Time-Varying Performance Analysis of Multihop Wireless Networks with CBR Traffic

Kunjie Xu, *Student Member, IEEE*, David Tipper, *Senior Member, IEEE*, Yi Qian, *Senior Member, IEEE*, Prashant Krishnamurthy, *Member, IEEE*, and Siriluck Tipmongkonsilp

Abstract—In this paper, we develop a performance modeling technique for analyzing the time varying network layer queuing behavior of multihop wireless networks with constant bit rate (CBR) traffic. Our approach is a hybrid of a time varying adjacency matrix and a fluid flow queuing network model. Mobile network topology is modeled using time varying adjacency matrix, while node queues are modeled using fluid flow based differential equations which are solved using numerical methods. Numerical and simulation experiments show that this new approach can provide reasonably accurate results. Moreover, when compared to the computation time required in a standard discrete event simulator, the fluid flow based model is shown to be a more scalable tool. Finally, an illustrative example of our modeling technique application is given to show its capability of capturing the time varying network performance as a function of traffic load, node mobility and wireless link quality.

Index Terms—Multihop wireless networks, time varying performance, fluid flow model, mobility

I. INTRODUCTION

In recent years, there has been significant growth of interest in multihop wireless networks, such as wireless mesh networks (WMN) [1], vehicular ad-hoc networks (VANET) [2], wireless sensor networks, (WSN) [3] and mobile ad-hoc networks (MANET) [4]. Multihop wireless networks are expected to become an important part of the communications landscape and may work in a fully autonomous scenario or as an extension to an infrastructure network. In multihop wireless networks, the nodes must cooperate to dynamically establish routes using wireless links, and routes may involve multiple hops with each node acting as a router. In many cases (e.g., MANET or VANET), the network nodes can move and the network topology is expected to change often and sometimes unpredictably. Hence, a basic challenge in building multihop wireless networks is designing highly adaptive and failure recovery strategies to properly route traffic [5] [6] [7] [8]. Meanwhile, multihop wireless networks also inherit the traditional problems of wireless communications (e.g.,

broadcast communication channels, asymmetric channels and signal propagation, links that are poor quality in comparison to wired links, etc.) [1] - [8]. These problems combined with the unique dynamic topology feature make it challenging to accurately evaluate and predict the performance of multihop wireless networks.

In general, the performance of multihop wireless networks can be evaluated using measurements, simulations, or analytical models. Measurement studies on multihop wireless network involve running experiments on a real network [9] or a prototype testbed [10]. These studies show a variety of interesting behaviors (e.g., asymmetric links, grey zones, etc.) not observed in most simulation or analytical based multihop wireless network performance models [11]. The general criticisms of measurement studies are the expense, the great deal of effort required to consider all cases/parameter values and the difficulty in generalizing results. Furthermore, measurements are generally non-repeatable because wireless network environments can be very different depending on interference, frequency band, geography, etc.

In contrast to measurement based studies, simulation models do not require a testbed as the one models the system on the computer and experiments with the computer model. Popular network simulation tools used in multihop wireless network studies include OPNET, NS-2, NS-3, Qualnet, and GloMoSim. The basic simulation approach adopted in the majority of the literature is as follows. For a given scenario (i.e., geographic space, number of nodes, mobility model, transmission range, routing scheme, etc.), the network is simulated over a fixed time period. Multiple runs are simulated with different random number seeds and the collected data is averaged over the runs. In terms of simulation methodology, this approach is considered as steady-state simulation [12]. Additionally, observations gathered during the transient period in each run are usually eliminated to avoid initialization bias [13].

Due to the mobility of nodes and their limited battery life, link and node failures are common in multihop wireless networks. Thus, one would expect that the network may spend much of the time in a transient/nonstationary state [14]. Simulation studies of the time varying behavior for such networks are possible [12] [15], though computationally difficult. To study nonstationary behavior, the measurements of quantities observed over small intervals or at specific points in time are important. Therefore, the time average is not a proper approach and the *ensemble average* is utilized instead. The idea is to construct ensemble averaged quantities of interest as a function of time across a set of statistically identical

but distinct independent simulation runs (e.g. different random number seeds are used), along with the calculated confidence interval. This approach is based upon the well-established concept of generating an ensemble of simulation replications and calculating performance measures as ensemble quantities versus time. However, to assure the accurate portrayal of the actual system, a large number of runs are required resulting in massive amounts of CPU time.

Network performance can also be evaluated by defining a system model and solving the model using analytical techniques. In this work, we develop an approximate fluid flow modeling approach which can be used to model the mean behavior of multihop wireless networks. The principle of this modeling approach has been termed the *Pointwise Stationary Fluid Flow Approximation (PSFFA)* [16]. The basic idea is to model the ensemble average number of packets in a queuing system by a single nonlinear differential equation which is solved numerically using standard numerical integration techniques (e.g., Runge Kutta). The *PSFFA* approach derives the form of the fluid flow differential equation from a pointwise mapping of the steady-state queuing relationships for the model. As shown in [17], the *PSFFA* is quite general and capable of determine the nonstationary behavior of finite and infinite capacity queuing systems with general arrival and service process. Furthermore, *PSFFA* models can be coupled using flow conservation principles to study networks of queues.

Regarding the network traffic load, we assume the offered traffic at each node is CBR. Typically, CBR traffic is tailored for on-demand or real time networking services, where the end systems require predictable response time and continuously available bandwidth during the life-time of the connection. For CBR traffic, both the packet size of CBR traffic and the packet inter-arrival time are constant. Real-time CBR traffic usually has the deterministically-bounded delay requirement for one-hop packet service time. In this paper, we adopt CBR as our traffic model for benchmarking purposes, and propose a modeling framework to efficiently evaluate time varying performance for multihop wireless networks, which to the best of our knowledge has not been studied before. In multihop wireless networks, it is possible that CBR traffic offered on the source node might not be exactly CBR after being forwarded to the following nodes, since wireless network could induce distortion on CBR traffic including *delay jitter* and *packet loss*, which are considered in Section III-B-3 and Section III-A, respectively. The specific performance model proposed in this work for CBR traffic is applicable to the scenario that the probability of jitter and packet loss is tolerable, so that the packet arrival and service process will not deviate far away from our assumed deterministic process. In this work, we assume error-free propagation in PHY layer and dedicated control channel for route discovery/maintenance in routing layer. Hence, MAC layer protocol plays a key role in determining the data packet arrival and service process at each node. Here, the data packet service time at a node is defined as the interval from the time instant when a packet is at the head of its transmission queue and ready for transmission until the time instant of being successfully received by the next-hop node. So we specify the applicable scenarios as *"collision-rare"* and *"collision-free"*

scenario. *Collision-rare* transmission occurs in contention-based but sparse or lightly loaded networks. For example, energy conserving techniques in WSNs put most of the nodes in sleep mode so that the network becomes sparse. Meanwhile, each sensor node typically has light traffic to transmit in order to save energy, and traffic generation rate is much less than the channel capacity. In this scenario, CBR traffic is used to model the periodically generated data flow in WSN [6]. Another example is to consider periodic routine message forwarding for safety applications in VANET. Generally, routine messages are transmitted by all vehicles periodically and have a fixed packet size containing state information of vehicles (e.g. position, speed, and direction). Hence, CBR traffic could approximate the periodic routine messages, which constitute the majority of safety-related traffic in VANET. Also, the routine packet generation rate is low (e.g. 2-20 pkt/sec) and each packet is typically very short (e.g. 100-300 bytes). As a result, the traffic load (e.g. 1.6-48 Kbps) is much less than the channel capacity (e.g. 3-27 Mbps in IEEE 802.11p), and thus the packet collision probability is small in the case of low or moderate vehicle density [18]. *Collision-free* transmission can be achieved by using resource-reservation MAC protocols for each node to access the channel. Since this type of MAC protocols rely on deterministically quantified resource reservation, their packet arrival and service time could be roughly approximated as deterministic or deterministically bounded. For example, in [8], a multihop TDMA-based reservation MAC protocol is developed for CBR traffic to avoid packet collision and provide delay guaranteed data delivery.

Our contribution in this paper can be summarized as follows. First of all, we propose a fluid flow based queuing model for each node with CBR traffic by using the PSFFA approach. We then extend the queuing analysis by considering a large number of input traffic streams to the queue, and the utilization function of the queue is approximated in a computationally efficient way. Secondly, on the basis of a fluid flow queuing model for a node, we develop a novel time varying performance model for multihop wireless networks with CBR traffic. An adjacency matrix, representing topology changes, is integrated into the model using either deterministic or stochastic based network connectivity modeling techniques. We then codify our performance modeling procedure into an executable and efficient algorithm, and it is shown to be a more scalable tool than discrete event simulator. Finally, we carefully evaluate the performance of a sample network impacted by node mobility, traffic load and wireless link quality via our model, which can respond to the ongoing nonstationary conditions properly.

The rest of the paper is organized as follows. Section II discusses the related work. Section III provides the details of our modeling approach. In Section IV, we present a series of numerical results by comparing our model with discrete event simulation in terms of accuracy and computation time, and lastly illustrating the use of our model to examine network behavior impacted by node mobility, traffic load and wireless link quality. Our conclusions are given in Section V.

II. RELATED WORK

Here we are primarily interested in identifying techniques that can be used for multihop wireless network performance evaluation and the design of dynamic network controls. Since many network controls are designed and implemented on the basis of average quantities, such as the average delay on the links, we have focused on determining the mean transient/nonstationary behavior of networks. With the concept of PSFFA [16], we have developed an approximate fluid flow modeling approach to model the mean transient/nonstationary behavior of a variety of queuing systems in a series of papers [17], [19], [20], [21]. The idea of PSFFA is to model the average number in the queuing system as a function of time by a single nonlinear differential equation, which is solved numerically using standard numerical integration techniques (e.g., Runge Kutta). The fluid flow model can be generally applied for various queuing systems [17]. In fact, it is even possible to develop the fluid flow model from measurement data. In addition, fluid flow model could be used as the basic mathematical model for developing network dynamic routing and flow control mechanisms along the lines illustrated in [22] [23].

We note that fluid flow models have been proposed for constructing computationally efficient simulation models for both wired [24], [25] and wireless networks [26], [27], [28]. On the wired network side, the basic idea of [24] is to model a few network nodes (e.g., a source destination pair) in details with packet based discrete event simulation and enlarge part of a IP network by fluid flow models which interface with the discrete event simulation. This approach has been shown to be accurate at the IP level and scalable. In [25], the authors use fluid flow techniques in combination with discrete event simulation to model the dynamics of TCP traffic, which is adaptive to the available bandwidth on the network. They illustrate that the proposed fluid flow model produces very similar behavior as the packet-level model, but can provide significant computational savings. On the wireless network side, Kim and Hou in [26] develop a fluid flow based simulator for WLAN with the consideration of the characteristics of IEEE 802.11 protocol behavior, and examine fluid simulation performance in terms of events generated, execution time required, relative error incurred, and time step value adopted in the simulation. In [27], a fluid flow model is presented to analyze the performance of backlog-based CSMA policies in the wireless networks environment with multiple arrival rates. Most recently, the fluid flow approximation is applied in [28] to model the TPC connection with time division multiplexing and scheduling in WiMAX wireless networks. While all of these approaches are related to our work, little fluid-flow related work has appeared to model the time varying behavior of multihop wireless networks by considering their unique characteristics (e.g., node mobility, wireless links, dynamic routing, and scalability, etc.).

III. MODELING DYNAMIC BEHAVIOR

In this section, we introduce the network topology modeling first, and then the nonstationary modeling of each node with CBR traffic. Finally, we derive our fluid flow model for multihop wireless network by combining these two components.

A. Network Topology Modeling

Consider a network consisting of M nodes, the topology in terms of connectivity at any point in time t is modeled by a $M \times M$ adjacency matrix denoted as $A(t)$.

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \ldots & a_{1M}(t) \\ a_{21}(t) & a_{22}(t) & \ldots & a_{2M}(t) \\ \vdots & \vdots & & \vdots \\ a_{M1}(t) & a_{M2}(t) & \ldots & a_{MM}(t) \end{bmatrix}$$

where $a_{ij}(t)$ represents the binary link connectivity between node i and j (i.e., $a_{ij}(t) = 1$ if link from node i to j exists, otherwise $a_{ij}(t) = 0$). With the assumption that all radios have a perfect coverage on a two-dimensional space, the problem of link connectivity is simplified by judging whether the distance d_{ij} between node i and node j is within the circular coverage range R (i.e., if $d_{ij} \leq R$ at time t, $a_{ij}(t) = 1$; otherwise $a_{ij} = 0$). Moreover, it is widely understood that the actual radio link connectivity may differ from this simple model. Even though two nodes are in the radio range of each other, they cannot always hear each other without any data loss, and the bit error rate is typically a function of the signal to noise plus interference ratio. In order to represent real link quality as well as connectivity, we let $a_{ij}(t)$ be a real number between 0 and 1 (i.e., $a_{ij}(t) \in (0, 1]$ if link from node i to j exists, otherwise $a_{ij}(t) = 0$).

To model node mobility, our approach is to directly manipulating the elements of the adjacency matrix according to a planned experiment (e.g. in three-node network of Section IV-A) or a probabilistic model (e.g. in five-node network of Section IV-A and thirty-node network of Section IV-C). Such a probabilistic model can be developed either from the mobility model assumptions and analysis [29] or from fitting a statistical model to data gathered from a test bed or simulation (e.g., two-state MMPP [30]).

B. Node Queuing Model

In developing a performance model of the network, we start with modeling a single queue and then generalize it to an arbitrary queue in a network. We first give a description of fluid flow model background. Then, the fluid flow model is extended for multi-class traffic queue. Finally, for the specific case of a queue with the superposition of CBR streams, the fluid flow based node queuing model is derived in detail.

1) Fluid Flow Model Background: For a single server first-come-first-serve (FCFS) queuing system with nonstationary arrival process, $\lambda(t)$ denotes the *ensemble average* arrival rate at time t. The model is developed by focusing on the dynamics of the packet queue at a transmission link. Let $x(t)$ be defined as the state variable representing the *ensemble average* number in the system at time t, $\dot{x}(t) = dx/dt$ is the rate of change of the state variable with respect to time. According to the flow conservation principle, the rate of change of the *ensemble average* number in the system equals the difference between

the flow in and the flow out of the system at time t, denoted by $f_{in}(t)$ and $f_{out}(t)$:

$$\dot{x}(t) = -f_{out}(t) + f_{in}(t) \qquad (1)$$

For an infinite waiting space queue, the flow in equals to the arrival rate $f_{in}(t) = \lambda(t)$. The flow out can be related to the ensemble average utilization of the server as $f_{out}(t) = \mu C G(x(t))$, where $1/\mu$ refers to the average packet length (bits) and C represents the server capacity or link bandwidth (bps). Thus, μC is the service rate (pkt/s). $G(x)$ denotes the link utilization, which is a monotonically increasing function in the range $[0,1)$ passing through the origin $G(0) = 0$. The utilization function $G(x(t))$ is determined by the stochastic properties of the queue such as traffic arrival process and service time distribution. The fluid flow equation can then be written in terms of the *ensemble average* of arrival rate and departure rate as:

$$\dot{x}(t) = -\mu C G(x(t)) + \lambda(t) \qquad (2)$$

Given an initial condition of the state variable at time zero as $x(0)$ and an approximation of the arrival rate as a constant λ over a small time step $[0, \Delta t]$, we can determine the state variable at the end of the time interval $x(\Delta t)$ by numerically integrating Equation (2), and then set $x(\Delta t)$ as an initial condition for the next time step $[\Delta t, 2\Delta t]$. The arrival rate for the new time step is adjusted if necessary, and this procedure is repeated for each time interval along the time horizon.

2) Multi-class Traffic Fluid Flow Model: We now extend the fluid flow model in Equation (2) to model a queuing system with multi-class input traffic. As illustrated in Fig. 1, a single queue has S classes of input traffic flows with the arrival rate of $\lambda^1(t), \lambda^2(t), \ldots, \lambda^S(t)$, respectively.
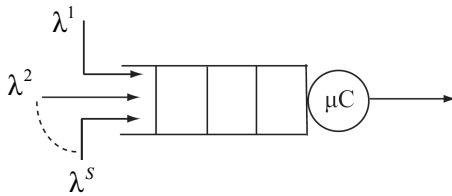

Fig. 1. Queuing model with S classes of traffic

The aggregated traffic can be considered as one arrival process $\lambda(t) = \sum_{l=1}^{S} \lambda^l(t)$. Let $x^l(t)$ represent the ensemble average number of class l packets in the system at time t, the total average number in the system is defined as $x(t) = \sum_{l=1}^{S} x^l(t)$, and the model in (2) becomes:

$$\dot{x}(t) = -\mu C (G(x(t))) + \lambda(t) \qquad (3)$$

We note that the flow conservation principle also applies to each traffic class. Therefore, a state model can be developed for each class with the average link utilization function of class l traffic $G(x_l(t), x(t))$, which is a function of the total average number in the system x and the average number of class l packets in the system x^l.

$$\dot{x}^l(t) = -\mu C (G^l(x^l(t), x(t))) + \lambda^l(t)$$
$$\forall l = 1, 2, \ldots, S \qquad (4)$$

Thus, the multi-class queuing system can be described by a set of S coupled differential equations, each representing the traffic behavior of its own class.

3) Modeling CBR Traffic: We focus on modeling a queue with constant bit rate input traffic streams and consider two cases: (1) where all streams have the same CBR data rate and (2) when each stream is Quasi-CBR traffic with different data rates. The two cases are studied in turn below.

Case I: Identical CBR traffic streams Following [31], we model a $N*D/D/1$ queuing system with the FCFS discipline. There are N input streams with the same packet size as well as the same arrival period D, which is measured in the unit of service period (i.e. D time slots). The first arrival of each flow is randomly phased and assumed to be independently and uniformly distributed over the first arrival period interval $[0, D]$. Since the server operates deterministically with the service rate of one packet per slot, the server utilization equals to $\rho = N/D$, under the constraint of $\rho < 1$ for stability. Let L denote the number of packets present in the system and $Q(r) = \Pr\{L > r\}$ is the survival function of the number of packets in the system. Then, let $A(t-s, t)$ be the number of arrivals in a time interval $(t-s, t)$ within the period D (i.e. $s \le [D]$, the integer part of D). As noted in [31], the survival function can be written as:

$$Q(r) = \sum_{s=1}^{[D]} p_s(r) \pi_0(r, s) \qquad (5)$$

where $p_s(r) = \Pr\{A(t-s, t) = r+s\}$ and $\pi_0(r, s) = \Pr\{\text{system empty at } t-s \mid r+s \text{ arrivals in } (t-s, t)\}$. Noting that the binomial distribution provides the probability of the number of arrivals during the time interval s. Then, the survival function $Q(r)$ [31] can be written as:

$$Q_t(r) = \sum_{s=1}^{N-r} \left[\binom{S}{r+s} \left(\frac{s}{D}\right)^{r+s} \left(1 - \frac{s}{D}\right)^{N-r-s} \right.$$
$$\left. \left(\frac{D-N+r}{D-s}\right) \right] \qquad \text{for } 0 \le r < N \qquad (6)$$

where the first three terms in the sum represent the number of arrivals and the last term represents the probability that the system is initially empty given $r+s$ arrivals. The total average number in the system x can be found using the survival function $Q(r)$, $x = \sum_{r=0}^{N-1} Q(r)$ [32]. Therefore, for the $N*D/D/1$ queue, x is given by:

$$x = \sum_{r=0}^{N-1} \sum_{s=1}^{N-r} \left[\binom{N}{r+s} \left(\frac{s}{D}\right)^{r+s} \left(1 - \frac{s}{D}\right)^{N-r-s} \right.$$
$$\left. \left(\frac{D-N+r}{D-s}\right) \right] \qquad \text{for } 0 \le r < N \qquad (7)$$

The above formula can be used to numerically determine x for a given $N*D/D/1$ queuing system (i.e. the values of N and D are known). Here, we assume that a $N*D/D/1$ queue has a varying number of input CBR streams N, but the CBR traffic period D is unique for all streams. Since the server utilization equals to $\rho = N/D$, the data set (ρ, x) can be obtained from (7) by varying the number of input streams N. Then, we apply a polynomial curve fitting approach using the data set (ρ, x) to

find the utilization function $\rho = G(x)$. The resulting $G(x)$ is in the form of a polynomial (i.e., $G(x) = ax^n + bx^{n-1} + ... + k$) and can be substituted back into the general fluid flow model (3). To determine the utilization function of class l at a queue, we follow the approach of steady state equilibrium matching with $\dot{x}(t) = 0$ and $\dot{x}^l(t) = 0$. Then, substituting them in Equation (3) and (4) respectively, results in

$$\lambda(t) = \mu C G(x(t)) \tag{8}$$

$$\lambda^l(t) = \mu C G^l(x^l(t), x(t)) \tag{9}$$

By combining the above two equations, we obtain the utilization function of class l traffic as

$$G^l(x^l(t), x(t)) = \frac{\lambda^l(t)}{\lambda(t)} G(x(t)) \tag{10}$$

According to Little's theorem, the average packet sojourn time in the queuing system W is equal to the steady-state number of packets x divided by the average arrival rate λ, i.e., $W = x/\lambda$. Because little's theorem also holds for the multi-class FIFO queue [31], it results in $W^l = x^l/\lambda^l$, where W^l is the average sojourn time of class l packets. Since all packets are served based on the FCFS discipline (i.e. $W = W^l$), we have $\lambda^l/\lambda = x^l/x$. Following the same approach of steady state equilibrium matching, we can write

$$\frac{\lambda^l(t)}{\lambda(t)} = \frac{x^l(t)}{x(t)} \tag{11}$$

After substituting Equation (11) in (10), $G^l(x^l(t), x(t))$ can be finally determined as

$$\begin{aligned} G^l(x^l(t), x(t)) &= \frac{x^l(t)}{x(t)} G(x(t)) \\ &= \frac{x^l(t)}{x(t)} \left[ax^n(t) + bx^{n-1}(t) + ... + k \right] \\ &\forall l = 1, 2, \ldots, S \end{aligned} \tag{12}$$

The resulting $G^l(x^l(t), x(t))$ can be substituted into (4) to provide the multi-class traffic fluid flow model. Notice that S represents the number of traffic classes in the queue in Equation (4) and (12), while N in Equation (6) and (7) denotes the number of input traffic streams into the queue. Since multiple input streams could be considered as a single class of traffic and buffered in the same subqueue, we have $N \geq S$.

Case II: Non-identical Quasi-CBR traffic streams. Consider the case where a group of heterogeneous Quasi-CBR (QCBR) traffic streams with different data rates are multiplexed on a transmission link under the condition that the total bit rate is less than the transmission capacity to ensure stability. In our study, each QCBR stream is expected to be transmitted at the requested constant bit rate, but delay jitter between successive arrival packets may occur due to either PHY layer propagation error or MAC layer collision in wireless networks. The packet size remains fixed, but the packet service time could also be quasi-deterministic with some jitter. We denote this type of queue as $Quasi - N * D/D/1$ queue. The exact formula for the queue length distribution in this type of queue cannot be obtained. Here, we propose a simple but effective approach

based on our analysis in case I. Specifically, the utilization function of the queue is bounded by assuming "homogeneous traffic". Suppose there are N input streams and the average packet inter-arrival time of stream i is denoted as D_i for $i = 1, 2, \ldots, N$. Here, the packet inter-arrival time is measured in the unit of service period. For a lower bound, all the input traffic streams are assumed to be fixed with the period of $D_{max} = \max\{D_i\}$, and for an upper bound, the traffic period equals to $D_{min} = \min\{D_i\}$. In addition, we use the average traffic period of all traffic streams with $D_{avg} = \frac{1}{N} \sum_{i=1}^{N} D_i$ to approximate the utilization function. Hence, we apply D_{max}, D_{min} and D_{avg} into $N * D/D/1$ steady-state formular (7) to obtain the data pair (ρ, x) in the cases of "lower and upper bounds" as well as "average approximation", and then determine the utilization function $G(x(t))$ for each case by curve fitting. After that, one can find the utilization function for each traffic stream using (12) and then substitute it back into the fluid flow model (4).

4) A Queue with a Large Number of CBR Traffic Streams: Since the implementation of Equation (6) requires $O(N^2)$ CPU operations to calculate $Q(r)$, a large number of input CBR traffic streams N results in considerable computation to calculate the average number of packets x using $x = \sum_{r=0}^{N-1} Q(r)$. Hence, an approximation is desirable to reduce the computation complexity when the transmission link carries a large number of streams.

For case I, when the traffic load in the queuing system has $\rho = N/D < 0.9$, the arrival process consisting of a superposition of a large number of periodic processes tends to a Poisson process, thus a $M/D/1$ approximation works reasonably well [31]. Then, the utilization function $G(x(t))$ of $N * D/D/1$ in this case can be obtained by referring to $M/D/1$ case in [17] (i.e., $G(x(t)) = x(t) + 1 - \sqrt{x(t)^2 + 1}$). In the heavy server utilization regime for $0.9 \leq \rho < 1$, the Poisson arrival approximation does not hold [31]. Hence, we derive an approximation of the utilization function as follows.

According to the definition of survival function, the value of $Q(r)$ is in the range of $[0, 1]$, and monotonically decreases with the increase of r. Let's consider $x = \sum_{r=0}^{N-1} Q(r)$ as the summation of $Q(r)$ over N steps. When $N \gg 1$, the step size is small compared with N, and the values of $Q(r)$ at two adjacent steps becomes very close. Then, x can be approximated as the integration of $Q(r)$, that is:

$$x \cong \int_0^{N-1} Q(r) dr \tag{13}$$

In the heavy traffic regime $\rho \to 1$, we have the following result from [31], based on a Brownian approximation.

$$Q(r) \approx e^{-2r\left(\frac{r}{N} + \frac{1-\rho}{\rho}\right)} \tag{14}$$

By substituting Equation (14) into (13) and integrating, we have

$$\begin{aligned} x \cong \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} &\left(\text{erf}\left(\frac{(\rho-1)\sqrt{2N}}{2\rho}\right) \right. \\ &\left. + \text{erf}\left(\frac{(N\rho - 2\rho + N)}{\sqrt{2N}\rho}\right) \right) \end{aligned} \tag{15}$$

Since $N \gg 1$ and $\rho \to 1$, the argument of the second erf term in Equation (15) satisfies $\frac{(N\rho - 2\rho + N)}{\sqrt{2N}\rho} \gg 1$. According to the property of erf function, this term can be approximated as 1. Hence, Equation (15) becomes:

$$x \cong \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} \left(\text{erf}\left(\frac{(\rho-1)\sqrt{2N}}{2\rho} \right) + 1 \right) \quad (16)$$

Since $\rho \leq 1$ and $\text{erfc}(X) = \text{erfc}(-X) + 1$, Equation (16) can be rewritten as

$$x \cong \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} \text{erfc}\left(\frac{(1-\rho)\sqrt{2N}}{2\rho} \right) \quad (17)$$

In [33], an elementary approximation is developed for $e^{X^2}\text{erfc}(X)$, with a maximum relative error less than 0.0033 for all $X \geq 0$, that is

$$\exp(X^2)\text{erfc}(X) \approx \frac{1}{AX + \sqrt{1 + BX^2}} \quad (18)$$

where $A = 377/324$, and $B = 314/847$. Since the server utilization is defined as $\rho = N/D$, we substitute N by ρD in (17). By using the approximation (18) in (17), we obtain the functional relationship between x and ρ for the queue with CBR traffic in the heavy server utilization $\rho = [0.9, 1)$ as

$$x \cong \frac{\sqrt{2\pi\rho D}}{\frac{2A(1-\rho)\sqrt{2\rho D}}{\rho} + 2\sqrt{4 + \frac{2B(1-\rho)^2 D}{\rho}}} \quad (19)$$

In Fig. 2, the approximation accuracy of (19) is evaluated by comparing the simulation results of a single queue as well as the exact results generated by (7) for different cases of period D. As shown in the figure, by solving (19), we can efficiently calculate x as a function of ρ with accuracy. After that, we apply polynomial curve fitting to obtain the utilization function $G^l(x^l(t), x(t))$ in the form of (12).
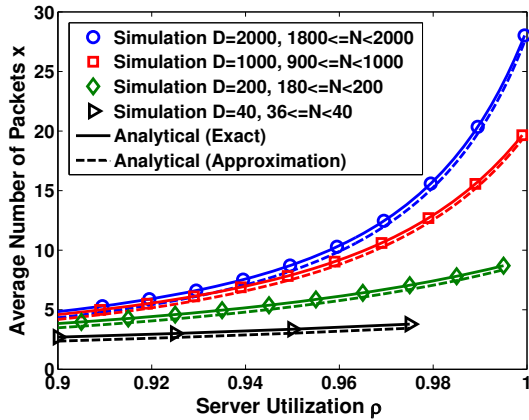

Fig. 2. Comparison of approximation (19) with simulation results and exact analytical results by (7) over $0.9 \leq \rho < 1$.

For case II with a large number of input traffic streams ($\sum_i N_i \gg 1$), we assume "homogeneous traffic" with the lower and the upper bound of traffic period D_{max} and D_{min} as well as the average traffic period D_{avg}, according to Section III-B-3 case II. Depending on the server utilization ρ, we then apply the $M/D/1$ or approximation (19) to efficiently compute the data pair (x, ρ) for the upper and lower bounds. Finally, the bounding utilization functions can be obtained in the same way as case I.

C. Modeling Multihop Wireless Networks with CBR traffic

Consider a network consisting of M nodes, an arbitrary node i is shown in Fig. 3. The data rates of the total incoming traffic flows and outgoing traffic flows at an arbitrary node i are denoted by f_{in_i} and f_{out_i}, respectively. The incoming traffic flows include the CBR traffic generated by node i as well as the forwarded traffic flows from different neighboring nodes. At each node, the packets are grouped into $M - 1$ classes according to their final destinations. We name the traffic destined for node j as class j traffic. Let x_i^j denote the average number of packets in the queuing system at node i destined for node j (class j) and the total number of packets in node i is $x_i = \sum_{\substack{j=1 \\ j \neq i}}^{M} x_i^j$. We represent the packet length as $1/\mu$ and the transmission capacity of node i as C_i. When considering the network as a whole, we must modify (4) to clearly identify the source node i and the destination node j for each variable $x_i^j(t)$, as well as to model the traffic being routed through intermediate nodes when a direct link is not accessible. We use $a_{ij}(t)$ to determine node connectivity, as described in Section III-A. In order to model network routing, we define the routing variable $r_{ik}^j(t)$ as a zero/one indicator variable determined by the routing algorithm, with $r_{ik}^j(t) = 1$ if class j traffic at node i is routed to node k at time t and $r_{ik}^j(t) = 0$ otherwise.

As we can see from Fig. 3, the outgoing traffic rate $f_{out\_i}^j$ at node i and destined for node j is composed of traffic flow to the next-hop node k, where $k = 1, 2, ..., M$ and $k \neq i$. The traffic flow $f_{out\_i}^j$ out of node i depends upon the existence of a direct link $a_{ik}(t)$ between node i and the next-hop node k as well as the routing variables $r_{ik}^j(t)$ for class j traffic. Hence, one must modify the flow out term of (4) to incorporate $a_{ik}(t)$ and $r_{ik}^j(t)$, resulting in

$$f_{out\_i}^j(t) = \mu C_i G_i^j(x_i^j(t), x_i(t)) \sum_{\substack{k=1 \\ k \neq i}}^{M} a_{ik}(t) r_{ik}^j(t) \quad (20)$$

The incoming traffic rate $f_{in\_i}^j$ at node i destined for node j consists of traffic generated at node i with rate $\gamma_i^j(t)$ as well as forwarded traffic flow from the neighboring node l, where $l = 1, 2, ..., M$ and $l \neq i, j$, as shown in Fig. 3. By considering link connectivity as well as routing, we then have

$$f_{in\_i}^j(t) = \gamma_i^j(t) + \sum_{\substack{l=1 \\ l \neq i, j}}^{M} \left(\mu C_l G_l^j(x_l^j(t), x_l(t)) a_{li}(t) r_{li}^j(t) \right) \quad (21)$$

To interconnect queues, the literature [34] indicates that the output from a queuing system with deterministic service time should be treated as a delayed input to the next stage. This idea is applicable to our model, where the input to the next stage is a superposition of the delayed input streams from the nearby nodes plus any external arriving traffic. We illustrate the concept by considering a simplified two-stage tandem queuing model as in Fig. 4(a)-(b). Let $x_i(t)$, $\lambda_i(t)$ and $G_i(t)$ be the average number of packets, the total arrival rate and the utilization function of node i at time t, respectively. Then, $\lambda_1(t) = \gamma_1(t)$ is the arrival rate to the first queue, and $\mu C G_1(t)$ is the departure rate from the first queue. The
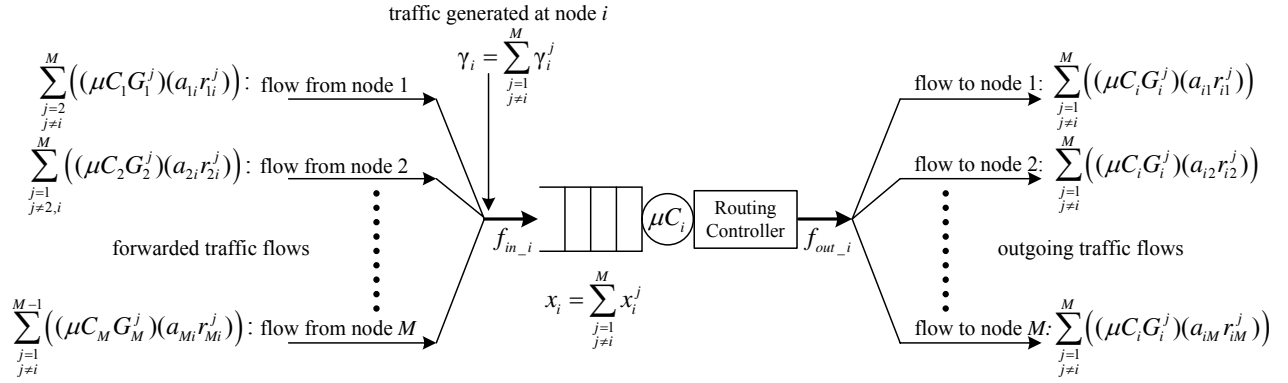
Fig. 3. An arbitrary node i queuing model

departure rate then becomes the input to the second queue after a deterministic forwarding delay δ_1 in the first queue, that is $\lambda_2(t) = \mu C G_1(t - \delta_1) + \gamma_2(t)$. We can then write a set of fluid flow equations at node 1 and node 2 for Fig. 4 as:

$$\dot{x}_1(t) = -\mu C G_1(t) + \gamma_1(t) \qquad (22)$$

$$\dot{x}_2(t) = -\mu C G_2(t) + \gamma_2(t) + \mu C G_1(t - \delta_1) \qquad (23)$$
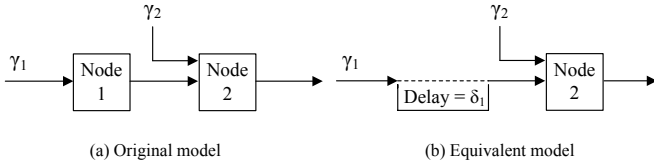

Fig. 4. A two-node deterministic service system with its equivalent model.

The general model of a M node network with deterministic service time is obtained by combining the fluid-flow model with connectivity, routing and the delayed output model for each traffic class at a node. Specifically,

$$\dot{x}_i^j(t) = -\mu C_i G_i^j(x_i^j(t), x_i(t)) \sum_{\substack{k=1 \\ k \neq i}}^{M} a_{ik}(t) r_{ik}^j(t) + \gamma_i^j(t)$$

$$+ \sum_{\substack{l=1 \\ l \neq i,j}}^{M} \left(\mu C_l G_l^j(x_l^j(t - \delta_l), x_l(t - \delta_l)) a_{li}(t) r_{li}^j(t) \right)$$

$$\forall i, j = 1, 2, \ldots, M \quad (24)$$

In (24), the first term to the right of the equal sign represents the flow of class j traffic out of node i, the second term denotes the type j traffic entering the network at node i, and the last term characterizes the flow of class j traffic being routed into node i from other nodes. For a queue with the superposition of periodic arrival streams, the server utilization function $G(.)$ can be written in the form of polynomial expression given in

(12). Therefore, the fluid flow model is

$$\dot{x}_i^j(t) = -\mu C_i \frac{x_i^j(t)}{x_i(t)} \left[a x_i^n(t) + b x_i^{n-1}(t) + \ldots + k \right] \times$$

$$\sum_{\substack{k=1 \\ k \neq i}}^{M} a_{ik}(t) r_{ik}^j(t) + \gamma_i^j(t) + \sum_{\substack{l=1 \\ l \neq i,j}}^{M} \left(\mu C_l \frac{x_l^j(t - \delta_l)}{x_l(t - \delta_l)} \times \right.$$

$$\left. \left[a x_l^n(t - \delta_l) + b x_l^{n-1}(t - \delta_l) + \ldots + k \right] a_{li}(t) r_{li}^j(t) \right)$$

$$\forall i, j = 1, 2, \ldots, M \quad (25)$$

Given a routing algorithm, connectivity model and traffic information, this model can be solved numerically using any standard numerical integration technique.

D. Additional Performance Metrics

The fluid flow modeling approach can be used to determine a variety of performance metrics. First of all, we discuss the estimation of the end-to-end delay. Typically, a packet is forwarded from the source via a path which may include several intermediate nodes until it reaches the destination. As a result, the end-to-end delay is the sum of delays experienced at each node along the way. The packet delay at a node consists of the queuing delay, the transmission delay and the propagation time over a link. Usually, the queuing and transmission delays are considered as the main factors. From Little's theorem, the average number in the system is equivalent to the product of the average arrival rate and the average sojourn time in the system, which includes the queuing an transmission delay. If x denotes the average number of packets in the system, λ the average arrival rate and W the average sojourn time, then $x = \lambda W$. With the assumption of a constant mean arrival rate over a small step, the change in average sojourn time can be related to the rate of change in the average number of packets in the system $\dot{W} = \dot{x}/\lambda$. Now consider a path for stream (s, d) from source node s to destination node d selected by routing algorithm. We define $P^{(s,d)}$ as the set of all nodes on this path except destination node d. Let W_i denote the average node delay at node i on this path, and $W^{(s,d)}$ represents the end-to-end delay of path $P^{(s,d)}$. The rate of change of this path delay is obtained by

$$\dot{W}^{(s,d)}(t) = \sum_{i \in P^{(s,d)}} \dot{W}_i = \sum_{i \in P^{(s,d)}} \frac{\dot{x}_i(t)}{\lambda_i(t)} \qquad (26)$$

where $\lambda_i(t)$ is the total arrival rate into node i at time t,(i.e. $\lambda_i(t) = \sum_{\substack{d=1 \\ d \neq i}}^{M} f_{in\_i}^d(t)$, with $f_{in\_i}^d(t)$ determined by (21)). We denote ε as the link propagation delay, which is assumed to be fixed and almost equal for each hop on the path. According to the definition of set $P^{(s,d)}$, the number of hops along the path of traffic stream (s,d) is equal to the cardinality (size) of the set, i.e. $|P^{(s,d)}|$. Hence, after adding the link propagation delays to Equation (26), we can finally write the end-to-end delay of path $P^{(s,d)}$ at time t as

$$D^{(s,d)}(t) = W^{(s,d)}(t) + |P^{(s,d)}|\varepsilon \qquad (27)$$

In addition, the fluid flow model can also estimate the following global performance metrics. The average number of packets per node at time t is obtained by dividing the total number of packets in all nodes at time t by the total number of nodes M in network, i.e.

$$x_{avg}(t) = \frac{1}{M} \sum_{i=1}^{M} x_i(t) = \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} x_i^j(t) \qquad (28)$$

The average end-to-end delay per traffic stream in network can also be determined. Let υ be the total number of traffic streams (s,d) in network. Thus, the average end-to-end delay per traffic stream in network at time t is given by

$$ETE_{avg}(t) = \frac{1}{\upsilon} \sum_{s=1}^{M} \sum_{\substack{d=1 \\ d \neq s}}^{M} D^{(s,d)}(t) \qquad (29)$$

Similarly, we can obtain the average utilization per link in network at time t as

$$G_{avg}(t) = \frac{1}{M} \sum_{i=1}^{M} G_i(t) = \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} G_i^j(t) \qquad (30)$$

Finally, one can also determine the instantaneous network throughput in bit per second as

$$T(t) = \sum_{d=1}^{M} \sum_{\substack{i=1 \\ i \neq d}}^{M} \left(C_i G_i^d(x_i^d(t), x_i(t)) a_{id}(t) r_{id}^d(t) \right) \qquad (31)$$

Here, the network throughput is measured by the traffic received by all the destination nodes (i.e. $d = 1, 2, \ldots, M$), and the traffic received by destination node d is calculated by summing up the traffic successfully sent from all its neighboring nodes (i.e. $i = 1, 2, \ldots, M$, and $i \neq d$).

E. Numerical Solution Algorithm

We codify our modeling procedure into the following algorithm to estimate the time verying network performance over a desired time interval $[t_0, t_f]$.

1. Configure network parameters including the link capacity C, the packet length $1/\mu$, the traffic load γ and D the service period.
2. Compute the data pair (ρ, x) off-line by using (7) for strict CBR traffic, or the bounded and averaged data pair (ρ, x) for QCBR traffic as discussed in Section III-B-3. If the number of input traffic streams N is large (e.g. $N \geq 30$), approximate the data pair (ρ, x) by (19).
3. Find the utilization functions off-line in the form of the polynomial (12) by curve fitting the data pair (ρ, x).
4. Set the current time $t = t_0$ as well as a time step Δt and initialize $x_i^j(t) = x_i^j(t_0)$, which is node i's initial occupancy by the packets destined for node j.
5. At time t, determine the traffic routes $r_{ik}^j(t)$ according to the routing protocol and the adjacency matrix $A(t)$. Also, update the offered traffic $\gamma_i^j(t)$ at each node, if necessary.
6. Numerically solve the fluid flow network model (25) and get the new $x_i^j(t+\Delta t)$ at the end of the time interval $[t, t+\Delta t]$, which becomes the initial condition for $[t + \Delta t, t + 2\Delta t]$.
7. Estimate the end-to-end delay $D^{(s,d)}(t + \Delta t)$ by summing up the link propagation delays ε with the node queuing delays $W^{(s,d)}(t + \Delta t)$ along the path $P^{(s,d)}$, given by (26) and (27). Here, $W^{(s,d)}(t + \Delta t)$ is obtained by numerically solving the differential equation over the time interval $[t, t + \Delta t]$, which is the initial condition for $[t + \Delta t, t + 2\Delta t]$.
8. Evaluate the global performance metrics including $x_{avg}(t + \Delta t)$, $ETE_{avg}(t + \Delta t)$, $G_{avg}(t + \Delta t)$ and $T_{avg}(t + \Delta t)$, according to (28)-(31).
9. Increment time $t = t + \Delta t$. If $t < t_f$, go back to step 5; else stop.

Any standard numerical integration method can be used to solve differential equations such as (25). Here, we use the *Runge-Kutta* algorithm in Matlab to generate numerical results.

IV. NUMERICAL RESULTS

A. Model Validation by Discrete Event Simulation

In this section, our model is compared with an equivalent discrete event simulation model built in OPNET [35]. In the simulation model, each queue of a node is configured as a FCFS queue with infinite size buffer. The output traffic out of the queue with the same destination is multiplexed on a single link, but can be virtually separated based on different streams. The simulation results are the ensemble average of 5000 replications with 98% confidence intervals using the nonstationary simulation approach discussed in [12] [15].

As illustrated in Fig. 5(a) - (f), a simple scenario of three nodes with pre-determined connectivity change between nodes is studied here. In this setup, when the direct link is no longer available, traffic must be rerouted through relay nodes and uses some portion of the shared link capacity. We set the link capacity for all nodes to $C = 10^4$ bps with the packet length $1/\mu = 1250$ bytes, so that the average service rate is normalized to one packet per second. The forwarding delay δ is assumed to be equal to the service period, (i.e. $\delta = 1$ second). The link connectivity a_{ij} is set to be binary and the link propagation delay ε is set to be 0.1 microsceond. In addition, we use minimum hop routing to find a path for each traffic flow. This three-node network has the corresponding queuing model, as shown in Fig. 6.

We first study case I denoted by $N * D/D/1$ queue. The rate of external traffic arrival streams at each node have the same average rate, e.g. $\gamma_1^2 = \gamma_1^3 = \gamma_2^1 = \gamma_2^3 = \gamma_3^1 = \gamma_3^2 = 0.2$ packets per second but are not synchronized (i.e., first packet
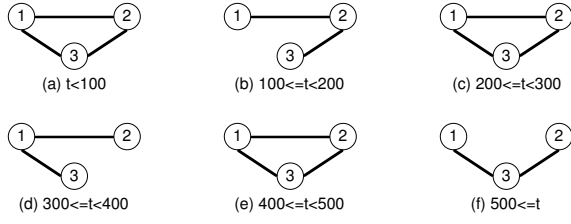
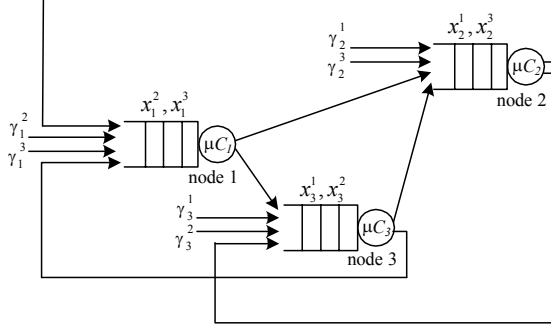Fig. 5.   Three node network connectivity scenario.

Fig. 6.   Three node network queuing model.

Fig. 7.   Average number of packets x_1^2 and end-to-end delay of D_{1-2}.

arrival time of each stream is a uniformly distributed random variable over $[0, 5]$ sec). After curve fitting (ρ, x) determined from (7), the server utilization function is determined as $G(x) = 0.0832x^3 - 0.4353x^2 + 1.0843x$, which is then used in (25) to model the network. Fig. 7 shows the results of the effect of topology change on the average number of packets and the end-to-end delay for the traffic at node 1 destined for node 2. For the time interval $t < 100$s, the network is fully-connected. All nodes go through an initial transient period and then reach steady state. The end-to-end delay D_{1-2} comes from the propagation delay in link 1-2 and the queuing delay at node 1. During the time $100 \leq t < 200$s, the link between node 1 and 3 breaks and traffic going through this link has to go through the relay node 2. But the average number of packets x_1^2 at node 1 is not affected. For the time interval $300 \leq t < 400$s, the link between node 2 and node 3 breaks, leading to traffic re-routing and an increase in x_1^2 at node 1. Due to higher server utilization of node 1, the average queuing time of each packet at node 1 increases and we can see the rise in D_{1-2}. During the time $t \geq 500$s, link 1-2 breaks and the traffic of x_1^2 has to go through the relay node 3 to reach the destination. Hence, D_{1-2} consists of the propagation delay on links 1-3 and 3-2 as well as the queuing delay at node 1 and node 3. The behavior of other traffic streams is similar and not discussed here for the sake of brevity.

Next we consider a five node network with the random waypoint mobility (RWM) model. The stochastic properties of the RWM model were studied in [30]. It was observed that the link connectivity of two nodes is shown to be a memoryless stochastic process that can be modeled as a two-state Markov process with up-down (connected-disconnected) transition. The Markov process based link connectivity model can be used to greatly reduce the computation load in comparison of a detailed node mobility simulation. In this experiment, the up and down durations of each link are exponentially distributed
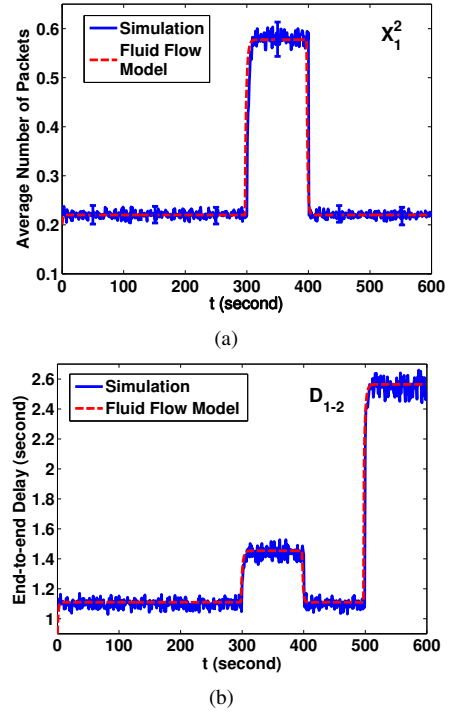
with the mean of T_{up} = 50s, T_{down} = 10s, respectively. The link capacity and the packet length remain the same as the ones in the three node scenario. The external arrival rates of QCBR traffic are: $\gamma_1^3 = 0.24, \gamma_1^5 = 0.16, \gamma_2^5 = 0.18, \gamma_3^5 = 0.22, \gamma_4^5 = 0.24$ packet per second. We conduct the experiment for a total duration of 6000s, and show the time varying link connectivity during the time interval $[2100, 2200]$s in Fig. 8. In the following discussion, we focus on the traffic buffered at node 1 and destined for node 5 (i.e. x_1^5). The routes of this traffic are marked by dotted lines in Fig. 8.
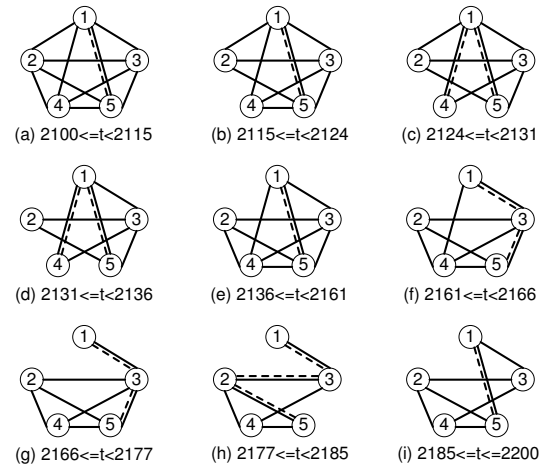

Fig. 8.   Typical RWM model connectivity scenario for five node network.

To obtain the results from fluid flow model, we calculate the utilization function of lower and upper bounds by assuming "homogeneous traffic". Due to delay jitter, the average period of each QCBR stream is slightly deviated from the requested one. For the lower bound, all sources are assumed

to have homogeneous traffic with the period of $D_{max} = \max\{D_i\} = 1/0.162 \approx 6.173$s, while for the upper bound, the homogeneous traffic period becomes $D_{min} = \min\{D_i\} = 1/0.243 \approx 4.115$s. The average traffic period is given by $D_{avg} = \text{avg}\{D_i\} = (1/0.243+1/0.162+1/0.184+1/0.221+1/0.242)/5 \approx 4.88$s. The utilization function for each case is obtained as $G_{lower}(x) = 0.0884x^3 - 0.4485x^2 + 1.0712x$, $G_{upper}(x) = 0.0747x^3 - 0.4175x^2 + 1.0971x$ and $G_{avg}(x) = 0.0757x^3 - 0.4261x^2 + 1.0842x$, which is then used in (25) to form the fluid flow model of the network. In addition, we plot the results obtained from the Pointwise Stationary Approximation (PSA) [36] modeling approach which approximates the nonstationary queuing system by using steady-state formula at each time point. Fig. 9 shows the results for x_1^5 and D_{1-5} when it is affected by the topology changes. Initially, every packet goes through the direct link. Then, during the time interval $2124 \leq t < 2136$s, link 4-5 breaks and the traffic x_4^5 needs to go through node 1 to reach the destination. Hence, a large transient increase of x_1^5 occurs at node 1 due to traffic rerouting. This event also results in the increase of D_{1-5}, because of the higher utilization of node 1. After that, link 4-5 is recovered and the traffic x_4^5 reroutes back to the direct link. Starting from $t = 2161$s, link 1-5 breaks and the routing protocol redirects the traffic x_1^5 to node 3, until this direct link restored at $t = 2185$s. Notice that, at $t = 2177$s, link 3-5 is disconnected, which causes the traffic x_1^5 to take one more hop from node 3 to node 2 and a further increase in D_{1-5}. At $t = 2185$s, link 1-5 is restored and the traffic x_1^5 is rerouted to the direct link resulting in a decrease of D_{1-5}. As seen in the figure, PSA method cannot capture the transient/nonstationary behavior of the network. Instead, our proposed fluid flow model can provide fairly accurate instantaneous results or tight bounds, all which match well with the discrete event simulation results in Fig. 7 and Fig. 9.

B. Computational Complexity

We first analyze the computation complexity of our fluid flow model according to the algorithm in Section III-E. Initially, the utilization functions are pre-computed based on the given network parameters, as seen from steps 1 to 3. This one-time off-line computation process is time efficient regardless of network size, thus this process is not counted in the following on-line complexity analysis. From step 4 to 8, the network performance metrics are estimated by integrating a set of differential equations with a specific adjacency matrix and routing variable at each time interval Δt. The exact number of arithmetic operations required for solving the differential equations over one step time is hard to determine [37]. However, an upper bound on the computation time complexity can be obtained. Let T refer to the desired simulation time interval, then $T/\Delta t$ represents the total number of steps. Let K represent the average time to execute one arithmetic operation on a CPU. Following [37], $C(n, p, \alpha)$ denotes the upper bound on the number of arithmetic operations required for each step time, so that n differential equations can be solved by pth order explicit Runge-Kutta algorithm with maximum error $e^{-\alpha}$. Hence, an upper bound of the model computation time turns out to be $K \cdot (T/\Delta t) \cdot C(n, p, \alpha)$. According to the
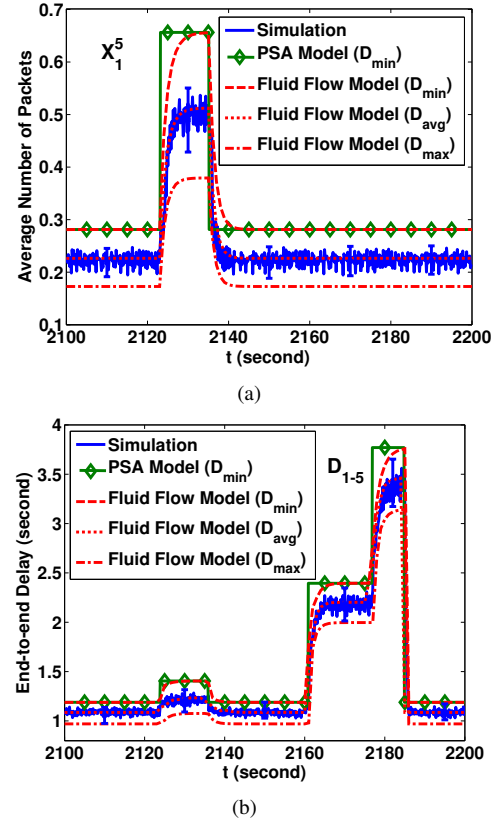

(a)

(b)

Fig. 9.   Average number of packets x_1^5 and end-to-end delay of D_{1-5}.

expression of $C(n, p, \alpha)$ in [37] with the predefined value of p and α, C increases linearly with n. As a result, only considering the increase of n in $K \cdot (T/\Delta t) \cdot C(n, p, \alpha)$, the computational time complexity of our model is upper bounded by $O(n)$. For an M-node network with full-mesh traffic load, the number of differential equations n equals to the number of traffic flows $M(M - 1)$. Therefore, the computation time complexity of our fluid flow modeling algorithm is upper bounded by $O(M(M - 1))$.

To further evaluate the computational complexity, we conducted numerical experiments on a series of sample networks. In the experiments, each node generates CBR traffic to all the other nodes resulting in a full mesh of identical source traffic. All links switch between on/off randomly following the two-state Markov process. Table I shows the computation time of fluid flow model and simulation for the sample networks over the time $[0, 600]$s, which are run on a PC with Intel i5-450M 2.4GHz processor and 4GB memory. The reported computation time by our model excludes the short time spent on the off-line stage. In addition, the minimum hop routes are pre-determined based on the network connectivity and stored in the routing table of each node, so the route discovery time of both approaches is not counted in Table I. In all sample networks, we set the CBR traffic rate for each node as 0.02pkt/s and the service rate is assigned to be 1pkt/s. Then, the utilization function is obtained as $0.0023x^5 - 0.0329x^4 + 0.1898x^3 - 0.5756x^2 + 1.025x$. To study the time varying behavior of the network via simulation, we perform the nonstationary simulation, which is to average

over an ensemble of statistically identical results generated by distinct independent runs with different random number seeds [12] [15]. Here, we execute 5000 independent runs in OPNET to observe the ensemble averaged system behavior versus time. As seen from the table, the computation time of nonstationary simulation grows dramatically, which is generally a complex function of number of nodes, traffic load, topology change, accuracy desired, etc. For the fluid flow model results, we curve fit the computation time data versus network size M and obtain the growth rate as $\Theta(0.87M^2 - 7.9M + 20)$, which is within our expected upper bound above.

TABLE I
COMPUTATION TIME COMPARISON

# of Nodes	# of Traffic Flows	Simulation (second)	Fluid Flow Model (second)
3	6	138.19	**2.23**
5	20	683.52	**4.29**
7	42	3421.59	**10.01**
9	72	17025.38	**20.73**
11	110	81092.42	**37.93**
20	380	489723.92	**212.56**
30	870	1046392.75	**564.72**

C. Example Network Performance Analysis

In this section, we illustrate the application of the fluid flow model approach with the study a 30-node network with full mesh CBR traffic loads. We focus on two local performance metrics (i.e. x_1^{30}, and D_{1-30}) plus four global or network-averaged metrics (i.e. x_{avg}, G_{avg}, ETE_{avg}, and T), as we derived in Section III-C and D. Here, we configure the network parameters as follows: the fixed packet size $1/\mu$ of 1250 bytes, the link capacity C of 10^4 bps, the forwarding delay δ of 1 second, the link propagation delay ε of 0.01 seconds and single-path minimum hop routing. The full mesh offered traffic at source node has the same requested data rate, but there exist slight delay jitter and possible packet loss in the forwarding traffic flows. We denote γ as the averaged arrival rate of all traffic flows (i.e. $\gamma = 1/D_{avg}$). To obtain the utilization function in this 30 node network, we compute the average queue length x in an efficient way by adopting the approximation for large number of input streams, as discussed in Section III.B.4. When the server utilization ρ stays in the light or moderate regime, and the utilization function is given by $G(x) = x + 1 - \sqrt{x^2 + 1}$ based on a $M/D/1$ approximation. When the link utilization ρ reaches 0.9 or above at certain node due to traffic forwarding, the utilization function has to change to the polynomial $G(x)$, which is determined by using the curve fitting to Equation (19). In addition, we use the two-state Markov model of RWM with the average link up lifetime T_{up} and the average link down lifetime T_{down} to represent the network mobility.

We first study the effect of traffic load on network performance. The growth of offered traffic load at a node inevitably results in an increase of packets in its buffer due to the limited link transmission capacity and a corresponding increase in the delay. In Fig. 10(a), x_1^{30} for the scenarios of

$(T_{up}, T_{down}, \gamma, a_{ij})$ = (50s, 20s, 0.015 pkt/s, $\{0, 1\}$) and (50s, 20s, 0.02 pkt/s, $\{0, 1\}$) is shown. Similar behavior is shown in Fig. 10(b) for D_{1-30}, which is determined by the queuing delays of all the nodes along the path. From the perspective of the whole network, link utilization, average number of packets at a node and the end-to-end delay all increase with the load of full-mesh traffic, as seen in Fig. 10(c)-(e). Fig. 10(f) plots the instantaneous network throughput, which fluctuates around the constant network load due to node mobility and traffic rerouting. For the full-mesh traffic, the averaged network load L can be calculated by $L = M(M-1) \times 1/\mu \times \gamma$. Since the average link utilization is always operated in the moderate regime (i.e. $\rho < 0.9$ in Fig. 10(c)), the steady-state (time-average) network throughput is shown to be roughly equal to the network load for both cases.

Next, we investigate the impact of node mobility on network performance. Here, we change the mobility model (T_{up}, T_{down}) from (50s, 20s) to (35s, 35s) to represent different cases of network connectivity and keep the offered traffic as 20×10^{-3} pkt/s. Comparing (50s, 20s, 0.02 pkt/s, $\{0, 1\}$) with (35s, 35s, 0.02 pkt/s, $\{0, 1\}$) in Fig. 10(a)-(b), we observe that shorter link uptime with longer link downtime for each link results in higher levels of nonstationarity, since nodes will get less overall connectivity time and be more likely to redirect the traffic to others. Moreover, the smaller ratio between link up and down time in the case of (T_{up}, T_{down}) = (35s, 35s) brings longer routes with more forwarding traffic at each node and greater end-to-end delay than the case of (50s, 20s). Fig. 10(c) shows that the average link utilization of (35s, 35s, 0.02 pkt/s, $\{0, 1\}$) is much higher. Due to limited link capacity, the number of packets accumulated in the buffer of each node rises up in Fig. 10(d). Meanwhile, the average end-to-end delay per traffic stream climbs up in Fig. 10(e) due to large queuing delay and long routes. All these phenomena demonstrate the occurrence of network congestion in the scenario of (35s, 35s, 0.02 pkt/s, $\{0, 1\}$). As a result, Fig. 10(f) illustrates that network throughput mainly lies below the offered load.

Finally, we consider a more realistic link quality and incorporate it into our fluid flow based model. Actual radio communication is not always symmetric and may exhibit diverse link quality in terms of error rate. Hence, the adjacency matrix is not necessary a binary matrix, and the connectivity a_{ij} can be any real number between 0 and 1 to indicate the effect of link errors. Here, we assume that the link connectivity $a_{ij}(t)$ is assigned a random number between 0.9 and 1, if the distance between two nodes $d_{ij}(t)$ is within the radio range R (i.e. $a_{ij}(t) \in [0.9\ 1]$ if $d_{ij}(t) \le R$, otherwise $a_{ij}(t) = 0$). Due to node mobility, the connectivity $a_{ij}(t)$ is updated for all links at each time instant. By comparing scenario (35s, 35s, 0.02 pkt/s, $\{0, 1\}$) and (35s, 35s, 0.02 pkt/s, $\{0, [0.9\ 1]\}$) in Fig. 10(a) and (d), we can see that after the link-level error is incorporated into the model, a portion of packets cannot be successfully delivered to the next hop and the packet service rate is reduced. Hence, more packets have to be buffered in the transmitting node in Fig. 10(d). Also, the increased queuing delay at each node prolongs the end-to-end delay of the traffic in Fig. 10(e). Due to the link-level error, Fig. 10(c) shows that the average utilization per link in the scenario (35s, 35s,
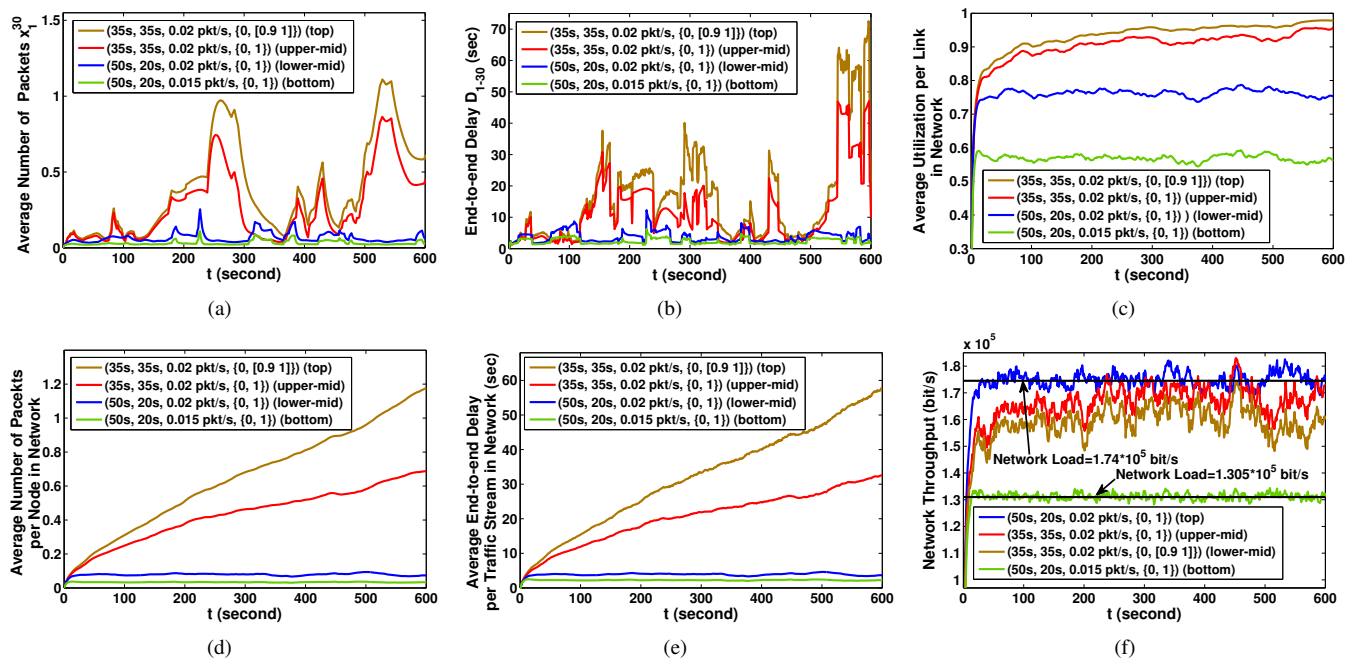
Fig. 10. Various network performance measures impacted by traffic load, node mobility and link quality (i.e. $(T_{up}, T_{down}, \gamma, a_{ij})$).

0.02 pkt/s, $\{0, [0.9 \ 1]\}$) is higher than the one in the scenario (35s, 35s, 0.02 pkt/s, $\{0, 1\}$). Since the network becomes more congested, network throughout in scenario (35s, 35s, 0.02 pkt/s, $\{0, [0.9 \ 1]\}$) is further degraded in Fig. 10(f).

V. CONCLUSIONS

In this paper, we propose a performance modeling technique to study the time varying behavior of multihop wireless networks with CBR traffic by numerical method based queuing analysis. Network queues are modeled using fluid flow based differential equations and solved by numerical integration routines, while topology change is integrated into the model using a time varying adjacency matrix determined from either trace data, a mobility model based simulation, or a deterministic/stochastic model. Numerical results for sample networks using the proposed model were given in comparison with results from discrete event simulations showing the accuracy and the tremendous computational advantage of the fluid flow based approach. Furthermore, we applied the fluid flow model to examine a variety of performance metrics of a sample network. We believe this modeling approach is potentially a valuable tool for evaluating the time-varying behavior of multihop wireless networks. With the computation time saved by the fluid flow based modeling technique, it is a tremendous gain in modeling complex networks with nonstationary effects or exploring design alternative with a quick insight into network performance.

REFERENCES

[1] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. 23–30, 2005.

[2] H. Moustafa and Y. Zhang, *Vehicular networks, Techniques, Standards, and Applications*. CRC Press, 2009.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.

[4] I. Chlamtac, M. Conti, and J. Liu, "Mobile ad hoc networking: Imperatives and challenges," *Ad Hoc Networks*, vol. 1, no. 1, pp. 13–64, 2003.

[5] J. Yang and Z. Fei, "Broadcasting with prediction and selective forwarding in vehicular networks," *International Journal of Distributed Sensor Networks*, vol. 2013, 2013.

[6] J. Li and P. Mohapatra, "Analytical modeling and mitigation techniques for the energy hole problems in sensor networks," *Pervasive and Mobile Computing*, vol. 3, no. 8, pp. 233–254, 2007.

[7] J. Yang and Z. Fei, "HDAR: Hole detection and adaptive geographic routing for ad hoc networks," in *Proc. of IEEE International Conference on Computer Communications and Networks (ICCCN)*, 2010, pp. 1–6.

[8] B. Tavli and W. Heinzelman, "MH-TRACE: multihop time reservation using adaptive control for energy efficiency," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 5, pp. 942–953, 2004.

[9] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," in *Proc. of ACM SIGCOMM*, Sep. 2004, pp. 121–132.

[10] J. Wang, Y. Liu, M. Li, W. Dong, and Y. He, "Qof: Towards comprehensive path quality measurement in wireless sensor networks." in *Proc. of IEEE INFOCOM*, Apr. 2011, pp. 775–783.

[11] C. Newport, D. Kotz, Y. Yuan, R. S. Gray, J. Liu, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," *Simulation: Transactions of the Society for Modeling and Simulation International,*, vol. 83, no. 9, pp. 643–661, 2007.

[12] J. Banks, J. Carson, B. Nelson, and D. Nicol, *Discrete Event System Simulation*, 5th ed. Prentice-Hall, 2010.

[13] K. Pawlikowski, H.-D. Jeong, and J.-S. Lee, "On credibility of simulation studies of telecommunication networks," *Communications Magazine, IEEE*, vol. 40, no. 1, pp. 132–139, 2002.

[14] D. Tipper, Y. Qian, and X. Hou, "Modeling the time varying behavior of mobile ad-hoc networks," in *Proc. of ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2004, pp. 12–19.

[15] W. Lovegrove, J. Hammond, and D. Tipper, "Simulation methods for studying nonstationary behavior of computer networks," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 9, pp. 1696–1708, 1990.

[16] D. Tipper and M. K. Sundareshan, "Numerical methods for modeling computer networks under nonstationary conditions," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 9, pp. 1682–1695, 1990.

[17] W. Wang, D. Tipper, and S. Banerjee, "A simple approximation for modeling nonstationary queues," in *Proc. of IEEE INFOCOM*, 1996, pp. 255–262.

[18] M. Hassan, H. Vu, and T. Sakurai, "Performance analysis of the IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2013.2297382, IEEE Transactions on Vehicular Technology

13

802.11 MAC protocol for DSRC safety applications," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3882–3896, 2011.

[19] S. Sharma and D. Tipper, "Approximate models for the study of non-stationary queues and their applications to communication networks," in *Proc. of IEEE International Conference on Communications (ICC)*, May 1993, pp. 352–358.

[20] K. Xu, S. Tipmongkonsilp, D. Tipper, P. Krishnamurthy, and Y. Qian, "A time dependent performance model for multihop wireless networks with CBR traffic," in *Proc. of IEEE International Performance Computing and Communications Conference (IPCCC)*, Dec. 2010, pp. 271–280.

[21] K. Xu, D. Tipper, P. Krishnamurthy, and Y. Qian, "An efficient hybrid model and dynamic performance analysis for multihop wireless networks," in *Proc. of IEEE International conference on Computing, Networking and Communications (ICNC)*, Jan. 2013.

[22] A. Pitsillides, P. Ioannou, M. Lestas, and L. Rossides, "Adaptive nonlinear congestion controller for a differentiated-services framework," *IEEE/ACM Transactions on Networking*, vol. 13, no. 1, pp. 94–107, 2005.

[23] K. Xu, D. Tipper, P. Krishnamurthy, and Y. Qian, "An framework of efficient hybrid model and optimal control for multihop wireless networks," in *Poster Papers of ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2012.

[24] Y. Gu, Y. Liu, and D. Towsley, "On integrating fluid models with packet simulation," in *Proc. of IEEE INFOCOM*, March 2004, pp. 2856–2866.

[25] D. M. Nicol and G. Yan, "Discrete event fluid modeling of background TCP traffic," *ACM Transactions on Modeling and Computer Simulation*, vol. 14, no. 3, pp. 211–250, Jul. 2004.

[26] H. Kim and J. Hou, "How good is fluid model-based simulation for simulating IEEE 802.11 operated WLANs," in *Proc. of Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS)*, Jan. 2003.

[27] A. Eryilmaz, P. Marbach, and A. Ozdaglar, "A fluid-flow model for backlog-based CSMA policies," in *Proc. of Annual International ICST Conference on Wireless Internet (WICON)*, Jan. 2008, pp. 77:1–77:9.

[28] K. Grochla, T. Czachorski, and J. Klamka, "Modelling TCP connection in WiMAX network using fluid flow approximation," in *Proc. of IEEE/IPSJ 11th International Symposium on Applications and the Internet (SAINT)*, July 2011, pp. 502–507.

[29] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communication & Mobile Computing (WCMC)*, vol. 2, no. 5, pp. 483–502, 2002.

[30] S. K. Hwang and D. S. Kim, "Markov model of link connectivity in mobile ad hoc networks," *Telecommunication System*, vol. 34, no. 1-2, pp. 51–58, 2007.

[31] P. V. Meighem, *Performance Analysis of Communications Networks and Systems*. Cambridge University Press, 2006.

[32] E. T. Lee and J. Wang, *Statistical Methods for Survival Data Analysis*. John Wiley and Sons, 1992.

[33] F. G. Lether, "An elementary approximation for $\exp(x^2)\text{erfc}(x)$," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 43, no. 6, pp. 511–513, Jun. 1990.

[34] M. J. Neely, "Equivalent models for queueing analysis of deterministic service time tree networks," *IEEE Transaction on Information Theory*, vol. 51, no. 10, pp. 3576–3584, 2005.

[35] *OPNET Simulaiton Tool*, http://www.opnet.com/.

[36] E. Gelenbe and J. A. Barria, *Communication Networks and Computer Systems*. Imperial College Press, 2006.

[37] A. G. Werschulz, "Computational complexity of one-step methods for systems of differential equations," *Mathematics of Computation*, vol. 34, no. 149, pp. 155–174, 1980.

David Tipper (S'78-M'88-SM'95) is the Director of the Graduate Telecommunications and Networking Program and a faculty member at the University of Pittsburgh, Pittsburgh, PA. He is a graduate of the University of Arizona (Ph.D. EE, MS SIE) and Virginia Tech (BS EE). His current research interests are survivable networks, performance analysis techniques, wireless/wired network design and information assurance. Professor Tipper's research has been supported by grants from various government and corporate sources such as NSF, DARPA, NIST, IBM, ARO and AT&T. Professional activities include serving as the General Chair of the 7th Design of Reliable Communication Networks Workshop (DRCN2009), co-guest editor of a special issue of the journal Telecommunication Systems on Reliable Networks Design and Modeling which appeared February, 2013. He is the co-author of the textbook The Physical Layer of Communication Systems, which was published by Artech House in 2006. Also, he is the co-editor and a contributor to Information Assurance: Dependability and Security in Networked Systems, which was published by Morgan Kaufmann in 2008.

Yi Qian (M'95, SM'07) received a Ph.D. degree in electrical engineering from Clemson University. He is an associate professor in the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln (UNL). Prior to joining UNL, he worked in the telecommunications industry, academia, and the government. Some of his previous professional positions include serving as a senior member of scientific staff and a technical advisor at Nortel Networks, a senior systems engineer and a technical advisor at several start-up companies, an assistant professor at University of Puerto Rico at Mayaguez, and a senior researcher at National Institute of Standards and Technology. His research interests include information assurance and network security, network design, network modeling, simulation and performance analysis for next generation wireless networks, wireless ad-hoc and sensor networks, vehicular networks, broadband satellite networks, optical networks, high-speed networks and the Internet. He has a successful track record to lead research teams and to publish research results in leading scientific journals and conferences. Several of his recent journal articles on wireless network design and wireless network security are among the most accessed papers in the IEEE Digital Library. Dr. Yi Qian is a member of ACM and a senior member of IEEE.

Prashant Krishnurmathy is an associate professor in the School of Information Sciences, University of Pittsburgh, Pennsylvania, where he regularly teaches courses on wireless networks and cryptography. From 2000 to 2005, he also served as the chair of the IEEE Communications Society Pittsburgh Chapter. His research interests include wireless network security, wireless data networks, position location in indoor wireless networks, and radio channel modeling for indoor wireless networks. He is a member of the IEEE.

Kunjie Xu (S'07) received the B.E. degree in electrical engineering from Southeast University, Nanjing, China in 2006, and the M.S. degree also in electrical engineering from University of North Carolina at Charlotte, NC, USA, in 2009. He is currently a PhD candidate in Telecommunications and Networking Program at School of Information Science, University of Pittsburgh, PA, USA. His research interests include network performance analysis, network simulation techniques and wireless network design.

Siriluck Tipmongkonsilp received the B.E. degree from Chulalongkorn University, Thailand, in 1997, the M.S. degree in electrical engineering from Carnegie Mellon University, PA, USA, in 2001, and the Ph.D. degree in Telecommunications and Networking Program at School of Information Science, University of Pittsburgh, PA, USA in 2009. She is currently with CAT Telecom in Thailand. Her research interests include mobile network simulation and modeling.