (12)$$

The absolute value of the negative state represents the system idle time. Denote the steady-state probability by π_j . The interdeparture time T is the sum of the packet service time S and system idle time. Upon a packet departing moment, $T = S - j$ if the system is in negative states $j < 0$ and $T = S$ if the system is busy with probability $\tilde{\pi}_B = \sum_{j < 0} \pi_j$. Given independent arrival and service processes, the pgf of the interdeparture time T is $G_T(z) = \tilde{\pi}_B S(z) + \sum_{j=1}^{\infty} \pi_{-j} z^j S(z)$. From (12), we obtain $\pi_{-j} = (1 - a_{01})^{j-1} \pi_{-1}$ for $j \geq 1$ and $\tilde{\pi}_l = 1 - \tilde{\pi}_B = \sum_{j < 0} \pi_j = \pi_{-1}/a_{01}$. For stable systems, the average departure rate equals to the average arrival rate, i.e., the average interdeparture time is $\bar{T} = 1/\lambda = (a_{01} + a_{10})/a_{01}$, from which we can calculate $\tilde{\pi}_B$ and $\tilde{\pi}_l$. Plugging these parameters into $G_T(z)$ yields (11). \square

Recall that at equilibrium, the node busy probability is $\pi_B = \rho$. Compared to the conditional busy probability $\tilde{\pi}_B$ upon the departure moment, $\tilde{\pi}_B = \pi_B$ only if $a_{01} = \lambda$ or $a_{10} + a_{01} = 1$, i.e., the arrival is Bernoulli.

Note that the second part of $G_T(z)$ (11) is a convolution of two geometric distributions \mathcal{G}_μ and $\mathcal{G}_{a_{01}}$. Then, the departure processes would exhibit a state explosion problem if it were fed into a tandem network [15]. Approximation is needed for tractable analysis. Since the on-off model can capture both the correlation and burstiness property of a traffic flow, in Lemma 5, the output process of a GI/Geo/1 system is approximated as an on-off process.

Lemma 5. Consider a GI/Geo/1 system with arrival rate λ , whose interdeparture time is T . The output process can be approximated as an on-off process with transition probabilities $a'_{11} = \Pr\{T = 1\}$ and $a'_{01} = \lambda(1 - a'_{11})/(1 - \lambda)$.

The proof is based on the fact that if $T = 1$, then the modulated state transits from ON to ON, with probability a'_{11} . For GI/Geo/1 with on-off arrivals, we have

$$a'_{11} = \Pr\{T = 1\} = \tilde{\pi}_B \mu = \mu - \frac{a_{01}(1 - \rho)}{\rho}. \quad (13)$$

In the following sections, the arrival to n_i ($i > 0$), also known as the departure from n_{i-1} , is approximated by an on-off process with $(a_{10}^{(i)}, a_{01}^{(i)})$.

4. Single-node analysis for TDMA

A TDMA node is modeled as a GI/Geo/1 system with service rate μ_τ at the frame level, at which the arrival is an accumulated version of the original flow over m slots. The traffic intensity is $\rho \triangleq m\lambda/\mu_\tau < 1$.

4.1. Source node: CBR traffic

In a TDMA node with CBR traffic of rate $\lambda = 1/r$, for general m and r , in Theorem 2, the conventional queueing model (Section 3.1) provides an accurate delay description, which, however, can be solved only numerically. For $r < 2m$, the delay can be approximated as a geometric distribution with parameter approximated as in Lemma 3. The approximation (8) is tight when $\Delta = r - m$ is large. For small Δ , say $\Delta = 1$, which represents the heaviest stable traffic that can be accommodated by the system, we use the delay model to calculate the distribution in Theorem 6.

Theorem 6. Consider a D/Geo/1 system with interarrival time r/m and service rate μ_τ . If $r = m + 1$, the pgf of the delay D_0 is

$$G_{D_0}(z) = \frac{(1 - z^m)z}{(1 - \mu_\tau)z^{m+1} - z + \mu_\tau} \cdot \frac{1 - \rho}{\rho}, \quad \rho = \frac{m}{r\mu_\tau} \quad (14)$$

Proof. Let the system state be the delay of the HOL packet in terms of slots. All state transitions occur at the frame boundaries. The state transition probabilities are

$$P_{kj} = \begin{cases} \mu & k \geq 0, \quad j = k - \Delta, \\ 1 - \mu & k \geq 0, \quad j = k + m, \quad \Delta = r - m. \\ 1 & k < 0, \quad j = k + m, \end{cases} \quad (15)$$

For $r = m + 1$, the steady-state probabilities for non-negative states are

$$\pi_k = \begin{cases} \mu\pi_{k+1} & 0 \leq k < m - 1 \\ \mu\pi_{k+1} + \mu\pi_0 & k = m - 1 \\ \mu\pi_{k+1} + (1 - \mu)\pi_{k-m} & k \geq m, \end{cases} \quad (16)$$

Based on $d_{k+1} = \pi_k / \sum_{k \geq 0} \pi_k$, the pgf of the delay is

$$G_{D_0}(z) = \frac{(z^m - 1)z\mu_\tau}{z - \mu_\tau - (1 - \mu_\tau)z^{m+1}} d_1^{(0)}, \quad (17)$$

which contains one unknown parameter $d_1^{(0)}$. Following $G_{D_0}(1) = 1$, we have

$$d_1^{(0)} = \frac{1 - (m + 1)(1 - \mu_\tau)}{m\mu_\tau} = 1 + \frac{1}{m} - \frac{1}{\mu_\tau}. \quad (18)$$

Plugging (18) into (17) leads to (14). \square

The delay mean and variance can be calculated through the first two derivatives of $G_{D_0}(z)$ at $z = 1$ as follows:

$$\bar{D}_0 = \frac{1}{2(1 - \rho)}, \quad \sigma_0^2 = \frac{1}{4(1 - \rho)^2} - \frac{m + 2}{6(1 - \rho)}. \quad (19)$$

Note that (15) hold for all $r > m$. But for $r > m + 1$, the pgf contains $\Delta > 1$ unknown parameters and cannot be solved like $r = m + 1$. So the accurate delay description (14) is only for $r = m + 1$. Comparing (9) and (19) shows that the approximation (9) is tight when $\rho \rightarrow 1$.

The departure process of this D/Geo/1 system is studied at the frame level since all packet depart at the boundary of frames. Lemma 7 proves that for $m < r < 2m$, the departure is an on–off process.

Lemma 7. Consider a D/Geo/1 system with service rate μ_T and interarrival time r/m ($m < r < 2m$). Then the departure process is an on–off process with transition probabilities $a_{01}^{(1)} = \mu_T$ and $a_{10}^{(1)} = \Delta\mu_T/m$, where $\Delta = r - m$.

Proof. The proof is similar to Lemma 4. The difference is that upon a packet departing moment, if the queue is empty, which happens with probability $\tilde{\pi}_I = 1 - \tilde{\pi}_B$, the interdeparture time is $T_0 = 1 + S$ since the system idle time is one frame for $r < 2m$. Besides, $\tilde{\pi}_I$ and $\tilde{\pi}_B$ can be directly deduced from the stability condition $\bar{T}_0 = r/m$ as $\tilde{\pi}_B\mu_T = 1 - \frac{\Delta\mu_T}{m}$. The pgf $G_{T_0}(z) = \tilde{\pi}_B S(z) + (1 - \tilde{\pi}_B)zS(z)$ gives rise to a closed-form pmf $\{t_0(k)|k \geq 1\}$ of T_0 :

$$t_0(k) = \begin{cases} 1 - \frac{\Delta\mu_T}{m} & k = 1 \\ (\mu_T(1 - \mu_T)^{k-2}) \frac{\Delta\mu_T}{m} & k > 1 \end{cases} \quad (20)$$

which corresponds to an on–off process (1) with transition probabilities $a_{10}^{(1)} = 1 - \tilde{\pi}_B\mu_T = \Delta\mu_T/m$ and $a_{01}^{(1)} = \mu_T$. \square

If $r > 2m$, the system idle time may exceed one frame, and the departure process is more complex than an on–off process. For a tractable e2e analysis, we approximate the departure process as an on–off process with transition probabilities $\{a_{01}^{(1)}, a_{10}^{(1)}\}$ derived from (15) as follows:

$$a_{11}^{(1)} = \Pr\{D(t+1) \geq 0, S(t+1) = 1 | D(t) \geq 0, S(t) = 1\} \\ = \frac{\mu_T \sum_{k=\Delta}^{\infty} \pi_k}{\sum_{l \geq 0} \pi_l} = \frac{\mu_T (\rho - \sum_{k=0}^{\Delta-1} \pi_k)}{\rho} = \mu_T - \frac{1-\rho}{\rho} = 1 - \frac{\Delta\mu_T}{m}. \quad (21)$$

The numerator excludes states 0 through $\Delta - 1$ since these states transit to negative states after a successful transmission. Besides, from (15), we obtain $\sum_{k=0}^{\Delta-1} \pi_k = \frac{\tilde{\pi}_I}{\mu_T} = \frac{1-\rho}{\mu_T}$. Then, based on $a_{01}^{(1)}/(a_{01}^{(1)} + a_{10}^{(1)}) = m/r$, we have $a_{10}^{(1)} = \Delta\mu_T/m$ and $a_{01}^{(1)} = \mu_T$, consistent with the result given in Theorem 7.

4.2. Source node: on–off traffic

For bursty on–off traffic, due to arrival accumulation, there may be multiple arrivals during one frame, constituting a batch arrival process. In Theorem 8, we use the delay model (Section 3.2) to analyze this GI/Geo/1 system with batch arrivals.

Theorem 8. Consider a GI/Geo/1 system with service rate is μ_T and batch arrivals, which are generated by an on–off source (a_{01}, a_{10}) in m time slots. Then, the pgf of the delay D_0 is

$$G_{D_0}(z) = \frac{(1 - \rho)H_0(z)/\rho}{1 - \mu_T(1 - \frac{a_{01}}{\lambda})z^{m-1} - (1 - \mu_T)z^m - H_0(z)}, \quad \rho = m\lambda/\mu_T, \quad (22)$$

where $H_0(z) = (a_{01}(1 - z^m))/(1 - z)$.

Proof. Let the system state be the delay of the HOL packet in terms of slot while all transitions occur at the frame boundaries. The transition probabilities are:

$$P_{ij} = \begin{cases} \mu_T a_k, & j = l + m - k, \quad l \geq 0 \\ 1 - \mu_T, & j = l + m, \quad l \geq 0 \\ 1, & j = l + m, \quad l < 0. \end{cases} \quad (23)$$

The steady-state probabilities $\{\pi_k\}$ are derived from the balance equations

$$\pi_k = \begin{cases} \pi_{k-m} + \mu_T \sum_{j=0}^{\infty} a_{j+m-k} \pi_j, & 0 \leq k < m \\ (1 - \mu_T)\pi_{k-m} + \mu_T \sum_{j=k-m+1}^{\infty} a_{j+m-k} \pi_j, & k \geq m \\ \frac{\mu_T}{1 - a_{00}^m} \sum_{j=0}^{\infty} a_{j+m-k} \pi_j, & k < 0 \end{cases}, \quad (24)$$

which leads to $\pi_k = a_{00}^{k|} \pi_0$ for $k < 0$ with $\pi_0 = \frac{\mu_T a_m}{1 - a_{00}^m} \sum_{j=0}^{\infty} a_{00}^j \pi_j$. Since the delay distribution $\{d_k^{(0)}|k \geq 1\}$ involves only the non-negative states $\{\pi_k|k \geq 0\}$, the pgf $G_{D_0}(z)$ can be calculated by multiplying both sides of (24) by z^k for $k \geq 0$ and plugging $\pi_k = a_{00}^{k|} \pi_0$. The obtained pgf contains only one unknown parameter $d_1^{(0)}$, which is deduced from $G_{D_0}(1) = 1$ as $d_1^{(0)} = \frac{a_{01}}{1 - a_{01}} \cdot \frac{1-\rho}{\rho}$ and then leads to (22). \square

Differentiating $G_{D_0}(z)$ gives rise to the mean and variance as follows:

$$\bar{D}_0 = \frac{1}{1 - \rho} \left(\frac{\rho - \lambda}{a_{01}} - \rho - \frac{m - 3}{2} \right), \quad (25) \\ \sigma_0^2 = \frac{1}{(1 - \rho)^2} \left(\frac{m^2 - 1}{12} + \frac{(m - 1)(m - 2)\rho}{6} - \frac{(1 - \mu_T)\rho^2 + (m - 2)\rho + \lambda}{a_{01}} + \frac{(\rho - \lambda)^2}{a_{01}^2} \right).$$

The pmf $\{d_k^{(0)}|k \geq 1\}$ can be derived from $G_{D_0}(z)$ using the inverse z-transform. The departure process is characterized in Lemma 9.

Lemma 9. Consider a GI/Geo/1 system with service time $S \sim \mathcal{G}_{\mu_T}$ and batch arrivals, which is generated by an on–off source (a_{01}, a_{10}) in a frame of m time slots. Then, the interdeparture time T_0 has the pgf

$$G_{T_0}(z) = \left(\tilde{\pi}_B + (1 - \tilde{\pi}_B) \frac{(1 - a_{00}^m)z}{1 - a_{00}^m z} \right) S(z), \\ \tilde{\pi}_B = 1 - \frac{(1 - a_{00}^m)(1 - \rho)}{m\lambda}. \quad (26)$$

The proof is similar to Lemma 4 except for replacing l in (12) by km . The system idle probability viewed by a departure is $\tilde{\pi}_I = \sum_{k=-1}^{\infty} \pi_k = \frac{a_{00}\pi_0}{1 - a_{00}}$. The interdeparture time T_0 involves only the negative states,

$$G_{T_0}(z) = \tilde{\pi}_B S(z) + S(z) \sum_{k=-1}^{\infty} z^{|k|} \sum_{j=1-(k+1)m}^{-km} \pi_k$$

$$= \left(\tilde{\pi}_B + \frac{(1 - a_{00}^m)z}{1 - a_{00}^m z} \cdot \frac{a_{00} \pi_0}{1 - a_{00}} \right) S(z). \quad (27)$$

Plugging in $\tilde{\pi}_I = 1 - \tilde{\pi}_B = \frac{a_{00} \pi_0}{1 - a_{00}}$ and deducing $\tilde{\pi}_B$ from the stability condition lead to (26). Note that the difference between batched on-off arrivals (26) and non-batched on-off arrivals (11) lies in a_{00} replaced by a_{00}^m and λ replaced by $m\lambda$. At the frame level, a_{00}^m is the probability that no packet arrives during one frame and $m\lambda$ is the average arrival rate, as the same as a_{00} and λ at the slot level. In other words, TDMA results in arrival accumulation but does not change the departure process characterization.

For tractable analysis, the departure process (27) is simplified to an on-off process with transition probabilities $a_{10}^{(1)}$ and $a_{01}^{(1)}$ calculated from (24) based on the principle in (21), i.e.,

$$a_{11}^{(1)} = 1 - a_{10}^{(1)} = \frac{\mu_T \sum_{k=0}^{\infty} \pi_k \sum_{j=1}^{m+k} a_j}{\rho} = \mu_T - (1 - a_{00}^m) \frac{1 - \rho}{\rho}, \quad (28)$$

and $a_{01}^{(1)} = m\lambda a_{10}^{(1)} / (1 - m\lambda)$.

The analysis for Bernoulli traffic is obtained by simply plugging $a_{01} + a_{10} = 1$ into the on-off analysis. Since TDMA changes the characteristic value of the output process, which implies $a_{01}^{(1)} + a_{10}^{(1)} \neq 1$, the departure process is no longer Bernoulli while conventionally, a Bernoulli arrival guarantees the same Bernoulli departure as the arrival [12,14]. In other words, TDMA regulation produces a batched MMBP arrival process at the frame level, destroys the memoryless property, and makes the departure different from the arrival.

For all three traffic models, the output from the source node n_0 is approximated as an on-off process with transition probabilities ($a_{01}^{(1)}, a_{10}^{(1)}$). So the analysis of the relay nodes is identical for all traffic models.

4.3. Relay nodes

The arrival process to the first relay node n_1 is an on-off ($a_{01}^{(1)}, a_{10}^{(1)}$). So n_1 is modeled as a GI/Geo/1 system, whose delay distribution is geometric with parameter α (10). At the frame level, the pmf $d_{km+1}^{(1)} = (1 - \alpha)\alpha^k$ ($k \geq 0$) leads to mean and variance as follows:

$$\bar{D}_1 = 1 + \frac{m\alpha}{1 - \alpha} = 1 + m\varepsilon, \quad \sigma_1^2 = \frac{m^2\alpha}{(1 - \alpha)^2} = m^2\varepsilon(1 + \varepsilon), \quad (29)$$

where $\varepsilon \triangleq \frac{\rho}{1 - \rho} \cdot \frac{1 - \mu}{a_{01}^{(1)}}$. The departure process of such a GI/Geo/1 system can be approximated as another on-off process with ($a_{01}^{(2)}, a_{10}^{(2)}$) calculated as in (13) by replacing λ with $m\lambda$. The remaining relay nodes are analyzed in the same way by iteratively calculating ($a_{01}^{(i+1)}, a_{10}^{(i+1)}$) from ($a_{01}^{(i)}, a_{10}^{(i)}$).

5. Single-node analysis for ALOHA

m -Phase TDMA achieves a high throughput but incurs a substantial amount of overhead to establish the frame

structure and requires a complete cooperation between all nodes involved. Moreover, in networks with multi-directional flows, TDMA favors the flows that have the same direction as the TDMA order while the flows in the opposite directions would experience much longer delays. In wireless networks, slotted ALOHA may be more practical since every node operates in a completely independent way. Besides, ALOHA is insensitive to a flows' direction. The disadvantage of ALOHA is its random and independent transmission pattern that generally results in poor performance unless the traffic load is light. This section analyzes ALOHA nodes as GI/Geo/1 systems at the slot level, in which all interarrival times are integer. The conventional queueing model (Section 3.1) is used for the delay analysis. Note that the service rate is defined as $\mu_s = \mu_\lambda p_m$ and the traffic intensity is $\rho = \lambda / \mu_s$.

For CBR traffic, the source node n_0 is a D/Geo/1 system with an interarrival time r , corresponding to the case of $m = 1$ in Theorem 2. Therefore, inside the unit circle, there is a unique root ξ of the polynomial $\mu_s y^r - y + 1 - \mu_s$. Based on Theorem 2, the delay is $D_0 \sim \mathcal{G}_{1-\xi}$. However, if r is large, $\mu_s y^r - y + 1 - \mu_s = 0$ can be solved only numerically. Using a similar approach as in Lemma 3, we approximate ξ as follows:

$$\xi \approx 1 - \frac{2(1 - \rho)}{(r - 1)\rho}. \quad (30)$$

The mean and variance of D_0 are

$$\bar{D}_0 = \frac{1}{1 - \xi} \approx \frac{(r - 1)\rho}{2(1 - \rho)}, \quad \sigma_0^2 = \frac{\xi}{(1 - \xi)^2}$$

$$\approx \frac{(r - 1)\rho}{2(1 - \rho)} \left(\frac{(r - 1)\rho}{2(1 - \rho)} - 1 \right). \quad (31)$$

For the departure process, we use the delay model to derive the interdeparture time distribution in Lemma 10.

Lemma 10. Consider a D/Geo/1 queueing system with interarrival time $r \in \mathbb{N}$ and service rate μ_s . Then, the departure process can be approximated as on-off with transition probabilities $a_{01}^{(1)} = (1 - \mu_s) / ((r - 1)\xi)$ and $a_{10}^{(1)} = (1 - \mu_s) / \xi$, where ξ is the unique root of $\mu_s y^r - y + 1 - \mu_s = 0$ in the region $(0, 1)$.

As a special case of $m = 1$ in Theorem 6, plugging $m = 1$ into the transition probabilities (15) and applying the delay's pmf $d_k = (1 - \xi)\xi^{k-1}$ ($k \geq 1$) and $d_{k+1} = \pi_k / (1 - \rho)$ to (21), we have

$$a_{11}^{(1)} = \frac{\mu_s \sum_{k=r-1}^{\infty} \pi_k}{\sum_{j \geq 0} \pi_j} = \frac{\mu_s \left(\rho - \sum_{k=0}^{r-2} \pi_k \right)}{\rho} = \mu_s \xi^{r-1} = 1 - \frac{1 - \mu_s}{\xi}, \quad (32)$$

where the last part is obtained from $\mu_s \xi^r - \xi + 1 - \mu_s = 0$. Then $a_{10}^{(1)} = 1 - a_{11}^{(1)}$ and $a_{01}^{(1)} = a_{10}^{(1)} \lambda / (1 - \lambda)$. Then, the analysis of relay nodes follows the principle for GI/Geo/1 systems with on-off arrivals (Section 4.3).

For on-off traffic, the delay is $D_0 \sim \mathcal{G}_{1-\alpha}$ (10). The departure process is approximated as an on-off process as in (13). Like in TDMA, all the departure processes of the source node n_0 are approximated as on-off so that all the relay nodes are analyzed in the same way as in TDMA (Section 4.3).

6. Comparison of TDMA and ALOHA nodes

Sections 4 and 5 present the analysis of single nodes in TDMA and ALOHA, respectively, considering three traffic models, CBR, on-off and Bernoulli. For the on-off process, the correlation and burstiness can be characterized through the burst size $B = 1/a_{10}$. Using the burst size $B_r = 1/(1 - \lambda)$ of the Bernoulli process as a reference, an on-off process with longer (shorter) burst size than B_r is referred to as heavy (light). For instance, let $a_{10} = (1 - \lambda)/2$ for heavy on-off and $a_{10} = 1 - \lambda/2$ for light on-off.

For the source node, the ratio of the delay means \bar{D}_0 for on-off and CBR is

$$\eta_{\text{TDMA}} = \frac{\bar{D}_{0, \text{on-off}}}{\bar{D}_{0, \text{cbr}}} = 2 \left(\frac{B}{B_r} \cdot \frac{m - \mu_T - \rho}{\mu_T} \right) - m + 3 > 1 \quad (33)$$

$$\eta_{\text{ALOHA}} = \frac{\bar{D}_{0, \text{on-off}}}{\bar{D}_{0, \text{cbr}}} = (1 - \xi) \left(1 + \frac{1}{1 - \rho} \cdot \frac{1 - \mu_s}{\mu_s} \cdot \frac{B}{B_r} \right) > 1, \quad (34)$$

where ξ is the unique root of $\mu_s y^r - y + 1 - \mu_s$ in the region $(0, 1)$. In both TDMA and ALOHA, CBR traffic (with burst size $B = 1$) always causes the smallest delay. For on-off traffic, the longer the burst size B , the longer the delay (mean) and delay jitter (variance).

For relay nodes, the delays are approximated as geometric distribution. Let $p_m = 1/m$, the ratio of the delay means \bar{D}_i ($i \leq 1$) for TDMA and ALOHA is

$$\eta = \frac{\bar{D}_{i, \text{TDMA}}}{\bar{D}_{i, \text{ALOHA}}} = 1 - \frac{\mu(m - 1)}{m(1 - \lambda)} < 1. \quad (35)$$

Obviously, TDMA outperforms ALOHA in the delay. As a matter of fact, from the perspective of traffic shaping, TDMA acts as a leaky bucket regulator, while ALOHA behaves like a Bernoulli regulator. So TDMA-regulated traffic is smoother than ALOHA-regulated traffic and thus causes smaller delays.

The arrivals to relays are approximated as on-off. Simulation results are provided to justify such an on-off

approximation. In simulations, we assume the same rate λ for all flows and the same success probability $\mu = \mu_T = \mu_A$ for all channels. Moreover, we set the ALOHA transmit probability as $p_m = 1/m$ such that TDMA and ALOHA have the average number of transmission opportunities. Delays are measured in the number of time slots that the packet stays in the system.

Figs. 4 and 5 compare the simulated per-node delay mean and variance with our analysis. As the number of nodes increases, the simulated per-node delay mean and variance converge to the analytical results \bar{D}_i and σ_i^2 . In other words, our analysis (the dash-dotted lines in Figs. 4 and 5) represents the limiting delay performance.

Traffic burstiness affects the delay performance as usual but its influence becomes trivial as the number of nodes increases, e.g., for TDMA, the \bar{D}_i 's ($i \geq 5$) are almost identical for all four traffic flows (Fig. 4a). The reason is that the delays spatially converge to the same value regardless of the original traffic burstiness. This convergence phenomenon appears in our analysis with respect to the transition probabilities $(a_{01}^{(i)}, a_{10}^{(i)})$. As shown in Fig. 6, the approximate on-off processes converge in such a way that $a_{01}^{(i)} \rightarrow m\lambda$ for TDMA and $a_{01}^{(i)} \rightarrow \lambda$ for ALOHA. Note that an on-off flow with a_{01} equal to the arrival rate is reduced to Bernoulli. In other words, our analysis reveals that the arrivals to relays converge to Bernoulli at the frame level for TDMA and at slot level for ALOHA. More importantly, simulation results confirm that the convergence value of delay mean and variance in Figs. 4 and 5 are indeed the mean and variance of a Geo/Geo/1 system with Bernoulli arrivals. For traffic flows with lighter burstiness than the eigentraffic process, our analysis provides upper bounds on the delay and vice versa.

In [30], entropy theory was used to prove that passing an arbitrary arrival process through a series of independent and identically distributed GI/Geo/1 queues will generate an invariant Bernoulli distribution, i.e., the flows to relays converge to Bernoulli. We denote this Bernoulli process

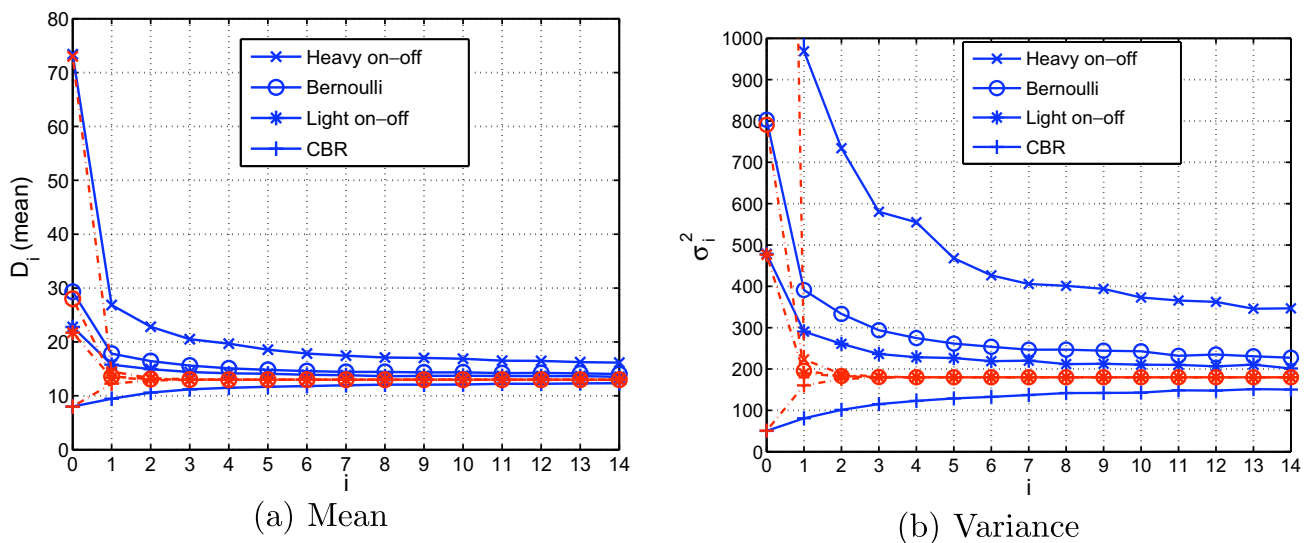


Fig. 4. The mean \bar{D}_i and variance σ_i^2 of single-node delays D_i at n_i in TDMA networks with $m = 3, \lambda = 0.25, \mu_T = 0.8, N = 15$. For light on-off, $a_{01} = 0.292, a_{10} = 0.875$; for heavy on-off, $a_{01} = 0.125, a_{10} = 0.375$. The heavy on-off flow causes a delay variance $\sigma_0^2 = 5176$ at n_0 . The dash-dotted lines represent analytical results while the solid lines are for simulation results.

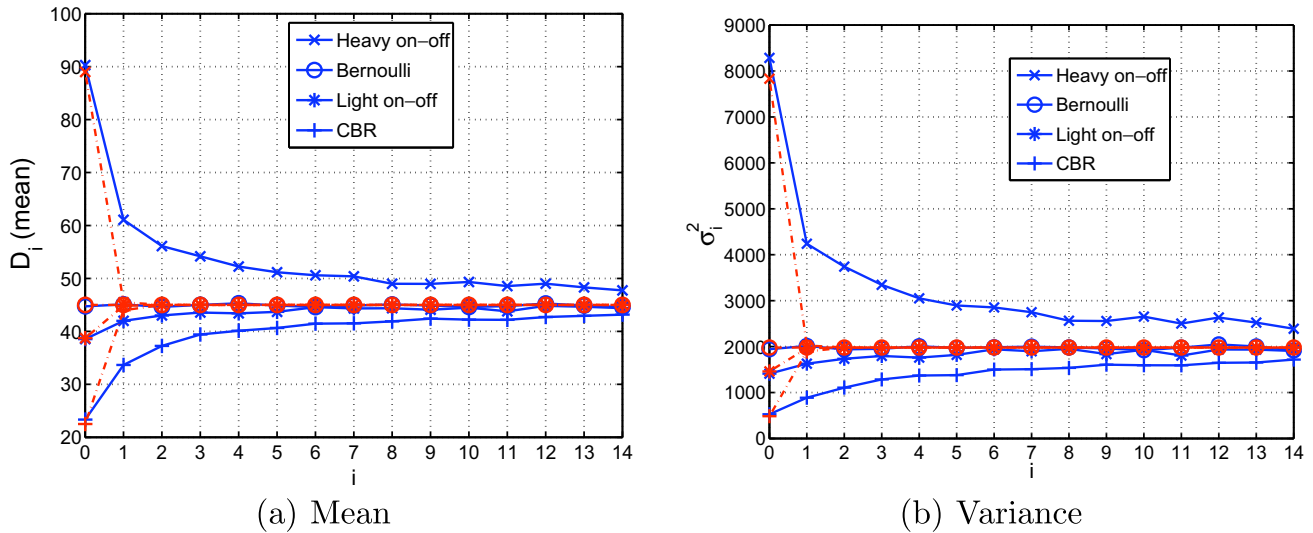


Fig. 5. The mean \bar{D}_i and variance σ_i^2 of single-node delays D_i at n_i in ALOHA networks with $p_m = 1/3, \lambda = 0.25, \mu_A = 0.8, N = 15$. For light on-off, $a_{01} = 0.292, a_{10} = 0.875$; for heavy on-off, $a_{01} = 0.125, a_{10} = 0.375$. The dash-dotted lines represent analytical results while the solid lines are for simulation results.

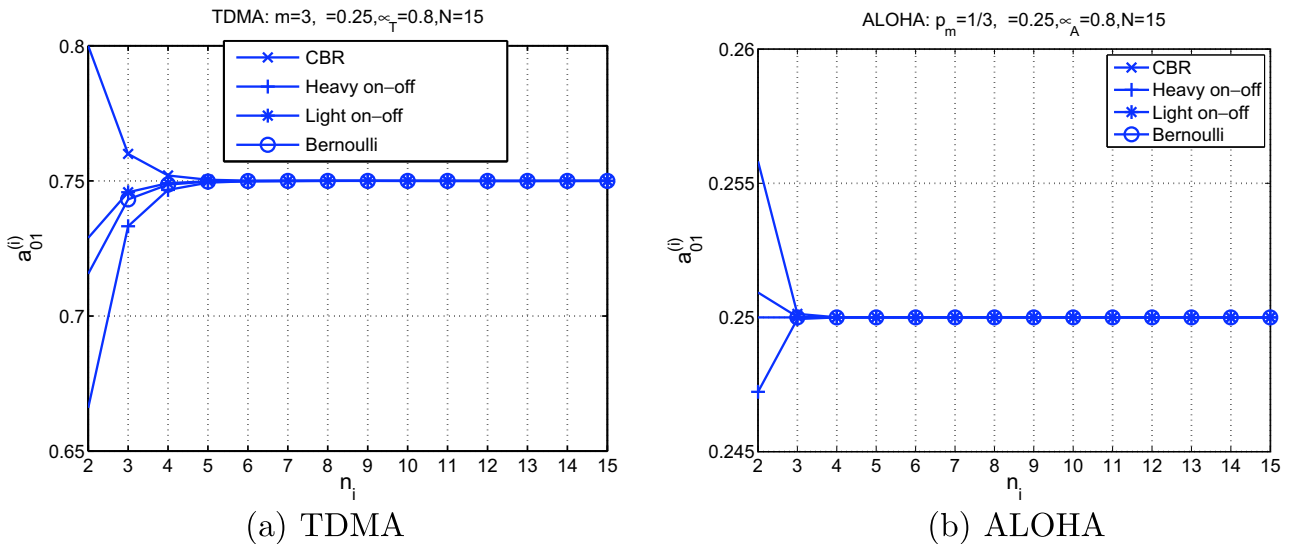


Fig. 6. The convergence of the analytical $a_{01}^{(i)}$ to $m\lambda$ and λ in TDMA and ALOHA networks, respectively, with $m = 3, p_m = 1/3, \lambda = 0.25, \mu_T = \mu_A = 0.8, N = 15$. For light on-off, $a_{01} = 0.292, a_{10} = 0.875$; for heavy on-off, $a_{01} = 0.125, a_{10} = 0.375$.

as an “eigenprocess” or “eigenprocess” since it represents the “eigenvalue” towards which the arrival traffic properties are tending to transform [31].

In this paper, using queueing theory, our analysis shows not only the convergence phenomenon but also the *direction of convergence*. In Figs. 4–6, the departure processes converge to the Bernoulli eigenprocess from different directions, in accordance with the relative burstiness of the eigenprocess. For instance, CBR and heavy on-off have different burstiness and therefore converge to the Bernoulli eigenprocess from opposite directions. More importantly, MAC plays an important role in determining the direction of convergence. In TDMA, the accumulated versions of both on-off and Bernoulli sources are more bursty than the Bernoulli eigenprocess and hence converge from the same direction. On the other hand, in ALOHA, without arrival accumulation, heavy and light on-off con-

verge from opposite directions because of their different burstiness compared to the Bernoulli eigenprocess.

The dependence of the direction of convergence on traffic burstiness can be explained from the viewpoint of traffic. Regard the geometric server as a Bernoulli regulator that regulates the traffic flows by randomly inserting “holes” into the arrival flows [32]. The insertion limits the maximum burstiness that the traffic flow can sustain as it traverses through the network. In the multihop network, after hop-by-hop regulation, the flow is turned into Bernoulli that possesses the “natural” level of burstiness favored by the network under a given traffic load. As such, a heavy bursty flow will converge with the burstiness decreasing while a smooth flow will converge with the burstiness increasing. Notice that here the traffic burstiness is the one after MAC regulation. In other words, the direction of convergence is burstiness- and MAC-dependent [33].

The rate at which the flows converge to the Bernoulli eigenprocess depends on the relative burst size B/B_r and the channel quality μ . Generally, heavy burstiness results in fast convergence (Figs. 4 and 5). A good channel is able to maintain the original traffic statistics, and it takes a very long path for the flows to converge to the eigenprocess. In contrast, a bad channel causes multiple retransmissions and the interdeparture time is mainly determined by the geometric service time. Then, the arrivals converge to the Bernoulli eigenprocess very quickly.

7. End-to-end delay in multihop networks

Our analysis shows that the arrival processes to relays converge to the Bernoulli eigenprocess. However, these arrival processes are not independent and cause correlations in the delays D_i 's. The accurate calculation of the e2e delay variance σ^2 should take into account the correlations, which makes the analysis intractable as the number of nodes N grows. In this section, we first study the correlation between neighboring nodes and then proceed to the e2e correlation based on the direction of convergence.

The correlation between n_i and n_{i+1} can be reflected through the queueing activities of n_i and n_{i+1} when a packet departs from n_i and arrives at n_{i+1} . The interdeparture time T_i depends on the node backlog state and the idle period upon a packet departing moment, *i.e.*, $\tilde{\pi}_B$ and $\tilde{\pi}_i$. Denote $\theta = \tilde{P}_B - P_B$. Previous work proved that for memoryless Bernoulli traffic, not only $\theta = 0$, but also there is no spatial correlation. On the other hand, for temporally correlated traffic like on–off and CBR, the spatial correlation exists and $\theta \neq 0$. Naturally, θ could be used to evaluate the spatial correlation. Between n_i and n_{i+1} , if $\theta > 0$, upon the departure moment, n_i is more backlogged than usual, which will lead to increasing queueing delays at n_i . Meanwhile, because of the non-zero idle period, the packets depart a backlogged n_i in a more bursty manner than departing an idle n_i . Based on queueing theory, a bursty flow results in a longer delay in n_{i+1} than a smooth flow. Therefore, $\theta > 0$ indicates an increase in both D_i and D_{i+1} , *i.e.*, D_i and D_{i+1} are positively correlated. Similarly, if $\theta < 0$, n_i is less backlogged at the packet departure moment than usual and D_i and D_{i+1} are negatively correlated. To start with, we calculate θ for n_0 and n_1 with $\tilde{\pi}_B$. For TDMA,

$$\theta = \begin{cases} -\frac{(r-m)(1-\rho)}{m} < 0, & \text{for CBR,} \\ (1-\rho)\frac{m\lambda-(1-a_{00}^m)}{m\lambda} > 0, & \text{for on–off,} \\ (1-\rho)\frac{m\lambda-(1-\lambda^m)}{m\lambda} > 0, & \text{for Bernoulli.} \end{cases} \quad (36)$$

In contrast to conventional queueing theory, even if the original flow is Bernoulli, spatial correlation exists and $\theta \neq 0$ since TDMA, as a deterministic regulator, destroys the memoryless property of the Bernoulli source. For ALOHA,

$$\theta = \begin{cases} -\frac{1-r\mu\xi^{r-1}}{r\mu} < 0, & \text{for CBR,} \\ (1-\rho)(1-a_{01}-a_{10}) < 0, & \text{for light on–off,} \\ (1-\rho)(1-a_{01}-a_{10}) > 0, & \text{for heavy on–off,} \\ 0, & \text{for Bernoulli,} \end{cases} \quad (37)$$

where $\xi \in (0, 1)$ is the root of $f(x) = \mu_s x^r - x + 1 - \mu_s$. $f'(\xi) = \mu r \xi^{r-1} - 1 < 0$, leading to $\theta < 0$. As a Bernoulli regulator, ALOHA does not change the temporal correlation property and hence $\theta = 0$ for Bernoulli.

Like the direction of convergence, Bernoulli and light on–off traffic flows cause different θ in TDMA and ALOHA while heavy on–off (resp. CBR) is always more (resp. less) bursty than the Bernoulli eigenprocess and thus has consistent $\theta > 0$ (resp. $\theta < 0$). Therefore, θ and the corresponding spatial correlation depend on the MAC-regulated traffic burstiness.

Similar correlations exist in (n_{i+1}, n_{i+2}) , (n_{i-1}, n_i) , and so on. As a result, n_i is correlated with all n_j 's. To determine the e2e correlation, recall that in Section 6, we reveal that if the source flow is more bursty than the Bernoulli eigenprocess, then the relayed flows will converge with the burstiness decreasing, *i.e.*, all the relayed flows are more or equally bursty than the Bernoulli eigenprocess. Then, all the neighboring nodes are positively correlated with $\theta > 0$. This correlation will extend to nodes more than one hop away, say n_i and n_{i+2} , and so on so forth. Overall, the e2e correlation is positive as well and vice versa. As a result, the sign of the correlation between n_i and n_{i+1} can be used as the sign of the e2e correlation.

Though we have derived the sign of the correlation, it is still difficult to explicitly derive $\text{cov}(D_i, D_j)$, especially if $|j - i| > 1$. Even in a simple tandem system of two D/M/1 nodes, the calculation involves of partitioning the state space into four parts and solving them individually [34]. Instead, we use simulation to explore the degree of the e2e correlation.

The e2e correlation is evaluated by the difference between σ^2 and $\beta = \sum_{i=0}^{N-1} \sigma_i^2$. In Fig. 7, the solid lines are for the simulated delay variance σ^2 and the dash-dotted lines represent β , the variance as if the nodes were spatially uncorrelated as assumed in the previous works. Obviously, $\sigma^2 = \beta$ occurs only when the arrival process is Bernoulli in the established GI/Geo/1 model, *e.g.*, Bernoulli in ALOHA, Fig. 7b. Otherwise, a gap exists between σ^2 and β . Sometimes, this gap is too large to ignore the spatial correlation, *e.g.*, for heavy on–off.

Consistent with our analysis, if the source flow is more bursty (smooth) than the limiting Bernoulli eigenprocess, then $\theta > 0$ (< 0) and the correlation is positive (negative). For example, in TDMA, CBR results in $\theta < 0$ and negative correlation, which is confirmed by $\sigma^2 < \beta$ in Fig. 7a. Similarly, all other three flows cause $\theta > 0$, meaning a positive correlation supported by $\sigma^2 > \beta$. In ALOHA, both CBR and light on–off are less bursty than the Bernoulli eigenprocess giving $\theta < 0$ and hence $\sigma^2 < \beta$ as expected. The only flow with a heavier burstiness is heavy on–off that has $\sigma^2 > \beta$ to give $\theta > 0$. Therefore, the sign of θ is sufficient to indicate the sign of the e2e correlation. Remarkably, unlike [35], our results have revealed that with MAC, on–off, as a special MMBP flow, could give rise to both positive and negative correlations.

Smooth traffic causes not only a small per-node delay, but also a negative correlation and a decreased e2e delay variance compared to the uncorrelated case. In contrast, bursty traffic incurs both large per-node delays and a positive e2e correlation. That is why a huge gap in σ^2 exists

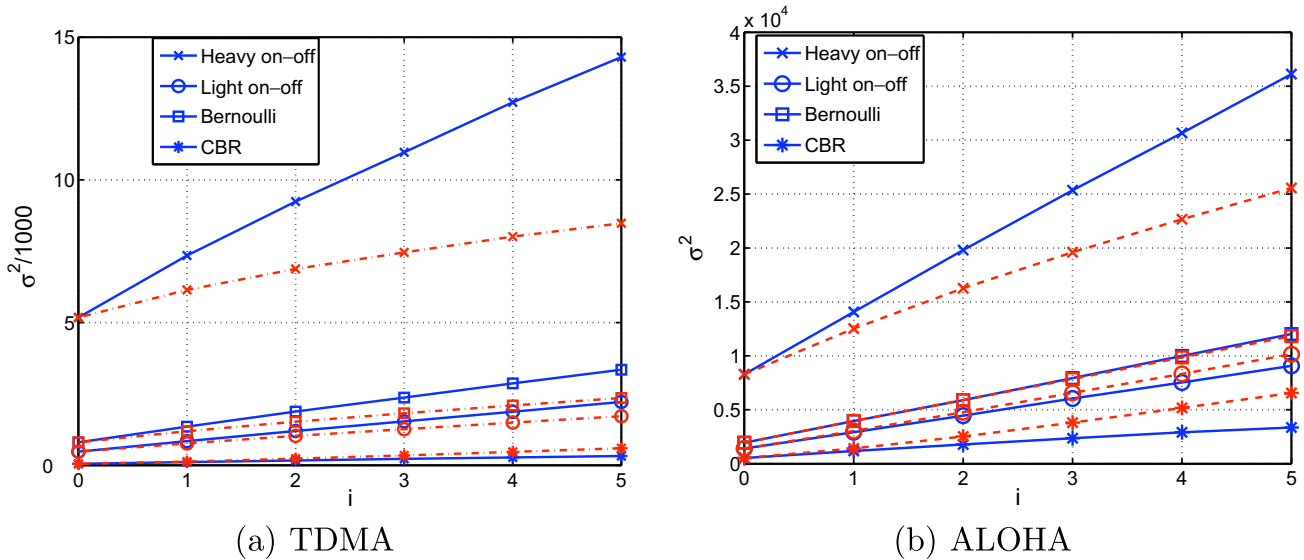


Fig. 7. The e2e delay variance in TDMA and ALOHA networks with $m = 3, p_m = 1/m, \lambda = 0.25, \mu = \mu_T = \mu_A = 0.8, N = 15$. For light on-off, $a_{01} = 0.292, a_{10} = 0.875$; for heavy on-off, $a_{01} = 0.125, a_{10} = 0.375$. The solid lines represent the simulated e2e variance σ^2 with D_i 's correlated while the dash-dotted lines would be the variance if all D_i 's were independent.

between CBR and heavy on-off in Fig. 7, e.g., in TDMA, $\sigma^2_{\text{heavy on-off}} \approx 14\sigma^2_{\text{CBR}}$ and in ALOHA, $\sigma^2_{\text{heavy on-off}} \approx 11\sigma^2_{\text{CBR}}$. For delay-sensitive applications, the heavy bursty flow should be regulated before entering the network.

Though θ itself is not sufficient to determine the degree of the e2e correlation, it still provides an insight. To show this, we define $\eta = \sigma^2/\beta$. If $\eta \rightarrow 1$, then the correlation coefficient decreases to zero. Simulation results show that for bursty traffic, η is non-increasing while for smooth traffic, η is non-decreasing. A similar relationship can be found in the analytical quantity $\theta' = \frac{\partial \theta}{\partial \mu}$ where θ' is decreasing and increasing with μ for bursty and smooth traffic, respectively. More importantly, the separation between θ' for different traffic models is consistent to that between η , showing a great potential of analyzing the e2e delay correlation degree by θ' .

It is interesting to observe that even with the correlations, the e2e delay variance is almost linear with the number of nodes (Fig. 7). Then it is reasonable to assume that the impact of the correlations is uniform in a line network and a product-form joint distribution of all D_i 's could be possible. Moreover, a huge η , say $\eta > 2$, implies that strong correlations exist not only between neighboring nodes, but also between nodes that are more than one hop away (Fig. 7a for heavy on-off). In this case, the assumption used in previous work that the correlation mainly exists between neighboring nodes does not hold.

8. Conclusions

This paper uses queueing theory to analyze the delay performance of two MAC schemes, TDMA and ALOHA, in a wireless line network. The queueing models are established in such a way that the service time is geometric and the access delay is incorporated into the service process for both TDMA and ALOHA. Both delay and departure process of each node are analyzed. For a tractable analysis,

we approximate the departure process of each node by a correlated and bursty on-off process, which is verified to be accurate as the number of nodes increases. We confirm the convergence behavior discussed in [30]. More importantly, we reveal that although all relayed flows converge to the same Bernoulli eigenprocess, they converge from different directions depending on the original traffic burstiness and the underlying MAC.

The departure process characterization gives rise to a parameter θ that can be used to measure the degree of the e2e spatial correlation. θ is consistent with the direction of convergence, depending on both MAC and traffic burstiness. Generally, smooth traffic results in a smaller per-node delay and negative correlation, which leads to much smaller e2e delay variance. In contrast, bursty traffic causes a positive correlation and large e2e delay variance. TDMA, as a deterministic regulator, outperforms ALOHA, which acts as a random regulator. Therefore, a MAC scheme should be designed together with a traffic regulator to optimize the e2e delay performance.

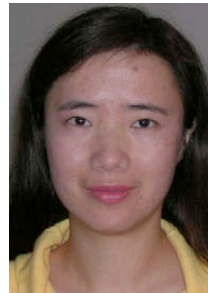
Acknowledgements

The support of the Center for Applied Mathematics (CAM) Fellowship of the University of Notre Dame, NSF (Grants CCF05-15012 and CAREER CNS04-47869), and the DARPA/IPTO IT-MANET program (Grant W911NF-07-1-0028) are gratefully acknowledged.

References

- [1] F.P. Kelly, Blocking probabilities in large circuit-switched networks, *Advances in Applied Probability* 18 (1986) 473–505.
- [2] W. Cheong Lau, S. Qi Li, Traffic analysis in large-scale high-speed integrated networks: validation of nodal decomposition approach, in: *IEEE INFOCOM*, 1993, pp. 1320–1329.
- [3] M. Conti, E. Gregori, I. Stavrakakis, Large impact of temporal/spatial correlations on per-session performance measures: single and multiple node cases, *Performance Evaluation* 41 (2–3) (2000) 83–116.

- [4] R. Nelson, L. Kleinrock, Spatial TDMA: a collision-free multihop channel access protocol, *IEEE Transactions on Communications* 33 (9) (1985) 934–944.
- [5] J.C. Arnbak, W.V. Blitterswijk, Capacity of slotted ALOHA in Rayleigh-fading channels, *IEEE Journal on Selected Areas in Communications* 5 (2) (1987) 261–269.
- [6] F. Borgonovo, M. Zorzi, Slotted ALOHA and CDPA: a comparison of channel access performance in cellular systems, *ACM Wireless Networks* 3 (1) (1997) 43–51.
- [7] Y. Yang, T.-S.P. Yum, Delay distributions of slotted ALOHA and CSMA, *IEEE Transactions on Communications* 51 (11) (2003) 1846–1857.
- [8] J.A. Morrison, Two discrete-time queues in tandem, *IEEE Transactions on Communications* 27 (3) (1979) 563–573.
- [9] J. Hsu, P.J. Burke, Behavior of tandem buffers with geometric input and Markovian output, *IEEE Transactions on Communications* 24 (3) (1976) 358–361.
- [10] M. Sidi, Tandem packet-radio queueing systems, *IEEE Transactions on Communications* 35 (2) (1987) 246–248.
- [11] M.J. Neely, Exact queueing analysis of discrete time tandems with arbitrary arrival processes, in: *IEEE International Conference on Communications (ICC'03)*, vol. 4, 2003, pp. 2221–2225.
- [12] J.J. Hunter, *Mathematical Techniques of Applied Probability*, Academic Press, 1983, ISBN 0123618029.
- [13] H.-W. Ferng, J.-F. Chang, The departure process of discrete-time queueing systems with Markovian type inputs, *Queueing Systems* 36 (1–3) (2000) 201–220.
- [14] H. Bruneel, B.G. Kim, *Discrete-time models for communication systems including ATM*, Kluwer Academic Publishers, 1993, ISBN 0792392922.
- [15] G. Hablinger, Waiting Time, Busy periods and output models of a server analyzed via Wiener–Hopf factorization, *Performance Evaluation* 40 (2000) 3–26.
- [16] D. Park, H.G. Perros, H. Yamashita, Approximate analysis of discrete-time tandem queueing networks with bursty and correlated input traffic and customer loss, *Operations Research Letters* 15 (1994) 95–104.
- [17] P. Jacquet, A.M. Naimi, G. Rodolakis, Routing on asymptotic delays in IEEE 802.11 wireless ad hoc networks, in: *First Workshop on Resource Allocation in Wireless NETWORKS (RAWNET) 2005*, 2005.
- [18] N. Gulpinar, P. Harrison, B. Rustem, Mean-variance optimization of response time in a tandem M/GI/1 router network with batch arrivals, in: *The Third International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks*, 2005.
- [19] F. Eshghi, A.K. Elhakeem, Y.R. Shayan, Performance evaluation of multihop ad hoc WLANs, *IEEE Communications Magazine* (2005) 107–115.
- [20] L.G. Roberts, ALOHA packet system with and without slots and capture, *ACM Sigcomm Computer Communication Review* 5 (2) (1975) 28–42.
- [21] M. Xie, M. Haenggi, A study of the correlations between channel and traffic statistics in multihop networks, *IEEE Transactions on Vehicular Technology* 56 (2007) 3552–3562.
- [22] A. Heindl, Decomposition of general tandem queueing networks with MMPP input, *Performance Evaluation* 44 (2001) 5–23.
- [23] G. Hasslinger, Waiting time, busy periods and output models of a server analyzed via Wiener–Hopf factorization, *Performance Evaluation* 40 (2000) 3–26.
- [24] T.G. Robertazzi, *Computer Networks and Systems: Queueing Theory and Performance Evaluation*, Springer-Verlag, 1994, ISBN 0387973931.
- [25] K.K. Lee, S.T. Chanson, Packet loss probability for bursty wireless real-time traffic through delay model, *IEEE Transactions on Vehicular Technology* 53 (3) (2004) 929–938.
- [26] I. Elhanany, M. Kahane, D. Sadot, On uniformly distributed on/off arrivals in virtual output queued switches with geometric service times, in: *IEEE International Conference on Communications (ICC'03)*, vol. 1, 2003, pp. 173–177.
- [27] M.L. Chaudhry, U.C. Gupta, J.G.C. Templeton, On the relations among the distributions at different epochs for discrete-time GI/Geom/1 queues, *Operations Research Letters* 18 (1996) 247–255.
- [28] I. Mitrani, R. Chakka, Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method, *Performance Evaluation* 23 (3) (1995) 241–260.
- [29] <http://functions.wolfram.com/ElementaryFunctions/Log/29/>.
- [30] B. Prabhakar, R. Gallager, Entropy and the timing capacity of discrete queues, *IEEE Transactions on Information Theory* 49 (2) (2003) 357–370.
- [31] M. Shell, Cascade all-optical shared-memory architecture packet switches using channel grouping under bursty traffic, Ph.D. Dissertation, 2004, http://smartechnology.gatech.edu/bitstream/1853/4892/1/shell_michael_d_200412_phd.pdf.
- [32] N.T. Plotkin, C. Roche, The entropy of cell streams as a traffic descriptor in atm networks, in: *IFIP Performance of Communication Networks*, 1995.
- [33] M. Xie, M. Haenggi, K.-K. Wong, On the end-to-end delay performance of spatially correlated wireless line networks, in: *IEEE ICC'08*, 2008.
- [34] C. Knessl, An explicit solution to a tandem queueing model, *Queueing Systems* 30 (1998) 261–272.
- [35] N. Ryoki, K. Kawahara, T. Ikenaga, Y. Oie, Performance analysis of queue length distribution of tandem routers for QoS measurement, in: *IEEE Symposium on Applications and the Internet (SAINT'02)*, 2002.



Min Xie received the B.S. and M.S. degrees in electrical engineering from the Xidian University, Xi'an, China, in 1996 and 1999, respectively, the M.S. degree in electrical and computer engineering from the National University of Singapore, Singapore, 2001, and the Ph.D. degree in electrical engineering from the University of Notre Dame, IN, USA in 2007. Currently she is a research associate at University College London Adastral Park Campus, UK. Her research interests include queueing theory, wireless communications and networking with emphasis on wireless ad hoc and cognitive radio networks.



Martin Haenggi is an Associate Professor of Electrical Engineering at the University of Notre Dame, Indiana, USA. He received the Dipl. Ing. (M.Sc.) and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology in Zurich (ETHZ) in 1995 and 1999, respectively. After a postdoctoral year at the Electronics Research Laboratory at the University of California in Berkeley, he joined the faculty of the electrical engineering department at the University of Notre Dame in 2001. For both his M.Sc. and his Ph.D. theses, he was awarded the ETH medal, and he received a CAREER award from the U.S. National Science Foundation in 2005. He is a member of the Editorial Board of the Elsevier Journal on Ad Hoc Networks. His scientific interests include networking and wireless communications, with an emphasis on ad hoc and sensor networks.