

Source Delay in Mobile Ad Hoc Networks

Juntao Gao, Yulong Shen, *Member, IEEE*, Xiaohong Jiang, *Senior Member, IEEE*
and Jie Li, *Senior Member, IEEE*

Abstract—Source delay, the time a packet experiences in its source node, serves as a fundamental quantity for delay performance analysis in networks. However, the source delay performance in highly dynamic mobile ad hoc networks (MANETs) is still largely unknown by now. This paper studies the source delay in MANETs based on a general packet dispatching scheme with dispatch limit f (PD- f for short), where a same packet will be dispatched out up to f times by its source node such that packet dispatching process can be flexibly controlled through a proper setting of f . We first apply the Quasi-Birth-and-Death (QBD) theory to develop a theoretical framework to capture the complex packet dispatching process in PD- f MANETs. With the help of the theoretical framework, we then derive the cumulative distribution function as well as mean and variance of the source delay in such networks. Finally, extensive simulation and theoretical results are provided to validate our source delay analysis and illustrate how source delay in MANETs are related to network parameters.

Index Terms—MANETs, packet dispatch, source delay, mean, variance.

I. INTRODUCTION

MOBILE ad hoc networks (MANETs) represent a class of self-configuring and infrastructureless networks with mobile nodes. As MANETs can be rapidly deployed, reconfigured and extended at low cost, they are highly appealing for a lot of critical applications, like disaster relief, emergency rescue, battle field communications, environment monitoring, etc [1], [2]. To facilitate the application of MANETs in providing delay guaranteed services in above applications, understanding the delay performance of these networks is of fundamental importance [3], [4].

Source delay, the time a packet experiences in its source node, is an indispensable behavior in any network. Since the source delay is a delay quantity common to all MANETs, it serves as a fundamental quantity for delay performance analysis in MANETs. For MANETs without packet redundancy [5], [6] and with one-time broadcast based packet redundancy [7], the source delay actually serves as a practical lower bound for and thus constitutes an essential part of overall delay in those networks. The source delay is also an indicator of packet lifetime, i.e., the maximum time a packet could stay in a network; in particular, it lower bounds the lifetime of a packet and thus serves as a crucial performance metric for MANETs with packet lifetime constraint.

Despite much research activity on delay performance analysis in MANETs (see section VI for related works), the source delay performance of such networks is still largely unknown by now. The source delay analysis in highly dynamic MANETs is challenging, since it involves not only complex network dynamics like node mobility, but also issues related to medium contention, interference, packet generating and

packet dispatching. This paper devotes to a thorough study on the source delay in MANETs under the practical scenario of limited buffer size and also a general packet dispatching scheme with dispatch limit f (PD- f for short). With the PD- f scheme, a same packet will be dispatched out up to f times by its source node such that packet dispatching process can be flexibly controlled through a proper setting of f . The main contributions of this paper are summarized as follows.

- We first apply the Quasi-Birth-and-Death (QBD) theory to develop a theoretical framework to capture the complex packet dispatching process in a PD- f MANET. The theoretical framework is powerful in the sense it enables complex network dynamics to be incorporated into source delay analysis, like node mobility, medium contention, interference, packet transmitting and packet generating processes.
- With the help of the theoretical framework, we then derive the cumulative distribution function (CDF) as well as mean and variance of the source delay in the considered MANET. By setting $f = 1$ in a PD- f MANET, the corresponding source delay actually serves as a lower bound for overall delay.
- Extensive simulation results are provided to validate our theoretical framework and the source delay models. Based on the theoretical source delay models, we further demonstrate how source delay in MANETs is related to network parameters, such as packet dispatch limit, buffer size and packet dispatch probability.

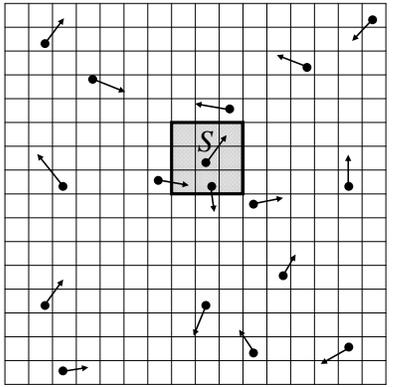
The rest of this paper is organized as follows. Section II introduces preliminaries involved in this source delay study. A QBD based theoretical framework is developed to model the source delay in Section III. We derive in Section IV the CDF as well as mean and variance of the source delay. Simulation/numerical studies and the corresponding discussions are provided in Section V. Finally, we introduce related works regarding delay performance analysis in MANETs in Section VI and conclude the paper in Section VII.

II. PRELIMINARIES

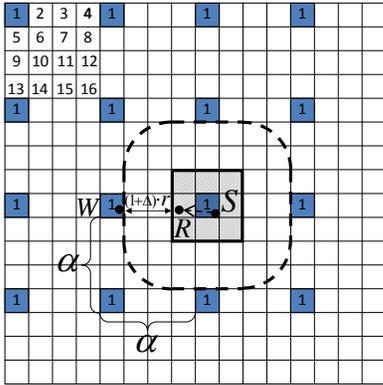
In this section, we introduce the basic system models, the Medium Access Control (MAC) protocol and the packet dispatching scheme involved in this study.

A. System Models

Network Model and Mobility Model: We consider a time slotted torus MANET of unit area. Similar to previous works, we assume that the network area is evenly partitioned into $m \times m$ cells as shown in Fig. 1a [8]–[11]. There are n mobile nodes in the network and they randomly move around following the



(a) A snapshot of a cell partitioned MANET.



(b) Illustration of MAC-EC protocol.

Fig. 1. An example of a cell partitioned MANET with a MAC protocol.

Independent and Identically Distributed (IID) mobility model [6], [12], [13]. According to the IID mobility model, each node first moves into a randomly and uniformly selected cell at the beginning of a time slot and then stays in that cell during the whole time slot.

Communication Model: We assume that all nodes transmit data through one common wireless channel, and each node (say S in Fig. 1a) employs the same transmission range $r = \sqrt{8}/m$ to cover 9 cells, including S 's current cell and its 8 neighboring cells. To account for mutual interference and interruption among concurrent transmissions, the commonly used protocol model is adopted [10], [12], [14], [15]. According to the protocol model, node i could successfully transmit to another node j if and only if $d_{ij} \leq r$ and for another simultaneously transmitting node $k \neq i, j$, $d_{kj} \geq (1 + \Delta) \cdot r$, where d_{ij} denotes the Euclidean distance between node i and node j and $\Delta \geq 0$ is the guard factor to prevent interference. In a time slot, the data that can be transmitted during a successful transmission is normalized to one packet.

Traffic Model: We consider the widely adopted permutation traffic model [10], [12], [13], where there are n distinct traffic flows in the network. Under such traffic model, each node acts as the source of one traffic flow and at the same time the destination of another traffic flow. The packet generating process in each source node is assumed to be a Bernoulli process, where a packet is generated by its source node with probability λ in a time slot [6]. We assume that each source node has a first-come-first-serve queue (called local-queue

hereafter) with limited buffer size $M > 0$ to store its locally generated packets. Each locally generated packet in a source node will be inserted into the end of its local-queue if the queue is not full, and dropped otherwise.

B. MAC Protocol

We adopt a commonly used MAC protocol based on the concept of Equivalent-Class to address wireless medium access issue in MANETs [10]–[12], [15]. As illustrated in Fig. 1b that an Equivalent-Class (EC) is consisted of a group of cells with any two of them being separated by a horizontal and vertical distance of some integer multiple of α ($1 \leq \alpha \leq m$) cells. Under the EC-based MAC protocol (MAC-EC), the whole network cells are divided into α^2 ECs and ECs are then activated alternatively from time slot to time slot. We call cells in an activated EC as active cells, and only a node in an active cell could access the wireless channel and do packet transmission. If there are multiple nodes in an active cell, one of them is selected randomly to have a fair access to wireless channel.

To avoid interference among concurrent transmissions under the MAC-EC protocol, the parameter α should be set properly. Suppose a node (say S in Fig. 1b) in an active cell is transmitting to node R at the current time slot, and another node W in one adjacent active cell is also transmitting simultaneously. As required by the protocol model, the distance d_{WR} between W and R should satisfy the following condition to guarantee successful transmission from S to R ,

$$d_{WR} \geq (1 + \Delta) \cdot r \quad (1)$$

Notice that $d_{WR} \geq (\alpha - 2)/m$, we have

$$(\alpha - 2)/m \geq (1 + \Delta) \cdot r \quad (2)$$

Since $\alpha \leq m$ and $r = \sqrt{8}/m$, α should be set as

$$\alpha = \min\{\lceil (1 + \Delta)\sqrt{8} + 2 \rceil, m\}, \quad (3)$$

where the function $\lceil x \rceil$ returns the least integer value greater than or equal to x .

C. PD- f Scheme

Once a node (say S) got access to the wireless channel in a time slot, it then executes the PD- f scheme summarized in Algorithm 1 for packets dispatch.

Remark 1: The PD- f scheme is general and covers many widely used packet dispatching schemes as special cases, like the ones without packet redundancy [5], [6], [8] when $f = 1$ and only unicast transmission is allowed, the ones with controllable packet redundancy [12], [16], [17] when $f > 1$ and only unicast transmission is allowed, and the ones with uncontrollable packet redundancy [7], [18] when $f \geq 1$ and broadcast transmission is allowed.

III. QBD-BASED THEORETICAL FRAMEWORK

In this section, a QBD-based theoretical framework is developed to capture the packet dispatching process in a PD- f MANET. This framework will help us to analyze source delay in Section IV.

Algorithm 1 PD- f scheme

```

1: if  $S$  has packets in its local-queue then
2:    $S$  checks whether its destination  $D$  is within its trans-
     mission range;
3:   if  $D$  is within its transmission range then
4:      $S$  transmits the head-of-line (HoL) packet in its local-
       queue to  $D$ ; {source-destination transmission}
5:      $S$  removes the HoL packet from its local-queue;
6:      $S$  moves ahead the remaining packets in its local-
       queue;
7:   else
8:     With probability  $q$  ( $0 < q < 1$ ),  $S$  dispatches the
       HoL packet;
9:     if  $S$  conducts packet dispatch then
10:       $S$  dispatches the HoL packet for one time; {packet-
        dispatch transmission}
11:      if  $S$  has already dispatched the HoL packet for  $f$ 
        times then
12:         $S$  removes the HoL packet from its local-queue;
13:       $S$  moves ahead the remaining packets in its
        local-queue;
14:      end if
15:    end if
16:  end if
17: else
18:    $S$  remains idle;
19: end if

```

A. QBD Modeling

Due to the symmetry of source nodes, we only focus on a source node S in our analysis. We adopt a two-tuple $\mathbf{X}(t) = (L(t), J(t))$ to define the state of the local-queue in S at time slot t , where $L(t)$ denotes the number of packets in the local-queue at slot t and $J(t)$ denotes the number of packet dispatches that have been conducted for the current head-of-line packet by slot t , here $0 \leq L(t) \leq M$, $0 \leq J(t) \leq f-1$ when $1 \leq L(t) \leq M$, and $J(t) = 0$ when $L(t) = 0$.

Suppose that the local-queue in S is at state (l, j) in the current time slot, all the possible state transitions that may happen at the next time slot are summarized in Fig. 2, where

- $I_0(t)$ is an indicator function, taking value of 1 if S conducts source-destination transmission in the current time slot, and taking value of 0 otherwise;
- $I_1(t)$ is an indicator function, taking value of 1 if S conducts packet-dispatch transmission in the current time slot, and taking value of 0 otherwise;
- $I_2(t)$ is an indicator function, taking value of 1 if S conducts neither source-destination nor packet-dispatch transmission in the current time slot, and taking value of 0 otherwise;
- $I_3(t)$ is indicator function, taking value of 1 if S locally generates a packet in the current time slot, and taking value of 0 otherwise.

From Fig. 2 we can see that as time evolves, the state transitions of the local-queue in S form a two-dimensional

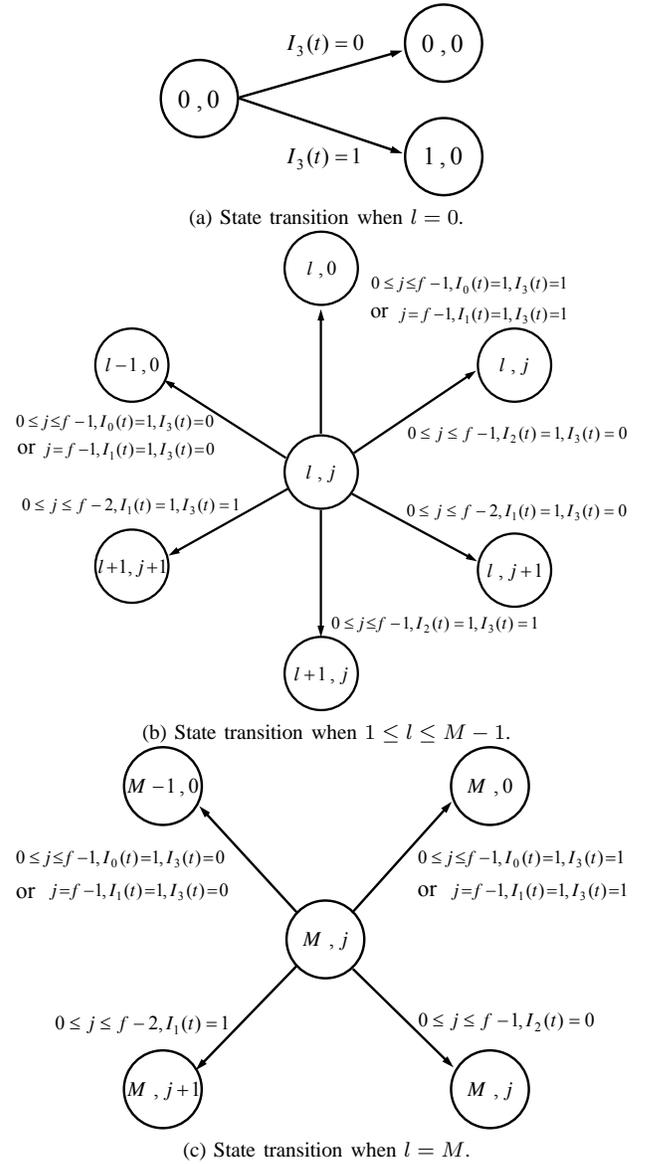


Fig. 2. State transitions from state (l, j) of the local-queue.

QBD process [19]

$$\{\mathbf{X}(t), t = 0, 1, 2, \dots\}, \quad (4)$$

on state space

$$\{(0, 0)\} \cup \{(l, j)\}; 1 \leq l \leq M, 0 \leq j \leq f-1\}. \quad (5)$$

Based on the transition scenarios in Fig. 2, the overall transition diagram of above QBD process is illustrated in Fig. 3.

Remark 2: The QBD framework is powerful in the sense it enables main network dynamics to be captured, like the dynamics involved in the packet generating process and these involved in the source-destination and packet-dispatch transmissions (i.e., node mobility, medium contention, interference and packet transmitting).

B. Transition Matrix and Some Basic Results

As shown in Fig. 3 that there are in total $1 + M \cdot f$ two-tuple states for the local-queue in S . To construct the transition

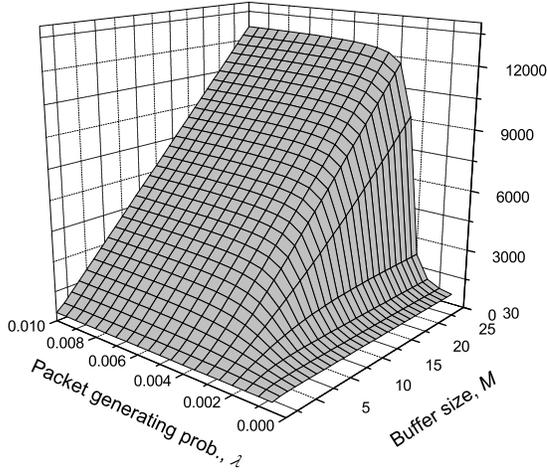
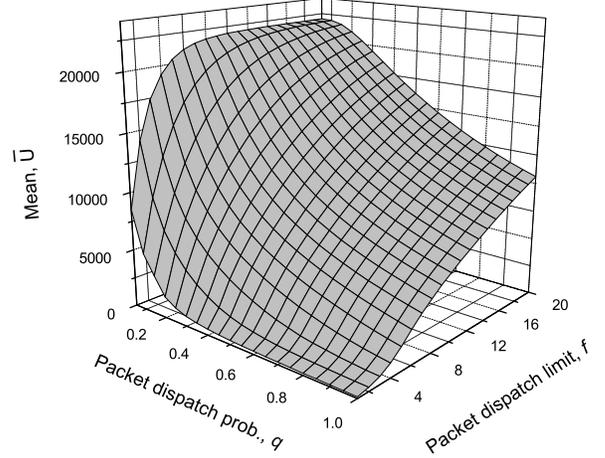
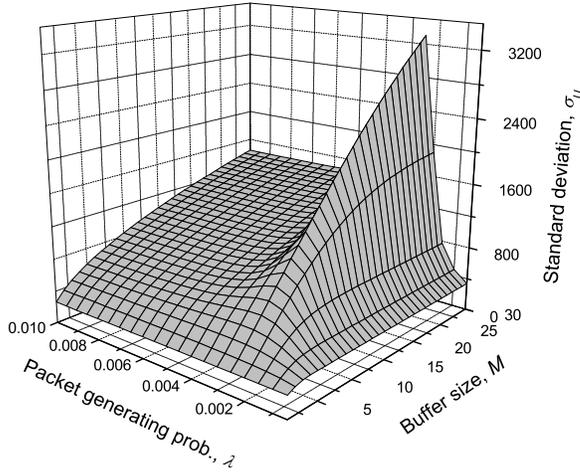
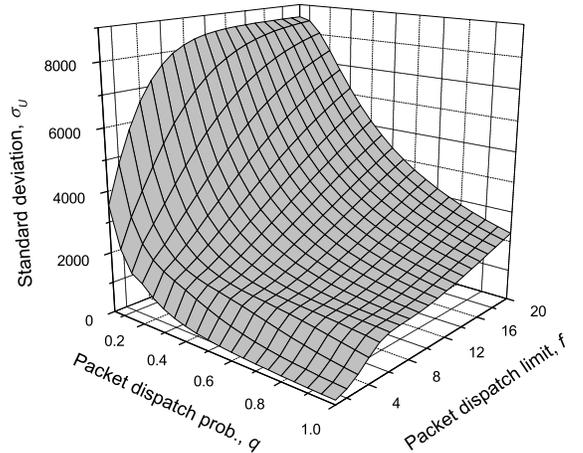
(a) \bar{U} versus (λ, M) (a) \bar{U} versus (q, f) (b) σ_U versus (λ, M) (b) σ_U versus (q, f)

Fig. 5. Source delay performance versus packet generating probability λ and local-queue buffer size M .

then remains constant, while for a given $\lambda \in [0.002, 0.01]$, \bar{U} always increases as M increases. Regarding the standard deviation σ_U of source delay, we see from Fig. 5b that for given M , as λ increases from 0.0005 to 0.01 σ_U first increases sharply to a peak value, then decreases sharply, and finally converges to a constant. It is interesting to see that the peak values of σ_U under different settings of M are all achieved at the same $\lambda = 0.0025$. The results in Fig. 5b further indicate that for fixed λ , as M increases σ_U always first increases and then gradually converges to a constant.

We then illustrate in Figs. 6 how \bar{U} and σ_U vary with packet dispatch parameters q and f under the network scenario of $n = 300, m = 16, M = 7$ and $\lambda = 0.002$. From Fig. 6a and Fig. 6b we can see that although both \bar{U} and σ_U always

Fig. 6. Source delay performance versus packet dispatch probability q and packet dispatch limit f .

decrease as q increases for a fixed f , their variations with q change dramatically with the setting of f . On the other hand, for a given $q \in [0.05, 0.2]$, as f increases both \bar{U} and σ_U first increase and then tend to a constant, while for a given $q \in [0.2, 1.0]$, both \bar{U} and σ_U always monotonically increase as f increases.

VI. RELATED WORKS

A substantial amount of works have been devoted to the study of delay performance in MANETs, which can be roughly divided into partial delay study and overall delay study.

A. Partial Delay Study

The available works on partial delay study in MANETs mainly focus on the delivery delay analysis [12], [16], [21]–

[26] and local delay analysis [27]–[29], which constitutes only a part of the overall packet delay.

The delivery delay, defined as the time it takes a packet to reach its destination after its source starts to deliver it, has been extensively studied in the literature. For sparse MANETs without channel contentions, the Laplace-Stieltjes transform of delivery delay was studied in [24]; later, by imposing lifetime constraints on packets, the cumulative distribution function and n -th order moment of delivery delay were examined in [21], [25]; the delivery delay was also studied in [16], [22], [26] under different assumptions on inter-meeting time among mobile nodes. For more general MANETs with channel contentions, closed-form results on mean and variance of delivery delay were recently reported in [12]. Regarding the local delay, i.e. the time it takes a node to successfully transmit a packet to its next-hop receiver, it was reported in [27] that some MANETs may suffer from a large and even infinite local delay. The work [28] indicates that the power control serves as a very efficient approach to ensuring a finite local delay in MANETs. It was further reported in [29] that by properly exploiting node mobility in MANETs it is possible for us to reduce local delay there.

B. Overall Delay Analysis

Overall delay (also called end-to-end delay), defined as the time it takes a packet to reach its destination after it is generated at its source, has also been extensively studied in the literature. For MANETs with two-hop relay routing, closed-form upper bounds on expected overall delay were derived in [6], [17]. For MANETs with two-hop relay routing and its variants, approximation results on expected overall delay were presented in [13], [30]. For MANETs with multi-hop relay routing, upper bounds on the cumulative distribution function of overall delay were reported in [31], [32], and approximations on the expected overall delay were derived in [33]. Rather than studying upper bounds and approximations on overall delay, some recent works explored the exact overall delay and showed that it is possible to derive the exact results on overall delay for MANETs under some special two-hop relay routings [6], [7].

VII. CONCLUSION

This paper conducted a thorough study on the source delay in MANETs, a new and fundamental delay metric for such networks. A QBD-based theoretical framework was developed to model the source delay behaviors under a general packet dispatching scheme, based on which the cumulative distribution function as well as the mean and variance of source delay were derived. As validated through extensive simulation results that our QBD-based framework is highly efficient in modeling the source delay performance in MANETs. Numerical results were also provided to illustrate how source delay is related with and thus can be controlled by some key network parameters, like local-queue buffer size, packet dispatch limit, and packet dispatch probability. It is expected that our source delay analysis and the related QBD-based theoretical framework will solidly contribute to the study of overall delay behavior in MANETs.

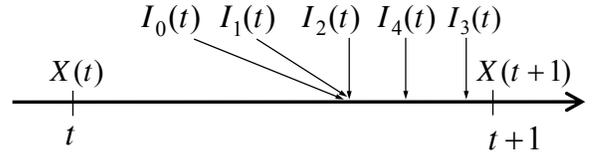


Fig. 7. Illustration for state transition from $\mathbf{X}(t)$ to $\mathbf{X}(t+1)$ during time slot $[t, t+1)$.

APPENDIX A PROOF OF LEMMA 1

The proof process is similar to that in [13], [17]. We omit the proof details here and just outline the main idea of the proof. To derive the probability p_0 (resp. p_1), we first divide the event that S conducts a source-destination transmission (resp. packet-dispatch transmission) in a time slot into following sub-events: 1) S moves into an active cell in the time slot according to the IID mobility model; 2) S successfully accesses the wireless channel after fair contention according to the MAC-EC protocol; 3) S selects to conduct source-destination transmission (resp. packet-dispatch transmission) according to the PD- f scheme. We can then derive probability p_0 (resp. p_1) by combining the probabilities of these sub-events.

APPENDIX B PROOF OF LEMMA 2

To derive the conditional steady state distribution π_{Ω}^* of the local-queue under the condition that a packet has just been inserted into the queue, we first study its corresponding transient state distribution $\pi_{\Omega}(t+1)$ at time slot $t+1$.

Similar to the definition of π_{Ω}^* , we can see that the $(2+(l-1)f+j)$ -th entry of row vector $\pi_{\Omega}(t+1)$, denoted by $[\pi_{\Omega}(t+1)]_{2+(l-1)f+j}$ here, corresponds to the probability that the local-queue is in state $\mathbf{X}(t+1) = (l, j)$ in time slot $t+1$ under the condition that a packet has just been inserted into the local-queue in time slot t , $1 \leq l \leq M, 0 \leq j \leq f-1$. The basic state transition from $\mathbf{X}(t)$ to $\mathbf{X}(t+1)$ is illustrated in Fig. 7, where $I_0(t)$ through $I_3(t)$ are indicator functions defined in Section III.A, and $I_4(t)$ is a new indicator function, taking value of 1 if the local-queue is not full in time slot t (i.e. the local-queue is in some state in $\{(0, 0)\} \cup \{(l, j)\}; 1 \leq l \leq M-1, 0 \leq j \leq f-1\}$), and taking value of 0 otherwise.

From Fig. 7 we can see that $[\pi_{\Omega}(t+1)]_{2+(l-1)f+j}$ is evaluated as

$$[\pi_{\Omega}(t+1)]_{2+(l-1)f+j} \quad (68)$$

$$= Pr\{\mathbf{X}(t+1) = (l, j) | I_4(t) = 1, I_3(t) = 1\} \quad (69)$$

$$= \frac{Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j)\}}{Pr\{I_4(t) = 1, I_3(t) = 1\}} \quad (70)$$

$$= \frac{Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j)\}}{\lambda \cdot Pr\{I_4(t) = 1\}}, \quad (71)$$

where (71) follows because the packet generating process is a Bernoulli process independent of the state of the local-queue.

For the probability $Pr\{I_4(t) = 1\}$ in (71), we have

$$Pr\{I_4(t) = 1\} \quad (72)$$

$$= \sum_{(l',j')} Pr\{I_4(t) = 1, \mathbf{X}(t) = (l', j')\} \quad (73)$$

$$= \sum_{(l',j')} Pr\{\mathbf{X}(t) = (l', j')\} Pr\{I_4(t) = 1 | \mathbf{X}(t) = (l', j')\} \quad (74)$$

where $Pr\{I_4(t) = 1 | \mathbf{X}(t) = (l', j')\}$ is actually the transition probability from state $\mathbf{X}(t) = (l', j')$ to states in $\{(0, 0) \cup \{(l, j); 1 \leq l \leq M-1, 0 \leq j \leq f-1\}\}$. The matrix \mathbf{P}_1 of such transition probabilities can be determined based on (6) by setting the corresponding sub-matrices according to (11)-(21). With matrix \mathbf{P}_1 and (74), we have

$$Pr\{I_4(t) = 1\} = \pi_\omega(t) \cdot \mathbf{P}_1 \cdot \mathbf{1}, \quad (75)$$

where $\pi_\omega(t) = (\pi_{\omega,l,j}(t))_{1 \times M \cdot f}$ with $\pi_{\omega,l,j}(t)$ being the probability $Pr\{\mathbf{X}(t) = (l', j')\}$.

For the numerator of (71), we have

$$Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j)\} \quad (76)$$

$$= \sum_{(l',j')} Pr\{\mathbf{X}(t) = (l', j'), I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j)\} \quad (77)$$

$$= \sum_{(l',j')} Pr\{\mathbf{X}(t) = (l', j')\} \cdot Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j) | \mathbf{X}(t) = (l', j')\}, \quad (78)$$

where $Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j) | \mathbf{X}(t) = (l', j')\}$ represents the transition probability from state $\mathbf{X}(t) = (l', j')$ to state $\mathbf{X}(t+1) = (l, j)$, with the condition that events $\{I_4(t) = 1\}$ and $\{I_3(t) = 1\}$ also happen simultaneously. The matrix \mathbf{P}_2 of such transition probabilities is determined based on (6) by setting the corresponding sub-matrices according to (25)-(35). With matrix \mathbf{P}_2 and (78), we have

$$Pr\{I_4(t) = 1, I_3(t) = 1, \mathbf{X}(t+1) = (l, j)\} = [\pi_\omega(t) \mathbf{P}_2]_{2+(l-1)f+j}. \quad (79)$$

After substituting (75) and (79) into (71), we get

$$\begin{aligned} & [\pi_\Omega(t+1)]_{2+(l-1)f+j} \\ &= \frac{[\pi_\omega(t) \mathbf{P}_2]_{2+(l-1)f+j}}{\lambda \pi_\omega(t) \mathbf{P}_1 \mathbf{1}}. \end{aligned} \quad (80)$$

Thus, in vector form

$$\pi_\Omega(t+1) = \frac{\pi_\omega(t) \mathbf{P}_2}{\lambda \pi_\omega(t) \mathbf{P}_1 \mathbf{1}}. \quad (81)$$

Taking limits on both sides of (81), we get the steady state distribution π_Ω^* as

$$\pi_\Omega^* = \lim_{t \rightarrow \infty} \pi_\Omega(t+1) \quad (82)$$

$$= \lim_{t \rightarrow \infty} \frac{\pi_\omega(t) \mathbf{P}_2}{\lambda \pi_\omega(t) \mathbf{P}_1 \mathbf{1}} \quad (83)$$

$$= \frac{\pi_\omega^* \mathbf{P}_2}{\lambda \pi_\omega^* \mathbf{P}_1 \mathbf{1}}, \quad (84)$$

where

$$\pi_\omega^* = \lim_{t \rightarrow \infty} \pi_\omega(t). \quad (85)$$

This completes the proof of Lemma 2.

APPENDIX C PROOF OF LEMMA 3

Recall that as time evolves, the state transitions of the local-queue form a QBD process shown in Fig. 3. From Fig. 3, we can see that the QBD process has finite states and all states communicate with other states, so the Markov chain is recurrent. We also see from Fig. 3 that every state could transition to itself, indicating that the Markov chain is aperiodic. Thus, the concerned QBD process is an ergodic Markov chain and has a unique limit state distribution π_ω^* defined in (85).

Notice that π_ω^* must satisfy the following equation

$$\pi_\omega^* = \pi_\omega^* \mathbf{P}_0, \quad (86)$$

where \mathbf{P}_0 is the transition matrix of the QBD process, which can be determined based on (6) by setting the corresponding sub-matrices according to (47)-(53). In particular, for $M = 1$ and $M = 2$, the transition matrix \mathbf{P}_0 is given by the following (87) and (88), respectively.

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_0 \\ \mathbf{B}_2 & \mathbf{A}_M \end{bmatrix}, \quad (87)$$

$$\mathbf{P}_0 = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_0 & \\ \mathbf{B}_2 & \mathbf{A}_1 & \mathbf{A}_0 \\ & \mathbf{A}_2 & \mathbf{A}_M \end{bmatrix}. \quad (88)$$

Thus, under the cases of $M = 1$ and $M = 2$, π_ω^* could be easily calculated by equations (36)-(38) and (39)-(42), respectively. Due to the special structure of the matrix \mathbf{A}_2 , which is the product of a column vector \mathbf{c} by a row vector \mathbf{r} [19], π_ω^* under the case $M \geq 3$ could be calculated by equations (43)-(46).

APPENDIX D PROOF OF THEOREM 1

Suppose that the local-queue is in some state according to the steady state distribution π_Ω^* , then the source delay of a packet (say Z) is independent of the packet generating process after Z is inserted into the local-queue and is also independent of the state transitions of the local-queue after Z is removed from the local-queue. Such independence makes it possible to construct a simplified QBD process to study the source delay of packet Z , in which new packets generated after packet Z are ignored, and once Z is removed from the local-queue (or equivalently the local-queue transits to state $(0, 0)$), the local-queue will stay at state $(0, 0)$ forever.

For the above simplified QBD process, its transition matrix \mathbf{P}_3 can be determined based on (6) by setting the corresponding sub-matrices as follows:

For $M = 1$,

$$\mathbf{B}_0 = \mathbf{0}, \quad (89)$$

$$\mathbf{B}_1 = [1], \quad (90)$$

$$\mathbf{B}_2 = \mathbf{c}, \quad (91)$$

$$\mathbf{A}_M = \mathbf{Q}. \quad (92)$$

For $M \geq 2$,

$$\mathbf{B}_0 = \mathbf{0}, \quad (93)$$

$$\mathbf{B}_1 = [1], \quad (94)$$

$$\mathbf{B}_2 = \mathbf{c}, \quad (95)$$

$$\mathbf{A}_0 = \mathbf{0}, \quad (96)$$

$$\mathbf{A}_1 = \mathbf{Q}, \quad (97)$$

$$\mathbf{A}_2 = \mathbf{c} \cdot \mathbf{r}, \quad (98)$$

$$\mathbf{A}_M = \mathbf{Q}. \quad (99)$$

By rearranging \mathbf{P}_3 as

$$\mathbf{P}_3 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{c}^+ & \mathbf{T} \end{bmatrix}, \quad (100)$$

we can see that matrices \mathbf{c}^+ and \mathbf{T} are determined as (60)-(63). With matrices \mathbf{c}^+ , \mathbf{T} and π_{Ω}^* , the probability mass function (56) and CDF (57) of the source delay follow directly from the theory of Phase-type distribution [19].

Based on the probability mass function (56), the mean \bar{U} of the source delay can be calculated by

$$\begin{aligned} \bar{U} &= \sum_{u=1}^{\infty} u \cdot Pr\{U = u\} \\ &= \sum_{u=1}^{\infty} u \pi_{\Omega}^- \mathbf{T}^{u-1} \mathbf{c}^+ \\ &= \pi_{\Omega}^- \left(\sum_{u=1}^{\infty} u \mathbf{T}^{u-1} \right) \mathbf{c}^+ \end{aligned} \quad (101)$$

Let

$$f(\mathbf{T}) = \sum_{u=1}^{\infty} u \mathbf{T}^{u-1}, \quad (102)$$

and use $f(x)$ to denote its corresponding numerical series

$$f(x) = \sum_{u=1}^{\infty} u x^{u-1} \quad (103)$$

$$= (1-x)^{-2}, \quad \text{for } x < 1. \quad (104)$$

Since above simplified QBD process is actually an absorbing Markov Chain with transition matrix \mathbf{P}_3 , we know from Theorem 11.3 in [34] that

$$\lim_{k \rightarrow \infty} \mathbf{T}^k = \mathbf{0}. \quad (105)$$

Based on the property (105) and the Theorem 5.6.12 in [35], we can see that the spectral radius $\rho(\mathbf{T})$ of matrix \mathbf{T} satisfies following condition

$$\rho(\mathbf{T}) < 1. \quad (106)$$

From (102), (104) and (106), it follows that the matrix series $f(\mathbf{T})$ converge as

$$f(\mathbf{T}) = \lim_{g \rightarrow \infty} \sum_{u=1}^g u \mathbf{T}^{u-1} \quad (107)$$

$$= (\mathbf{I} - \mathbf{T})^{-2} \quad (108)$$

After substituting (108) into (101), (58) then follows.

The derivation of the variance of source delay (59) could be conducted in a similar way and thus is omitted here.

REFERENCES

- [1] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. J. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, December 2008.
- [2] A. Goldsmith, M. Effros, R. Koetter, M. Mdard, and L. Zheng, "Beyond shannon: the quest for fundamental performance limits of wireless ad hoc networks," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 195–205, May 2011.
- [3] L. H. II and R. Tafazolli, "A survey of qos routing solutions for mobile ad hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 9, no. 2, pp. 50–70, 2nd Quarter 2007.
- [4] L. Chen and W. B. Heinzelman, "A survey of routing protocols that support qos in mobile ad hoc networks," *IEEE Network*, vol. 21, no. 6, pp. 30–38, November–December 2007.
- [5] M. Grossglauser and D. N. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, August 2002.
- [6] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917–1936, June 2005.
- [7] J. Gao and X. Jiang, "Delay modeling for broadcast-based two-hop relay manets," in *11th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2013.
- [8] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks-part i: The fluid model," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2568–2592, June 2006.
- [9] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and capacity tradeoffs for mobile ad hoc networks: A global perspective," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 981–992, October 2007.
- [10] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1745–1758, December 2011.
- [11] P. Li, Y. Fang, J. Li, and X. Huang, "Smooth trade-offs between throughput and delay in mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 427–438, March 2012.
- [12] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Generalized two-hop relay for flexible delay control in manets," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1950–1963, December 2012.
- [13] J. Liu, J. Gao, X. Jiang, H. Nishiyama, and N. Kato, "Capacity and delay of probing-based two-hop relay in manets," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 4172–4183, November 2012.
- [14] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [15] S. R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.
- [16] T. Small and Z. Hass, "Resource and performance tradeoffs in delay-tolerant wireless networks," in *Proc. ACM SIGCOMM workshop on Delay-tolerant networking (WDTN)*, 2005, pp. 260–267.
- [17] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Delay and capacity in ad hoc mobile networks with f -cast relay algorithms," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2738 – 2751, August 2011.
- [18] B. Williams and T. Camp, "Comparison of broadcasting techniques for mobile ad hoc networks," in *MobiHoc*, 2002.

- [19] G. Latouche and V. Ramaswamy, *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability, 1999.
- [20] C++ simulator for PD-f MANETs. [Online]. Available: <http://researchplatform.blogspot.jp/>.
- [21] A. A. Hanbali, P. Nain, and E. Altman, "Performance of ad hoc networks with two-hop relay routing and limited packet lifetime (extended version)," *Performance Evaluation*, vol. 65, no. 6-7, pp. 463–483, June 2008.
- [22] A. Panagakis, A. Vaios, and I. Stavrakakis, "Study of two-hop message spreading in dtns," in *WiOpt*, 2007.
- [23] M. Ibrahim, A. A. Hanbali, and P. Nain, "Delay and resource analysis in manets in presence of throwboxes," *Performance Evaluation*, vol. 64, no. 9-12, pp. 933–947, October 2007.
- [24] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1-4, pp. 210–228, October 2005.
- [25] A. A. Hanbali, A. A. Kherani, and P. Nain, "Simple models for the performance evaluation of a class of two-hop relay protocols," in *Proc. IFIP Networking*, 2007, pp. 191–202.
- [26] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77–90, February 2008.
- [27] F. Baccelli and B. Blaszczyszyn, "A new phase transitions for local delays in manets," in *INFOCOM*, 2010.
- [28] M. Haenggi, "The local delay in poisson networks," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1788–1802, March 2013.
- [29] Z. Gong and M. Haenggi, "The local delay in mobile poisson networks," [Online]. Available: <http://www3.nd.edu/~mhaenggi/pubs/>.
- [30] J. Liu, X. Jiang, H. Nishiyama, N. Kato, and X. Shen, "End-to-end delay in mobile ad hoc networks with generalized transmission range and limited packet redundancy," in *WCNC*, 2012.
- [31] O. H. Florin Ciucu and P. Hui, "Non-asymptotic throughput and delay distributions in multi-hop wireless networks," in *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010.
- [32] F. Ciucu, "Non-asymptotic capacity and delay analysis of mobile wireless networks," in *SIGMETRICS*, 2011.
- [33] A. Jindal and K. Psounis, "Contention-aware performance analysis of mobility-assisted routing," *IEEE Transactions on Mobile Computing*, vol. 8, no. 2, pp. 145–161, February 2009.
- [34] C. M. Grinstead and J. L. Snell, *Introduction to Probability: Second Revised Edition*. American Mathematical Society, 1997.
- [35] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

