# Towards an End-to-End Delay Analysis of Wireless Multihop Networks

Min Xie [*,1]

*Department of Electrical Engineering,*

*University of Notre Dame, Notre Dame, IN 46556, USA*

Martin Haenggi [1]

*Department of Electrical Engineering,*

*University of Notre Dame, Notre Dame, IN 46556, USA*

**Abstract**

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. Unlike general two-dimensional networks where there exists sufficient traffic multiplexing that would permit the arrival processes to be approximated as independent, in linear networks with multihop communication, the arrival processes are correlated due to the lack of traffic multiplexing. This paper studies an extreme scenario, a linear network fed with a single flow. A decomposition approach is used to decouple the whole network into isolated nodes. Each node is modeled as a GI/Geo/1 queueing system. We derive the complete per-node delay distribution, accounting for both the queueing delay and access delay. Additionally, we characterize the departure processes by a correlated and bursty on-off traffic model. The per-node analysis provides the e2e delay mean while for the e2e delay variance, the strong correlations between the arrival processes need to be considered. Our study shows that the sign of the correlation coefficients depends on both the MAC scheme and the traffic burstiness, both of which determine the relative burst size of the source flow compared to a Bernoulli process, which constitutes an "eigentraffic" process. There is a wide gap in the e2e delay variances for the source flows with different burst sizes even if they have identical average rates. The relative burst size also determines from which direction and at which rate the departure processes converge to the eigentraffic process after traversing multiple relay nodes.

*Key words:* Multihop, MAC, TDMA, ALOHA, Delay, Correlation

## 1 Introduction

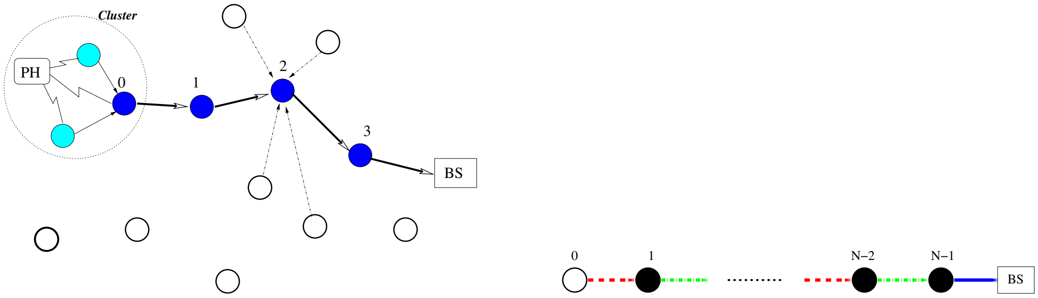
With the growing demand for real-time applications over wireless networks, increasing attention is paid to the delay analysis of transmissions over error-prone channels. In multihop networks, like ad hoc, mesh, and multihop cellular networks, the analysis is more challenging than in single-hop networks due

* Corresponding author.
*Email addresses:* mxie@nd.edu (Min Xie), mhaenggi@nd.edu (Martin Haenggi).

[1] 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 DARPA IT-MANET program (Grant W911NF-07-1-0028) are gratefully acknowledged.

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. We assume a single active path and FIFO as the local packet scheduling discipline. Then, the two-dimensional (2-D) topology (Fig. 1(a)) is reduced to one dimension (1-D), which, in an ideal case, can be further simplified to a regular line network (Fig. 1(b)). Due to the zero inter-flow interference assumption and the equal node spacing, the analysis of the regular line network provides an upper performance bound for general 2-D networks.

From the perspective of queueing theory, this linear topology causes more complications than the 2-D topology. In general network topologies, there are usually multiple flows so that the arrival process to a node is an aggregation of multiple flows (*e.g.*, node 2 in Fig. 1(a)). Such multiplexing would eliminate or weaken the correlations between the arrival processes, permitting these processes to be approximated as independent, which would greatly simplify the analysis. In linear networks, at most two traffic flows (relayed flow and local flow) are multiplexed at each node, and it is hardly possible to assume that the arrival processes are independent. Consider the extreme scenario where there is a single source (Fig. 1(b)). Then, the departure process of node $i$ is exactly the arrival to node $i + 1$ and so forth. In this sense, node $i + 1$ is correlated not only with its immediate neighbor node $i$ but also all other nodes. Such correlations not only affect the network performance but also substantially complicate the e2e analysis.



(a) General network (PH: phenomenon)

(b) Line network

Fig. 1. Wireless multihop networks

Due to the multihop transmission pattern, MAC schemes are needed to efficiently schedule node transmission orders to mitigate interference and achieve spatial reuse, *i.e.*, allowing multiple nodes to transmit simultaneously (*e.g.*, node 2 and $N-2$ in Fig. 1(b)). MAC incurs extra access delays, which should be accounted for in the calculation of the packet delay at each node together with the queueing delay during which the buffer is cleared up. On the other hand, the MAC control also changes the correlations between nodes. In this paper, we take into account the impact of the MAC scheme on the correlations between arrival processes.

### *1.1 Previous work*

The throughput and single-hop delay of many MAC schemes have been comprehensively studied in the literature [1, 2]. 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 [2–4]. 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 rather than the access delay [5–8].

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 [9]. 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.*, [10]. Moreover, for non-Bernoulli and non-Poisson arrivals, it is known that the departure process is correlated with the queue length and arrival process [11], which results in cumbersome expressions [10, 12] that prohibit a scalable e2e analysis. Closed-form solutions for the delay of wireless regular line networks with a single source (like Fig. 1(b)) are available only if the arrival is Bernoulli [6] or the channels are error-free [8]. For other cases, approximations are needed. [13] 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 [14]. Similarly, in [15], 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 linear 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 linear network (Fig. 1(b)) with a single source, considering both the access delay and queueing delay. For a tractable queueing analysis, we consider two simple but typical MAC schemes, $m$-phase spatial TDMA [1] 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 independent) represent the two extremes in terms of the level of the node coordination and are expected to provide upper and lower performance bounds for other meaningful MAC schemes. The arrival processes to every node are all relayed versions of the original traffic flow generated at the source node. Traffic models under investigation include CBR (for voice data [16] and periodic traffic in sensor networks), correlated on-off and Bernoulli (for bursty data).

Our contributions are two-fold. First, we use discrete-time queueing theory to analyze the MAC-controlled nodes, deriving a complete delay and departure process characterization. This analysis provides the e2e delay mean and shows that TDMA outperforms ALOHA in terms of not only throughput, but also delay. It also proves that as the number of hops increases, the departure processes inside the network spatially converge to a MAC-dependent reference Bernoulli process, regardless of the original traffic statistics. Second, we use simulation results to reveal the impact of the MAC schemes and the traffic burstiness on the correlations in the single-node delays and on 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 first present two approaches to derive the delay and departure process characterization of GI/Geo/1 systems. Then we establish and analyze GI/Geo/1 models for each node 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 convergence of the departure processes. Section 7 extends the analysis to the e2e delay and studies the correlation property. Section 8 concludes the paper.

## 2 System Model

The regular line network under consideration (Fig. 1(b)) is composed of $N$ transmitting nodes and a sink or base station (BS). Denote node $i$ by $n_i$ ($i = 0, 1, 2, \ldots, N-1$) and the delay experienced at $n_i$ by $D_i$ with mean $\overline{D}_i$ and variance $\sigma_i^2$. The e2e delay is given by $D = \sum\limits_{i=0}^{N-1} D_i$ with mean $\overline{D}$ and variance $\sigma^2$. A FIFO discipline is used at $n_i$. A flow of fixed-length packets is generated at the source $n_0$ at rate $\lambda$, 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 $\sigma^2 \neq \sum\limits_{i=0}^{N-1} \sigma_i^2$ while previous work usually assumed $\sigma^2 = \sum\limits_{i=0}^{N-1} \sigma_i^2$. If $D_i$'s are positively (negatively) correlated, then $\sigma^2 > \sum\limits_{i=0}^{N-1} \sigma_i^2$ ($\sigma^2 < \sum\limits_{i=0}^{N-1} \sigma_i^2$).

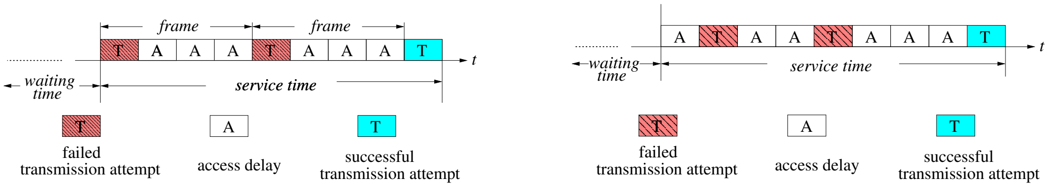
The channel is characterized by a "capture" model [17] with a capture probability $\mu \triangleq \Pr(\text{SNIR} \geq \Theta)$, *i.e.*, a transmission is successful with probability $\mu$. It is assumed that the channels are subject to independent errors (*e.g.*, AWGN or block fading channels). To guarantee 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 $\mu$, denoted by $\mathcal{G}_\mu$. Note that in practice, TDMA and ALOHA result in different capture probabilities [18]. So, we denote the capture probability of TDMA and ALOHA by $\mu_{\mathrm{T}}$ and $\mu_{\mathrm{A}}$, respectively.

The traffic flow to $n_0$ is characterized by the interarrival time $A$, whose probability mass function (pmf) is $a_k = \Pr\{A = k\}$) and probability generating function (pgf) is $A(z) = \sum_{k=0}^{\infty} a_k z^k$. The arrival and departure processes of $n_i$ ($i > 0$) are characterized by the interarrival time $A_i$ and interdeparture time $T_i$, respectively. We consider three typical traffic models, i) CBR, where the packet interarrival time is an integer constant $r = 1/\lambda$; ii) Bernoulli, where a packet is generated with probability $\lambda$ in each time slot; iii) On-off, where the arrival process is modulated by a two-state Markov chain that alternates between ON (1) and OFF (0) states. One packet is generated when the Markov chain is in state ON. The transition probabilities between ON and OFF are $a_{01}$ and $a_{10}$, respectively. The pmf is

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

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 $\mu_{\mathrm{T}}$. So the service time is $S \sim \mathcal{G}_{\mu_{\mathrm{T}}}$ and a TDMA node can be modeled as a GI/Geo/1 system at the frame level, where 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 $\mu_{\mathrm{s}} \triangleq \mu_{\mathrm{A}} p_m$ (given that the arrival and the channel state are independent [2] ). 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 $S \sim \mathcal{G}_{\mu_{\mathrm{s}}}$ 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.


frame
frame
T
A
A
A
T
A
A
A
T
t
waiting time
service time
T
failed transmission attempt
A
access delay
T
successful transmission attempt
A
T
A
A
T
A
A
A
T
t
waiting time
service time
T
failed transmission attempt
A
access delay
T
successful transmission attempt

(a) TDMA

(b) ALOHA

Fig. 2. Packet transmission procedure in TDMA and ALOHA

We use a decomposition approach to analyze the tandem queueing network that decomposes the network into single nodes in isolation [19, 20]. The e2e analysis is based on the single node analysis and thus requires not only the node delay performance but also node departure process 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 $\mu_{\mathrm{A}}$ is the conditional capture probability given that the receiver is listening.

## 3 GI/Geo/1 Queueing Systems

The delay distribution of GI/Geo/1 systems can be derived in two ways. The first one is to use a conventional queueing model that provides the queue length distribution [6, 21]. The second one is to use a delay model [22] that directly tracks the evolution of the delay of the Head-of-Line (HOL) packet in the queue. Both approaches will be used in this paper depending on the arrival process characterization.

### *3.1 Queueing Models*

In a conventional queueing model, the system state is denoted by the queue length [6, 21]. For a GI/Geo/1 system, the pmf of the queue length is [9, 23]

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

where $\rho$ is the traffic intensity, the ratio of the average arrival rate to the service rate. $\gamma$ 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 derived in a geometric form $q_k = (1-\gamma)\gamma^k$ $(k \geq 0)$ [24]. However, (2) holds under the condition that the interarrival times are integers, *i.e.*, there are no bulk or batch arrivals (multiple arriving epochs during a single time unit).

In TDMA GI/Geo/1 systems, since the system is analyzed at the frame level and packet arrivals occur at the slot level, more than one packet may arrive during one frame. So the interarrival times are no longer integer. Specially, consider a TDMA node with CBR arrivals with frame length of $m$ and interarrival time of $r$ 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 $\mathcal{G}_\mu$ 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*

$$\pi_k = \begin{cases} 1-\rho & k=0 \\ \sum_{j=1}^{m} C_j \gamma_j^{k-1} & k>0 \end{cases} \tag{3}$$

*where $\rho = m/r\mu$ is the traffic intensity, $C_j$ is a normalizing constant and $\{\gamma_j \,|\, j = 1, 2, \ldots, m\}$ 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 $Y(t) = 1, 2, \ldots, r$ is the number of *slots* to the next packet arrival. Divide the set $\{1, 2, \ldots, r\}$ into two parts $\mathbb{Y}_0 \triangleq \{1, \ldots, \Delta\}$ and $\mathbb{Y}_1 \triangleq \{\Delta + 1, \ldots, r\}$

($\Delta \triangleq r-m$), where the subscript represents the number of packets arriving during one frame. Denote the steady-state system probability by $Q(k,y) := \lim_{t\to\infty} \Pr\{Q(t)=k, Y(t)=y\}$. The balance equations are

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

Define the row vector $\vec{v}_k := \{Q(k,1),\ldots,Q(k,r)\}$ ($k\geq 0$). 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

$$\mathbf{M_0} = \begin{bmatrix} \mathbf{0} & (1-\mu)\mathbf{I}_m \\ \\ \mathbf{0}_\Delta & \mathbf{0} \end{bmatrix}, \quad \mathbf{M_1} = \begin{bmatrix} \mathbf{0} & \mu\mathbf{I}_m \\ \\ (1-\mu)\mathbf{I}_\Delta & \mathbf{0} \end{bmatrix} - \mathbf{I}, \quad \mathbf{M_2} = \begin{bmatrix} \mathbf{0} & \mathbf{0}_m \\ \\ \mu\mathbf{I}_\Delta & \mathbf{0} \end{bmatrix}.$$

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 [25], $\vec{v}_k$ is solved as $\vec{v}_k=\mathbf{C}\boldsymbol{Z}^k\mathbf{\Phi}$, where the diagonal matrix $\boldsymbol{Z}=\text{diag}(z_j)$ and the matrix $\mathbf{\Phi}=[\vec{\phi}_j]^T$ 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 $\vec{\phi}_j=\{\phi_j(1),\phi_j(2),\ldots,\phi_j(r)\}$. 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

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

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

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* $\mathcal{G}_\mu$ *and constant interarrival time* $r/m$ *(*$r > m$, $r, m \in \mathbb{N}$ *and* $r/m$ *is irreducible). The pmf of the delay is*

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

*where* $\{\gamma_j \mid j = 1, 2, \ldots, m\}$ *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* $\mu_T x^r - x^m + 1 - \mu_T = 0$.

**PROOF.** In terms of slots, the packet delay $D_0$ is composed of three independent parts, the access delay $D_A \in \{0, \ldots, m-1\}$, the waiting time $D_W$, and the service time $D_S$. If a packet arrives in the middle of frame $t$, then the access delay is $D_A = m - Y(t)$ $(Y(t) < m)$ and $Y(t+1) = Y(t) + \Delta = r - D_A \in \mathbb{Y}_1$.

The probability that this packet sees $k-1$ packets in the buffer is

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

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

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

Inverse z-transform yields (6).

Theorem 1 and 2 can be regarded as a generalized analysis of D/Geo/1 systems with non-integer interarrival times. In fact, (2) represents a special case of $m = 1$ in Theorem 1. With non-integer interarrival times, among the $m > 1$ roots $\{\gamma_j \mid j = 1, 2, \ldots, m\}$ inside the unit circle, there exist complex and negative real roots. In this case, the pmf calculation of the queue length and delay is only possible numerically if $m$ and $r$ are large.

If $r < 2m$, we simplify the results by ignoring the complex and negative roots and considering the unique real positive root $\xi_1 \in (0,1)$. Then, (6) is reduced to $d_k \approx (1-\xi_1)\xi_1^{k-1}$ $(k \geq 1)$, which, however, is still difficult to calculate for large $m$ and $r$ since $\xi_1$ is the root of a high degree polynomial $\mu x^r - x^m + 1 - \mu = 0$. Lemma 3 gives an approximate calculation of $\xi_1$.

**Lemma 3** *Consider the 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 $\xi_1$ in the region $(0,1)$ can be well approximated by*

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

**PROOF.** Based on Descartes' Sign Rule, there are exactly two real positive roots, one of which is 1 and the other is $\xi_1 \in (0,1)$. A single local minimum $x_{min} = \rho^{\frac{1}{\Delta}} < 1$ lies between $\xi_1$ and 1. Using two inequalities, $-\frac{1-\rho}{\rho} < \ln\rho \leq \Delta(\rho^{\frac{1}{\Delta}} - 1)$ [26], $x_{min}$ is lower bounded by $x_{min} \gtrapprox 1 - (1-\rho)/(\Delta\rho)$. Assuming an equal distance from $x_{min}$ to 1 and $\xi_1$, *i.e.*, $\xi_1 \approx 2x_{min} - 1$, leads to (8).

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

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

### *3.2 Delay Models*

In a delay model, the system state is denoted by the current delay of the HOL packet [22]. The advantage of the delay model is the convenience to trace both the evolution of the packet delay and the interdeparture time. Consider a GI/Geo/1 system with on-off arrivals $(a_{01}, a_{10})$ (1). The delay distribution is shown to be geometric in [27], but the mean is not calculated. In Lemma 4, we use the delay model to derive the mean.

**Lemma 4** *Consider a discrete-time GI/Geo/1 queueing system with service rate* $\mu$ *and on-off arrival with transition probabilities* $a_{01}$ *and* $a_{10}$*. Then, the delay is geometrically distributed with parameter* $1-\alpha$ *where*

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

**PROOF.** Let the system state be the delay of the HOL packet. Negative states indicate an idle server. All probabilities of going beyond a delay $-1$ are included in the state $-1$. The transition probabilities $P_{jk}$ are

$$P_{jk} = \begin{cases} \mu b_l \quad j \geq 0,\ k = j+1-l \\ 1-\mu\ j \geq 0,\ k = j+1 \\ a_{01} \quad j = -1,\ k = 0, \end{cases} \tag{11}$$

where $b_l = a_l$ for $k \geq 0$ and $b_l = \sum_{h=l}^{\infty} a_h = a_{10}(1-a_{01})^{l-2}$ for $k = -1$. From (11), we obtain the steady-state probability $\pi_j = \pi_0 \alpha^j$ for non-negative states $j \geq 0$. Since the pmf $\{d_j \,|\, j \geq 1\}$ of the delay involves only non-negative states

$$d_{j+1} \triangleq \frac{\mu \pi_j}{\sum_{k=0}^{\infty} \mu \pi_k} = \frac{\pi_j}{\sum_{k=0}^{\infty} \pi_k}, \tag{12}$$

where $\sum_{k=0}^{\infty} \mu \pi_k$ is the normalization constant and the factor $\mu$ is needed to account for successful packet transmissions, the delay is geometrically distributed with parameter $1-\alpha$.

From (11), we calculate the system busy probability $\pi_B = \sum_{k \geq 0} \pi_k = m\lambda/\mu = \rho$ and idle probability $\pi_I = 1-\rho$ at any frame. The delay model can also be used to derive the departure process characterization.

**Lemma 5** *Consider a discrete-time GI/Geo/1 queueing system with service rate $\mu$ and on-off arrival with transition probabilities $(a_{01}, a_{10})$. The interdeparture time $T$ has the pgf*

$$G_T(z) = \tilde{\pi}_B S(z) + (1-\tilde{\pi}_B)\frac{a_{01} z}{1-(1-a_{01})z} S(z), \tag{13}$$

*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 modified from (11) to

$$P_{jk} = \mu(1-\mu)^{l-1} a_h, \quad \begin{cases} k = j + l - h, \; j \geq 0 \\ k = l - h, \qquad j < 0. \end{cases} \tag{14}$$

The absolute value of the negative state represents the system idle time. Denote the steady-state probability by $\pi_j$. The interdeparture time $T$ is the sum of the packet service time $S$ and system idle time, *i.e.*, $T = S - j$ if the system is in negative states $j < 0$ and $T = S$ if the system is busy with probability $\widetilde{\pi}_B = \sum_{j<0} \pi_j$ when the packet departs. Given independent arrival and service processes, the pgf of the interdeparture time $T$ is $G_T(z) = \widetilde{\pi}_B S(z) + \sum_{j=1}^{\infty} \pi_{-j} z^j S(z)$. From (14), we obtain $\pi_{-j} = (1-a_{01})^{j-1}\pi_{-1}$ for $j \geq 1$ and $\widetilde{\pi}_I = 1 - \widetilde{\pi}_B = \sum_{j<0} \pi_j = \pi_{-1}/a_{01}$. For stable systems, the average departure rate equals to the average arrival rate, *i.e.*, the average interdeparture time is $\overline{T} = 1/\lambda = (a_{01} + a_{10})/a_{01}$, from which we can calculate $\widetilde{\pi}_B$ and $\widetilde{\pi}_I$. Plugging these parameters into $G_T(z)$ yields (13).

Recall that $\pi_I = 1 - \rho$ while $\widetilde{\pi}_I = (a_{10} + a_{01})(1-\rho)$. The conditional idle probability $\widetilde{\pi}_I$ upon the departure moment is identical to the system idle probability $\pi_I$ at any moment only if $a_{10} + a_{01} = 1$, *i.e.*, the arrival is Bernoulli, which is consistent with previous work.

Note that the second part of $G_T(z)$ (13) is a convolution of two geometric distributions $\mathcal{G}_\mu$ and $\mathcal{G}_{a_{01}}$. Unsurprisingly, the departure processes would exhibit a state explosion problem if it were fed into a tandem network [12]. A natural approximation is to model the departure process as an on-off process, which captures both the correlation and burstiness property of a traffic flow. The corresponding transition probabilities are calculated as follows

$$a_{11}^{'} = \Pr\{T = 1\} = \widetilde{\pi}_B \mu = \mu - \frac{a_{01}(1-\rho)}{\rho}, \quad a_{01}^{'} = \frac{\lambda}{1-\lambda}(1 - a_{11}^{'}). \qquad (15)$$

The conventional queueing model can be used for the delay analysis of conventional GI/Geo/1 systems with integer interarrival times while the delay model is convenient for the delay analysis of systems with non-integer interarrival times and the departure process characterization. In the following sections, every node analysis consists of two parts, the delay and the departure characterization. The pmf of $D_i$ is denoted by $\{d_j^{(i)}\}$ and the arrival to $n_i$ (or the departure of $n_{i-1}$) is characterized by parameters like $(a_{10}^{(i)}, a_{01}^{(i)})$ or $r^{(i)}$.

## 4 Single Node Analysis for TDMA

A TDMA node is modeled as a GI/Geo/1 system with service rate $\mu_{\mathrm{T}}$ at the frame level. Note that the average arrival rate should be cumulated over the frame of $m$ slots. Therefore, given the arrival rate $\lambda$ packets/slot, the traffic intensity of a TDMA GI/Geo/1 system is $\rho \triangleq m\lambda/\mu_{\mathrm{T}} < 1$.

### *4.1 Source Node: CBR Traffic*

For a TDMA node with CBR traffic of rate $1/r$, using the conventional queueing model (Section 3.1), Theorem 2 shows that an accurate calculation of the delay's pmf is possible only numerically. For $r < 2m$, Lemma 3 provides an approximate geometric expression for $D_0$'s distribution. For the heaviest stable traffic load $r = m + 1$ that can be accommodated by the system, we can use the delay model to derive the exact delay distribution.

**Theorem 6** *Consider a D/Geo/1 system with interarrival time $r/m$ and service rate $\mu_T$. If $r = m + 1$, the pgf, mean and variance of the delay $D_0$ are*

$$G_{D_0}(z) = \frac{(1-z^m)z}{(1-\mu_T)z^{m+1} - z + \mu_T} \cdot \frac{1-\rho}{\rho}, \qquad \rho = \frac{m}{r\mu_T} \tag{16}$$

$$\overline{D}_0 = \frac{1}{2(1-\rho)}, \quad \sigma_0^2 = \frac{1}{4(1-\rho)^2} - \frac{m+2}{6(1-\rho)}. \tag{17}$$

**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

$$P_{kj} = \begin{cases} \mu & k \geq 0,\, j = k - \Delta, \\ \\ 1-\mu & k \geq 0,\, j = k + m, \\ \\ 1 & k < 0,\, j = k + m, \end{cases} \tag{18}$$

where $\Delta = r - m$. The $\Delta$ negative states represent an idle system. For $r = m + 1$, the steady-state probabilities for non-negative states are

$$\pi_k = \begin{cases} \mu\pi_{k+1} & 0 \leq k < m - 1 \\ \\ \mu\pi_{k+1} + \mu\pi_0 & k = m - 1 \\ \\ \mu\pi_{k+1} + (1-\mu)\pi_{k-m} & k \geq m, \end{cases} \tag{19}$$

which can be transformed to the pmf $\{d_k^{(0)} \mid k \geq 1\}$ based on (12). Then, the pgf of the delay is

$$G_{D_0}(z) = \frac{(z^m - 1)z\mu_{\mathrm{T}}}{z - \mu_{\mathrm{T}} - (1 - \mu_{\mathrm{T}})z^{m+1}} d_1^{(0)}, \tag{20}$$

which contains one unknown parameter $d_1^{(0)}$. From $G_{D_0}(1) = 1$, it follows that

$$d_1^{(0)} = \frac{1 - (m+1)(1 - \mu_{\mathrm{T}})}{m\mu_{\mathrm{T}}} = 1 + \frac{1}{m} - \frac{1}{\mu_{\mathrm{T}}}. \tag{21}$$

Plugging (21) into (20) leads to (16). The delay mean and variance (17) are calculated through the first two derivatives of $G_{D_0}(z)$ at $z = 1$.

Note that (18) hold for all $r > m$. But for $r > m + 1$, the pgf contains $\Delta > 1$ unknown parameters and cannot be solved as for $r = m + 1$. So we need the approximate delay analysis (9) if $r > m + 1$. Plugging $\Delta = 1$ into (9) and comparing with the exact analysis (17) justify that the approximation is tight if $\rho \to 1$.

The departure process is studied at the frame level since according to the TDMA policy, the packet departs only at the boundary of frames. Theorem 7 proves that for $m < r < 2m$, the departure is an on-off process.

**Theorem 7** *Consider a D/Geo/1 system with service rate $\mu_T$ 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.** Consider the packet departure moment. With probability $\tilde{\pi}_B$, the queue is non-empty, and the interdeparture time is $T_0 = S$; while with probability $\tilde{\pi}_I = 1 - \tilde{\pi}_B$, the queue is empty, and the interdeparture time is $T_0 = 1 + S$, the service time $S$ plus the system idle time, which is exactly one frame for $r < 2m$. The system idle and busy probabilities $\tilde{\pi}_I$ and $\tilde{\pi}_B$ viewed by the departing packet can be deduced from the stability condition $\overline{T}_0 = r/m$. That is, $\tilde{\pi}_B\mu_\mathrm{T} = 1 - \dfrac{\Delta\mu_\mathrm{T}}{m}$. The pgf $G_{T_0}(z) = \tilde{\pi}_B S(z) + (1-\tilde{\pi}_B)zS(z)$ gives rise to a closed-form pmf $\{t_0(k) \mid k \geq 1\}$ of $T_0$:

$$t_0(k) = \begin{cases} 1 - \dfrac{\Delta\mu_\mathrm{T}}{m} & k = 1 \\ \left(\mu_\mathrm{T}(1-\mu_\mathrm{T})^{k-2}\right)\dfrac{\Delta\mu_\mathrm{T}}{m} & k > 1, \end{cases} \tag{22}$$

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

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. The corresponding transition probabilities $\{a_{01}^{(1)}, a_{10}^{(1)}\}$ can be derived from (18). More specifically, $a_{11}^{(1)} = 1 - a_{10}^{(1)}$ is the probability that two packet depart the system consecutively, *i.e.*, the system stays in non-negative states (since negative states imply an empty buffer) in two consecutive frames and is accompanied with a successful transmission with probability $\mu_{\mathrm{T}}$,

$$\begin{aligned} a_{11}^{(1)} &= \Pr\{D(t+1) \geq 0, S(t+1) = 1 \mid D(t) \geq 0, S(t) = 1\} \\ &= \frac{\mu_{\mathrm{T}} \sum\limits_{k=\Delta}^{\infty} \pi_k}{\sum\limits_{l \geq 0} \pi_l} = \frac{\mu_{\mathrm{T}}(\rho - \sum\limits_{k=0}^{\Delta-1} \pi_k)}{\rho} = \mu_{\mathrm{T}} - \frac{1-\rho}{\rho} = 1 - \frac{\Delta \mu_{\mathrm{T}}}{m}. \end{aligned} \tag{23}$$

The numerator excludes states 0 through $\Delta - 1$ since these states transit to negative states after a successful transmission. Besides, from (18), we obtain $\sum\limits_{k=0}^{\Delta-1} \pi_k = \frac{\pi_I}{\mu_{\mathrm{T}}} = \frac{1-\rho}{\mu_{\mathrm{T}}}$. Then, based on $a_{01}^{(1)}/(a_{01}^{(1)} + a_{10}^{(1)}) = m/r$, we have $a_{10}^{(1)} = \Delta\mu_{\mathrm{T}}/m$ and $a_{01}^{(2)} = \mu_{\mathrm{T}}$, consistent with the result given in Theorem 7.

### *4.2 Source Node: On-off traffic*

For CBR traffic with constant interarrival time $r > m$, there is at most one packet arrival during a frame. However, for bursty on-off traffic, due to arrival cumulation, there may be multiple arrivals during one frame, constituting a batch arrival process. The conventional queueing model (Section 3.1) is not convenient to solve the delay distribution problem for batch arrivals. So, we use the delay model (Section 3.2) instead.

**Theorem 8** *Consider a GI/Geo/1 system with batch arrivals, which are generated by an on-off source $(a_{01}, a_{10})$ in a frame of $m$ time slots. The service rate is $\mu_T$. Then, the pgf, mean and variance of the delay $D_0$ are*

$$G_{D_0}(z) = \frac{(1-\rho)H_0(z)/\rho}{1-\mu_T(1-\frac{a_{01}}{\lambda})z^{m-1}-(1-\mu_T)z^m - H_0(z)}, \quad \rho = m\lambda/\mu_T \tag{24}$$

$$\overline{D}_0 = \frac{1}{1-\rho}\Big(\frac{\rho-\lambda}{a_{01}} - \rho - \frac{m-3}{2}\Big), \quad H_0(z) = \frac{a_{01}(1-z^m)}{1-z} \tag{25}$$

$$\sigma_0^2 = \frac{1}{(1-\rho)^2}\Big(\frac{m^2-1}{12} + \frac{(m-1)(m-2)\rho}{6} - \frac{(1-\mu_T)\rho^2+(m-2)\rho+\lambda}{a_{01}} + \frac{(\rho-\lambda)^2}{a_{01}^2}\Big). \tag{26}$$

**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:

$$P_{lj} = \begin{cases} \mu_{\mathrm{T}} a_k, & j = l+m-k,\ l \geq 0 \\ 1-\mu_{\mathrm{T}}, & j = l+m, \quad l \geq 0 \\ 1, & j = l+m, \quad l < 0. \end{cases} \tag{27}$$

The negative states indicate an empty buffer. Denoting the steady-state probabilities by $\{\pi_k\}$, we derive the balance equations

$$\pi_k = \begin{cases} \pi_{k-m} + \mu_{\mathrm{T}} \sum_{j=0}^{\infty} a_{j+m-k} \pi_j, & 0 \le k < m \\ (1-\mu_{\mathrm{T}})\pi_{k-m} + \mu_{\mathrm{T}} \sum_{j=k-m+1}^{\infty} a_{j+m-k} \pi_j, & k \ge m \\ \dfrac{\mu_{\mathrm{T}}}{1-a_{00}^m} \sum_{j=0}^{\infty} a_{j+m-k} \pi_j, & k < 0, \end{cases} \tag{28}$$

which leads to $\pi_k = a_{00}^{|k|}\pi_0$ for $k < 0$ with $\pi_0 = \dfrac{\mu_{\mathrm{T}} a_m}{1-a_{00}^m} \sum_{j=0}^{\infty} a_{00}^j \pi_j$. Since the delay distribution $\{d_k^{(0)} \mid k \ge 1\}$ involves only the non-negative states $\{\pi_k \mid k \ge 0\}$ as in (12), the pgf $G_{D_0}(z)$ of $D_0$ can be calculated by multiplying both sides of (28) by $z^k$ for the non-negative states $k \ge 0$ and plugging $\pi_k = a_{00}^{|k|}\pi_0$. The obtained pgf contains only one unknown parameter $\pi_0$ (or $d_1^{(0)}$), which is deduced from $G_{D_0}(1) = 1$ to be $d_1^{(0)} = \dfrac{a_{01}}{1-a_{01}} \cdot \dfrac{1-\rho}{\rho}$ and then leads to (24). Differentiating $G_{D_0}(z)$ gives rise to the mean (25) and variance (26).

The system idle probability is $\pi_I = \sum_{k<0} \pi_k = 1 - \rho$. The pmf $\{d_k^{(0)} \mid k \ge 1\}$ can be derived from $G_{D_0}(z)$ using the inverse z-transform. Comparing (24) with Lemma 4 reveals the impact of TDMA on the single node delay distribution.

Though TDMA results in a completely different and more complex delay distribution from that without TDMA (Lemma 4), an analogous characterization of the departure process exists (Lemma 9) as in Lemma 5 without TDMA.

**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*

$$G_{T_0}(z) = \Big(\widetilde{\pi}_B + (1-\widetilde{\pi}_B)\frac{(1-a_{00}^m)z}{1-a_{00}^m z}\Big)S(z), \quad \widetilde{\pi}_B = 1 - \frac{(1-a_{00}^m)(1-\rho)}{m\lambda}. \tag{29}$$

**PROOF.** Let the system state be the delay of the HOL packet at a packet departing moment in terms of *slots*. The transition probabilities are

$$P_{lj} = \mu_{\mathrm{T}}(1-\mu_{\mathrm{T}})^{k-1}a_h, \begin{cases} j = l + km - h, \, l \geq 0 \\ \\ j = km - h, \qquad l < 0, \end{cases} \tag{30}$$

Negatives states represent the system idle time. Denote the steady state probability by $\pi_k$. We obtain $\pi_k = a_{00}^{|k|}\pi_0$ for $k < 0$ and $\widetilde{\pi}_I = \sum_{k=-1}^{-\infty} \pi_k = \dfrac{a_{00}\pi_0}{1-a_{00}}$ that is the system idle probability viewed by a departure. Unlike the delay distribution, the interdeparture time $T_0$ involves only the negative states,

$$\begin{aligned} G_{T_0}(z) &= \widetilde{\pi}_B S(z) + S(z) \sum_{k=-1}^{-\infty} z^{|k|} \sum_{j=1-(k+1)m}^{-km} \pi_k \\ &= \Big(\widetilde{\pi}_B + \frac{(1-a_{00}^m)z}{1-a_{00}^m z} \cdot \frac{a_{00}\pi_0}{1-a_{00}}\Big)S(z). \end{aligned} \tag{31}$$

Plugging $\widetilde{\pi}_I = 1 - \widetilde{\pi}_B = \dfrac{a_{00}\pi_0}{1-a_{00}}$ and deducing $\widetilde{\pi}_B$ from the stability condition lead to (29).

Note that (29) differs from (13) in that $a_{00}$ is replaced by $a_{00}^m$ and $\lambda$ is 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, just like $a_{00}$ and $\lambda$ at the slot level. Therefore, the MAC control does not change the departure process characterization. Like in Section 3.2, the departure process (31) needs to be simplified for a tractable e2e analysis. As usual, we approximate it as an on-off process with transition probabilities $a_{10}^{(1)}$ and $a_{01}^{(1)}$ calculated from (28) based on the principle provided in (23), *i.e.*,

$$a_{11}^{(1)} = 1 - a_{10}^{(1)} = \frac{\mu_{\mathrm{T}} \sum\limits_{k=0}^{\infty} \pi_k \sum\limits_{j=1}^{m+k} a_j}{\rho} = \mu_{\mathrm{T}} - (1 - a_{00}^m)\frac{1-\rho}{\rho}, \tag{32}$$

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

The delay and departure process of Bernoulli traffic are analyzed by setting $a_{01} + a_{10} = 1$. The impact of TDMA is reflected through the deviation of the departure process from the arrival Bernoulli process since without TDMA, a geometric server and a Bernoulli arrival guarantees the equivalence of the departure process to the arrival process [9, 11]. In other words, the TDMA control generates a batched MMBP arrival process at the frame level and destroys the memoryless property.

In summary, for all three traffic models, smooth CBR, memoryless Bernoulli, and bursty and correlated on-off, all the departure processes of the source node $n_0$ can be characterized by an on-off process with transition probabilities

$(a_{01}^{(1)}, a_{01}^{(1)})$. 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 $(a_{01}^{(1)}, a_{01}^{(1)})$. So $n_1$ is modeled as a GI/Geo/1 system, whose delay distribution is geometric as proved in Lemma 4. Here a modification is required since the delay should be evaluated at the slot level while the system is analyzed at the frame level, *i.e.*, the pmf is $d_{km+1}^{(1)} = (1-\alpha)\alpha^k$ $(k \geq 0)$, where $\alpha$ is given in (10). The pgf is $G_{D_1}(z) = \dfrac{(1-\alpha)z}{1-\alpha z^m}$. The mean and variance are

$$\overline{D}_1 = 1 + \frac{m\alpha}{1-\alpha} = 1 + m\varepsilon, \quad \sigma_1^2 = \frac{m^2\alpha}{(1-\alpha)^2} = m^2\varepsilon(1+\varepsilon), \tag{33}$$

where $\varepsilon \triangleq \dfrac{\rho}{1-\rho} \cdot \dfrac{1-\mu}{a_{01}^{(1)}}$. The departure process of such a GI/Geo/1 system can be approximated as another on-off process with $(a_{01}^{(2)}, a_{01}^{(2)})$ calculated as in (15), in which $\lambda$ replaced by $m\lambda$. The remaining relay nodes are analyzed in the same way by iteratively calculating $(a_{01}^{(i+1)}, a_{01}^{(i+1)})$ from $(a_{01}^{(i)}, a_{01}^{(i)})$.

## 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 that are modeled as GI/Geo/1 but analyzed at the slot level. So all interarrival times are integer and there are no batch arrivals. The conventional queueing model (Section 3.1) is used for the delay analysis. Note that the service rate is defined as $\mu_{\mathrm{s}} = \mu_{\mathrm{A}} p_m$ and the traffic intensity is $\rho = \lambda/\mu_{\mathrm{s}}$.

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 $\xi$ of the polynomial $\mu_{\mathrm{s}} y^r - y + 1 - \mu_{\mathrm{s}}$. Based on Theorem 2, the delay $D_0 \sim \mathcal{G}_{1-\xi}$. However, if $r$ is large, $\mu_{\mathrm{s}} y^r - y + 1 - \mu_{\mathrm{s}} = 0$ can be solved only numerically. Using a similar approach as in Lemma 3, we approximate $\xi$ as follows

$$\xi \approx 1 - \frac{2(1-\rho)}{(r-1)\rho}. \tag{34}$$

The mean and variance of $D_0$ are

$$\overline{D}_0 = \frac{1}{1-\xi} \approx \frac{(r-1)\rho}{2(1-\rho)}, \quad \sigma_0^2 = \frac{\xi}{(1-\xi)^2} \approx \frac{(r-1)\rho}{2(1-\rho)}\Big(\frac{(r-1)\rho}{2(1-\rho)} - 1\Big). \tag{35}$$

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* $\mu_s$*. 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* $\xi$ *is the unique root of* $\mu_s y^r - y + 1 - \mu_s = 0$ *in the region* $(0, 1)$*.*

**PROOF.** Let the system state be the delay of the HOL packet. Given the delay's pmf $d_k^{(0)} = (1-\xi)\xi^{k-1}$ $(k \geq 1)$ and the relationship between $\{\pi_k | k \geq 0\}$ and $\{d_k^{(0)} \mid k \geq 1\}$(12), plugging $m = 1$ into the transition probabilities (18), it can be proved that $\pi_B = \sum_{j\geq 0} \pi_j = 1 - \rho$ and

$$a_{11}^{(1)} = \frac{\mu_{\mathrm{s}} \sum_{k=r-1}^{\infty} \pi_k}{\sum_{j\geq 0} \pi_j} = \frac{\mu_{\mathrm{s}}(\rho - \sum_{k=0}^{r-2} \pi_k)}{\rho} = \mu_{\mathrm{s}}\xi^{r-1} = 1 - \frac{1-\mu_{\mathrm{s}}}{\xi}, \tag{36}$$

where the last part is obtained from $\mu_{\mathrm{s}}\xi^r - \xi + 1 - \mu_{\mathrm{s}} = 0$. Then $a_{10}^{(1)} = 1 - a_{11}^{(1)}$ and $a_{01}^{(1)} = a_{10}^{(1)}\lambda/(1-\lambda)$.

For on-off traffic, Lemma 4 proves that the delay $D_0 \sim \mathcal{G}_{1-\alpha}$ (10). The departure process is approximated as an on-off process as in (15). 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.

## 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. The on-off process is featured by its correlation and burstiness, both of which 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). Specially, we set $a_{10} = (1-\lambda)/2$ for heavy on-off and $a_{10} = 1-\lambda/2$ for light on-off. We compare the delay of single nodes from the following aspects:

- *Source node: CBR vs. on-off.* The ratio of the delay means $\overline{D}_0$ for on-off and CBR is

$$\eta_{\text{TDMA}} = \frac{\overline{D}_{0,\text{on-off}}}{\overline{D}_{0,\text{cbr}}} = 2\Big(\frac{B}{B_r} \cdot \frac{m-\mu_{\text{T}}}{\mu_{\text{T}}} - \rho\Big) - m + 3 > 1 \tag{37}$$

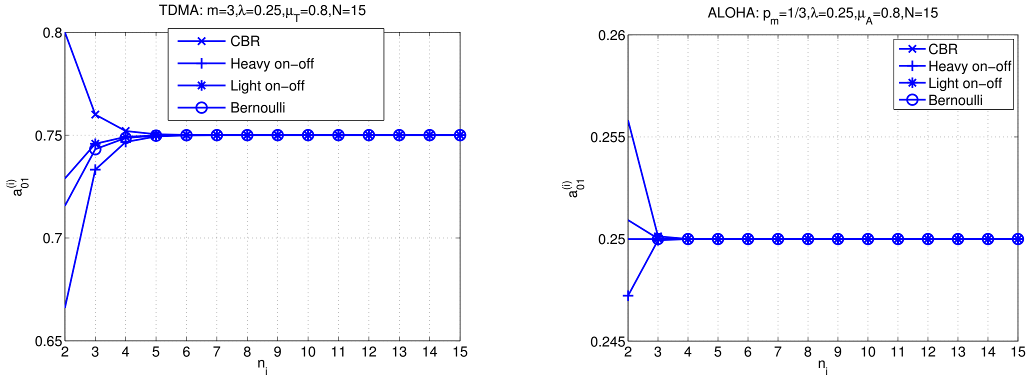
$$\eta_{\text{ALOHA}} = \frac{\overline{D}_{0,\text{on-off}}}{\overline{D}_{0,\text{cbr}}} = (1-\xi)\Big(1 + \frac{1}{1-\rho} \cdot \frac{1-\mu_s}{\mu_s} \cdot \frac{B}{B_r}\Big) > 1, \tag{38}$$

  where $\xi$ is the unique root of $\mu_{\text{s}} y^r - y + 1 - \mu_{\text{s}}$ 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).

- *Relay nodes: TDMA vs. ALOHA.* The departure processes in the linear network are all approximated as on-off. So the delays in relay nodes are geometrically distributed. Fig. 3 shows that such on-off departure processes converge to Bernoulli rapidly with $a_{01}^{(i)} \to m\lambda$ for TDMA and $a_{01}^{(i)} \to \lambda$ for ALOHA. If we set $p_m = 1/m$, the ratio of the delay means $\overline{D}_i$ in relay nodes for TDMA and ALOHA is

$$\eta = \frac{\overline{D}_{i,\mathrm{TDMA}}}{\overline{D}_{i,\mathrm{ALOHA}}} = 1 - \frac{\mu(m-1)}{m(1-\lambda)} < 1, \tag{39}$$

*i.e.*, TDMA outperforms ALOHA in the delay in terms of both mean and variance. From the perspective of traffic shaping, TDMA acts as a leaky bucket regulator, while ALOHA behaves like a Bernoulli regulator. So TDMA-shaped traffic is more regular than ALOHA-shaped traffic and it is well known that smooth traffic causes smaller delays than bursty traffic [9].



(a) TDMA

(b) ALOHA

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

To justify the on-off approximation for the departure processes, we use simulations to show the tightness of the approximations. In the simulations, all traffic flows have the same rate $\lambda$. All channels have the same success probability $\mu = \mu_{\mathrm{T}} = \mu_{\mathrm{A}}$. Moreover, we let the transmit probability of ALOHA be $p_m = 1/m$ such that the average number of transmission opportunities are equivalent for TDMA and ALOHA. Delays are measured in the number of time slots that the packet stays in the system.

Fig. 4 and Fig. 5 compare the simulated per-node delay mean and variance with our analysis and lead to the following observations:

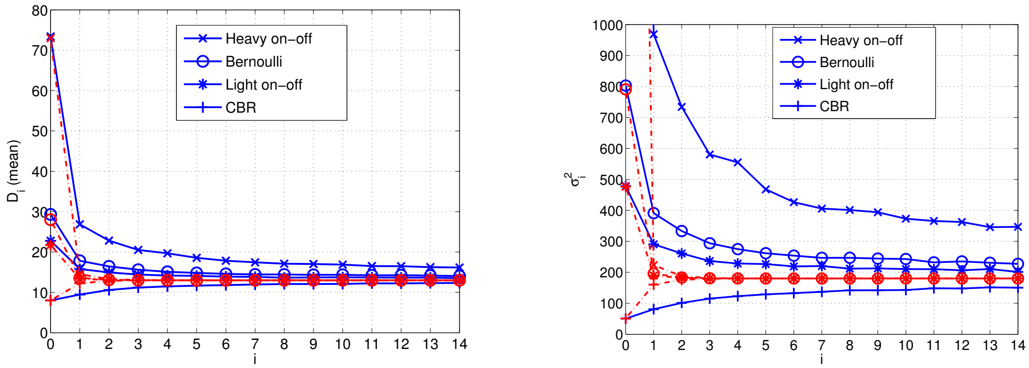
- as the node index $i$ increases, the simulated per-node delay mean and variance converge to the analytical results $\overline{D}_i$ and $\sigma_i^2$. In other words, the longer the chain of nodes that the traffic flow traverses, the tighter the on-off approximation;
- the delay depends on traffic burstiness, *i.e.*, the heavier the traffic burstiness, the larger the delay and delay jitter;
- traffic burstiness mainly affects the single node delays in the first few nodes, like at the source node $n_0$, as shown previously in the comparison (37) and (38). After the traffic flow traverses a long path, the influence of the traffic burstiness on the delays at the relay nodes diminishes (*e.g.*, for TDMA, the $\overline{D}_i$'s ($i \geq 5$) are almost identical for all four traffic flows (Fig. 4(a)). Finally, the delay mean and variance at the relay nodes converge to the same value regardless of the original traffic burstiness;
- our analysis (the dash-dotted lines in Fig. 4 and Fig. 5) represents the limiting delay performance. It also shows in Fig. 3 the convergence of the approximate on-off processes to Bernoulli. In [27], a GI/Geo/1 queue is viewed as

a "Markov operator" that produces a departure process distribution from an arrival process distribution. Passing an arbitrary arrival process through a series of independent and identically distributed GI/Geo/1 queues is like observing the evolution of a discrete-time Markov chain, which establishes a connection to the existence of invariant distributions. Using entropy theory, the invariant distribution was shown to be Bernoulli [27], which justifies the on-off approximation theoretically. This Bernoulli process constitutes an "eigentraffic" process since it represents the "eigenvalue" towards which the arrival traffic properties are tended to transform [28];

- the departure processes converge to Bernoulli from different directions, depending on the relative burstiness $B/B_r$ of the original traffic flow. Assuming that the eigentraffic process has the reference burst size $B_r$, a traffic flow with a longer burst size ($B > B_r$) causes a longer delay and thus will converge from above. Similarly, a flow with a shorter burst size ($B < B_r$) will converge from below. Note that $B_r$ depends on the established GI/Geo/1 system. In TDMA, the nodes are analyzed at the frame level. So $B_r$ should be the burst size of a Bernoulli process at the *frame level*. Therefore, even if the original traffic flow is Bernoulli at the *slot level*, it is still regarded to have a longer burst size at the frame level due to the packet cumulation during one frame. The similar principle is applied to both heavy and light on-off traffic. Accordingly, in TDMA, only CBR traffic has $B < B_r$ while all three remaining flows (Bernoulli, heavy and light on-off) have $B > B_r$. That is why the asymptotic value lies between CBR and all three bursty flows

(Fig. 4). In ALOHA, the nodes are analyzed at the slot level. So the original Bernoulli process itself is the eigentraffic process. Then, both light on-off and CBR have $B < B_r$ and only heavy on-off has $B > B_r$. Accordingly, the asymptotic value lies between the heavy on-off and the light on-off.

Therefore, our analysis provides the limiting delay performance and thus gives delay bounds. That is, for traffic flows with lighter burstiness than the eigentraffic process, our analysis provides upper bounds on the delay and vice versa. The eigentraffic process is obtained from the established GI/Geo/1 model.

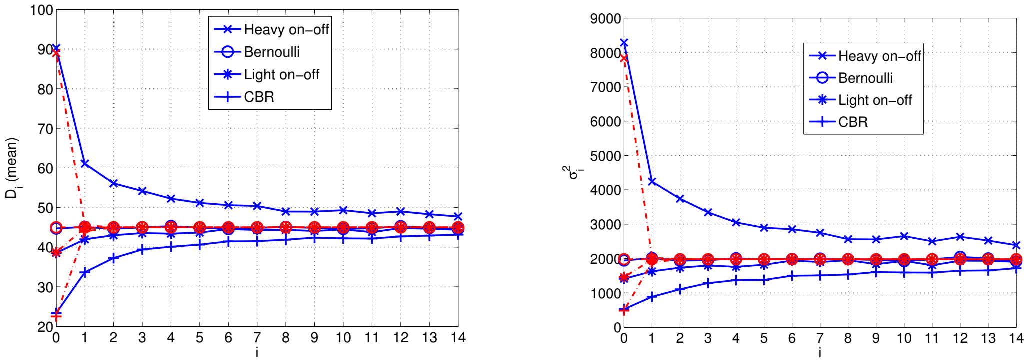


(a) Mean

(b) Variance

Fig. 4. The mean $\overline{D}_i$ and variance $\sigma_i^2$ of single node delays $D_i$ at $n_i$ in TDMA networks with with $m = 3, \lambda = 0.25, \mu_{\mathrm{T}} = 0.8, N = 15$. 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.

In short, the linear network of GI/Geo/1 queues turns the flows with different correlation and burstiness into the same memoryless Bernoulli process. The error-prone wireless channel behaves as an "entropy booster" [29] that inserts "holes" into the arrival flows randomly (with probability $\mu$). This inserting operation limits the maximum burstiness the traffic flow can sustain as it tra-



(a) Mean

(b) Variance

Fig. 5. The mean $\overline{D}_i$ and variance $\sigma_i^2$ of single node delays $D_i$ at $n_i$ in ALOHA networks with with $p_m = 1/3, \lambda = 0.25, \mu_{\mathrm{A}} = 0.8, N = 15$. 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.

verses through the network. In other words, Bernoulli possesses the “natural” level of burstiness that is favored by the network under a given traffic load.

The rate at which the flows converge to Bernoulli depends on the relative burst size $B/B_r$ and the channel quality $\mu$. Generally, the longer the burst size, the faster the convergence (Fig. 4 and Fig. 5). The relative burstiness is determined by both the original traffic statistics and the MAC scheme. For example, TDMA incurs arrival cumulations. Thus the relative burstiness is higher than in ALOHA and all traffic flows cause a sharp decrease in $a_{01}^{(i)}$ (Fig. 3). On the other hand, a good channel is able to maintain the original traffic statistics in that most packets can be sent out without retransmissions. So the interdeparture time is almost equivalent to the interarrival time and it takes a very long path for the flows to converge to Bernoulli. In contrast, a bad channel causes multiple retransmissions and the interdeparture time is mainly determined by the geometric service time. Therefore, the traffic flows converge to Bernoulli very quickly.

## 7 End-to-End Delay in Multihop Networks

Our analysis shows that the arrival processes to the nodes of a linear network converge to Bernoulli and the delays at each node converge to a geometric distribution. However, these arrival processes are not independent so the delays $D_i$'s are not independent either. In general, delay $D_i$ experienced at $n_i$ depends on both the service and the arrival process, the latter of which is the departure of $n_{i-1}$ and, as stated before, except for Bernoulli processes, is correlated with the delay $D_{i-1}$. Iteratively, $D_i$ is correlated with all $D_j$'s ($j \neq i$). The accurate calculation of the e2e delay variance $\sigma^2$ should take into account all these correlations and thus quickly becomes intractable as $N$ grows. In this section, we use simulations to reveal the influence of the correlations.

Fig. 6 show the e2e delay variance $\sigma^2$ and $\sum_{i=0}^{N-1} \sigma_i^2$ in TDMA and ALOHA, respectively. Previous work assumed "independent" $D_i$'s, meaning $\sigma^2 = \sum_{i=0}^{N-1} \sigma_i^2$, which holds only if the original traffic flow is Bernoulli (or Poisson). For other arrival processes, due to the existence of the correlations, $\sigma^2 \neq \sum_{i=0}^{N-1} \sigma_i^2$. As shown in Fig. 6, a gap exists between $\sigma^2$ and $\sum_{i=0}^{N-1} \sigma_i^2$ except for Bernoulli traffic in ALOHA (Fig. 6(b)). For some traffic flows such as heavy on-off and

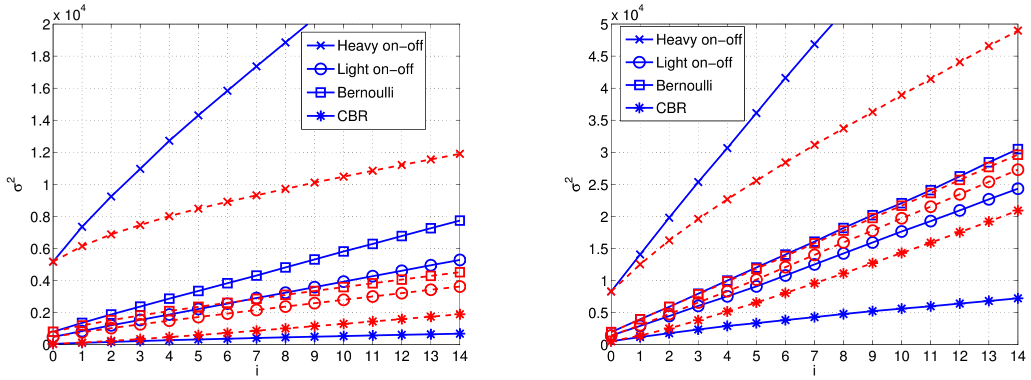
CBR, the gap is too large to permit the assumption $\sigma^2 = \sum_{i=0}^{N-1} \sigma_i^2$. Therefore, it is critical to study the impact of the correlations in $D_i$'s in the multihop topology.

We observe in Fig. 6 that different traffic flows cause different correlations:

- in TDMA, the CBR source results in a negative correlation while all three on-off sources including Bernoulli result in a positive correlation;
- in ALOHA, both CBR and light on-off cause a negative correlation while heavy on-off causes a positive correlation;
- in ALOHA, the Bernoulli source causes zero correlation.

Recall that in TDMA, the nodes are modeled as GI/Geo/1 at the frame level. All traffic flows except CBR cause converge to the eigentraffic process from *above* because they have a longer relative burst size. On the other hand, in ALOHA, the nodes are GI/Geo/1 systems at the slot level, at which the Bernoulli flow itself represents the asymptotic process and only heavy on-off has a longer relative burstiness and converges to Bernoulli from above. Therefore, it is natural to connect the type of the correlation with the relative burst size $B/B_r$. That is, i) a flow with $B/B_r < 1$ causes a negative correlation; ii) a flow with $B/B_r > 1$ leads to a positive correlation; iii) a flow with $B/B_r = 1$ causes no correlation at all.

The *relative* burst size is MAC-dependent. With TDMA, a Bernoulli flow generated at the slot level is not Bernoulli at the frame level and thus loses the



(a) TDMA

(b) ALOHA

Fig. 6. The e2e delay variance in TDMA and ALOHA networks with with $m = 3, p_m = 1/m, \lambda = 0.25, \mu = \mu_{\mathrm{T}} = \mu_{\mathrm{A}} = 0.8, N = 15$. 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 $\sigma^2$ with $D_i$'s correlated while the dash-dotted lines would be the variance as if all $D_i$'s were independent.

memoryless property and causes a positive correlation in $D_i$'s, which leads to $\sigma^2 > \sum_{i=0}^{N-1} \sigma_i^2$ (Fig. 6(a)). Similarly, a light on-off flow at the slot level is transformed into a heavy on-off flow at the frame level and cause a positive correlation. It is because TDMA acts like a leaky bucket regulator that destroys the correlation and burstiness of the on-off flows. On the other hand, ALOHA makes transmission decisions independently and acts like a random geometric (or Bernoulli) regulator that can preserve i) the memoryless property so $\sigma^2 = \sum_{i=0}^{N-1} \sigma_i^2$ for Bernoulli; and ii) the burstiness property so $\sigma^2 < \sum_{i=0}^{N-1} \sigma_i^2$ for light on-off in ALOHA (Fig. 6(b)).

The relative burst size also determines the correlation coefficients. The larger the relative burst size, the larger the correlation. For instance, in Fig. 6, for heavy on-off, $\sigma^2 \approx 4 \sum_{i=0}^{N-1} \sigma_i^2$ in TDMA and $\sigma^2 \approx 2.5 \sum_{i=0}^{N-1} \sigma_i^2$ in ALOHA, while for light on-off, $\sigma^2 \approx 2 \sum_{i=0}^{N-1} \sigma_i^2$ in TDMA and $\sigma^2 \approx 1.2 \sum_{i=0}^{N-1} \sigma_i^2$ in ALOHA. Besides, if $\sigma^2 \geq 2 \sum_{i=0}^{N-1} \sigma_i^2$, strong correlations exist not only between neighboring nodes, but also between nodes that are more than one hop from each other. Therefore, if the arrival processes are correlated, the linear network evolves in a more complex way than a Markov chain [27], in which non-neighboring nodes are conditionally independent.

The regular spacing introduced by TDMA results in not only a small single node delay but also a small e2e delay than ALOHA. Furthermore, smooth traffic results in a negative correlation that decreases the e2e delay variance while bursty traffic causes a positive correlation that increases the e2e delay variance. Therefore, the gap in $\sigma^2$ between CBR and heavy on-off is huge, *e.g.*, in TDMA, $\sigma^2_{\text{heavy on-off}} \approx 14\sigma^2_{\text{CBR}}$ and in ALOHA, $\sigma^2_{\text{heavy on-off}} \approx 11\sigma^2_{\text{CBR}}$. In order to guarantee the e2e delay bound for delay-sensitive applications, the traffic flow should be shaped, that can be implemented by both traffic regulation and MAC control.

The other interesting observation from Fig. 6 is that even with the existence of the correlations, the e2e delay variance is almost linear with the number of nodes (Fig. 6). It means that the impact of the correlations is uniform in the linear network.

## 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. For the e2e analysis, we calculate the pmf of the delays at each node (including the source node and relay nodes) and derive the departure process. For a tractable analysis, we approximate the departure process by a correlated and bursty on-off process, which is proved to be accurate as the network length increases. Regardless of the original traffic statistics, all departure processes converge to an identical eigentraffic process as the number of nodes $N$ increases – albeit from different directions and at different rates, all depending on the relative burst size with respect to the asymptotic Bernoulli process.

The traffic burstiness also affects the correlations in the per-node delays $D_i$. A flow with a long burst size $B/B_r > 1$ causes a positive correlation, leading to an increased e2e delay variance while a flow with a short burst size $B/B_r < 1$ causes a negative correlation, leading to a decrease in the e2e delay variance. It is also shown that the MAC policy may change the relative burst size and, in turn, change the type of the correlation. For example, TDMA destroys the memoryless property of the original Bernoulli flow and causes a positive correlation while ALOHA preserves the memoryless property. Simulation results reveal the significance of the correlations in the $D_i$'s. So simply assuming independent $D_i$'s and $\sigma^2 = \sum_{i=0}^{N-1} \sigma_i^2$ would lead to very optimistic and conservative e2e delay variances for bursty and smooth traffic, respectively.

Generally, smooth traffic leads to smaller delays than bursty traffic. TDMA outperforms ALOHA since it introduces a more regular spacing between packet arrivals than ALOHA. Similarly, CBR traffic results in a much smaller e2e delay variance than bursty traffic. Therefore, a MAC scheme should be designed together with a traffic regulator to optimize the e2e delay performance. Note that although TDMA achieves a better delay performance, it also incurs a substantial overhead to establish and maintain the frame structure, which may be impractical in certain wireless multihop networks. Furthermore, TDMA favors traffic flows that have the same direction as the TDMA transmission order and incurs a large access delay for flows with opposite direction. Therefore, an ideal delay-guaranteed MAC scheme should be able to smooth the traffic flows like TDMA but also be able to operate more independently and with less sensitivity to the traffic flow's direction like ALOHA.

## References

[1] R. Nelson, L. Kleinrock, Spatial TDMA: A Collision-Free Multihop Channel Access Protocol, IEEE Transactions on Communications 33 (9) (1985) 934–944.

[2] 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.

[3] 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.

[4] Y. Yang, T.-S. P. Yum, Delay Distributions of Slotted ALOHA and CSMA, IEEE Transactions on Communications 51 (11) (2003) 1846–1857.

[5] J. A. Morrison, Two Discrete-Time Queues in Tandem, IEEE Transactions on Communications 27 (3) (1979) 563–573.

[6] J. Hsu, P. J. Burke, Behavior of Tandem Buffers with Geometric Input and Markovian Output, IEEE Transactions on Communications 24 (3) (1976) 358 – 361.

[7] M. Sidi, Tandem Packet-Radio Queueing Systems, IEEE Transactions on Communications 35 (2) (1987) 246–248.

[8] 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.

[9] J. J. Hunter, Mathematical Techniques of Applied Probability, Academic Press, 1983, ISBN:0123618029.

[10] 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.

[11] H. Bruneel, B. G. Kim, Discrete-Time Models for Communication Systems Including ATM, Kluwer Academic Publishers, 1993, iSBN: 0792392922.

[12] G. Hablinger, Waiting Time, Busy Periods and Output Models of A Server Analyzed via Wiener-Hopf Factorization, Performance Evaluation 40 (2000) 3–26.

[13] 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.

[14] P. Jacquet, A. M. Naimi, G. Rodolakis, Routing on Asymptotic Delays in IEEE 802.11 Wireless Ad Hoc Networks, in: 1st Workshop on Resource Allocation in Wireless NETworks (RAWNET) 2005, 2005.

[15] 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 3rd International Working Conference on Performance Modeling and Evaluation of Heterogeneous Networks, 2005.

[16] F. Eshghi, A. K. Elhakeem, Y. R. Shayan, Performance Evaluation of Multihop Ad Hoc WLANs, IEEE Communications Magazine (2005) 107–115.

[17] L. G. Roberts, ALOHA Packet System With and Without Slots and Capture, ACM Sigcomm Computer Communication Review 5 (2) (1975) 28–42.

[18] M. Xie, M. Haenggi, A Study of the Correlations between Channel and Traffic

Statistics in Multihop Networks, accepted by IEEE Transactions on Vehicular Technology (2007).

[19] A. Heindl, Decomposition of General Tandem Queueing Networks with MMPP Input, Performance Evaluation 44 (2001) 5–23.

[20] G. Hasslinger, Waiting Time, Busy Periods and Output Models of a Server Analyzed via Wiener-Hopf Factorization, Performance Evaluation 40 (2000) 3–26.

[21] T. G. Robertazzi, Computer Networks and Systems: Queueing Theory and Performance Evaluation, Springer-Verlag, 1994, iSBN: 0387973931.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] Http://functions.wolfram.com/ElementaryFunctions/Log/29/.

[27] B. Prabhakar, R. Gallager, Entropy and the Timing Capacity of Discrete Queues, IEEE Transactions on Information Theory 49 (2) (2003) 357–370.

[28] M. Shell, Cascade All-Optical Shared-Memory Architecture Packet Switches Using Channel Grouping Under Bursty Traffic, PhD Dissertation, available at http://smartech.gatech.edu/bitstream/1853/4892/1/shell_michael_d_200412_phd.pdf (2004).

[29] N. T. Plotkin, C. Roche, The Entropy of Cell Streams as a Traffic Descriptor in ATM Networks, in: IFIP Performance of Communication Networks, 1995.