TIME DEPENDENT PERFORMANCE ANALYSIS OF WIRELESS NETWORKS

by

Kunjie Xu

B.Eng. in EE, Southeast University, China, 2006

M.S. in EE, University of North Carolina at Charlotte, NC, 2009

Submitted to the Graduate Faculty of

the Graduate Telecommunications and Networking Program in

partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCE

This dissertation was presented

by

Kunjie Xu

It was defended on

June 20th 2014

and approved by

David Tipper, Ph.D., Associate Professor, University of Pittsburgh

Prashant Krishnamurthy, Ph.D., Associate Professor, University of Pittsburgh

Vladimir Zadorozhny, Ph.D., Associate Professor, University of Pittsburgh

Konstantinos Pelechrinis, Ph.D., Assistant Professor, University of Pittsburgh

Yi Qian, Ph.D., Associate Professor, University of Nebraska-Lincoln

Dissertation Director: David Tipper, Ph.D., Associate Professor, University of Pittsburgh

TIME DEPENDENT PERFORMANCE ANALYSIS OF WIRELESS NETWORKS

Kunjie Xu, PhD Candidate

University of Pittsburgh, 2014

Many wireless networks are subject to frequent changes in a combination of network topology, traffic demand, and link capacity, such that nonstationary/transient conditions always exist in packet-level network behavior. Although there are extensive studies on the steady-state performance of wireless networks, little work exists on the systematic study of their packet-level time varying behavior. However, it is increasingly noted that wireless networks must not only perform well in steady state, but must also have acceptable performance under nonstationary/transient conditions. Furthermore, numerous applications in today's wireless networks are very critical to the real-time performance of delay, packet delivery ratio, etc, such as safety applications in vehicular networks and military applications in mobile ad hoc networks. Thus, there exists a need for techniques to analyze the time dependent performance of wireless networks.

In this dissertation, we develop a performance modeling framework incorporating queuing and stochastic modeling techniques to efficiently evaluate packet-level time dependent performance of vehicular networks (single-hop) and mobile ad hoc networks (multi-hop). For vehicular networks, we consider the dynamic behavior of IEEE 802.11p MAC protocol due to node mobility and model the network hearability as a time varying adjacency matrix. For mobile ad hoc networks, we focus on the dynamic behavior of network layer performance due to rerouting and model the network connectivity as a time varying adjacency matrix. In both types of networks, node queues are modeled by the same fluid flow technique, which follows flow conservation principle to construct differential equations from a pointwise mapping

of the steady-state queueing relationships. Numerical results confirm that fluid-flow based performance models are able to respond to the ongoing nonstationary/transient conditions of wireless networks promptly and accurately. Moreover, compared to the computation time of standard discrete event simulator, fluid-flow based model is shown to be a more scalable evaluation tool. In general, our proposed performance model can be used to explore network design alternatives or to get a quick estimate on the performance variation in response to some dynamic changes in network conditions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Firstly and most importantly, I would like to express my greatest gratitude to my advisor Dr. David Tipper for his enlightening guidance and consistent support. His knowledgeable, wise and inspiring discussions have guided me throughout my whole Ph.D. career. It was such a pleasure to work with him for all these years. Facing so many obstacles, I am very fortunate that he has always been there to show me the right direction and influence me as an active thinker.

I am also truly grateful to my dissertation committee members: Dr. Prashant Krishnamurthy, for giving me advice in our research project and offering kind support since I came to Pitt; Dr. Yi Qian, for his deliberate revisions on my research papers as well as countless valuable suggestions throughout the entire course of my Ph.D. study; Dr. Konstantinos Pelechrinis and Dr. Vladimir Zadorozhny, for their great comments and suggestions to make this dissertation better.

My fellow students made my life at Pitt cheerful and memorable. I thank all the enjoyable discussion. Your friendship is my best fortune. Please accept my sincere gratitude for accompanying me through such a long journey. I thank you all.

This dissertation is dedicated to my parents whose love, sacrifice, and support have always been the greatest inspiration for me in my pursuit for betterment. My deepest acknowledgment goes to my sincere wife for her dedicated support and encouragement. This dissertation could not be completed without her presence beside me.

Finally, my dissertation research is supported by U.S. Army Research Office (ARO) grant W911NF-07-1-0318 and graduate student assistantships in the Telecommunications and Networking Program at Pitt. I am very grateful for these kindly supports.

1.0 INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

With the explosion of demand for wireless communication services, wireless networking has received significant attention over the past decades. Based on how packets are forwarded, wireless networks can be divided into two categories: single-hop and multi-hop wireless networks. In single-hop networks, packets are simply delivered to the destination via direct wireless connection. In multi-hop wireless networks, packets are forwarded by multiple wireless nodes. Since a wireless channel is inherently lossy and shared by mobile users, wireless networks exhibit time varying behavior by nature.

Vehicular network is one example of single-hop wireless networks. In recent years, there have been a dramatic growth in research and development in the dedicated short range communication (DSRC) applications in vehicular networks. DSRC employs the 5.9 GHz frequency band to support reliable and timely delivery of safety-related messages, which includes current status of a vehicle (e.g., location, speed and direction) as well as the event-driven emergency information. The dissemination of safety messages among vehicles is based on the single-hop broadcast service by each vehicle. IEEE 802.11p adopts enhanced distributed coordination function (EDCF) as the MAC protocol to broadcast safety messages with different QoS supports in vehicular networks [1]. For 802.11p EDCF, the CSMA/CA mechanism plays a central role in MAC layer functionality. Due to high-speed vehicle mobility, the number of exposed/hidden terminals of a target vehicle changes over time, so that non-stationary/transient performance behavior often exists and at times dominates in vehicular networks.

Unlike single-hop, multi-hop wireless networks use two or more wireless hops to convey

information from a source to a destination. Typical example includes wireless mesh networks (WMNs) [2], wireless sensor networks (WSNs) [3] and mobile ad-hoc networks (MANETs) [4]. Multi-hop 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 multi-hop wireless networks, the mobile 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), the network nodes can move arbitrarily and the network topology is expected to change often and 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], [9], [10]. Meanwhile, multi-hop wireless networks also inherit the traditional problems of wireless communications (e.g., broadcast communication channels, asymmetric channels and signal propagation, energy constraints in mobile nodes, links that are of poor quality in comparison to wired links, etc). Therefore, all of these problems combined with the unique dynamic topology feature make it challenging to develop and deploy multi-hop wireless networks.

Fundamental to the design of wireless networks is the ability to estimate and predict the network performance. Traditionally, the performance of wireless networks is evaluated using discrete event simulations. Popular network simulation tools used in wireless network studies include OPNET [11], NS-2 [12], NS-3 [13], Qualnet [14], and GloMoSim [15]. The basic simulation approach adopted in the majority of the literature [2], [16], [3], [4] 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 necessary with different random number seeds and the collected data has to be averaged over the runs. The observations gathered during the transient period in each run are usually eliminated to avoid initialization bias. In terms of simulation methodology, this approach is considered as steady-state simulation [17].

However, the nature of wireless networks suggests that transient conditions are likely to occur and possibly be common due to node mobility and nonstationary traffic. For single-hop wireless networks, a node may move in/out of the carrier sensing coverage of another node, causing the node pair to be exposed/hidden from each other. Consider IEEE 802.11 wireless

networks as an example, since each node accesses the channel essentially via CSMA/CA, its channel capacity depending on the number of exposed/hidden nodes changes over time. For multi-hop wireless networks, node mobility can frequently result in link failure and traffic rerouting. Since a significant factor in network performance after a link failure is the transient congestion period, one would expect that transient conditions can dominate the behavior of multi-hop wireless networks. Besides that, the on-demand traffic could be bursty and embedded with variable idle periods. As a result, studying time varying performance of wireless networks is important and meaningful.

Simulation study of the time varying performance of wireless networks is possible, though computationally difficult [18]. To study the nonstationary behavior of a network, the measurements of quantities observed over small intervals or at specific points in time are important. Hence, the time average used by steady-state simulation is not a proper approach, while ensemble averages are more appropriate and this simulation approach is named nonstationary simulation. The idea is to construct ensemble average curves of quantities of interest across a set of statistically identical but distinct independent simulation runs, along with the calculated confidence interval. With many such points collected at different time instants, the behavior of the system can be shown as a function of time. However, the principle difficulty in conducting simulation studies of this type is the large number of runs (typically thousands) that must be generated in order to get a representative ensemble from which a statistically accurate portrayal of the system behavior can be determined. Hence, very large amounts of CPU time are required for even small sized networks and this approach is quite difficult to scale. It is worth noting that parallel and distributed simulation techniques [19], [20], [21] have been applied to the development of wireless networks simulators by scheduling tasks and distributing the execution of those tasks to independent computing platforms that operate in parallel. However, the scalability of wireless network simulations even for steady state behavior is still a major problem by considering the trade-offs between execution time and fidelity.

Network performance evaluation can also be achieved by defining the model using analytical techniques. The standard analytical model used in network performance evaluation mainly deals with steady state conditions by using queuing and stochastic techniques. Since

such a great amount of work exists in the solution of steady state queuing or stochastic models and their application to wireless networks, an obvious issue is how steady state results can be used to model a network undergoing nonstationary conditions.

In summary, while significant progress has been made towards developing simulation tools [11], [12], [13], [14], [15] and models [22], [23], [24], [25], [26] to estimate the steady-state performance of wireless networks, relatively little work has appeared on the performance models to capture their time varying behavior. Finally, it is important to remark that our proposed analytical modeling framework is a viable option, if the goal is to design or tune the parameters of the wireless network under study. To evaluate the performance of a wireless network that cannot be measured, for example, during the design and development stages, it is necessary to use analytical models, which make predictions about network behavior and give quick results to eliminate inadequate and bad designs. Analytical models are always approximate but tractable, and the mathematical expressions allow us to gain insight into the interaction of different parameters. Since conducting the measurements is generally expensive or catastrophic, the analytical model has its advantage of flexibility and computation efficiency [27], [28], [29].

1.2 PROBLEM STATEMENT

Wireless networks consist of time-varying demand (source) and time-varying capacity (channels) in nature. The traffic load produced by sources is typically in time varying bursts. This is clearly evident for both real-time and non real-time traffic. For example, voice conversations have periods of silence and speech interspersed, clearly de-marking periods of high information followed by low-information segments. In fact, the traffic in many networks is bursty at multiple time-scales, evident from abundance of traffic models studied in wireline and wireless networks. For multi-hop wireless networks, the traffic could be rerouted dynamically due to node or link failure, so that the forwarded traffic to the next-hop node may vary over time. Analogous to time-varying traffic load is the well-known property of wireless channels, whose capacity is also nonstationary. The time-varying nature of the wire-

less channels can be attributed to the physical environment and node mobility. The radio signal is subject to large-scale and small-scale fading as well as multi-path effect, which will result in signal power fluctuations in both time and frequency domains. In addition, the bandwidth of the wireless medium shared by each node is strongly coupled with each other and varies over time due to node mobility.

Although there are extensive studies on packet-level steady-state performance of wireless network, little work exists on the systematic study of their time varying behavior. As a result of this dissertation work, we expect to answer the following questions: *What does the packet-level time varying performance of wireless networks look like? How can one construct an efficient analytical model to capture the time varying behavior?*

1.3 FRAMEWORK

In this dissertation, we aim to develop an integrated performance modeling framework to analyze both time varying and steady state behavior of wireless networks. As shown in Figure 1, there are three key components needed to construct such a model and evaluate the various network performance metrics.

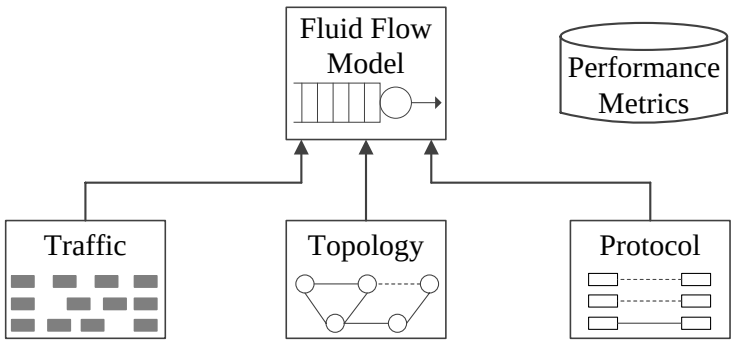


Figure 1: Performance analysis framework.

- *Traffic*: Data traffic is typically bursty in nature while some types of streaming traffic such as voice operates in an on-off manner with variable idle periods. Then, the nonstationary

traffic load results in the time varying behavior of each queue in the network.

- *Topology*: In wireless networks, the topology can change depending on link connectivity between transmitter and receiver. When nodes in a networks are allowed to move arbitrarily, it will lead to frequent changes in the connectivity of a queuing network.
- *Protocol*: A protocol is a set of rules that governs the communications between network nodes. Since nodes must understand and use the protocol to exchange information over the network, the service process of the queue at each node depends on the underlying protocol. According to the protocol, we can gain insights into the performance of the network.

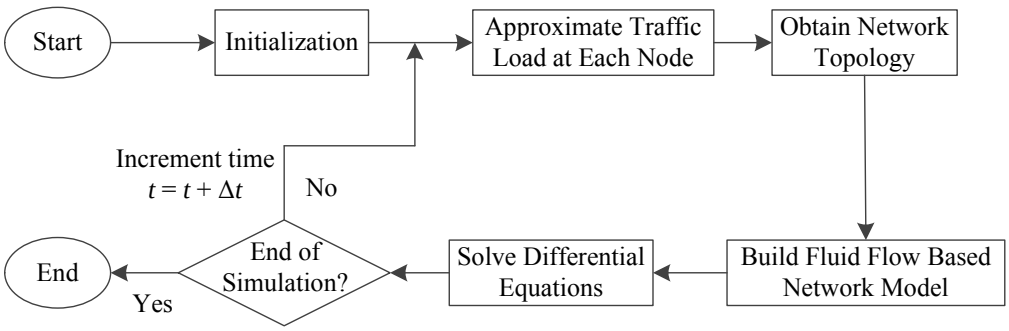


Figure 2: Performance analysis framework.

With the key components in the performance analysis already identified, Figure 2 depicts the flow chart about how all the pieces can work together to obtain the time dependent performance metrics of interest. We start by identifying initial network conditions. Using the pointwise stationary approximation method [30], the nonstationary arrival rate is estimated to be constant over a small time interval Δt. The information about network topology is used to determine the hearability/connectivity between any two nodes and how the traffic should be routed towards the destination. Then, the fluid flow based network model consisting of a set of differential equations is established by incorporating the underlying protocols with traffic and topology. After that, we apply the Runge-Kutta algorithm to solve the differential equations at the end of the time interval $t + \Delta t$, which then becomes the initial condition for the next time step $[t + \Delta t, t + 2\Delta t]$. If necessary, the traffic load, the topology, the link

capacity and all other network parameters will be adjusted to the new time step. Then, the solution to the fluid flow model is calculated based on the updated parameters, and the procedure will be repeated until the end of the simulation process. Please note that any standard numerical integration method can be used to solve such mathematical problem numerically; however, the Runge-Kutta algorithm is one of the widely used methods.

1.4 CONTRIBUTIONS

The contribution of this dissertation work is summarized as follows.

Chapter 3 (Vehicular Networks)

- We propose a fluid flow based model of vehicular networks to describe the dynamic queuing behavior for both routine traffic and emergency traffic. Based on this approach, we also develop the time varying model to evaluate the real-time packet delay and packet delivery ratio, which are the most critical performance metrics for safety applications in vehicular networks.
- We study the nonstationarity of vehicular network performance impacted by the traffic load, vehicle velocity as well as vehicle density in a two-way highway scenario. We point out that the transient period could dominate the network behavior in cases of imbalanced vehicle density, high-speed mobility and heavy traffic load.

Chapter 4 (Multi-hop Wireless Networks)

- We propose fluid flow based queuing models for a single node with Poisson, CBR and general traffic loads, respectively. For the case of CBR traffic load, we 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.
- We develop a novel time varying performance model for multihop wireless networks with Poisson, CBR and general traffic loads, on the basis of single-node fluid flow model. An adjacency matrix, representing topology change, 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,

which is shown to be a more scalable than an equivalent discrete event simulator.

- We carefully evaluate the performance of a sample network impacted by node mobility, traffic load and wireless link quality by our model. It shows that the fluid flow model can respond to the ongoing nonstationary conditions properly. In addition, we study the network performance with a series of parametric configurations, which result in the same steady steady results but distinct time varying behavior. We then propose a measure of "instantaneous variation" to quantify the nonstationarity of network behavior.

In general, our proposed time varying performance model can be used to explore wireless network design alternatives or to get a quick estimate on the performance variation in response to some dynamic changes in network conditions.

1.5 DISSERTATION OUTLINE

The remainder of this dissertation is organized as follows: Chapter 2 provides a literature review of network performance evaluation methods followed by the related queuing and stochastic modeling survey. Chapter 3 provides the details of the time varying performance model for IEEE 802.11p vehicular networks. Following the same modeling principle, we introduce our model to evaluate the dynamic behavior of multihop wireless networks in Chapter 4. Chapter 5 presents conclusions and future work.

2.0 LITERATURE REVIEW

The performance of wireless networks is generally evaluated using measurement, simulation or analytical model. Measurement involves running experiments on an existing system or prototype and gathering data on metrics of interest. Simulation is a process of conducting experiments on a computer model of a system over a range of scenarios and parameter values. The third one, analytical model is a description of a system using mathematical concepts and analysis techniques. Quantitative characterizations of the traffic, topology and protocol are crucial in the creation and validation of analytical models of wireless networks. Beside steady-state measures, the time dependent performance modeling techniques are also necessary to capture the dynamics of the network. In the following, we will discuss the above issues based on an extensive literature review.

2.1 MEASUREMENT-BASED PERFORMANCE EVALUATION

Measurement studies on wireless network involve running experiments on a real network [31], [32] or a prototype testbed [33], [34]. The key strength of measurement is to provide accurate performance results including details of all network components. Hence, measurement-based network performance evaluation studies can reveal some useful discoveries that might be hidden in simulation or analytical-based methods.

The measurement results in [35], [36] are used to challenge six common assumptions in wireless simulation studies, namely: (1) the world is two dimensional, (2) a radio's transmission area is circular, (3) all radios have equal range, (4) channels are symmetric (i.e., if node A can hear B, then B can hear A), (5) perfect transmission channels (i.e., if node A can hear

node B then A can hear B perfectly) and (6) signal strength is a simple function of distance. To study these axioms and their impact on simulation studies, they use data collected from a large MANET experiment in which 33 laptops with WiFi network cards roamed a field for over an hour while exchanging broadcast beacons and operating different ad hoc routing protocols. The measurement results demonstrate the weakness of these assumptions, and show how these assumptions cause simulation results to differ from the reality. However, in the case of the best-effort model, the simulation produced good results that were reasonably matched to their particular outdoor experiment scenario. It was suggested that the axiomatic assumptions are undoubtedly invalid in many situations; however, it is possible for the realistic stochastic model that carefully describes the chosen target environment to return acceptable results in the context of those conditions and assumptions. In [37], Tala et al. proposed an experimental methodology for empirical studies in wireless networks, in order to detect measurement problems early, increase reliability and obtain reproducible results. The stages of the proposed methodology include: experimental design, description of scenarios, sanity check, validation test, multiple runs and capture, traces processing, analysis, packing and storage, and documentation and reports.

For single-hop wireless network, such as 802.11 network, Bianchi et al. [38] shows in an experimental assessment of six widespread commercial 802.11 cards that the commercially available wireless cards often do not comply with the IEEE 802.11 standards. These six PCMCIA commercial cards include ASUS WL-107g (Ralink RT2500 chipset), Intel Centrino (2200BG chipset), Digicom Palladio (Realtek RTL8180 chipset), Dlink DWL-650 (Intersil PRISM II chipset), Dlink DWL-G650 Air-Plus (Atheros chipset), and Linksys WPC54G (Broadcom chipset). The measurement study shows that neither one performs exactly as expected in terms of backoff operation, and they experience different performance either when accessing the channel alone as well as when competing against each other. In some cases, implementation issues seem to affect the proper card operation. In other cases, manufacturers rely on backoff parameters different from the standard specification, this perhaps being done on purpose to provide an indeed unfair advantage of these cards with respect to the competitors. Therefore, non-standard behavior at commercial wireless cards makes it very difficult to set up the experiments for standard-based model validation.

An example of multihop wireless network measurement testbed is at the Quail Ridge Nature Reserve, called QuRiNet [39], [40]. The network is originally used for environmental research study of the flora and fauna in the region by the Department of Ecology at UC Davis. This QuRiNet claims to be different from other measurement-based simulation in terms of its location and its usage, so that the validity of theoretical ideas under practical situation can be tested, when complicated issues such as hilly terrain, forest growth, or long distance are involved. Preliminary results on the network utilization, round trip time performance, signal strength and network capacity were provided. Due to geographical factors, they found out that the same setting from indoor laboratory testbed did not work on QuRiNet. In order to achieve the expected performance, several parameters in QuRiNet such as the power level of wireless cards, the antenna type, or the placement of antennas should be adjusted accordingly. One observation of the project was asymmetric signal strengths and throughput values between the same two nodes. Even on the same link, there are large variations on both signal strength and throughput at different points of time during a day. It is however important to point out that many communication problems observed from measurement testbed cannot be easily modeled by commonly used simulation tools.

The general criticism of measurement studies is 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 the environments can be very different, especially for wireless networks. Constructing a wireless network testbed for a given scenario remains limited in terms of the experimental scenarios that can be studied. For these reasons, protocol scalability, sensitive to user mobility patterns and speeds are difficult to investigate on a real testbed.

2.2 SIMULATION-BASED PERFORMANCE EVALUATION

In contrast to measurement based studies, simulation models do not require a testbed, since one models the system on the computer and experiments with a computer model. Simulation permits the study of system behavior over a range network scenarios and parameter

values and allows the modeler control of the time scale. Popular network simulation tools used in wireless network studies include OPNET [11], NS-2 [12], NS-3 [13], Qualnet [14], and GloMoSim [15]. They all provide advanced simulation environments to test and debug different networking protocols.

The use of simulation techniques in the performance evaluation of communication networks is a well studied research area [17]. As noted by Pawlikowski and his team in [41], the statistical nature of simulation results is often ignored in communication network studies calling into question the credibility of the conclusions in much of the literature. Specifically, as noted in [41] recommended simulation methodology (i.e., independent multiple runs, deletion of the initial transient period, confidence intervals on results, avoiding repetition of the random numbers generated, etc.) is not followed in the majority of simulation studies reported in the literature. Note that, even with the existence of many simulation tools, a recent analysis of wireless network research literature where simulation was used as the main analysis technique showed that the proper simulation methodology was rarely followed [42], thus as noted above calling into question the credibility of the results. Most recently, Sarkar and Gutiérrez revisited the credibility issue of simulation study in telecommunication networks in [43] based on a comprehensive survey of IEEE publications. They found that a significant amount of authors did not provide enough information about the confidence interval (CI) and confidence level (CL) or relative statistical error (SE) of the results presented in the survey paper, some of them even did not mention the name of simulation tools used. Thus, there is no significant change since [41] with respect to quality and credibility of the simulation studies revised and the deep crisis of credibility still remains. Furthermore, some of the literature questions the validity/fidelity of simulations based on simulation tools. In particular, [44] presents the simulation results of the simple flooding algorithm using OPNET, NS-2, and GloMoSim. Important divergences between the simulator results were measured. The observed differences are not only quantitative (not the same numerical value), but also qualitative (not the same general behavior) making some past observations of many wireless network simulation studies based on these tools an open issue.

Related to simulation models in recent research studies, excellent discussions of the modeling issues in building more accurate models are given in [45], [46]. The major issues in the

design and modeling of wireless networks include the geographic space in which the mobiles move, geographic boundary policy, number of nodes and their position distribution, signal propagation models, signal interference models, mobility models, network protocols, and the traffic workload characteristics. In order to construct a model that matches the goal of the simulation project, all issues should be carefully considered to sufficiently include relevant elements that capture the major effects and prevent any unnecessary detail that will introduce overheads and a delay in the simulation process. Since mobility plays a central role in the wireless network model design, much of the effort in the literature has focused on the mobility and boundary policy. Typical simulation studies assume a fixed number of active mobile nodes implemented in a closed two-dimensional limited area of interest with some boundary policies defined for users arriving at the edge of the simulated area.

Packet-level simulation of computer networks becomes prohibitively expensive as link speeds, workloads, and network size increase. In addition, the performance measures determined from a simulation are random in exactly the same manners as measurements; hence confidence intervals must be used when discussing performance measures estimated by simulations. To assure independently and identically distributed data generated in simulation studies and increase credibility of statistical simulation results [41], large amount of computer run time for multiple independent runs with confidence intervals on results are required. In [47], Vasan et al. proposed time-stepped stochastic simulation by generating a sample path of the system state at discrete time steps rather than at each packet transmission. This method can achieve the modeling accuracy of packet-level simulation in a fraction of the computational cost. Kim and Hou in [48] develop a time-efficient 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 addition, parallel simulation and distributed simulation approaches exploit the idea of concurrency among events to achieve a reduction in execution time [20]. Researchers have also tried to apply the concept of parallel and distributed simulation techniques to the development of a more realistic simulation tool for nonstationary network behavior, however the scalability is still believed an open issue.

In summary, the drawbacks of simulation are the accuracy of simulation model and the time and effort involved in considering all cases/parameter values and analyzing the corresponding output data. Similar to measurement, it is difficult to generalize the results and evaluate its sensitivity by simulation studies. Furthermore, the accuracy and fidelity of simulation models might be questionable, such as the radio propagation model in wireless networks. Another concern is the time to develop and execute the discrete event simulation.

2.3 ANALYTICAL-BASED PERFORMANCE EVALUATION

Network performance evaluation can also be achieved by defining a system model and solving the model using analytical techniques. Analytical models are usually computationally inexpensive, and expressions can be obtained in a fast manner. Analytical modeling is a viable option if the goal is to explore design alternatives, and it is sufficient to have approximate estimates of the expected behavior and performance [46], [22]. The standard analytical network model used in network performance evaluation mainly deals with the steady state conditions of queuing theory models [49]. Performance metrics must be carefully defined to evaluate and understand the critical features of the considered system. A good performance metric should have the following characteristics: (a) the performance metric should allow an unambiguous comparison to be made between systems, (b) it should be possible to develop models to estimate the metric, (c) it should be relevant or meaningful, and (d) the model used to estimate the metric should not be difficult to estimate [50].

It is worth noting that the fidelity of an analytical model depends on the quality of input data and on the modeling accuracy of the involved mechanisms. In the area of wireless network performance analysis, there are three main components need to be carefully studied, those are *traffic*, *topology* and *protocol*. From queuing theory perspective, traffic load defines the arrival process of each queue, while topology and protocol determine the service process in single-hop networks and also affect the next stage arrival process in multihop networks. From the view of OSI seven-layer stack model, traffic comes from the application layer and topology defines the network connectivity in physical layer. The protocol actually defines the

implementation of data communication at each layer and the interaction between different layers. In the following, we will give a comprehensive discussion on how to model each of them using statistical tools and characterize the individual effect on network performance.

2.3.1 Traffic Load

Regardless of how good the analytical model may be, it cannot give accurate results if the input data are inaccurate or not representative of the traffic load in the real world. Moreover, the essential features of traffic loads can be helpful in providing early predication about the design. One of the most intriguing aspects of the traffic modeling is its multi-level temporal nature. It is widely accepted that a model which characterizes traffic at the session level, connection level and packet level adequately captures the basic behavior of a real-time traffic source. The session level describes the user session between association and disassociation with the network infrastructure. Within a user session, the connection level is a traffic flow connection (TCP/UDP) from one IP address to another IP address. Packet level accounts for the correlation among successive arrival of packets, which is a finer level of modeling.

2.3.1.1 Session-Level Characterization The session level captures the interaction between users and the network. It is widely recognized that the arrival of user session can be modeled as a time varying Poisson process [51], [52], [53]. The preliminary data analysis of the real-word measurement in [52] revealed significant variation in the number of connection arrivals across different venues. To capture such variation, for instance, within the coffee shop venues, the authors observe the number of customer arrival over 96 time slots of a day in 60 different venues. Then they average the observed slot arrival counts over the last five weekdays and over all the venues to get an average intra-day profile. The empirically obtained result is very close the equivalent non-stationary Poisson model. In addition, the session duration, defined as the time between which a device associates and disassociates with the network, is also widely studied [51], [52], [53]. The measurement study in a public Wi-Fi network in [52] shows that 78% of the sessions are of duration at most 10 minutes, and a very small fraction of connections maintain for above 6 hours. Thus, the connection

durations are heavy tailed and can be expressed using a power law distribution.

2.3.1.2 Connection-Level Characterization The connection level characterizes the user behavior with the time series representing TCP/UDP connections established by a user. As such, the burstiness of the traffic generated by individual user can be modeled as an on/off process, where the duration of a connection corresponds to an on period and the idle interval between two consecutive connections corresponds to the off time. Naturally, this on/off process can be characterized by two traffic metrics, i.e. the connection duration and the inter-arrival time between two consecutive connections established by the same user. A general agreement among various measurement studies shows that both inter-arrival of connection and the connection duration follow the heavy tailed distribution [51], [52], [53], [54]. In [54], authors select 100 users and perform a study on the per-user behavior. They find that the user behavior among different users is independent when it is characterized by the use of time series. In addition, the probability distribution of the inter-arrival time of connections generated by a user is constructed based on the 120-day trace in campus-wide WLANs. The complementary cumulative distribution function (CCDP) of measurement data is then compared with the result generated from Pareto distribution.

2.3.1.3 Packet-level Characterization Packet level model characterizes the time between successive arrivals of packets generated by source. Depending on the type of traffic, different traffic models have been suggested, such as Poisson, constant bit rate, on-off traffic, and etc. [55].

The Poisson process has been shown to be a plausible representation of a number of physical phenomena, e.g. the occurrence of telephone calls and the arrival of customers at a service facility. Poisson process is also used as a simplified traditional traffic model for circuit-switched data and the general packet data [56], [55]. The Poisson process is actually a mathematical model of a completely random arrival pattern. Poisson traffic is characterized by assuming that the packet arrivals are independent and the inter-arrival time is exponentially distributed. Another reason for its popularity is the analytical tractability that this process provides. One important property is that the superposition of independent

Poisson processes is a Poisson process. There are several ways to verify whether a particular arrival process is Poisson [57]. An easy visual way consists in plotting the histogram of the inter-arrival times and verifying whether it is an exponentially decreasing function. One special case of the Poisson model is represented by time-dependent Poisson processes. This representation is suitable for situations where the mean rate varies over time.

The constant bit rate (CBR) traffic source sends fixed size packet at constant intervals. CBR is tailored for connections where the end-systems require predictable response time and a static amount of bandwidth continuously available for the lifetime of the connection [58]. Many of the real-time services by CBR traffic require an inherent reliance on time synchronization between the traffic source and destination [59]. In the case of streaming video as a CBR, the source could be under the CBR data rate target. So in order to complete the stream, it's necessary to add stuffing packets in the stream to reach the data rate wanted [55].

For on-off traffic model, the source generates packet during on time and keeps silent in off state. This model exhibits the burstiness nature of traffic in telecommunications [55], [60]. For example, in the traditional voice or VoIP applications, a user is not always talking and in fact there is a considerable length of "silence" between "bursts" or "talk-spurts" [61]. If we allow the inter-arrival time between packets to be a general distribution during on period, we then have an Interrupted Renewal Process (IRP). IRP degenerates into an Interrupted Poisson Process (IPP) when the general distribution becomes exponential. When the distribution is deterministic with a constant value, the IRP becomes an Interrupted Deterministic Process (IDP).

More sophisticated arrival process models such as the Markov arrival process (MAP), Markov modulated Poisson process (MMPP) and batch Markovian arrival process (BMAP) can capture the burstiness and correlation of network traffic at the packet level [62], [55]. These models are generally non-renewal process, but the numerical techniques such as the matrix geometric method are still available to solve them. However, the computational complexity required for solving queues characterized by a larger set of parameters often becomes prohibitive.

2.3.2 Network Topology

The network topology depends on the position of all nodes and the link connectivity between any two nodes. Given an initial placement of the mobile nodes in the network, their time varying positions are determined by node mobility trajectories. In addition, the wireless link connectivity between two nodes is a function of a variety of factors, such as the distance between the nodes, antenna radiation pattern, power level, geographic terrain, propagation environment, interference, receiver sensitivity, etc. In the following, we will discuss the wireless link connectivity and node mobility separately.

2.3.2.1 Wireless Link Connectivity It is well known in the wireless communication literature that radio signal propagation is subject to path-loss, multipath fading and shadowing [63]. The path loss is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through the medium, since the waves propagate outwards from the transmitter in an expanding sphere. Multipath fading comes from the destructive inference of the radio signals reaching the receiver by two or more paths. Shadowing occurs when large objects block paths of propagation. Both path loss and shadowing are considered as large-scale fading, which impacts the range of communications between nodes and the interference levels. Multipath fading is regarded as small-scale fading that impact the rates for transmission, bit error rates and packet delivery rates. In addition, the quality of the wireless link is also affected by the physical layer interference signal in the same frequency band as well as the MAC layer packet collision under the contention based access scheme.

2.3.2.2 Node Mobility In a wireless network with mobile nodes, nodes' movement speed, direction, and acceleration/deacceleration, can have a significant effect on the network design to support mobility. Unfortunately, movement in the physical world is often unrepeatable. Live use of a mobile system can provide meaningful insight, but cannot form the sole basis of experimental evaluation. Instead, the community has turned to simulating the movement of nodes and users. Of course, one must derive a model of movement to drive such a simulation. Typically, simulation studies assume a closed system, where the num-

ber of users moving inside a two dimensional simulated area (e.g., rectangular) is constant. Rules are defined for users arriving at the edges of the area. The random waypoint mobility model is the model most commonly used to define the way users' movement [44-46]. The random waypoint mobility model has become the most popular mobility models in the mobility research of single-hop and multi-hop wireless networks, because of its simplicity and wide availability [64], [65], [66].

In single-hop wireless networks, node mobility can affect the network performance in three ways. Since the mobile nodes could move in/out of the coverage area of each other, the dynamic number of associated nodes will impact the bandwidth shared by each user [67]. Moreover, if the IEEE 802.11 CSMA/CA-based MAC protocol is implemented in the network, node mobility can aggravate the detrimental effect of hidden nodes on the network performance. As the nodes move around, the network hearing graph varies over time and the current collision avoidance could become ineffective in the next second [68], [69].

In multihop wireless networks, since the mobile nodes must cooperate to dynamically establish routes using wireless links, traffic has to be rerouted if any link on the path is broken. Hence, the network performance heavily depends on the underlying network topology. The node mobility models used in multihop wireless network studies include the random walk mobility model, the random waypoint mobility model, the random direction mobility model, the reference point group mobility model, and others. [64]. Among these, the random waypoint mobility model is the "benchmark" mobility model to evaluate network performance. The shortcomings of simulating the random waypoint mobility model are the long warm up period and heavy computation required to compute node position and determine link connections between each node pair. A computationally simpler approach is to model mobility by directly manipulating the elements of the adjacency matrix according to a probabilistic model [70], [71]. The link connectivity model has fewer parameters, thus simplifying the design of experiments. To be specific, only the average link lifetime parameters, namely T_{up} and T_{down}, are sufficient to fully characterize the model. Moreover, the idea of using only these two variables seems to provide a better control in link stability characteristics than typical mobility models. The state of each link can be aggregated in an adjacency matrix to represent the topology of the network. The elements in the matrix are two-state random

variables that specify the status of the link which can either be symmetric bidirectional links or asymmetric ones. Though the link connectivity model based on a two-state Markov chain addresses the computation burden of the random waypoint model, it comes with one drawback that it does not support high-fidelity models of mobility and physical layer characteristics. However, it is believed that the practical effect of computational time reduction outweighs the potential loss of fidelity when compared to actual node mobility models.

2.3.3 MAC Protocol

Much work has been done to analyze different layers of the protocol stack on the performance of wireless networks. Since the wireless medium is inherently a shared resource, controlling channel access becomes a central theme that determines the fundamental capacity of the wireless network and has a dramatic impact on system complexity and cost. In this dissertation work, we focus on the performance study of MAC layer protocol in wireless networks. Wireless MAC protocols can be broadly categorized as contention-based and contention-free, depending on the channel access mechanisms.

2.3.3.1 Contention-based MAC The first contention based MAC protocol is ALOHA [72]. A node is permitted to transmit packet any time. To confirm the successful transmission, the intended receiver then sends an acknowledgement within a certain time-out period. Otherwise, the transmitter assumes a collision and has to retransmit the packet. Before retransmission, the transmitter waits for a random period of time to avoid continuously repeated conflicts. After ALOHA, "Slotted ALOHA" improves channel efficiency by slotting time into equal length pieces. Each node is synchronized and transmits only at the beginning of a slot [73]. Analytical results from [74], [75] show that the effective channel capacity can be increased by slotting channel access into time segments. Another approach, Carrier Sensing Multiple-Access (CSMA) is superior to both slotted and pure ALOHA [76]. In order to avoid collisions, a CSMA node senses the carrier to detect any transmission on the channel. Given a slotted implementation, if the channel is busy in a slot, the node would not transmit. If the channel becomes idle, the node schedules the transmission. CSMA-based MAC protocols

have two well known problems: hidden and exposed terminals. Tobagi and Kleinrock proposed a solution to eliminate hidden terminals using a busy-tone sent by a receiving node in a separate band [77]. To maintain in-band carrier sensing, Karn [78] proposed "virtual carrier sensing" in his Multiple Access with Collision Avoidance (MACA) protocol that employs Ready-to-Send (RTS) and Clear-to-Send (CTS) to reserve the channel and silence potential hidden terminals before the DATA-ACK exchange. In addition, a binary exponential backoff scheme is used in MACA. Recently, Cai et al. [79] incorporated reactive jamming scheme with the IEEE 802.11 DCF to improve network throughput of WLAN in the presence of hidden terminals.

The basic access mode of the IEEE 802.11 DCF [80] is essentially "physical carrier sensing" coupled with the binary exponential backoff mechanism for collision avoidance. The protocol employs a fixed carrier sensing (CS) threshold to determine whether the medium is free for transmission. There are two packet transmission schemes employed by DCF, namely, the basic access and the optional RTS/CTS access mechanism. For multihop wireless networks, the authors in [26] examine the interactions of the 802.11 DCF protocol and ad hoc forwarding. They found that the theoretical maximum capacity per node in a large random network with random traffic scales as $O(1/\sqrt{n})$, where n is the total number of nodes. In addition, the authors conclude that 802.11 DCF is more efficient for scheduling local traffic. In this work, we will focus on the performance analysis of 802.11 DCF in single-hop networks.

The performance of 802.11 MAC has been studied extensively. The seminal paper of Bianchi [22] is the first work to model the binary exponential backoff mechanism of the DCF protocol in saturation conditions by a two dimensional Discrete Time Markov Chain. To improve the accuracy of Bianchi's Markov chain model, Chatzimisios [81] and Wu [82] take into account the retransmission limit where a packet is dropped after reaching a maximum number of retransmissions. Zhang [83] and Xiao [23] include the backoff counter freezing probability into their models. Q.Ni et al [84] considered the saturation throughput for both congested and error-prone Gaussian channels. Parallel to the Markovian modeling efforts, other analytical models predict the saturation performance of 802.11 without fully describing the detailed behavior of the binary exponential backoff [85] and [86]. Because non-saturation conditions are more typical in WLAN, analysis of IEEE 802.11 DCF in non-saturated case has

also attracted remarkable attention. Zhai, etc. [87] and Xu, etc. [88] scale the transmission probability of a saturation model by the probability of having at least one packet in the queue (i.e., the queue utilization), and model the MAC layer system as $M/M/1/K$ and $M/G/1/K$ queue. The novelty of their approach is to derive the probability distribution of the MAC layer service time by using the Z-transform of the state transition diagram. In addition, some Markovian models like Engelstad [89] and Huang [90] use additional states to model the behavior of a tagged node when it does not have any packets to send. Tickoo [24] developed a non-Markovian $G/G/1$ queuing model by considering arbitrary traffic arrival patterns and packet size distribution. Garetto and Chiasserini [91] study the number of contending nodes using a multidimensional Markov chain which includes, in addition to the backoff process, the number of packets in the queue as well as the number of nodes in the network. Developing an analytical model is desirable not only because of the wide deployment of 802.11 devices, but also because the CSMA/CA mechanism continues to play a core role in new standard in 802.11 family, such as 802.11p.

2.3.3.2 Contention-free MAC In contention-free protocols, the nodes are following some particular schedule which guarantees collision-free transmission times. Generally, contention-free protocols are capable of providing QoS guarantees by Frequency Division Multiple Access (FDMA) [92], Time Division Multiple Access (TDMA) [92], Code Division Multiple Access (CDMA) [93] or various reservation-based schemes.

For single hop wireless network, the access point has centralized control on who and when they access the medium. The downlink transmissions from base station can be heard by all the wireless nodes in the network. For example, the point coordination function (PCF) defined in IEEE 802.11 [80] use centralized polling techniques to achieve contention-free transmission of each node. Coutras et [94] and Sikdar [95] analyze the performance of PCF in support of multimedia services.

For multihop wireless network, it is challenging to implement distributed channel access control in the presence of unpredictable channel conditions and node movements. TDMA-based MAC [96], [97] lies in the usage of time slots in a time frame structure. Since each frame has to start exactly at the same time at each node, network-wide synchronization in-

curs extra overhead and it is difficult to achieve mobility. For CDMA-based MAC [98], how to assign spreading codes in a distributed multihop wireless networks is a major challenge. In addition, Power controlled MAC protocols, such as GAF [99], SPAN [100] and ASCENT [101], have been considered by in settings that are based on collision avoidance and transmission scheduling. Since all the above contention-free protocols rely on deterministically quantified resource availability information and resource reservation, their one-hop packet transmission time could be roughly approximated as deterministic.

2.3.4 Steady-state Performance Modeling for Wireless Networks

2.3.4.1 IEEE 802.11p Vehicular Networks Several recent studies have proposed a variety of models to study the the steady-state performance of IEEE 802.11p based vehicular networks. In [102], Chen et al. constructed a continuous-time Markov chain (CTMC) to model the operation of the IEEE 802.11p MAC backoff counter for broadcast. This model was then extended to consider enhanced distributed channel access (EDCA) mechanism including both emergency and routine services with different priorities in [103]. In [104], Hassan et al. analyzed the MAC service process by decoupling the resulting packet collision probability into the exposed-terminal and hidden-terminal cases. Then, [105] developed a semi-Markov process (SMP) to approximate the service time of broadcast packet and used fixed point iteration to solve the SMP model. In summary, while extensive work has been made towards developing analytical models to estimate the steady state performance of vehicular networks, little work has appeared on the performance model to capture their nonstationary/transient behavior.

2.3.4.2 Multihop Wireless Networks The steady-state performance of multihop wireless network has been studied extensively via various models. In [106], Haan modeled a multihop wireless network as a polling system, where the server offered service to a sequence of queues one-by-one. However, this model cannot exactly describe the stochastic process of MAC protocol in the networks. To consider the actual packet service process in wireless networks, Vassis et al. in [107] evaluated the steady-state performance of IEEE 802.11 ad

hoc networks, which consist of only exposed terminals with finite traffic load. This paper adopts the CSMA/CA modeling technique in [108] to derive channel utilization and media access delay at each hop. The major limitation of this work is to assume that all nodes in the network are hearable to each other; however, such an assumption may not be valid in practice. In [109], Jin et al. developed an analytical model to evaluate the performance of 802.11 ad hoc networks in imperfect wireless channels due to fading channel and probabilistic carrier sensing capability. In this work, authors assume a saturated MAC queue and restrict the analysis within a fixed homogenous topology where all senders have the same probability to transmit to avoid further complications in the model. A more general performance analysis is provided in [110] to estimate the throughput of 802.11 ad hoc networks, with the considerations of hidden terminals, imperfect carrier sensing as well as the unsaturated traffic loads. The experimental measurements validates the accuracy of the analytical model. However, the end-to-end packet delay in multi-hop networks has not been addressed and only the performance of network throughput is analyzed in the work.

With the aid of queuing model, Ray et al. [111] derived analytical expression for end-to-end packet delay following IEEE 802.11 standard and approximate each wireless node as an $M/D/1$ queue. They then extended the analysis to model a static ad hoc network with a linear topology. Medepalli and Tobagi derived the average service time needed for a packet at each node by modeling the transmission queue of IEEE 802.11 MAC to be $M/M/1$ in [112]. They showed that an $M/M/1$ approximation compared favorably with a more detailed model while significantly simplifying the analysis. In [25], a more general $M/G/1$ queuing model was developed to obtain the single-hop media access delay distribution, which is then used to study the end-to-end delay in the static multihop scenario. Bisnik and Abouzeid proposed an open $G/G/1$ queuing network in [113] to model two-dimensional static ad hoc networks and the diffusion approximation technique is used to obtain the steady-state performance. In summary, while extensive work has been made towards developing analytical models to estimate the steady state performance of ad hoc wireless networks, little work has appeared on performance models of *mobile* ad hoc networks to capture their time varying queuing behavior.

2.4 TIME DEPENDENT BEHAVIOR MODELING TECHNIQUES

In comparison to the general literature on the steady state models of communication networks using queuing theory, there is relatively little work on modeling their transient or nonstationary behavior. Note that a distinction is made between *transient* and *nonstationary* behavior since transient behavior describes the system transition from one stationary state to another, whereas nonstatinary behavior occurs when the arrival and/or service rate vary continuously with time. In this dissertation work, we use the term *time dependent* to cover both of them.

2.4.1 General Techniques

In order to apply steady state queuing results to model a system under transient or nonstationary conditions, there are four different approximation techniques in the context of a simple Markovian queuing model with time varying arrivals, namely the simple stationary approximation (SSA), the peak approximation (PA), the quasi-static approximation (QSA) and the pointwise stationary approximation (PSA).

The simple stationary approximation (SSA) is to replace the time varying load by its mean value and then use the steady-state queuing results. The peak approximation (PA) is similar, but peak values for the load are used in the steady-state analysis. The quasi-static approximation (QSA) approach approximates time varying behavior by approximating the nonstationary load over a set of small time intervals by a constant during each time interval (usually the mean value over the time interval) and using steady-state results for each time interval. The pointwise stationary approximation (PSA) is similar except that the load is sampled at various time points and the steady-state value is used at each time point. In [114], it was shown that for even modest amounts of nonstationarity (i.e., deviations from the average load of 10%) the Mean Value and Peak Value approach can lead to large errors. The error of the QSA and the PSA have not been thoroughly studied, but results in [115] show that the accuracy will depend upon the rapidity and magnitude of changes in the load and is quite poor in many cases. In wireless networks, we note that both the arrival and service

rates of queues could vary rapidly as the load and topology changes, such that the QSA and PSA methods may not be sufficiently accurate and transient/nonstationary modeling techniques are desired.

2.4.2 Time Dependent Queuing System

The transient/nonstationary behavior of the queuing system is usually studied by solving differential/difference equation of a Markov process/chain model. The differential/difference equation is normally a function of $p^i(t)$, which is the time varying probability of i customers in the queuing system. However, only in some special simple cases are the transform expressions invertible to yield a closed form expression and even then the result is usually computationally complex to evaluate. In general, one must try to numerically invert the transform expression and this is known to be a difficult numerical analysis problem. There has been an effort to determine the transient behavior numerically rather than by deriving a closed form expression. Numerical approaches have largely focused on two methods: uniformization [116] and numerical analysis techniques (e.g., numerical integration of the underlying differential equation model) [117]. Note that these techniques provide accurate results and enable multiple types of performance measures (relaxation times, mean behavior and distributional behavior) to be studied. The primary disadvantage of both methods is that the computational complexity grows with the queue state space and one is limited to considering Markovian type systems.

In order to study the transient/nonstationary behavior of general queueing systems in an efficient manner, several approximate analysis methods have been proposed such as diffusion models [118], fluid flow models [30], and service time convolution [119]. The accuracy of these approximations along with the performance measures that can be determined vary according to the system under study. Note that very little work has been done on comparing the accuracy and computational complexity of these approximations.

2.4.3 Fluid Flow Background

We note that the fluid flow concept has been proposed for constructing computationally efficient simulation models for both wired [120], [121] and wireless networks [48], [122], [123]. On the wired network side, the basic idea of [120] is to model a few network nodes (e.g., a source-destination pair) in detail with packet-based discrete event simulation and enlarge part of an IP network by fluid flow models which interface with the discrete event simulation. This approach was shown to be accurate at the IP level and scalable. In [121], the authors use fluid flow techniques in combination with packet-level 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 computation speedups. On the wireless network side, Kim and Hou in [48] develop a fluid flow based simulator for a 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 [122], a fluid flow model is presented to analyze the performance of backlog-based CSMA policies in the wireless network environment with multiple arrival rates. Most recently, the fluid flow approximation is applied in [123] to model a TCP connection with time division multiplexing and scheduling in WiMAX wireless networks.

In this dissertation, a fluid flow based modeling approach is developed to efficiently approximate the mean transient/nonstationary packet-level behavior of wireless networks. The purpose of this work is to develop a technique that can be used for 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, this dissertation work focuses on determining the mean transient/nonstationary behavior of networks. With the concept of pointwise stationary fluid flow approximation (PSFFA) [30] and [30], this fluid flow modeling approach has bee used to approximate the mean transient/nonstationary behavior of a variety of queuing systems in a series of papers [115], [124], [125], [126], [127], [128]. 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 detailed description of PSFFA is given in Section 3.3.1. The use of the PSFFA to determine the transient/nonstationary behavior of finite and infinite capacity queuing systems with general arrival and service distributions is discussed in [115]. In fact, it is even possible to develop the fluid flow model from measurement data. The PSFFA is quite general in nature and shown to be reasonably accurate for the cases considered and a considerable improvement over the Pointwise Stationary Approximation (PSA) [115]. Furthermore, PSFFA models can be coupled using flow conservation principles to study a wide range of real queuing systems or networks (e.g., ATM [124], MPLS [129], and MANET [125], [127], [128]). The principal advantages of this approach are its generality, its simplicity in modeling large queuing networks and computational efficiency. Additionally, PSFFA model could be used as the basic mathematical model for developing network dynamic control mechanisms along the lines illustrated in [129], [126], [130], [131].

3.0 TIME DEPENDENT PERFORMANCE ANALYSIS OF IEEE 802.11P VEHICULAR NETWORKS

In recent years, there has been a dramatic increase in research and development activity in vehicular networks. In such networks, a challenging environment is presented for protocol and application design due to their low latency and high packet delivery ratio requirements in a highly mobile environment. The IEEE 1609 working group has defined 5.850-5.925 GHz band as the Dedicated Short Range Communications (DSRC) spectrum band. As a key enabling technology for the next generation inter-vehicle safety communications, the IEEE 802.11p protocol is currently attracting significant attention [1]. For IEEE 802.11p based DSRC applications, there are two types of safety messages including event-driven (emergency) messages and periodic (routine) messages. Event-driven messages are triggered by vehicle abnormal or environment hazards, such as hard brakes, car accidents, and hazardous road conditions. After receiving this type of messages, the approaching vehicles can take appropriate actions to avoid potential accidents or non-safe situations. Periodic (routine) messages contain the state of vehicles (e.g. position, speed, and direction), and vehicles actively broadcast these messages to the neighboring vehicles. The former messages require more stringent QoS requirements (e.g. reliability, latency, etc.) than the latter one, and thus are granted higher transmission priority in 802.11p enhanced distributed coordination function (EDCF) MAC.

One of the key characteristics of vehicular networks is the high speed node mobility. Mobile vehicles move into/out of the carrier sensing range of other vehicles and they result in dynamic changes of network hearing topology. Since the wireless channel is accessed via contention-based 802.11p MAC [1], the available bandwidth shared by the individual vehicle is strongly coupled with each other and becomes a time-varying function of the network

hearing topology. Consider the scenario of a eight-lane freeway as shown in Figure 3, one would expect that the transient/nonstationary effects on a tagged vehicle come from the high-speed movement of the vehicles in the opposite direction, since they continuously join the carrier sensing range of a tagged vehicle. Once the network hearing topology is updated, the performance metric might go through a transient period and then reach steady state. Since each vehicle moves fast, the highly-dynamic network hearing topology perturbates the stationary state frequently and the network might spend significant time in the transient state. Moreover, the steady-state simulation study on vehicular networks cannot capture their transient/nonstationary behavior. While nonstationary simulation method can be used to study the time varying behavior, they are computationally inefficient [17]. Therefore, building an efficient analytical model that considers the fine details of node mobility as well as 802.11p MAC operation to estimate the real-time performance of vehicular networks, presents a significant challenge as well as practical interest. The major notations used in this chapter are summarized in Table 1.

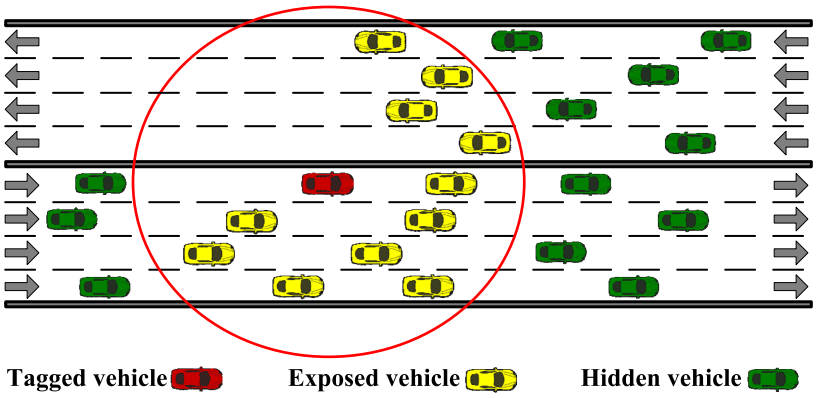


Figure 3: A vehicular ad hoc network on freeway.

This Chapter is organized as follows. Section 3.1 models the network hearing topology by an adjacency matrix. The modeling details for the packet service process of the IEEE 802.11p MAC is provided in Section 3.2. Section 3.3 presents our time dependent performance modeling approach for vehicular networks. In Section 3.4, our modeling approach is codified into an executable algorithm. The numerical solution of our differential equation based

Table 1: Notation list

Notation	**Definition**
$B_i^{j,k}(z)$	Vehicle i's backoff process at stage k of ACj class packet
c_i^j	Coefficient of variation of vehicle i's service time for ACj class packet
D_i^j	Vehicle i's ACj packet delay (sec)
G_i^j	Vehicle i's transmission link utilization function for ACj class traffic
h_{ij}	Vehicle i's hearability of vehicle j
M	Number of nodes in vehicular networks
N^1	Maximum times that the contention window of AC1 class can be doubled
PDR_i^j	Vehicle i's ACj packet delivery ratio
p_i^1	Internal collision probability of AC1 with AC0 class packets in vehicle i
$p_{ir}^{exposed}$	Vehicle i's packet collision probability with its exposed vehicle r
p_{ir}^{hidden}	Vehicle i's packet collision probability with vehicle r's hidden terminals
p_{ir}^c	Vehicle i's overall collision probability with vehicle r
$S_i^j(z)$	PGF of vehicle i's service time for ACj class packet
$T(z)$	PGF of one packet transmission time
T_v	Vulnerable period due to packet collision with hidden terminals (sec)
TS_i^j	A random variable of in vehicle i's service time for ACj class packet (sec)
$U_i^j(z)$	PGF of one-slot backoff decrement time for ACj class packet in vehicle i
$W^{i,k}$	Vehicle i's contention window size at stage k
x_i^j	Number of ACj class packets in vehicle i
λ_i^j	Arrival rate of ACj class packets into vehicle i (pkt/s)
μ_i^j	Average service rate for ACj class packet in vehicle i (pkt/s)
ε	Link propagation delay (sec)
ρ	Link utilization
τ_r	Transmission probability of vehicle r
σ_i^j	Standard deviation of service time for ACj class traffic in vehicle i (sec)
ϵ	One slot duration defined in IEEE.802.11p (sec)
θ	Initial transient period of a queueing system

model and the computation complexity are discussed in Section 3.5. Our model is then compared by discrete event simulation in Section 3.6. Section 3.7 analyzed the computation complexity of our model followed by numerical results on a series of sample networks. Section 3.8 applied the proposed model to study the time varying behavior of vehicular networks in a variety of scenarios. The summary of this chapter is given in Section 3.9.

3.1 MODELING NETWORK HEARABILITY

Consider a network with M vehicles, all vehicles are not necessarily within the carrier sensing range of each other. Then, the network hearing topology at any time t is modeled by a $M \times M$ square matrix as $H(t)$.

$$H(t) = \begin{bmatrix} h_{11}(t) & h_{12}(t) & \cdots & h_{1M}(t) \\ h_{21}(t) & h_{22}(t) & \cdots & h_{2M}(t) \\ \vdots & \vdots & & \vdots \\ h_{M1}(t) & h_{M2}(t) & \cdots & h_{MM}(t) \end{bmatrix} \tag{3.1}$$

where,

$$h_{ij}(t) = \begin{cases} 1, & \text{if vehicle } i \text{ can hear vehicle } j \text{ at time } t \ (i \neq j) \\ 0, & \text{otherwise} \end{cases}$$

By definition, vehicle i can "hear" vehicle j if vehicle j is within the carrier sensing range of vehicle i (i.e. $h_{ij} = 1$). Otherwise, vehicle j is a hidden node of vehicle i (i.e. $h_{ij} = 0$). Taking Figure 3 as an example, vehicles can be enumerated increasingly from left to right on a single lane, and then from bottom to top across different lanes (i.e. the leftmost vehicle on the bottom lane is No. 1 and its right-side neighboring vehicle on the same lane is No. 2, the rightmost vehicle on the top lane is No. M). Due to their mobility, vehicles might move out of/into the carrier sensing range of other vehicles and hence result in a dynamic network hearing topology. Hence, the time varying hearing matrix captures the mobility feature in vehicular networks. If all nodes have the same transmission power, carrier sensing threshold and propagation environment, then matrix $H(t)$ is symmetric (e.g.

$h_{ij}(t) = h_{ji}(t)$). Information about the node movement and hearability with respect to each other can be determined from experimentally gathered trace data, a discrete event simulation of a mobility model or stochastic/probabilistic models of mobility effects on link-pair hearability. The IEEE 802.11p essentially adopts the CSMA/CA with exponential back-off mechanism to control packet access. The exposed nodes can sense the channel busy once a packet is transmitted on the medium and then freeze the backoff, while the hidden nodes always consider the channel as idle and thereby cause packet collision much more often than the exposed one. The network hearing topology has been shown to play a significant role in determining the performance of wireless networks with CSMA/CA based MAC [69].

3.2 MODELING PACKET SERVICE PROCESS OF IEEE 802.11P MAC

3.2.1 An overview of IEEE 802.11p MAC

IEEE 802.11p adopts the enhanced distributed channel access (EDCA) mechanism proposed in IEEE 802.11e to support contention-based prioritized QoS. The EDCA mechanism defines four access categories (ACs) that provide support for data traffic with four priorities. Each AC queue works as an independent DCF station with EDCA mechanism to contend for Transmission Opportunities (TXOP) using its own EDCA parameters. Prioritization of transmission in EDCA is implemented by arbitration interframe space (AIFS), which can be considered as an extension of the backoff procedure in DCF. As shown in Figure 4, aside from the original short interframe spacing (SIFS), PCF IFS (PIFS), and DCF IFS (DIFS), new AIFS values for different ACs are introduced in EDCA. The frames from the higher layer arrive at the MAC layer with different priorities, and then enter different queues. The backoff instances in a node can be considered as being independent of each other without virtual internal collisions. For each AC, if the channel is idle for a period time equal to an AIFS, it transmits. Otherwise, if the channel is sensed busy, the AC will persist to monitor the channel until the idle duration up to the AIFS. At this point, the AC generates a random interval according to its CW value and starts a backoff procedure. The backoff counter de-

creases again only when the channel keeps idle for an AIFS time, which is a duration derived from the value AIFSN[AC] by

$$AIFS[AC] = AIFSN[AC] \times SlotTime + SIFS \tag{3.2}$$

where the value of AIFSN[AC] is set in the EDCA parameter table, SlotTime is the duration of a slot time, and SIFSTime is the length of SIFS. In Table 2 taken from the IEEE 802.11p standard [1], the AC with a smaller AIFSN has higher priority to access the channel. Assigning a shorter CW size to a higher priority AC ensures that a higher priority AC has a better chance to access the channel than a lower priority AC.

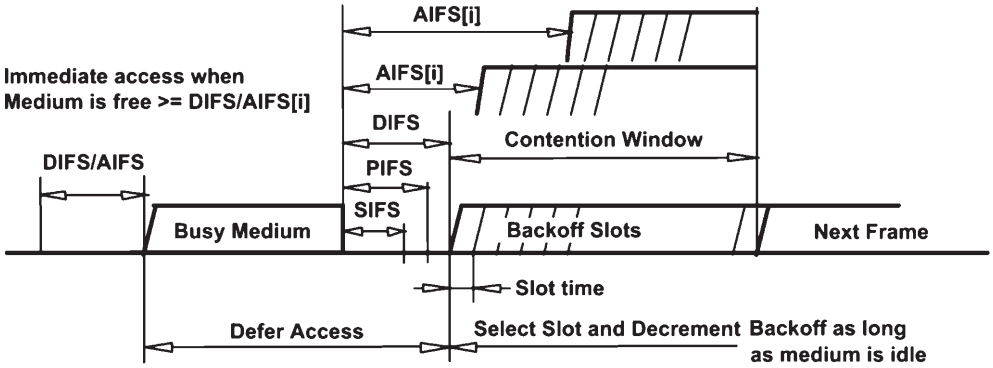


Figure 4: IEEE 802.11 EDCA procedure.

Table 2: IEEE 802.11p EDCA parameters for AC0 and AC1

AC	$CW_{min}[AC]$	$CW_{max}[AC]$	$AIFSN$
3	CW_{min}	CW_{max}	9
2	CW_{min}	CW_{max}	6
1	$(CW_{min}+1)/2-1$	CW_{min}	3
0	$(CW_{min}+1)/4-1$	$(CW_{min}+1)/2-1$	2

3.2.2 Markov Chain for IEEE 802.11p MAC service process

For safety applications in vehicular networks, there are typically two types of messages in each vehicle to be broadcasted including emergency messages and routine messages. Hence, in this paper, we only consider two types of messages which are set priorities with AC0 (emergency messages) and AC1 (routine messages), respectively. We assume the routine safety message is generated periodically with rate λ^0 packets per second, and the packet arrival rate λ^1 for emergency safety message is a Poisson process. Figure 5 shows the prioritization mechanism inside each transmitter, where there are two transmission queues associated with two independent traffic categories. Within one node, each ACi MAC queue behaves like a virtual transmitter. If two ACs of a transmitter try to access the channel in the same time slot, a virtual collision occurs. In this case, the packets with the highest priority will be transmitted and the lower priority packets enter another backoff stage with doubled CWs immediately. If the number of failed retransmissions reaches the retry limit, the packet will be discarded.

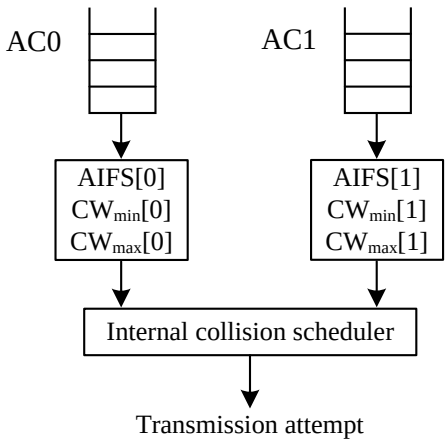


Figure 5: Prioritization mechanism inside a single transmitter.

To model the service process of 802.11p MAC, we apply the probability generating function (PGF) approach to transform the Markov chain into the z domain in Figure 6, by following the similar approach in [132]. Figure 6(a)-(b) illustrates the PGF of service time for AC0 and AC1 traffic. We denote $S_i^j(z)$ as the PGF of packet service time TS_i^j for vehicle

i's class j traffic ($j = 0, 1$). $S_i^j(z)$ includes the backoff time $B_i^{j,k}(z)$ as well as the transmission time of one packet $T(z)$. Because the higher-priority AC0 traffic is free from virtual internal collision, the PGF of its backoff time is simply $B_i^{0,0}(z)$ with constant contention window size $W^{0,0}$. However, the lower-priority AC1 may suffer from internal collisions with higher-priority AC0, and such a collision triggers another backoff stage of AC1 by doubling its contention window size. Here, we denote the virtual internal collision probability of vehicle i's AC1 traffic as p_i^1, which depends on the transmission attempt probability of AC0 with a certain network hearing topology. Let N^1 be the maximum number of times that the contention window of AC1 can be doubled. When N^1 is reached and more internal collisions occur, the transmitter will continue to trigger backoff stages with the maximum-size contention window until the retry limit L is reached (e.g. $B_i^{1,L}(z)$). Since p_i^1 is the virtual internal collision probability of AC1 traffic with AC0 of the tagged vehicle i, p_i^1 also represents the probability of a low-priority AC1 packet to enter another backoff stage with double-size contention window. Figure 6(c) describes vehicle i's backoff process $B_i^{j,k}(z)$ at stage k of ACj traffic with contention window size $W^{i,k}$. $U_i^j(z)$ refers to the PGF of the average time that a backoff counter decreases by one slot unit. From Figure 6, we can derive $B_i^{0,0}(z)$, $B_i^{1,k}(z)$, $S_i^0(z)$ and $S_i^1(z)$ as follows.

$$B_i^{0,0}(z) = \frac{1}{W^{0,0}} \sum_{n=0}^{W^{0,0}-1} \left[U_i^0(z)\right]^n \tag{3.3}$$

$$B_i^{1,k}(z) = \begin{cases} \frac{1}{W^{1,k}} \sum_{n=0}^{W^{1,k}-1} \left[U_i^1(z)\right]^n, k \in [0, N^1] \\ \frac{1}{W^{1,N^1}} \sum_{n=0}^{W^{1,N^1}-1} \left[U_i^1(z)\right]^n, k \in [N^1, L] \end{cases} \tag{3.4}$$

$$S_i^0(z) = T(z) B_i^{0,0}(z) \tag{3.5}$$

$$S_i^1(z) = (1 - p_i^1) T(z) \sum_{n=0}^{L} \left[(p_i^1)^n \prod_{k=0}^{n} B_i^{1,k}(z) \right] + (p_i^1)^{L+1} \prod_{k=0}^{L} B_i^{1,k}(z) \tag{3.6}$$

Given $S_i^j(z)$, we can obtain the arbitrary nth moment of TS_i^j, where the unit of TS_i^j is one-bit transmission time. For example, the mean and the variance of the service time of vehicle i are given by

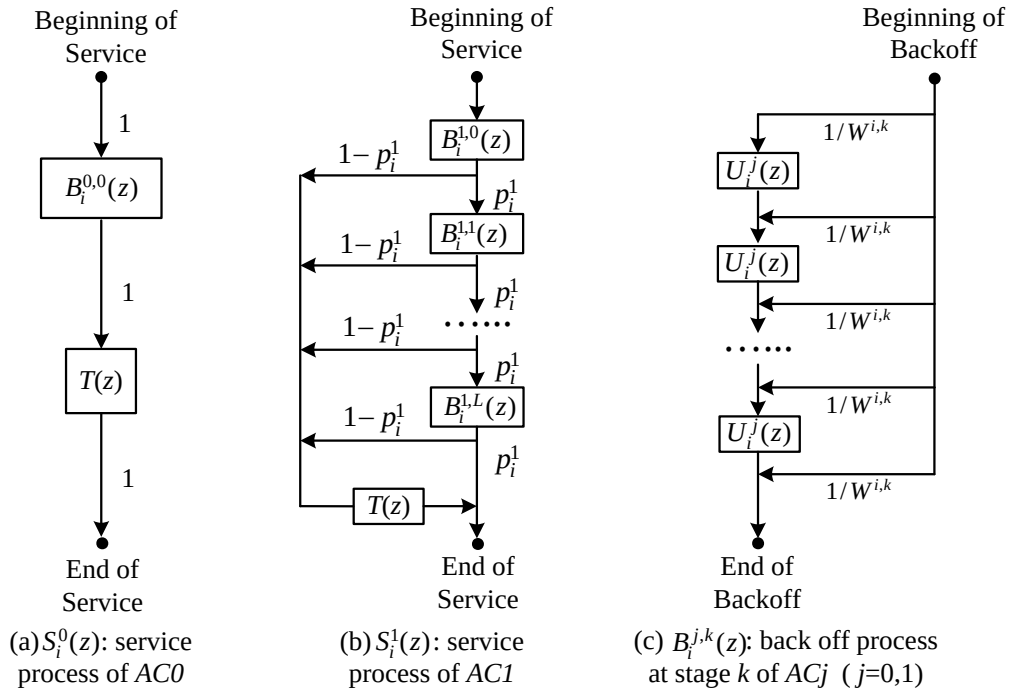


Figure 6: z-transformed 802.11p MAC service process for vehicle i.

$$(\mu_i^j)^{-1} = E[TS_i^j] = \frac{dS_i^j(z)}{dz}|_{z=1}, \quad j = 0, 1 \tag{3.7}$$

$$(\sigma_i^j)^2 = Var[TS_i^j] = \left[\frac{d^2S_i^j(z)}{dz^2} + \frac{dS_i^j(z)}{dz} - \left(\frac{dS_i^j(z)}{dz}\right)^2\right]|_{z=1}, \quad j = 0, 1 \tag{3.8}$$

For an IEEE 802.11p vehicular network, since packet transmissions are based on carrier sensing and access contention, the packet service times distribution at different vehicles in a vehicular network are strongly coupled and dependent on the number of vehicles in each other's carrier sensing range. Thus, given the network hearing topology $H(t)$, the first and second moment of packet service time $(\mu_i^j)^{-1}$ and $(\sigma_i^j)^2$ can be solved by coupling Equations (3.3)-(3.8) for all vehicles in a network. The details of the solution procedure can be found in [132]. The results for packet service time distribution provided in this section will be used as the basis to construct the time dependent queueing model in the next section.

3.3 MODELING DYNAMIC BEHAVIOR OF VEHICULAR NETWORKS

We now adopt a fluid-flow based approximation technique to describe the dynamic behavior of the transmission queue at each vehicle, with the help of the PSFFA approximation concept [30]. Specifically, the PSFFA method models the average number in the queuing system by a non-linear differential equation which is derived from a point-wise mapping of the steady-state queuing relationship and can be solved numerically using a standard numerical integration technique. To understand how the fluid flow model approach can be used to characterize the dynamic queuing behavior of each vehicle in IEEE 802.11p vehicular networks, we first give a brief discussion on the principle of fluid flow model.

3.3.1 The Principle of Fluid Flow Model

In this work, we consider a single server first-come-first-serve (FCFS) queuing system with nonstationary arrival process, where $\lambda(t)$ represents the *ensemble average* arrival rate at time t. The model is developed by focusing on the dynamics of the packet buffered 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 $f_{out}(t)$:

$$\dot{x}(t) = -f_{out}(t) + f_{in}(t) \tag{3.9}$$

For an infinite waiting space queue, the flow in equals to the arrival rate $f_{in}(t) = \lambda(t)$ in the unit of customers per time unit. The flow out can be related to the *ensemble average* utilization of the server as $f_{out}(t) = \mu(t)\rho(t)$, where μ refers to the average service rate in number of customers per time unit and ρ is the server utilization. The Equation (3.9), sometimes referred to as the fluid flow equation, is quite general and can model a wide range of queuing systems as shown in [115]. The fluid flow equation can be written in terms of the average arrival rate and the departure rate as:

$$\dot{x}(t) = -\mu(t)\rho(t) + \lambda(t) \tag{3.10}$$

The server utilization $\rho(t)$ in (3.10) depends on the stochastic modeling assumptions of the queue under study such as traffic arrival process and packet length distribution. In general, the exact expression of $\rho(t)$ is difficult to determine. In [115], an approximation approach was proposed by matching the equilibrium point in the differential equation to the corresponding steady state queuing theory result. To adopt the point-wise mapping approach, the server utilization function is approximated by the non-negative function $G(x(t))$, which represent the ensemble average utilization of the server at time t in the form of the state variable. Thus, Equation (3.10) can then be written as:

$$\dot{x}(t) = -\mu(t)G(x(t)) + \lambda(t) \tag{3.11}$$

Generally, the resulting function $G(x(t))$ is nonlinear and a closed form solution is not possible. However, one can solve for $x(t)$ in Equation (3.11) using numerical methods. 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 $[t, t + \Delta t]$, we can determine the state variable at the end of the time interval $x(t + \Delta t)$ by numerically integrating (3.11), and then set $x(t + \Delta t)$ as an initial condition for the next time step $[t + \Delta t, t + 2\Delta t]$. The arrival rate for the new time step can be adjusted if necessary, and this procedure is repeated for each time interval along the time horizon. Numerical studies in [115] have shown that results from PSFFA models are reasonably accurate. The extended fluid flow model for the queuing system with finite buffer size is derived in [115].

3.3.2 Fluid Flow Model for Vehicular Networks

We model the packet transmission process at each vehicle as a generally distributed "service" process, and the packet service time as an i.i.d variable with mean $1/\mu$ and variance σ^2. For the AC0 emergency traffic load at a vehicle, we assume the packet inter-arrival time follows an exponential distribution with mean $1/\lambda$. The queue utilization ρ is equal to λ/μ. Then, each vehicle in network is modeled as an $M/G/1$ queue. The most well-known result

of the $M/G/1$ steady-state queuing model is the Pollaczek-Khintchine (P-K) formula [49], which gives the average number of packet in the queuing system of a vehicle as

$$x = \rho + \frac{\rho^2(1+c^2)}{2(1-\rho)} \tag{3.12}$$

where c^2 is the squared coefficient variation of service time and $c^2 = \sigma^2\mu^2$. One can analytically invert the P-K formula (3.12) to find ρ as a function of x (i.e. $G(x)$), resulting in

$$G(x) = \frac{x+1-\sqrt{x^2+2c^2x+1}}{1-c^2} \tag{3.13}$$

For the AC1 routine safety message periodically generated in a vehicle, we use a $D/G/1$ queue to model the system, but the closed-form formula for the queue length distribution is not readily available. Instead, we adopt the well-known Kramer and Lagbenbach-Belz (KLB) $G/G/1$ formula [133] and assign the squared coefficient variation of the arrival process by zero to approximate the expected number of packets in a $D/G/1$ queuing system

$$x \approx \rho + \frac{\rho^2 c^2 e^{\frac{-2(1-\rho)}{3\rho c^2}}}{2(1-\rho)} \tag{3.14}$$

Numerical results in [133] indicate that KLB formula provides very good approximation for $D/G/1$ queue. Since it is extremely difficult to analytically invert the KLB equation (4.18) to get a closed form for server utilization $\rho = G(x)$. We thus numerically determine (x, ρ) from the KLB equation (4.18) for a given parameter c, and then apply a curve fitting approach based on the (x, ρ) data pair to find the utilization function $\rho = G(x(t))$ in the form of a polynomial, that is

$$G(x) = a_n(c)x^n + a_{n-1}(c)x^{n-1} + ... + a_0(c) \tag{3.15}$$

Given the models above, we now consider a network with M vehicles, an arbitrary vehicle i generates AC0 emergency (class 0) and AC1 routine (class 1) traffic with the mean rate $\lambda_i^0(t)$ and $\lambda_i^1(t)$, respectively (i.e. $f_{in\_i}^j(t) = \lambda_i^j(t)$, $\forall i = 1, 2, \ldots, M$, $\forall j = 0, 1$). Let $x_i^j(t)$ represent the average number of packets in the corresponding subqueuing system j of vehicle i at time t and $\dot{x}_i^j(t)$ is the change rate of $x_i^j(t)$. In a vehicular network, nodes move into/out of each other's carrier sensing range and result in the dynamic change of network hearing topology $H(t)$. As a result, vehicle i's class j packet service rate $\mu_i^j(t)$ and coefficient variation

of service time c_i^j are time varying functions of the number of vehicles in its carrier sensing range. Thus, we incorporate the network hearing topology $H(t)$ to write the flow out term of Equation (3.10) as

$$f_{out\_i}^j(t) = -\mu_i^j(H(t))G_i^j\Big(c_i^j(H(t)), x_i^j(t)\Big), \qquad \forall i = 1, 2, \ldots, M, \quad \forall j = 0, 1 \tag{3.16}$$

After substituting $G(x(t))$ (3.13) back into the general fluid flow model (3.11), we finally determine the fluid flow model for an arbitrary vehicle i with AC0 emergency traffic (class 0) as

$$\dot{x}_i^0(t) = -\mu_i^0(H(t))\frac{x_i^0(t) + 1 - \sqrt{x_i^0(t)^2 + 2c_i^0(H(t))^2 x_i^0(t) + 1}}{1 - c_i^0(H(t))^2} + \lambda_i^0(t), \quad \forall i = 1, 2, \ldots, M \tag{3.17}$$

Similarly, we can also obtain the fluid flow model for an vehicle i with AC1 routine traffic (class 1)

$$\begin{aligned}\dot{x}_i^1(t) = -\mu_i^1(H(t))\Big[a_n\Big(c_i^1(H(t))\Big)x_i^1(t)^n + a_{n-1}\Big(c_i^1(H(t))\Big)x_i^1(t)^{n-1} + \ldots + a_0\Big(c_i^1(H(t))\Big)\Big] \\ +\lambda_i^1(t), \qquad \forall i = 1, 2, \ldots, M\end{aligned} \tag{3.18}$$

where the service rate μ_i^j and the coefficient variation of service time c_i^j are calculated for all vehicles based on the network hearing topology $H(t)$ at every time step $[t, t + \Delta t]$. Finally, the differential equation based fluid flow models (3.17) and (3.18) can be solved numerically using standard numerical integration techniques (e.g., Runge-Kutta algorithm).

3.3.3 Performance Metrics

To study the performance of safety message broadcast in DSRC based vehicular communications, there are two important performance metrics: packet delay and packet delivery ratio (PDR), as analyzed below.

3.3.3.1 Packet Delay The packet delay is defined as the period between the time of generating a packet at sender vehicle and time of receiving that packet. Note that IEEE 802.11p operates in the broadcast mode, which does not support retransmission and packet response from the receiver. Since IEEE 802.11 EDCA essentially follows the non-preemptive priority scheduling, Little's law is applicable for each individual class of queueing system [134]. From Little's law, the average number in the queuing system is equivalent to the product of the average arrival rate and the average sojourn time in the system. We denote x_i^j as the average number of packets in vehicle i's subqueuing system j, λ_i^j is the corresponding arrival rate and W_i^j is the average sojourn time of class j packets, then $W_i^j = x_i^j/\lambda_i^j$. With the assumption of constant arrival and service rates over a small step, the change in average sojourn time can be related to the rate of change in average number of packets in the system $\dot{W} = \dot{x}/\lambda$. Let D_i denote the average packet delay at vehicle i on this path, and ε represents the link propagation delay. Hence, we can finally write the delay of vehicle i's class j packet at time t as

$$D_i^j(t) = \int \dot{W}_i^j(t)dt + \varepsilon = \int \frac{\dot{x}_i^j(t)}{\lambda_i^j(t)}dt + \varepsilon, \qquad \forall i = 1, 2, \ldots, M, \ \forall j = 0, 1 \tag{3.19}$$

3.3.3.2 Packet Delivery Ratio The packet delivery ratio is interpreted as the probability of the transmitted packet to reach all intended receiving vehicles within the coverage of a given sender vehicle. It illustrates the reliability of packet transmission over the wireless medium. Following the basic approach in [132], we derive the performance measure of packet delivery ratio as follow. In the following analysis, we assume that the network hearability between two vehicles is a binary value, as indicated in Section 3.1. Suppose the tagged vehicle i transmits a packet, we first calculate the exposed collision probability $p_{i,r}^{exposed}$ with the packet sent by vehicle r within vehicle i's carrier sensing range (i.e. $h_{ir} = 1$)

$$p_{ir}^{exposed} = \tau_r h_{ir} \tag{3.20}$$

where τ_r denotes the transmission probability of the exposed vehicle r for both AC0 and AC1 traffic. Then, we consider the packet collision probability due to hidden terminals during the vulnerable period. Here, the vulnerable period T_v is the time period when hidden nodes

can start a transmission that would collide with the packet from the tagged transmission node. We denote t_{data} as the transmission time of data packet. For IEEE 802.11p, this vulnerable time interval spans over $(-t_{\text{data}}, t_{\text{data}})$ with respect to the transmission start time of the tagged node and the vulnerable period T_v equals to $2t_{\text{data}}$. Here, we assume that all vehicles in the network are timely synchronized and packet size is identical for both ACs so that transmission time of one packet t_{data} is a constant. To avoid collision with the packet transmitted by the tagged vehicle i, we consider the condition that no packet is sent out by the hidden vehicle j during T_v with probability $(1-\tau_j)^{\frac{T_v}{\epsilon}}$, where ϵ denotes the duration of one slot time defined in the IEEE 802.11p standard. Since each vehicle can only possibly transmit a packet at the beginning of a time slot, T_v/ϵ represents the number of chances vehicle j could transmit. As a result, the packet collision probability between the tagged vehicle i and all its hidden nodes within the reception range of node r in an M-node vehicular network is given by

$$p_{ir}^{hidden} = h_{ir}\left(1 - \prod_{j=1, j\neq i}^{N} (1-\tau_j)^{\frac{T_v}{\epsilon}(1-h_{ij})h_{rj}}\right) \tag{3.21}$$

With the assumption that the events of packet collision with exposed terminals and hidden terminals are independent with each other, we have the overall collision probability of the transmitted packet from vehicle i to vehicle r as follow:

$$p_{ir}^{c} = 1 - (1 - p_{ir}^{exposed})(1 - p_{ir}^{hidden}) \tag{3.22}$$

Finally, the delivery ratio of the tagged vehicle i's class j packet at time t can be calculated by dividing the total class j traffic received at all exposed vehicles with the overall class j traffic broadcasted from the tagged vehicle i

$$PDR_i^j(t) = \frac{\sum_{r=1}^{N} f_{out,i}^{j}(t) h_{ir}(t)(1 - p_{ir}^{c}(t))}{\sum_{r=1}^{N} f_{in,i}^{j}(t) h_{ir}(t)}, \qquad \forall i = 1, 2, \ldots, M, \ \forall j = 0, 1 \tag{3.23}$$

3.4 PERFORMANCE MODELING ALGORITHM

We now codify our modeling procedure into the following algorithm to estimate the time dependent behavior of tagged vehicle i in a vehicular network over the desired time interval $[t_0, t_f]$.

1. Configure the vehicular network including packet length, traffic arrival rate, mobility model, vehicle density and PHY/MAC layer parameters.

2. At time t, determine the network hearability by a $M \times M$ square matrix $H(t)$. If $t = t_0$ or $H(t) \neq H(t - \Delta t)$, then go to step 3 to calculate the updated mean and variance of MAC service time. If $H(t) = H(t-\Delta t)$, retain the current MAC service time distribution and jump to step 4.

3. Derive the PGFs of vehicle i's MAC service time (i.e. $S_i^0(z)$ and $S_i^1(z)$) for AC0 routine traffic and AC1 emergency traffic, respectively, according to Equation (3.3)-(3.6). Then, calculate the first and the second moment of the MAC service time (i.e. μ_i^j and σ_i^j), according to Equation (3.7)-(3.8).

4. Numerically solve the fluid flow model (3.17) and (3.18) to 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 $[t+\Delta t, t+2\Delta t]$. Meanwhile, $f_{in,i}^j(t + \Delta t)$ and $f_{out,i}^j(t + \Delta t)$ are also obtained.

5. Estimate vehicle i's class j packet delay $D_i^j(t + \Delta t)$ by integrating the time-varying packet sojourn time $\dot{W}_i^j(t+\Delta t)$ plus the link propagation delays ε, according to Equation (3.19).

6. Evaluate vehicle i's class j packet delivery ratio $PDR_i^j(t + \Delta t)$ by dividing the overall delivered traffic with the total offered traffic, according to Equation (3.23).

7. Increment time $t = t + \Delta t$. If $t < t_f$, go back to step 2; else stop.

Any standard numerical integration method can be used to solve differential equations such as (3.17) and (3.18). Here, we use the fourth-fifth order *Runge-Kutta* algorithm in Matlab to generate numerical results in Section 3.6.

3.5 NUMERICAL SOLUTION OF ORDINARY DIFFERENTIAL EQUATIONS (ODE)

3.5.1 Order of Accuracy and Stiffness

Since our fluid flow model is essentially consisted of ODEs, we hereby provide a discussion on two important concepts involved in numerically solving ODEs, those are *order of accuracy* and *stiffness*.

Order of accuracy evaluates how well the numerical algorithm approximates the solution. A numerical solution to a differential equation is said to be nth-order accurate if the error is proportional to the step size h to the nth power. For example, the order of accuracy of fourth-order Runge-Kutta method is $O(h^n)$. Usually, the algorithm increases the step size to improve the computation efficiency when the solution varies less. Alternatively, when the solution curve displays much variation, step size has to be reduced to guarantee the local error per step is within the tolerance level.

An ordinary differential equation problem is stiff if the solution being sought is varying slowly, but there are nearby solutions that vary rapidly, so the numerical method must take small steps to obtain satisfactory results. The reason of stiffness is that the equation includes some terms that can lead to rapid variation in the solution. Stiffness is an computation efficiency issue. If one does not care about how much time a computation takes, stiffness is not an issue. Nonstiff methods can solve stiff problems, but a long computation time is required. Instead, stiff methods are able to compute solutions more efficiently.

3.5.2 ODE Solvers in MATLAB

In MATLAB, differential equations can be solved numerically with the commands listed in Table 3. The algorithms used in the ODE solvers vary according to order of accuracy and the type of ODE (stiff or nonstiff) they are designed to solve.

In general, *ode45* is the best function to apply as a "first try" for most problems [135]. *ode45* uses the fourth-fifth order Runge-Kutta algorithm, while *ode23* adopts the same algorithm with the second and third order formulas. Since *ode23* works with lower order

Table 3: ODE Solvers in MATLAB

Solver	Order of Accuracy	Stiffness	Algorithm
ode45	Medium	Nonstiff	Runge-Kutta
ode23	Low	Nonstiff	Runge-Kutta
ode113	Low to high	Nonstiff	Adams
ode15s	Low to medium	Stiff	NDFs (BDFs)
ode23s	Low	Stiff	Rosenbrock
ode23t	Low	Moderately Stiff	Trapezoidal

formulas, *ode23* generally gives less smooth solution than *ode45* [136]. *ode113* uses Adams integration algorithm, which is more efficient than *ode45* and *ode23* at stringent error tolerances or a computationally intensive ODE [137]. Regarding the stiff differential equations, the previous solvers will not be able to find an accurate solution, or may need excessive computation times for taking very small time steps. In that case, *ode15s* is a better choice, since it is a multistep variable-order solver based on the numerical differentiation formulas (NDFs) and optionally the backward differentiation formulas (BDFs) [138]. *ode23s* is a one-step solver, and it may be more efficient than *ode15s* at crude tolerances [139]. *ode23t* is used if the problem is only moderately stiff and you need a solution without numerical damping [138]. All of these numerical approaches are able to solve a single ODE or any size of an ODE system under the condition of Picard-Lindelöf theorem, which guarantees the existence and uniqueness of the solutions [140].

3.6 MODEL COMPARISON BY DISCRETE EVENT SIMULATION

In order to evaluate the accuracy of our analytical model, we compare the modeling results with the corresponding simulation results. The simulation of a vehicular network is conducted in OPNET while the equivalent fluid-flow based analytical model is solved by using

Runge-Kutta numerical integration technique in MATLAB. In the experiment, each vehicle is equipped with an IEEE 802.11p-compliance RF transceiver, and the traffic load is set to be Poisson. Due to low arrival rate of safety message with short packet size, we assume all the transmission queues have infinite buffer size which can keep all backlogged packets. All the PHY/MAC layer parameters used in both the simulation and analytical model are adopted from the IEEE 802.11p standard [1] and listed in Table 4. Here, the data rate is set to be 3 Mbps with BPSK modulation, which is a reasonable choice for a freeway scenario [141]. Such a simple modulation mechanism only requires low sensitivity, and thus enlarges vehicle's transmission coverage area and improve the receiver's robustness to multipath fading and Doppler shift. Also, we consider that every broadcast packet includes 400 bytes security fields containing a digital signature plus a certificate, which are very important for inter-vehicle communications in practice.

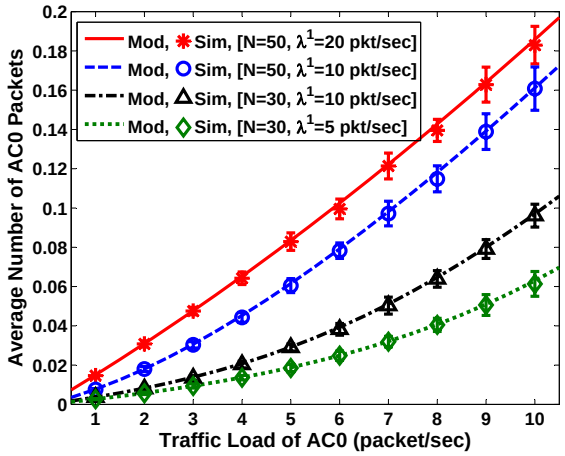
Table 4: IEEE 802.11p Parameters

Packet payload size	300 bytes
MAC header size	224 bits
PHY header size	48 bits
Data rate	3 Mbps
Short Inter-frame Space $SIFS$	32 μs
Time slot duration ϵ	13 μs
AC0 Inter-frame Space $AIFS[0]$	58 μs
AC1 Inter-frame Space $AIFS[1]$	71 μs
Minimum contention window size	16
Retry limit	1

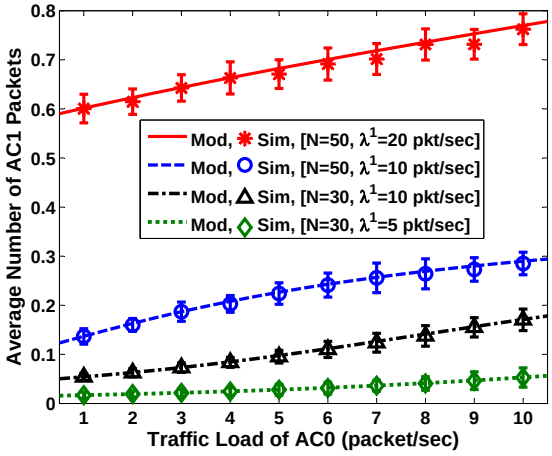
We first examine the accuracy of our steady state performance model for both AC0 and AC1 traffic in a vehicular network. The packet service time distribution for both traffic classes are calculated according to the derivation in Section 3.2. Then, we model the packet transmission queues for AC0 emergency traffic and AC1 routine traffic as $M/G/1$ and $D/G/1$, respectively. Figure 7(a)-(b) depicts the steady-state number of packets in both

AC0 and AC1 queueing systems as a function of various traffic load of AC0. As seen from the figure, the average number of packets in both AC0 and AC1 queueing systems is in the direct proportion to the increase of AC0 traffic load, but with different growth rate. Since the traffic load of AC0 enters its corresponding queue, the number of packets in AC0 queuing system is inevitably impacted by the arrival rate. Meanwhile, since 802.11p EDCF assigns AC1 traffic with lower priority than AC0 to access the channel, the intensity of AC0 traffic load can also influence the queue size of AC1 traffic. In Figure 8(a)-(b), the steady-state sojourn time for both queueing systems are evaluated according to Little's law. Due to the high-priority access of AC0, the growth of AC0 traffic load is alleviated by consuming more portion of the bandwidth so that the queueing delay of AC0 packets does not rise up significantly. At the same time, less bandwidth is shared by AC1 traffic and thus the delay of AC1 packets is prolonged. As shown by the legends of both figures, we parameterize each curve by the duplet [number of vehicle in the same carrier sensing range N, traffic load of AC1 λ^1 (packet/sec)]. By varying theses parameters in different scenario, we demonstrate the individual and combinational effects of all these parameters on the system performance. As we can see from Figures 7-8, the good match between the analytical results and the steady-state simulation results shows that our analytical steady-state model can accurately describe the IEEE 802.11p MAC.

Next we evaluate the tagged vehicle's time dependent performance due to node mobility in a vehicular network for both traffic classes. In this experiment, we consider a typical eight-lane freeway with four lanes in each direction, and each lane is 400 meters long and 3.5 meters wide, as shown in Figure 9. The packet transmission range of all vehicles is 100 m, and their speed is uniformly distributed within 35-40 meters/second (i.e. 78.3-89.5 miles/hour). All vehicles move according to the freeway mobility model given in [142]. We create three scenarios with various vehicle densities and traffic loads and examine their effect on tagged vehicle's performance (i.e. packet delay and PDR) in Figures 10-14. In these figures, the vehicle density of the lanes towards the same and opposite directions of tagged vehicle is denoted as $[d^s, d^o]$, and the generation rate of AC0 and AC1 class traffic is represented by $[\lambda^0, \lambda^1]$. By using fluid flow model, we are able to look into the real-time performance of the tagged vehicle, according to the instantaneous network topology. The

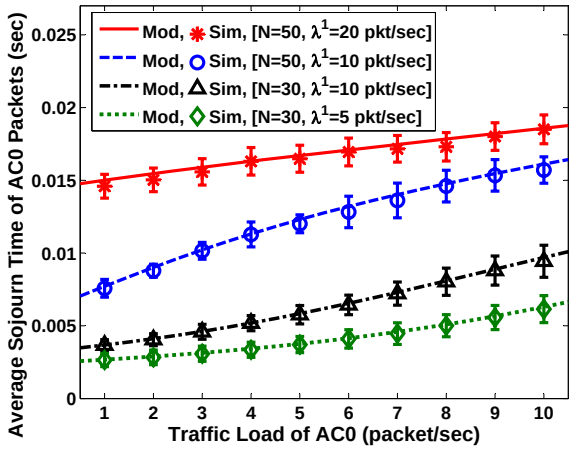


(a) Queuing system for AC0 emergency traffic

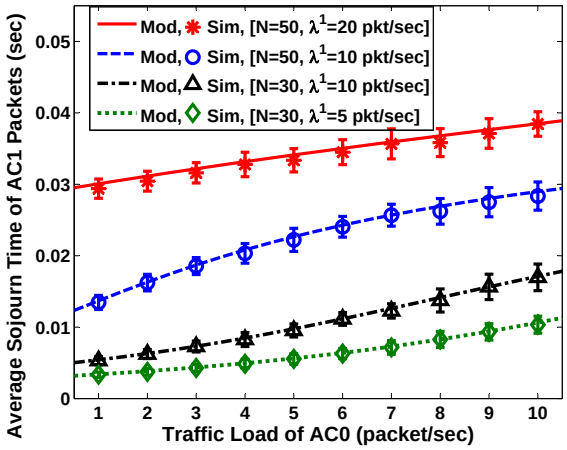


(b) Queuing system for AC1 routine traffic

Figure 7: Average number of packets in AC0 and AC1 queuing systems vs. traffic load of AC0 packet.



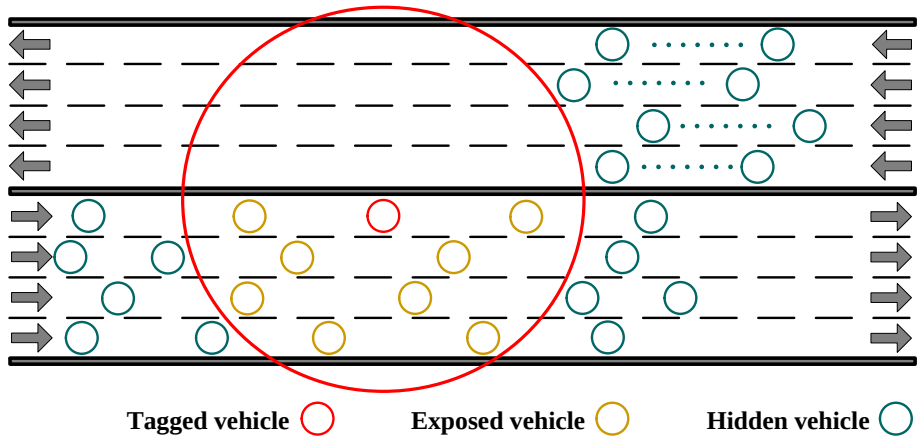
(a) Queuing system for AC0 emergency traffic



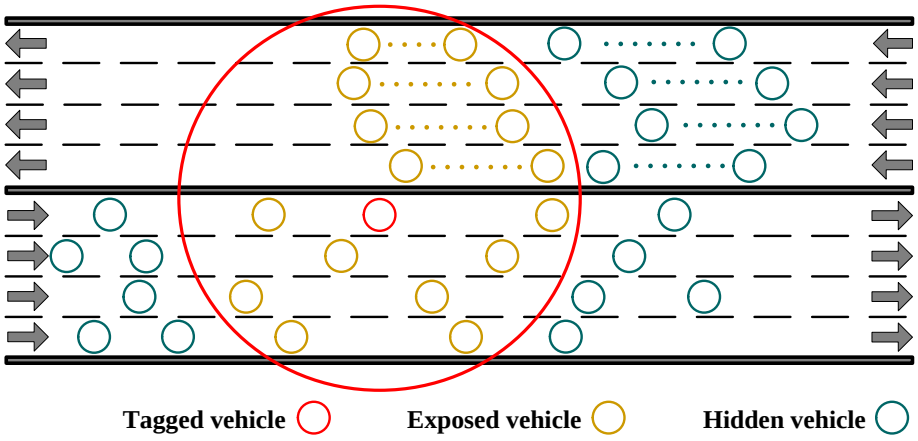
(b) Queuing system for AC1 routine traffic

Figure 8: Average sojourn time in AC0 and AC1 queuing systems vs. traffic load of AC0 packet.

equivalent simulation model is set up and conducted by following nonstationary simulation method [17]. The steady state results are also shown by averaging the simulation results over the time. In addition, we plot the results obtained from the pointwise stationary model [114], which approximates the nonstationary queuing system by using steady-state formula at each time step.



(a) $t = 0$ second



(b) $t = 1$ second

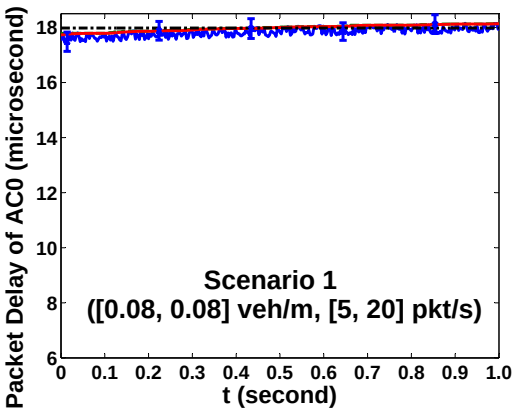
Figure 9: Network hearing topology of the tagged vehicle at sampled time points.

In Figures 10-11, the packet delay consists of the medium access delay and the packet transmission delay. Since the packet transmission delay is fixed due to a constant packet size in our experiment, the variation of packet delay comes from the change of medium access

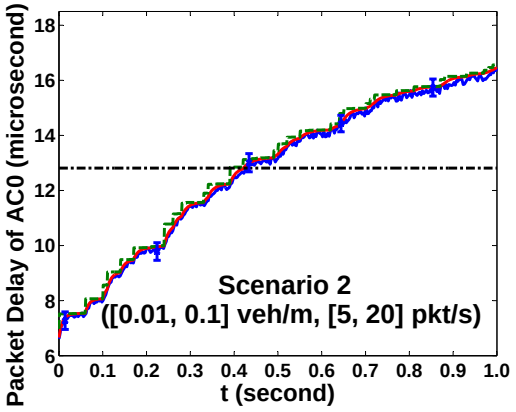
delay because of the dynamic hearing topology. As shown in Figure 3, when vehicles moving towards left continuously get into the carrier sensing range of the tagged vehicle, more and more vehicles contend to access the channel. As a result, the medium access delay of the tagged vehicle rises up every time when new vehicles move into its carrier sensing range. Due to the light traffic load and high-priority service for AC0 emergency traffic, its packet delay is about half of the delay experienced by AC1 routine traffic.

In addition, the packet delivery ratio (PDR) for both classes is evaluated in Figures 13-14. Initially, the packet collision probability steps up, once additional vehicles move into the carrier sensing range of tagged vehicle and compete to access the channel. But meanwhile, less hidden terminals of the tagged vehicle exist in the network. In another words, the tagged vehicle can sense more and more peers in the network, so that the collision probability is gradually decreased and thus improve the PDR after 0.3 second due to the elimination of hidden terminals. The dynamic behavior of the packet collision probability is plotted in Figure 12. In Figures 13-14, the dropping spike of PDR at each topology transition comes from the fact that a newly-joined vehicle cannot receive the packet immediately from the tagged vehicle. Because of the longer medium access delay of AC1 routine traffic, such spikes are more significant than AC0 emergency traffic. But still, the PDR of both classes are similar because all the transmitted packets by the tagged vehicle experience the similar collision probability, which depends on the transmission probability of the rest vehicles in the network.

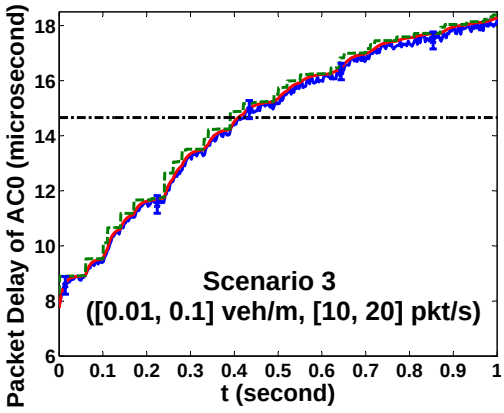
In Figures 10-14, we first examine the effect of vehicle density on the nonstationarity of network performance by comparing the numerical results in scenario 1 and 2. By changing the vehicle density from a balanced case [0.08 0.08] vehicle/m to an imbalanced case [0.1 0.01] vehicle/m, we observe that the network behavior becomes more nonstationary. In reality, such an imbalanced vehicle density phenomenon occurs often. For example, during the rush hour in the morning, the lanes towards downtown are much more crowded than the lanes in the opposite direction. With imbalanced vehicle density (i.e. low density on the tagged vehicle direction and high density on the opposite direction), the change of number of vehicles within carrier sensing range of the tagged vehicle is significant so that it triggers large variations of tagged vehicle's behavior. Then, we increase the AC0 traffic load of each



(a) Scenario 1

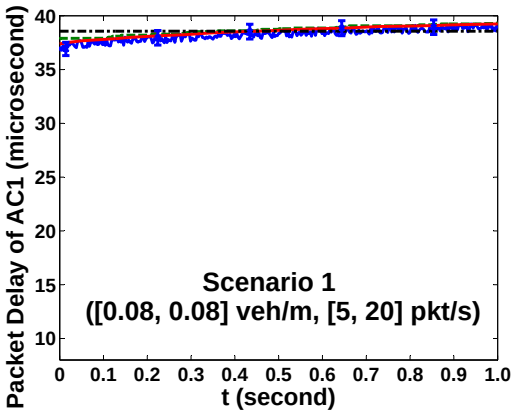


(b) Scenario 2

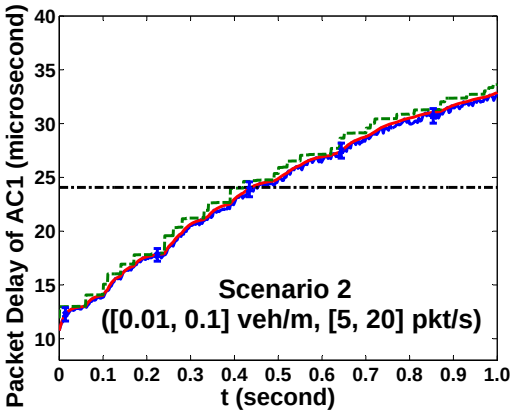


(c) Scenario 3

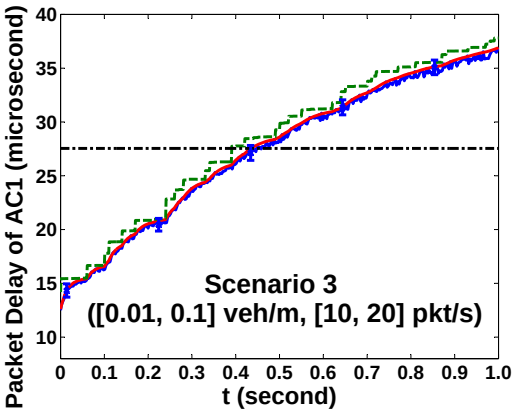
Figure 10: Delay of AC0 packets transmitted by the tagged vehicle for various scenarios of $([d^s, d^o], [\lambda^0, \lambda^1])$.



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Figure 11: Delay of AC1 packets transmitted by the tagged vehicle for various scenarios of $([d^s, d^o], [\lambda^0, \lambda^1])$.

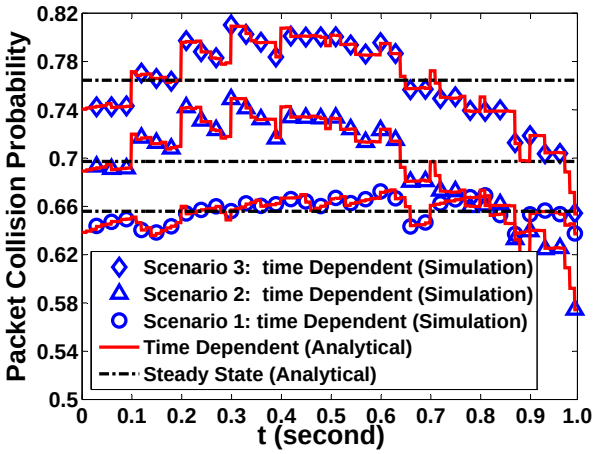
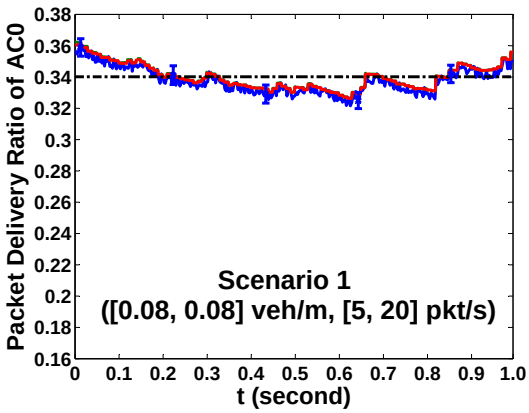


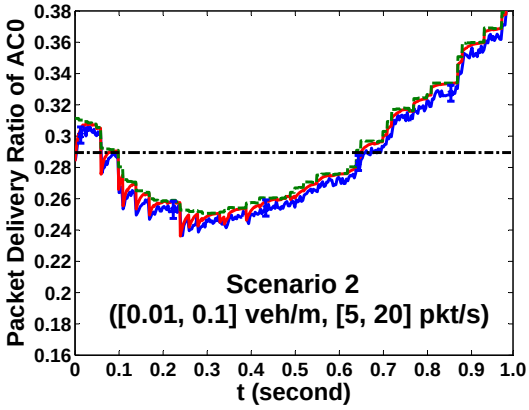
Figure 12: Collision probability for various scenarios.

vehicle from 5 pkt/s in scenario 2 to 10 pkt/s in scenario 3. Since the growth of transmission attempt of each vehicle inevitably increase the packet collision probability in network, both packet delay and packet delivery ratio of the tagged vehicle are degraded. As seen from Figure 10-14, both steady state model and pointwise stationary model cannot capture the transient behavior of the network. Instead, the fluid flow model can provide fairly accurate instantaneous results, all of which match well with the discrete event simulation results.

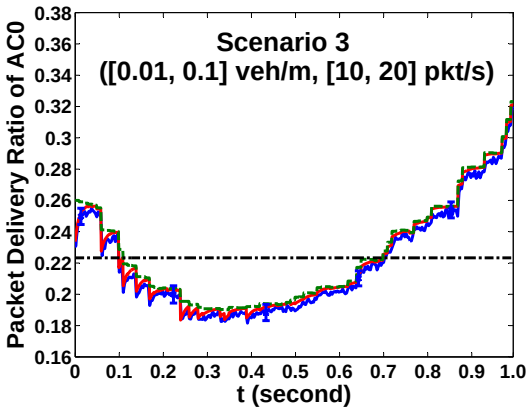
Finally, we compare the accuracy of steady state (SS) model, pointwise stationary (PS) model and fluid flow (FF) model by using the results obtained from nonstationary simulation as the benchmark. As listed in Table 8, the maximum deviation of all three models is calculated by using $Maximum\ Deviation = Max(\frac{|Simulation\ value - SS/PS/FF\ Model\ value|}{Simulation\ value} \times 100\%)$. Here, the *value* denotes the time-dependent result of the plots obtained by the corresponding technique. In the experiment, the network becomes more and more nonstationary from scenario 1 to 3 due to imbalanced vehicle density and high traffic load. Since the accuracy of SS model and PSA model depends greatly on the rate of time-varying changes, their maximum deviation of time dependent result can be as high as 49.24% and 156.39% respectively in scenario 3. Instead, the fluid flow model performs much better and remains the error of nonstationary response less than 5%.



(a) Scenario 1

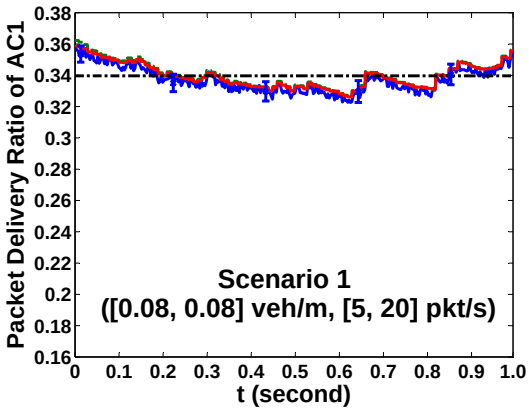


(b) Scenario 2

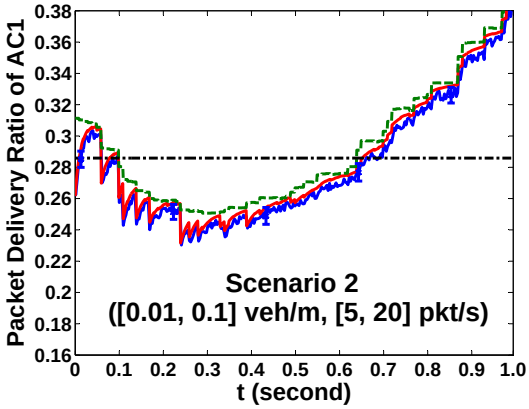


(c) Scenario 3

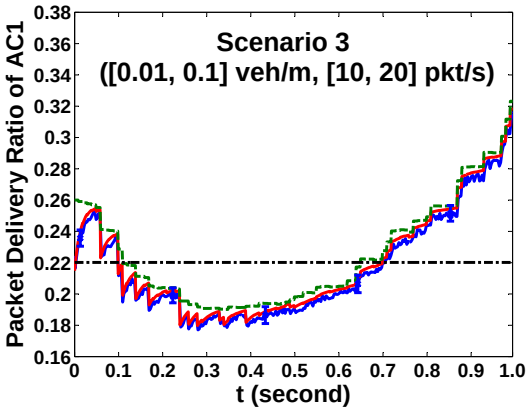
Figure 13: Delivery ratio of AC0 packets transmitted by the tagged vehicle for various scenarios of $([d^s, d^o], [\lambda^0, \lambda^1])$.



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Figure 14: Delivery ratio of AC1 packets transmitted by the tagged vehicle for various scenarios of $([d^s, d^o], [\lambda^0, \lambda^1])$.

Table 5: Accuracy Comparison of Various Modeling Techniques

Scenario No.	Performance Metric	Traffic Class	Max. Deviation (Sim vs. SS)	Max. Deviation (Sim vs. PS)	Max. Deviation (Sim vs. FF)
1	Delay	AC0	2.91%	1.92%	1.82%
1	Delay	AC1	4.82%	3.02 %	1.83 %
1	PDR	AC0	5.3%	1.89 %	1.72 %
1	PDR	AC1	5.78 %	2.5 %	1.73 %
2	Delay	AC0	89.66%	11.3%	1.82%
2	Delay	AC1	117.67%	17.35%	2.11%
2	PDR	AC0	25.55%	7.41%	2.55%
2	PDR	AC1	25.73%	12.07%	2.56%
3	Delay	AC0	83.71 %	11.48 %	1.82 %
3	Delay	AC1	112.35 %	18.98 %	2.13 %
3	PDR	AC0	29.38 %	9.04 %	2.54 %
3	PDR	AC1	29.78 %	12.61 %	2.58 %

3.7 COMPUTATION SCALABILITY

We first analyze the computation complexity of our fluid flow model according to the algorithm in Section 3.4. Initially, the first and second moments of packet service time are calculated based on the network hearing topology, as given by step 1 to 3. This computation process is time efficient and can be done off-time, thus this process is not counted in the following on-line computation complexity analysis. From step 4 to 7, various network performance are estimated by integrating a set of differential equations with a specific network hearing topology 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 [143]. 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 [143], $C(n, p, \alpha)$ 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 expression of $C(n, p, \alpha)$ in [143] 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 vehicular network with AC0 and AC1 traffic loads, the number of differential equations n equals to the number of traffic flows $2M$. Therefore, the computation time complexity of our fluid flow modeling algorithm is upper bounded by $O(M)$.

Then, we compare the computation time of the vehicular network fluid flow model with the nonstationary simulation on a sequence of sample networks. In the experiments, all vehicles broadcast both Poison traffic of AC0 as well as CBR traffic of AC1 with the rate of 5 pkt/s and 20 pkt/s, respectively. Vehicles' movement follows the freeway mobility model [142]. In Table 6, we list the computation time of the average number of packets at each vehicle in the sample networks over 100 seconds by both fluid flow model and simulation. All the computations are executed on a PC with Intel i5-450M 2.4GHz processor and 4GB

memory. The off-line computation time for the first and second moments of IEEE 802.11p packet service time is not counted in the reported time. To study the time varying behavior of the network via simulation, we perform the nonstationary simulation to average over an ensemble of statistically identical results generated by distinct independent runs with different random number seeds [17] [18]. 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 rises up dramatically with the network size. For the fluid flow model results, we curve fit the computation time data versus network size M ranging from 3 to 200 and obtain the growth rate as $\Theta(0.562M - 0.769)$, which is compliant with the expected upper bound discussed above.

In addition, the accuracy of fluid flow model is examined with the increase of network size by comparing with the results obtained from nonstationary simulation as the benchmark. The model accuracy is calculated by averaging time dependent deviation of the results between modelling and simulation. In the experiment, since the change of networking hearing topology triggers the variation of packet service time, the increase of network size inevitably raises up the frequency of such a variation and leads to higher level nonsationarity in the network. Table 6 shows that the fluid flow model can always provide fairly accurate performance evaluation with tolerable deviation from simulation results.

Table 6: Computation Time and Accuracy Comparison

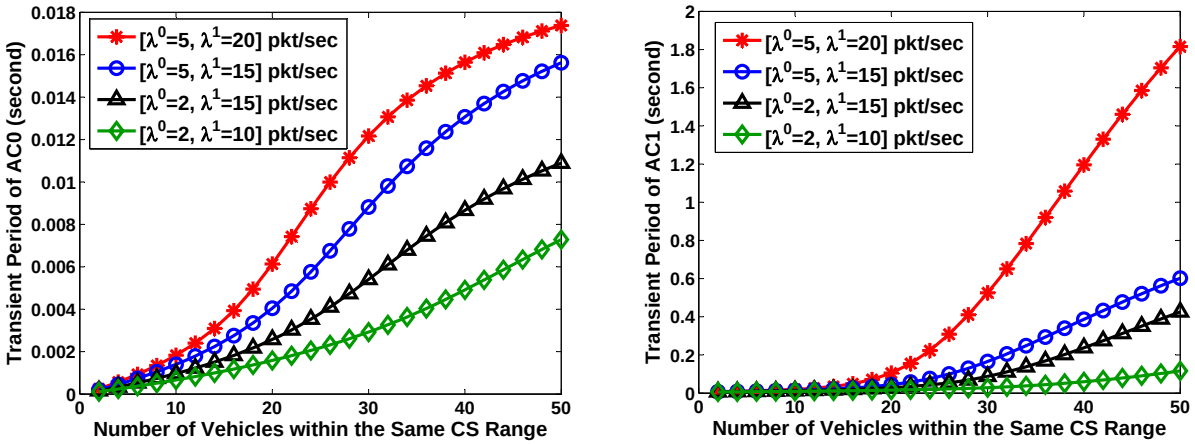
# of Nodes	# of Diff. Equations	Simulation (sec)	Fluid Flow Model (sec)	Model Accuracy
3	6	129.32	1.98	0.34%
5	10	284.97	2.16	0.84%
7	14	381.32	2.96	1.21%
9	18	593.94	4.01	1.53%
20	40	3095.28	8.95	1.97%
50	100	79682.41	26.73	2.83%

3.8 EVALUATION OF NONSTATIONARITY IN VEHICULAR NETWORKS

To evaluate the dynamics of the nonstationary queuing system in each vehicle, we can take advantage of our performance model to estimate the transient period, which is the time interval when system going from one stationary state to another. Hence, the transient period provides a measure of how much time does the system need to reach steady-state equilibrium after a perturbation. Several methods have been suggested for approximating the length of the transient period of a queuing system with an infinite buffer space. As a noteworthy work, Odoni and Roth [144] developed a closed-form approximate expression for the transient period θ that fits well with empirical results for extensive queuing systems

$$\theta \approx \frac{5.3\,(C_a^2 + C_s^2)}{2.8\mu\Big(1 - \sqrt{\rho}\Big)^2} \tag{3.24}$$

where the transient period θ is approximated by the time it takes for the system to stay within a 0.5% of the steady-state results. By using this formula, we compute the transient period of the transmission queue at each vehicle in a contention-based 802.11p vehicular network, as shown in Figure 15. We observe that the increase of packet arrival rate and the number of vehicles in the same carrier sensing (CS) range lead to the growth of transient period, since both of these two effects intensify the contention of channel access and deteriorate the packet service rate of each vehicle. Also, Figure 15 shows that the transient period of AC1 routine traffic is much longer than the one AC0 emergency traffic needs to reach the steady state. The reason is that AC1 routine traffic is offered at a high rate but served with low priority so that both its low service rate μ and its high utilization ρ make its transient period θ considerable, according to Equation (3.24). When the transient period is in the unit of one second or even more, the high-speed vehicle could move in/out of the carrier sensing range of the tagged vehicle during the transient period. As a result, the unsettled queue could be perturbed all the time by the dynamic hearing topology and remain nonstationary. Such a phenomenon can be observed in Figure 10-14, where the AC1 routine traffic shows longer transient period than AC0 emergency traffic once the hearing topology is updated.



(a) Queuing system for AC0 emergency traffic

(b) Queuing system for AC1 routine traffic

Figure 15: Estimation of initial transient period at various scenarios of $[\lambda^0, \lambda^1]$.

Furthermore, we study the nonstationarity of network performance impacted by the vehicle velocity as well as vehicle density at two opposite directions. It is intuitive to imagine that the transient/nonstationary effects on the tagged vehicle mainly comes from the high-speed movement of the vehicles in the opposite direction, since their relative speed with the tagged vehicle is much higher than the one with the vehicles in the same direction as the tagged one. In the experiments, we first set the homogeneous vehicle density with very low value (i.e. [0.01, 0.01] vehicles/meter) for both directions and the network performance is shown to be quite stationary, as shown by the bottom two curves in Figure 16. However, when the vehicle density becomes imbalanced, the percentage of transient period grows up, which indicates that the network performance becomes nonstationary. Meanwhile, the frequency of this variation depends on vehicle's speed, which we evaluate from 10 to 35 meters/second for all vehicles. In addition, we observe that the percentage of transient period of AC1 traffic is always higher than the one of AC0 traffic for all the scenarios due to AC1's high traffic load and low-priority service. For the imbalanced vehicle density with high-speed mobility, one would expect that transient period could dominate the network behavior in vehicular networks. For example, in the scenario of [0.1, 0.01] vehicles/meter, the percentage of transient period of AC1 routine traffic can rise up to 75% when the movement velocity

reach 35 meter/sec, as shown by the top curve in Figure 16.

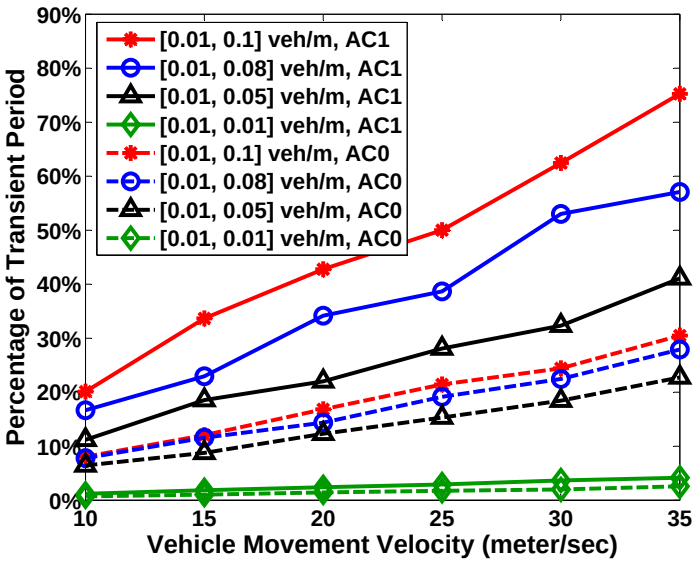


Figure 16: Percentage of transient period vs. vehicle velocity at various scenarios of $[d^s, d^o]$.

3.9 SUMMARY

In this chapter, we propose a fluid flow based performance model to study the time varying behavior of vehicular networks by using numerical method based queuing analysis. Numerical results, illustrating the application of our model, show that it can provide reasonably accurate performance results efficiently. We believe that the proposed performance modeling framework is a valuable tool to explore design alternatives of vehicular networks, and to get a quick estimate on the performance variation in response to some dynamic changes of network conditions.

4.0 TIME DEPENDENT PERFORMANCE ANALYSIS OF MULTIHOP WIRELESS NETWORKS

In multihop wireless network research, while experimentations have taken off during the last years, we are still relying on simulations for examining the performance of various protocols on large scale networks. Nevertheless, many of the existing simulation tools are known to be lacking scalability. Another weakness of most simulation studies of multihop wireless networks is that steady state analysis is used even though transient or nonstationary periods will occur often and likely dominant the network behavior.

In this chapter, we propose a time varying performance model for multihop wireless networks. We model the offered traffic at each node as a Poisson, CBR or general process to represent different types of traffic (e.g. data or video traffic). Our proposed performance model is a hybrid of time varying connectivity matrix and nonstationary network queues. Network connectivity is captured using stochastic modeling of adjacency matrix by considering node mobility. Nonstationary network queues are modeled using fluid flow based differential equations to approximate the time varying behavior. The notations adopted in this chapter are summarized as follows:

The outline of this chapter is as follows. Section 4.1 presents our modeling approach for time varying network topology. Section 4.2 provides the details of the node queueing model with nonstationary condition. These two components are combined together in Section 4.3 to model the multihop wireless networks. Section 4.4 codifies our performance modeling procedure into an executable and efficient algorithm. In Section 4.5, we validate our hybrid model by discrete event simulation with a series of numerical results. Section 4.6 studies the computational time complexity of our proposed modeling technique and gives numerical results illustrating the advantages of our method in comparison to simulation. Section 4.7

Table 7: Notation list

Notation	**Definition**
a_{ij}	Link connectivity between node i and j
C_i	Transmission link capacity of node i(bit/sec)
$D^{(s,d)}$	End-to-end delay of traffic flow from source s to destination d (sec)
ETE_{avg}	average end-to-end delay in network (sec)
G_i^j	Link utilization of class j traffic at node i
G_{avg}	Average link utilization in network
$P^{(s,d)}$	Set of all nodes on the path of stream (s,d) except destination node d
$f_{in\_i}^j$	Node i's incoming traffic flows of class j (pkt/s)
$f_{out\_i}^j$	Node i's outgoing traffic flows of class j (pkt/s)
L_i	Length of packet transmitted by node i
M	Number of nodes in a multihop wireless network
N	Number of input traffic flows into a queuing system
r_{ik}^j	routing indicator at node i routed to node k for class j traffic
R	Transmission range of a node (meter)
S	Number of traffic classes in a queuing system
T	Network throughput (bit/sec)
Q	Survival function of number of packets in a queuing system
W_i	Queueing delay at node i (sec)
$W^{(s,d)}$	Overall queueing delay along the path from source s to destination d (sec)
x_i^j	Number of class j packets in node i
λ_i^j	Arrival rate of class j traffic into node i (pkt/s)
μ_i	Packet service rate of node i (pkt/sec)
γ_i^j	Class j traffic generated by node i (pkt/s)
ε	Link propagation delay (sec)
δ_i	Forwarding delay at node i (sec)

illustrates the capability of our model to capture the network transient behavior due to node mobility and traffic load. Our summary in the future work are given in Section 4.8.

4.1 MODELING NETWORK CONNECTIVITY

The network topology in multihop wireless networks can change dynamically depending on the link connectivity between each node pair. When nodes in the network are allowed to move arbitrarily, it will lead to frequent changes in the topology of the queuing network model. In this section, we introduce the network topology modeling by using a time varying adjacency matrix.

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

$$A(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \dots & a_{1M}(t) \\ a_{21}(t) & a_{22}(t) & \dots & a_{2M}(t) \\ \vdots & \vdots & & \vdots \\ a_{M1}(t) & a_{M2}(t) & \dots & a_{MM}(t) \end{bmatrix} \tag{4.1}$$

where,

$$a_{ij}(t) = \begin{cases} 1, & \text{if node } i \text{ and } j \text{ are directly connected at time } t \ (i \neq j) \\ 0, & \text{otherwise} \end{cases}$$

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, the dynamic network topology can be reflected in the adjacency matrix by changing the value of $a_{ij}(t)$ with time. The topology change is based on a series of events, such as, links being broken and removed from the topology as two nodes move away from each other, or links being added to the topology when two nodes come into radio range. Information about the node movement and connectivity can be determined from experimentally gathered trace data, a discrete event simulation of a mobility model (e.g., random waypoint [64]) or stochastic/probabilistic models of mobility effects on link connectivity [70]. In the trace based approach, the data is mined for the link connectivity information versus time. In the simulation approach, a mobility model is used to create the network topology dynamics. Specifically, given a geographic space, a set of configured nodes and the propagation environment, every node pair is checked for the possible connectivity change based on their current speeds and directions. Note, that changes in speed, direction and power level are also considered events. The event times are placed in chronological order and as time evolves the pair-wise connectivity calculation is repeated for every event time and the matrix is changed accordingly. In this way, the adjacency matrix can reflect the topology change dynamically. A computationally simpler approach is to model mobility by directly manipulating the elements of the adjacency matrix according to a planned experiment (as illustrated in Figure 22-26) or a stochastic/probabilistic model (as shown in Figure 30-35). Note that, a probabilistic model (e.g., two-state MMPP [70]) can be developed either from the mobility model assumptions and analysis [64] or from fitting a statistical model to data gathered from a test bed or simulation.

4.2 NODE QUEUING MODEL

In multihop wireless networks, the traffic in the network is normally divided into a number of classes and the control actions (i.e. routing and flow control) are based on the class type. Hence, we now extend the single class fluid flow model in Equation (3.9)-(3.11) to model the

time varying behavior of the queue with multi-class input traffic at each network node. As illustrated in Figure 17, 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.

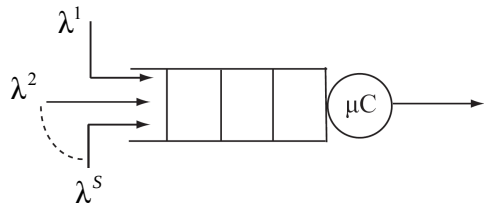


Figure 17: Queuing model with S classes of traffic.

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 $x(t) = \sum_{l=1}^{S} x^l(t)$. Then, the fluid flow model for the overall traffic is:

$$\dot{x}(t) = -\mu G(x(t)) + \lambda(t) \tag{4.2}$$

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.

$$\dot{x}^l(t) = -\mu G^l(x^l(t), x(t)) + \lambda^l(t) \qquad \forall l = 1, 2, \ldots, S \tag{4.3}$$

Thus, the multi-class queueing system can be described by a set of S coupled differential equations, each representing the traffic behavior of its own class. We have seen that the average server utilization of the fluid flow model depends on stochastic modeling assumptions of the queue under study. Therefore, various queue types with different traffic arrival process will lead to separate fluid models.

4.2.1 Modeling the Queue with Poisson Traffic

Now we consider a multihop wireless network consisting of nodes with deterministic service time and Poisson traffic load [127]. Hence, each node is modeled as an $M/D/1$ queue. From queueing theory, the average packet number in the system at steady state is given by $x(t) = \rho + \rho^2/(2(1-\rho))$, where $\rho = G(x(t)) = \lambda(t)/\mu$. We match the steady-state equilibrium point of the fluid flow model with the $M/D/1$ model to obtain the utilization function $G(x(t))$. Under steady state conditions (i.e. $\dot{x}(t) = 0$), the state model turns out to be:

$$\dot{x}(t) = -\mu(x(t) + 1 - \sqrt{x(t)^2 + 1}) + \lambda(t) \tag{4.4}$$

For the queue with S classes of traffic, $x^l(t)$, and $\lambda^l(t)$ represents the *ensemble average* number of packets and the arrival rate of class l traffic, while $x = \sum_{l=1}^{S} x_i$ and $\lambda = \sum_{l=1}^{S} \lambda_i$ denote the total *ensemble average* number of packets and the mean aggregate arrival rate into the system. Since the flow conservation principle still applies to each traffic class, a fluid flow model can also be developed for each class with $G(x^l(t), x(t))$ as the average utilization function of class l traffic in the multiclass queue.

$$\dot{x}^l(t) = -\mu G^l(x^l(t), x(t)) + \lambda^l(t) \quad \forall l = 1, 2, \ldots, S \tag{4.5}$$

Note that at steady state, the average number of total packets in the $M/D/1$ queuing system is

$$x(t) = \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2(1 - \frac{\lambda}{\mu})} \tag{4.6}$$

From multiclass queuing theory [49], we can write the steady-state average number of packets of class l traffic $x^l(t)$ as

$$x^l(t) \quad = \quad \frac{\lambda^l(2\mu - \lambda)}{2\mu(\mu - \lambda)} \tag{4.7}$$

Following the approach of steady state equilibrium matching with $\dot{x}(t) = 0$ and $\dot{x}^l(t) = 0$, we get $\lambda(t) = \mu(x(t)+1-\sqrt{x^2+1})$ and $\lambda^l(t) = \mu G^l(x^l(t), x(t))$ from (4.4) and (4.5), respectively.

Solving these two equations along with (4.13), we obtain the utilization function for class l traffic $G^l(x^l(t), x(t))$ as:

$$G^l(x^l(t), x(t)) = \frac{2x^l(t)(\sqrt{x^2(t)+1} - x(t))}{\sqrt{x^2(t)+1} - (x(t)-1)} \tag{4.8}$$

Substituting (4.14) back into (4.5), we get:

$$\dot{x}^l(t) = -\mu[\frac{2x^l(t)(\sqrt{x^2(t)+1} - x(t))}{\sqrt{x^2(t)+1} - (x(t)-1)}] + \lambda^l(t)$$
$$\forall l = 1, 2, \ldots, S \tag{4.9}$$

As a result, a node can be represented by a set of S nonlinear equations of the form of Equation (4.15) describing the queue length dynamics of each class separately. The multiclass fluid flow model developed here represents the dynamics of a single node with Poisson input traffic. Next we develop the time varying queuing model by considering CBR traffic load.

4.2.2 Modeling the Queue with CBR Traffic

Generally, 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, and thus the one-hop packet transmission (service) time is assumed to be deterministic here. In multihop wireless networks, one-hop deterministic packet transmission (service) times can be achieved by contention-free transmission with the technique of distributed scheduling [145] or service differentiation [146]. Deterministic packet transmission (service) times could also approximately occur in contention-based transmission networks which are sparse or lightly loaded. For example, energy conserving techniques in WSNs put most of the nodes in sleep modes so that the network becomes sparse. Moreover, each sensor node only has light traffic to transmit in order to save energy. 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 4.2.2.2 and Section 4.1, respectively. A detailed justification of CBR traffic is provided in [128]. In the following, we focus on modeling a queue with CBR traffic streams and consider two cases: (1) where all CBR streams have the same data rate and (2) when a group of heterogeneous Quasi-CBR (QCBR) traffic streams with different data rates and delay jitters are multiplexed. The two cases are studied in turn below.

4.2.2.1 Case I: Identical Sources Following [49], 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. $s \leq [D]$, the integer part of D). As noted in [49], the survival function can be written as:

$$Q(r) = \sum_{s=1}^{[D]} p_s(r)\pi_0(r, s) \tag{4.10}$$

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 $Q(r)$ [49] can be written as:

$$Q(r) = \sum_{s=1}^{N-r} \left[\binom{S}{r+s} \left(\frac{s}{D}\right)^{r+s} \left(1 - \frac{s}{D}\right)^{N-r-s} \left(\frac{D-N+r}{D-s}\right) \right] \quad \text{for } 0 \leq r < N \tag{4.11}$$

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 $Q(r)$, $x = \sum_{r=0}^{N-1} Q(r)$ [147]. Therefore, for the $N * D/D/1$ queue, x is given by:

$$x = \sum_{r=0}^{N-1} \sum_{s=1}^{N-r} \left[\binom{N}{r+s} \left(\frac{s}{D}\right)^{r+s} \left(1 - \frac{s}{D}\right)^{N-r-s} \left(\frac{D-N+r}{D-s}\right) \right] \quad \text{for } 0 \leq r < N \tag{4.12}$$

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 (4.12) 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 (4.2). 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 (4.2) and (4.3) respectively, results in

$$\lambda(t) = \mu G(x(t)) \tag{4.13}$$

$$\lambda^l(t) = \mu G^l(x^l(t), x(t)) \tag{4.14}$$

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

$$G^l(x^l(t), x(t)) = \frac{\lambda^l(t)}{\lambda(t)} G(x(t)) \tag{4.15}$$

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 [49], 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

$$\frac{\lambda^l(t)}{\lambda(t)} = \frac{x^l(t)}{x(t)} \tag{4.16}$$

After substituting Equation (4.16) in (4.15), $G^l(x^l(t), x(t))$ can be finally determined as

$$\begin{aligned} G^l(x^l(t), x(t)) &= \frac{x^l(t)}{x(t)} G(x(t)) \\ &= \frac{x^l(t)}{x(t)} \left[ax^n(t) + bx^{n-1}(t) + ... + k\right] \\ &\qquad \forall l = 1, 2, \ldots, S \end{aligned} \tag{4.17}$$

The resulting $G^l(x^l(t), x(t))$ can be substituted into (4.3) to provide the multi-class traffic fluid flow model. Notice that S represents the number of traffic classes in the queue in Equation (4.3) and (4.20), while N in Equation (4.11) and (4.12) 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 $N \geq S$.

4.2.2.2 Case II: Non-identical Sources Consider the case where a group of heterogeneous Quasi-CBR (QCBR) traffic streams with different data rates and delay jitters 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 (4.12) 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 class using (4.20) and then substitute it back into the fluid flow model (3.11).

4.2.2.3 Approximating the Queue with a Large Number of Input CBR Streams

Since the implementation of Equation (4.11) 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 desired to reduce the computation complexity when the transmission link carries a large number of CBR streams.

In case I, when the traffic load in queuing system has $\rho = N/D < 0.9$, an arrival process of $N * D/D/1$ consisting of a superposition of a large number of periodic processes tends to a Poisson arrival process, thus the $M/D/1$ approximation works reasonable well for engineering purpose [49]. 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 [115] (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 any more [49]. Hence, we derive an approximation of the utilization function in Appendix . The approximation accuracy of Equation (.7) is evaluated in Figure 18 by comparing the simulation results of a single queue as well as the exact results generated by Equation (4.12) in the different cases of period D. As shown in the figure, by solving Equation (.7), 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 Equation (4.20).

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

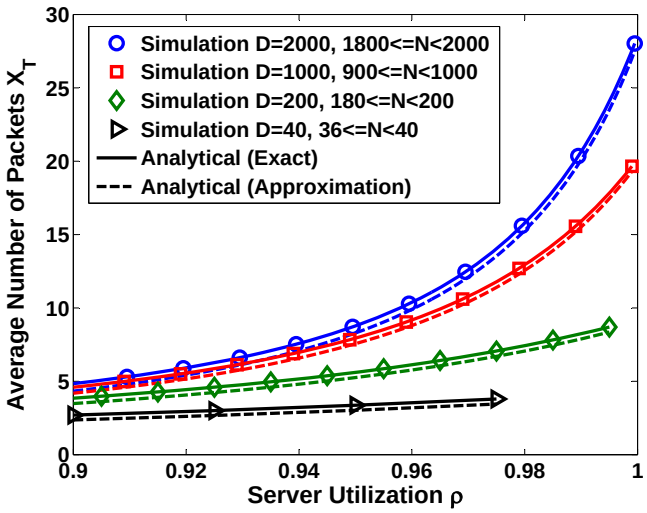


Figure 18: Comparison of the approximation by (.7) with the simulation results as well as the exact analytical results by (4.12) during heavy load regime $0.9 \leq \rho < 1$.

well as the average traffic period D_{avg}, according to Section 4.2.2.2. Depending on the server utilization ρ, we then apply the $M/D/1$ or Equation (.7) approximation 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.

4.2.3 Modeling the Queue with General Arrival and General Service Processes

We now model a queuing system with general arrival and service processes, i.e. $G/G/1$ queue. Such a model is quite general without considering the details of packet arrival and service processes (e.g. network protocol, wireless link characteristics and packet size distribution), as long as the first and second moments of both processes at each node are known. A well-known approximation to the expected number in the $G/G/1$ queuing system was developed by Kramer and Lagbenbach-Belz [133].

$$x \approx \rho + \frac{\rho^2 \cdot (c_a^2 + c_s^2) \cdot g(c_a^2, c_s^2, \rho)}{2(1-\rho)} \tag{4.18}$$

where

$$g(c_a^2, c_s^2, \rho) = \begin{cases} e^{-\frac{2(1-\rho)(1-c_a^2)}{3\rho(c_a^2+c_s^2)}} & c_a^2 \leq 1 \\ e^{-\frac{(1-\rho)(c_a^2-1)}{c_a^2+4c_s^2}} & c_a^2 \geq 1 \end{cases} \tag{4.19}$$

where c_a^2 and c_s^2 represent the squared coefficient of variation of the arrival and service process, respectively, and $\rho = \lambda/\mu$ is the server utilization. Since it is difficult to analytically invert the KLB equation (4.18)-(4.19) to get a closed form utilization function $\rho = G(x)$, we numerically determine x for a given parameter set (i.e., c_a^2, c_s^2, ρ). Then, we apply a curve fitting approach using the (ρ, x) data from the KLB equation to find the utilization function $G(x(t))$ in the form of a polynomial (i.e. $G(x(t)) = ax^n + bx^{n-1} + ... + k$).

For a multi-class queuing system, we determine the utilization function of class l at a queue by the approach of steady state equilibrium matching with $\dot{x}(t) = 0$ and $\dot{x}^l(t) = 0$. Then, substituting them in Equation (4.2) and Equation (4.3) respectively, results in $\lambda(t) = \mu G(x(t))$ and $\lambda^l(t) = \mu G^l(x^l(t), x(t))$. By combining these two equations, we obtain the utilization function of class l traffic as $G^l(x^l(t), x(t)) = \frac{\lambda^l(t)}{\lambda(t)} G(x(t))$. Following the same derivation discussed in Section 4.2.2.1, we can write $\lambda^l(t)/\lambda(t) = x^l(t)/x(t)$. Therefore, $G^l(x^l(t), x(t))$ can be determined as

$$G^l(x^l(t), x(t)) = \frac{x^l(t)}{x(t)} G(x(t)) \qquad \forall l = 1, 2, \ldots, S \tag{4.20}$$

The resulting $G^l(x^l(t), x(t))$ can be substituted into (4.3) to construct the multi-class traffic fluid flow model.

4.3 MODELING DYNAMIC BEHAVIOR OF MULTIHOP WIRELESS NETWORKS

4.3.1 Fluid Flow Model for multihop Wireless Networks

Consider a network consisting of M nodes, an arbitrary node i is shown in Figure 19. 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 traffic

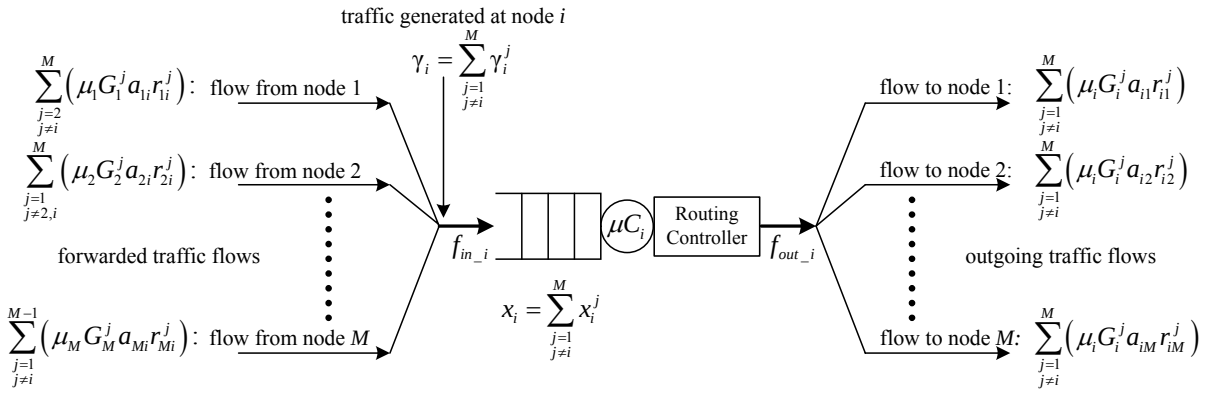


Figure 19: An arbitrary node i queuing model.

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 denote the packet length as L_i and the transmission capacity of node i as C_i, then the service rate of node i is represented by $\mu_i = C_i/L_i$. When considering the network as a whole, we must modify (3.11) 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 4.1. 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 Figure 19, 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 (3.11) to incorporate $a_{ik}(t)$ and $r_{ik}^j(t)$,

resulting in

$$f^j_{out\_i}(t) = \mu_i G^j_i(x^j_i(t), x_i(t)) \sum_{\substack{k=1 \\ k\neq i}}^{M} a_{ik}(t) r^j_{ik}(t) \tag{4.21}$$

The incoming traffic rate $f^j_{in\_i}$ at node i destined for node j consists of traffic generated at node i with rate $\gamma^j_i(t)$ as well as forwarded traffic flow from the neighboring node l, where $l = 1, 2, ..., M$ and $l \neq i, j$, as shown in Figure 19. By considering link connectivity as well as routing, we then have

$$f^j_{in\_i}(t) = \gamma^j_i(t) + \sum_{\substack{l=1 \\ l\neq i,j}}^{M} \left(\mu_l G^j_l(x^j_l(t), x_l(t)) a_{li}(t) r^j_{li}(t) \right) \tag{4.22}$$

To interconnect queues, the literature [148] 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 Figure 20(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 $\mu G_1(t)$ is the departure rate from the first queue. The departure rate then becomes the input to the second queue after a deterministic forwarding delay δ_1 in the first queue, that is $\lambda_2(t) = \mu G_1(t - \delta_1) + \gamma_2(t)$. We can then write a set of fluid flow equations at node 1 and node 2 for Figure 20 as:

$$\dot{x}_1(t) = -\mu G_1(t) + \gamma_1(t) \tag{4.23}$$

$$\dot{x}_2(t) = -\mu G_2(t) + \gamma_2(t) + \mu G_1(t - \delta_1) \tag{4.24}$$

To interconnect the $G/G/1$ queuing model of each node, we make use of the approximation approach proposed by Whitt in [149]. The arrival and departure processes are determined based on three basic network operations: decomposition (splitting), flow through a queue (serving), and superposition (merging), as seen in Figure 21. For the splitting network operation, consider a departure stream with the mean rate d and the squared coefficient of

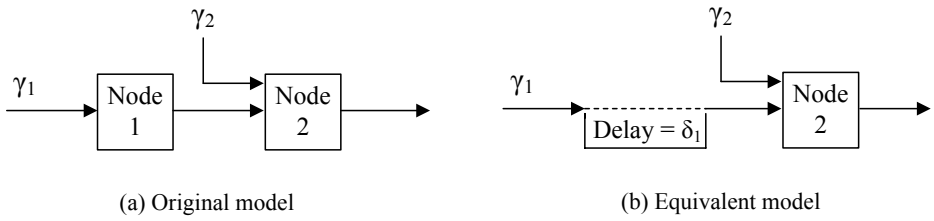


Figure 20: A two-node deterministic service system with its equivalent model.

variation c_d^2 is split into N streams, with each selected independently according to probabilities p_i (i.e. $d = \sum_{i=1}^{N} d_i$ and $p_i = d_i/d$, where $i = 1, 2, \cdots, N$). Then the ith stream obtained from the splitting has the squared coefficient of variation $c_{a\_i}^2$ given by

$$c_{a\_i}^2 = p_i c_d^2 + 1 - p_i \tag{4.25}$$

The squared coefficient of variation of an inter-departure c_d^2 for a single server node is approximated by

$$c_d^2 = \rho^2 c_s^2 + (1 - \rho^2) c_a^2 \tag{4.26}$$

For the merging of N input streams and self-generated stream (i.e. $\lambda = \sum_{i=1}^{N} d_i + \gamma_i$), we can obtain the coefficient of variation of an inter-arrival time c_a^2 by

$$c_a^2 = \frac{1}{\lambda} \left(\sum_{i=1}^{N} d_i c_{d\_i}^2 + \gamma_i c_{\gamma\_i}^2 \right) \tag{4.27}$$

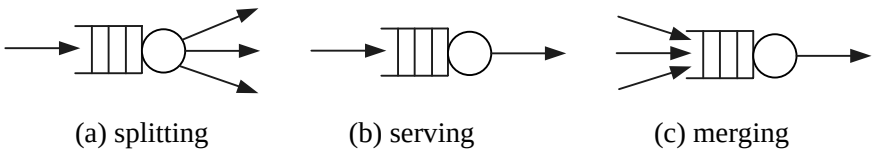


Figure 21: Basic network operations

According to [149], for a FIFO $G/G/1$ queuing system holding N streams, the mean service rate and the squared coefficient variation of service time are

$$\mu = \sum_{i=1}^{N} \left(\frac{\lambda_i}{\lambda} \mu_i \right) \tag{4.28}$$

$$c_s^2 = \sum_{l=1}^{N} \left(\frac{\lambda_i}{\lambda} c_{s\_i}^2 \right) \tag{4.29}$$

Now, having the first and second moments of both arrival and service processes, we are able to determine the utilization function $G(.)$ according to Equation (4.18)-(4.20).

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

$$\begin{aligned} \dot{x}_i^j(t) = & -\mu_i G_i^j(x_i^j(t), x_i(t)) \sum_{\substack{k=1 \\ k \neq i}}^{M} a_{ik}(t) r_{ik}^j(t) + \gamma_i^j(t) \\ & + \sum_{\substack{l=1 \\ l \neq i,j}}^{M} \left(\mu_l G_l^j(x_l^j(t-\delta_l), x_l(t-\delta_l)) a_{li}(t) r_{li}^j(t) \right) \\ & \forall i, j = 1, 2, \ldots, M \end{aligned} \tag{4.30}$$

In (4.30), 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 Poisson input traffic and the deterministic service rate (i.e. $M/D/1$ queue), the service utilization function $G(\cdot)$ can be written by Equation (4.14). Thus, the network based fluid flow model is

$$\begin{aligned} \dot{x}_i^j(t) = & -\mu_i \frac{2x_i^j(t)(\sqrt{x_i^2(t)+1} - x_i(t))}{\sqrt{x_i^2(t)+1} - (x_i(t)-1)} \sum_{\substack{k=1 \\ k \neq l}}^{M} a_{ik}(t) r_{ik}^j(t) + \gamma_i^j(t) \\ & + \sum_{\substack{l=1 \\ l \neq i,j}}^{M} \left(\mu_l \frac{2x_l^j(t-\delta_l)(\sqrt{x_l^2(t-\delta_l)+1} - x_l(t-\delta_l))}{\sqrt{x_l^2(t-\delta_l)+1} - (x_l(t-\delta_l)-1)} (a_{li}(t) r_{li}^j(t)) \right) \\ & \forall i, j = 1, 2, \ldots, M \end{aligned} \tag{4.31}$$

Also, the utilization function of a queue with CBR traffic and deterministic service rate (i.e. $D/D/1$ queue) can be written in the form of polynomial expression given in Equation (4.20). Therefore, the fluid flow model for the network with CBR traffic is

$$\dot{x}_i^j(t) = -\mu_i \frac{x_i^j(t)}{x_i(t)} \left[ax_i^n(t) + bx_i^{n-1}(t) + ... + k\right] \times \sum_{\substack{k=1 \\ k\neq i}}^{M} a_{ik}(t) r_{ik}^j(t) + \gamma_i^j(t)$$
$$+ \sum_{\substack{l=1 \\ l\neq i,j}}^{M} \left(\mu_l \frac{x_l^j(t-\delta_l)}{x_l(t-\delta_l)} \times \left[ax_l^n(t-\delta_l) + bx_l^{n-1}(t-\delta_l) + ... + k\right] a_{li}(t) r_{li}^j(t) \right)$$
$$\forall i, j = 1, 2, \ldots, M \qquad (4.32)$$

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

4.3.2 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 the 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

$$\dot{W}^{(s,d)}(t) = \sum_{i \in P^{(s,d)}} \dot{W}_i = \sum_{i \in P^{(s,d)}} \frac{\dot{x}_i(t)}{\lambda_i(t)} \tag{4.33}$$

where $\lambda_i(t)$ is the total arrival rate into node i at time t,(i.e. $\lambda_i(t) = \sum_{\substack{d=1 \\ d \neq i}}^{M} f^d_{in\_i}(t)$, with $f^d_{in\_i}(t)$ determined by (4.22)). We denote ε as the link propagation delay, which is assumed to be fixed and 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 (4.33), we can finally write the end-to-end delay of path $P^{(s,d)}$ at time t as

$$D^{(s,d)}(t) = \int \dot{W}^{(s,d)}(t)dt + |P^{(s,d)}|\varepsilon \tag{4.34}$$

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.

$$x_{avg}(t) = \frac{1}{M} \sum_{i=1}^{M} x_i(t) = \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} x_i^j(t) \tag{4.35}$$

The average end-to-end delay per traffic flow 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 in network at time t is given by

$$ETE_{avg}(t) = \frac{1}{\upsilon} \sum_{s=1}^{M} \sum_{\substack{d=1 \\ d \neq s}}^{M} D^{(s,d)}(t) \tag{4.36}$$

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

$$G_{avg}(t) = \frac{1}{M} \sum_{i=1}^{M} G_i(t) = \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} G_i^j(t) \tag{4.37}$$

Also, one can determine the instantaneous network throughput in bit per second (bps) as

$$T(t) = \sum_{d=1}^{M} \sum_{\substack{i=1 \\ i \neq d}}^{M} \Big(\mu_i L_i G_i^d(x_i^d(t), x_i(t)) a_{id}(t) r_{id}^d(t) \Big) \tag{4.38}$$

where L_i is node i's the packet length and $\mu_i L_i$ represents the service rate in the unit of bps. 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 $i \neq d$).

To estimate the energy consumption of a MANET node, we first construct the energy model of data packet communications. Here, we do not consider the energy consumption of baseline idle state or route discovery/maintenance. In addition, the energy cost of discarding packet is ignored, since non-destination nodes generally employ energy-conserving strategy based on the presence of uninteresting data on the media. In [150], experimental results confirm the accuracy of the linear energy model for IEEE 802.11 RTS/CTS point-to-point communication, which consists of fixed costs associated with channel acquisition and an incremental cost proportional to the size of data packet:

$$E_{TX} = b_{send-rts} + b_{recv-cts} + m_{send} \times size + b_{send} + b_{recv-ack} \tag{4.39}$$

$$E_{RX} = b_{recv-rts} + b_{send-cts} + m_{recv} \times size + b_{recv} + b_{send-ack} \tag{4.40}$$

In practice, the RTS packet may be retransmitted due to packet collision with probability p, and thus $b_{send-rts}$ should be multiplied by $1/(1-p)$ to account for retransmission. Following the similar approach in [151] [152], the power consumption for data packet communications at node i is finally given by

$$E_i(t) = E_{TX} \sum_{j=1, j\neq i}^{M} f^j_{out\_i}(t) + E_{RX} \left(\sum_{j=1, j\neq i}^{M} f^j_{fwd\_i}(t) + f^i_{des\_i}(t) \right) \tag{4.41}$$

where $f^j_{out\_i}(t)$ and $f^j_{fwd\_i}(t)$ are denoted by the first and last terms in the general fluid flow model (4.30). $f^i_{des\_i}(t)$ is the traffic received by node i as the final destination. This portion of traffic is not routed into the transmission queue of node i and thus not considered in the fluid flow model. Instead, $f^i_{des\_i}(t)$ can be calculated by summing up the traffic sent from all its neighboring nodes and destined for node i (e.g. $f^i_{des\_i}(t) = \sum_{l=1, l\neq i}^{M} \mu_l G^i_l(\cdot) a_{li}(t) r^i_{li}(t)$).

4.4 PERFORMANCE MODELING ALGORITHM

We summarize the hybrid model numerical solution procedure with the following algorithm to estimate the time dependent performance metrics over the desired time interval $[t_0, t_f]$.

1. Configure network parameters including the link capacity C (bps), the packet length L (bit), the link service rate μ packets per second (i.e. $\mu = C/L$), the offered traffic rate γ (pkt/s) and arrival period D in the unit of service period (i.e. $D = \mu/\gamma$).
2. Compute the data pair (ρ, x) off-line. For the Poisson input traffic, apply Equation (4.10). For the CBR input traffic, use Equation (4.12) if the number of input traffic streams N is small (e.g. $N <= 30$). Otherwise, approximate the data pair (ρ, x) as discussed in Section 4.2.2.3.
3. Find the utilization functions off-line. For the Poisson input traffic, apply Equation (4.14). For the CBR input traffic, curve fit the data pair (ρ, x) to obtain the polynomial in the form of Equation (4.20).
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 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 in the form of differential equations (4.30) and get the new $x_i^j(t + \Delta t)$ at the end of the time interval $[t, t + \Delta t]$, which then becomes the initial condition for the next time interval $[t + \Delta t, t + 2\Delta t]$.
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 (4.33) and (4.34). Here, $W^{(s,d)}(t + \Delta t)$ is obtained by numerically solving the differential equations 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 (4.35)-(4.38).
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 (4.30), (4.31) and (4.32). However, the *Runge-Kutta* algorithm is one of the widely used methods and the results in the next section show that outcomes using the fourth or the fifth order are reasonably accurate for the cases considered.

4.5 MODEL VALIDATION BY DISCRETE EVENT SIMULATION

In this section, our model is validated by comparing with an equivalent discrete event simulation model built in OPNET [11]. In the OPNET simulation model, each queue of the node is configured as a FIFO queue with infinite size buffer and each traffic class is buffered at the corresponding subqueue. The discrete event simulation results are the ensemble average of 5000 replications with 98% confidence intervals using the nonstationary simulation methodology discussed in [18].

A simple simulation scenario of three nodes with pre-determined connectivity change between nodes as illustrated in Figure 22(a)-(f), is studied here. This topology is used to evaluate the accuracy of our proposed model. In this setup, when the direct link is no longer available, traffic must be rerouted through relay nodes and uses some available portion of the shared link capacity. We set the link capacity for all nodes $C_i = 10^4$ bps with the packet length $1/\mu = 1250$ bytes, so that the average service rate is normalized to one packet per second. The propagation delay of each link is assumed to be 0.1 microseconds. In addition, we use minimum hop proactive routing to find one single path for each traffic flow in both the fluid flow model and the discrete event simulation. The three node network of Figure 22 has the corresponding queuing model as shown in Figure 23.

We first study the network with Poisson traffic load on each node. The traffic arrival rates are configured to be $\gamma_1^2 = 0.16, \gamma_1^3 = 0.2, \gamma_2^1 = 0.16, \gamma_2^3 = 0.2, \gamma_3^1 = 0.16, \gamma_3^2 = 0.2$ packets per second. Figure 24 shows the effect of the topology changes on the average number of packets and the end-to-end delay for the traffic at node 1 destined for node 2, as computed via our hybrid model and the nonstationary simulation. For the time interval $t < 100$ sec, the network is fully connected. All nodes go through an initial transient period and then

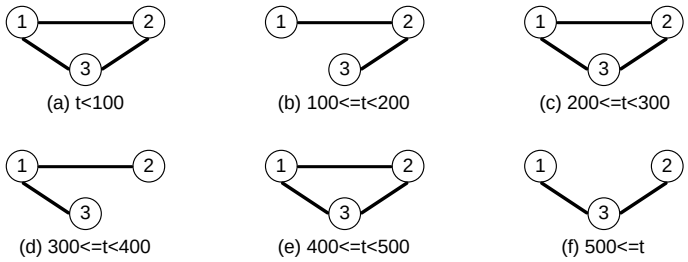


Figure 22: Three node network connectivity scenario.

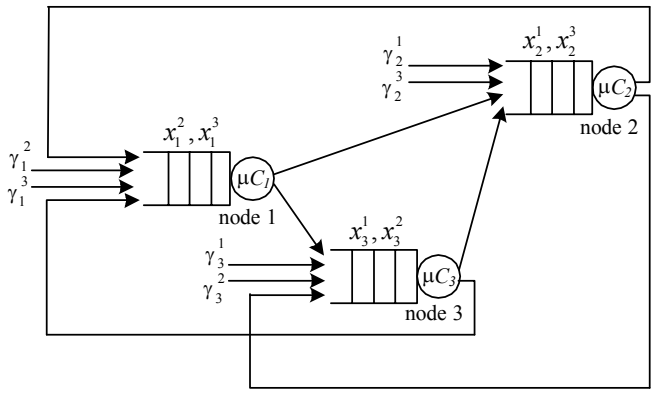
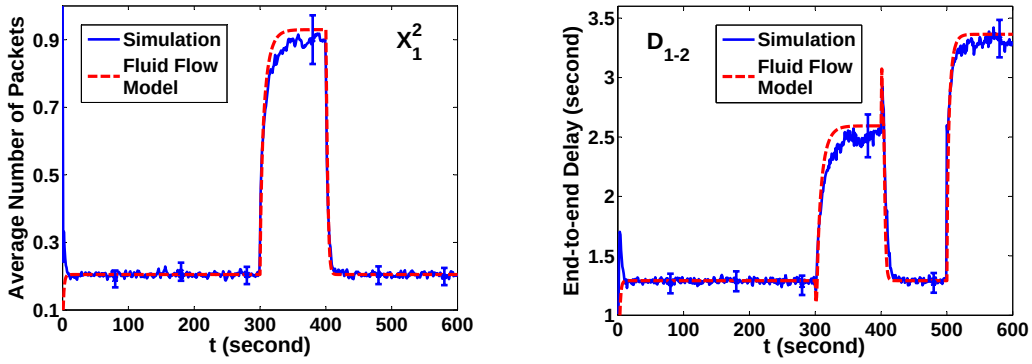


Figure 23: Three node network queuing model.

reach the steady state. For time $100 \leq t < 200$ sec, the link between node 1 and 3 breaks, so that the traffic between 1 and 3 has to be re-routed through the relay node 2. But the packets x_1^2 buffered in source node 1 are not affected. Then, the broken link 1-3 is restored during the time interval $200 \leq t < 300$ sec, and all nodes return to steady state. During the time interval $300 \leq t < 400$ sec, the link between node 2 and 3 breaks, leading to the traffic from node 2 and 3 re-routed through node 1. Thereby, more packets x_1^2 are buffered in node 1. Due to higher server utilization, the queueing delay in node 1 increases and the corresponding end-to-end delay D_{1-2} from node 1 destined for node 2 increases as well. Starting from time $t = 500$, the link between node 1 and 2 is broken and the traffic in x_1^2 has to go through the relay node 3 to reach the destination. Thus, D_{1-2} experiences the delay of two hops including the propagation delay of link 1-3 and 3-2 as well as the queuing delay at node 1 and 3. The behavior of other nodes and traffic streams are similar and not shown here for the purpose of brevity.



(a) (b)

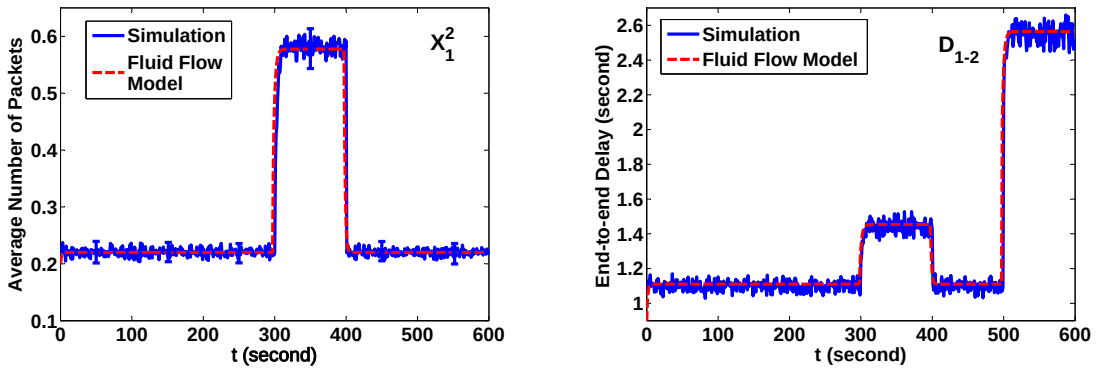
Figure 24: Dynamic behavior of the traffic destined for node 2 at node 1 buffer.

We then consider the network with CBR traffic and fouce on case I denoted by $N*D/D/1$ queue. The rate of externally arrival traffic stream of each node is set to be the same, 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., the first packet arrival time of each stream is a uniformly distributed random variable over $[0, 5]$ second). According to Section 4.2.2.1 case I, we first compute the data pair (x, ρ) from

(4.12). After curve fitting the data pair (ρ, x), the server utilization function is determined as $G(x) = 0.0832x^3 - 0.4353x^2 + 1.0843x$, which is then used in (4.32) to model the network. It is worthy to note that there exists a trade-off between the order of fitting accuracy and the fluid-flow model solving time. For utilization function, higher order polynomial can better fit the data set but make the fluid flow model more complex to solve. Hence, we gradually increase the polynomial fitting order until the averaged fitting error is within 10^{-5} of the data set. As we can see from Figure 25, the results show similar behavior in the sense that when direct link between two nodes is not available, traffic needs to be rerouted through relay nodes causing the large increase in the number of packets and queueing delay at those node buffers during such time interval until the broken link is restored.

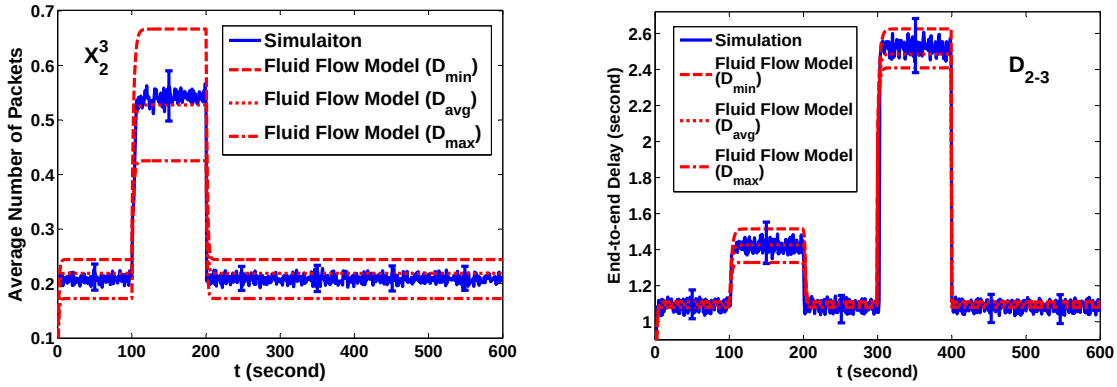
Finally, for case II referred to as $\sum_{i=1}^{m} N_i D_i/D/1$, the rates of externally arrival traffic streams are set as: $\gamma_1^2 = 0.16, \gamma_1^3 = 0.17, \gamma_2^1 = 0.18, \gamma_2^3 = 0.2, \gamma_3^1 = 0.22, \gamma_3^2 = 0.21$ packets per second. Following the "homogeneous traffic" approach described in Section 4.2.2.2 case II, we have $D_{max} = \max\{D_i\} = 1/0.22 \approx 4.545$s, $D_{min} = \min\{D_i\} = 1/0.16 = 6.25$s and $D_{avg} = \text{avg}\{D_i\} = (1/0.16 + 1/0.17 + 1/0.18 + 1/0.2 + 1/0.22 + 1/0.21)/6 \approx 5.333$s for "lower and upper bounds" as well as "average period approximation", respectively. After curve fitting the data pair (ρ, x), we obtain the utilization function for each case as $G_{lower}(x) = 0.0889x^3 - 0.4496x^2 + 1.0706x$, $G_{upper}(x) = 0.0755x^3 - 0.4227x^2 + 1.0894x$ and $G_{avg}(x) = 0.0849x^3 - 0.4395x^2 + 1.0801x$. Then we apply the utilization functions into (4.32) to form the hybrid models of the network. Figure 26 shows the average number of packets and the end-to-end delay of the traffic at node 2 destined for node 3 due to topology change. In Figure 26, we notice that when the direct link breaks, a wider gap occurs between upper and lower bounds. The reason is that traffic rerouting increases the server utilization of the relay nodes, and then the gap between bounds is enlarged at high utilization. The growth of utilization also leads to the increase of the relative precision of the confidence interval. We also observe that the D_{avg} approximation is quite accurate in this case.

As an example of using the general $G/G/1$ fluid flow model (4.30), we consider phase type distribution as an example to model more general packet arrival and service processes. In wireless networks, the coefficient variation of packet transmission time can be less than one, since the length of packet payload is exponentially distributed but the length of packet head-



(a) (b)

Figure 25: Average number of packets x_1^2 and end-to-end delay of D_{1-2} in the case of $N * D/D/1$ queue.



(a) (b)

Figure 26: Average number of packets x_2^3 and end-to-end delay of D_{2-3} in the case of $\sum_{i=1}^{m} N_i D_i/D/1$ queue.

er, inter-frame space and ACK are all generally fixed. In this case, the packet service time can be approximated by Erlangian distributions (a simple version of PH-type distribution), where the squared coefficient of variation is $1/k$ (i.e. $c_s^2 = 1/k$). The packet interarrival time is also assumed to follow Erlangian distribution (i.e. $c_a^2 = 1/k$). By assuming $k = 2$ in this numerical example, we can determine the first and second moments of both arrival and service processes from Eq.(4.25) to Eq.(4.29). Then, the data pair (ρ, x) obtained from $G/G/1$ formula (4.18)-(4.19) is used to fit the polynomial utilization function $G(x)$ (4.20). With the resulting $G(x) = -0.0002x^6 + 0.0029x^5 - 0.0263x^4 + 0.1445x^3 - 0.4708 + x^2 0.8857x + 0.0062$, we can construct the fluid flow model (4.30) and then evaluate the instantaneous power consumption of node 1 and 3 as an example, according to Eq.(4.41). In this experiment, the packet collision probability p is assumed to be 0.01 for all links, and the parameters used in our energy model (4.39)-(4.40) are assigned with the values from Lucent IEEE 802.11 WaveLAN card measurement in [150]. This energy model is then build into both simulation and analytical model for numerical experiments. As seen from Fig.27, the power consumption is inevitably increased at relay node due to traffic forwarding.

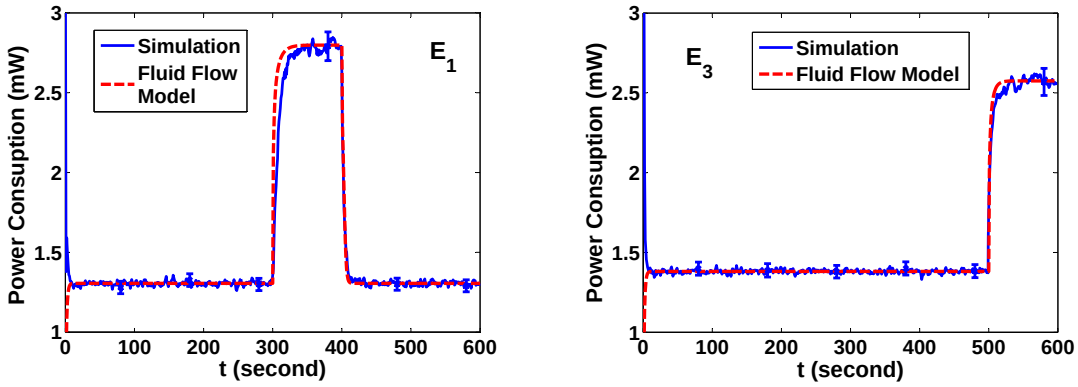


Figure 27: The power consumption of node 1 and node 3.

Next we consider a five node network with the random waypoint mobility (RWM) model. The stochastic properties of the RWM model were studied in [70]. 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) transi-

tion. 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 $T_{up} = 50$s, $T_{down} = 10$s, 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 Figure 28. 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 Figure 28.

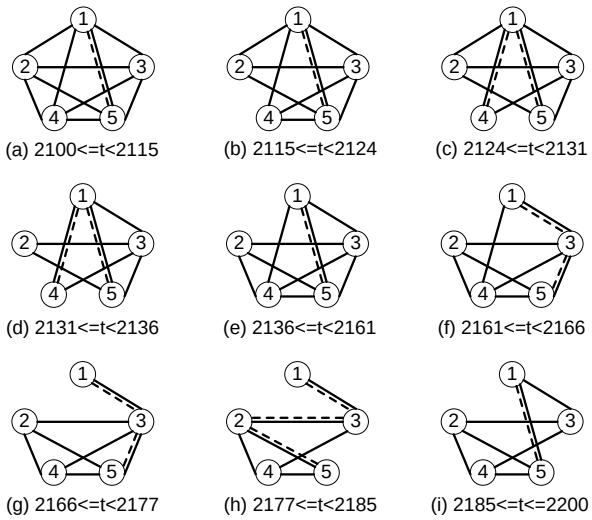


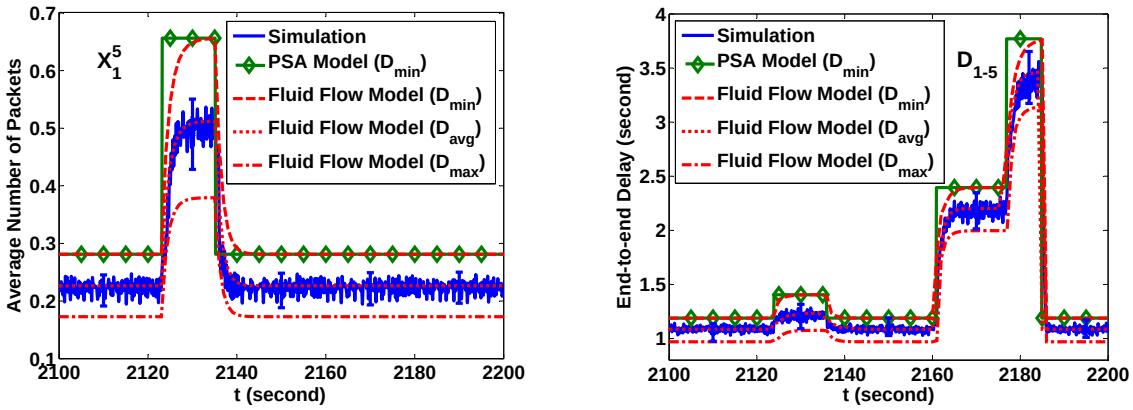
Figure 28: Typical RWM model connectivity scenario for five node network.

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 (4.32) to form the fluid flow model of the network. In addition, we plot the results obtained from the Pointwise Stationary Approximation (PSA) [114] modeling approach which approximates the nonstationary queuing system by using steady-state formula at each time point. Figure 29 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 Figures 25-29.

4.6 COMPUTATION SCALABILITY

According to the analysis in Section 3.7, the computation time complexity of fluid flow model for multihop wireless networks is upper bounded by the number of differential equations n (i.e. $O(n)$). For an M node network, the number of differential equations n equals to $M(M-1)$. Therefore, the computation time complexity of our fluid flow modeling algorithm is upper bounded by $O(M^2 - M)$. To further evaluate the computation complexity of our fluid flow model and compare it with the nonstationary simulation, we conducted numerical



(a) (b)

Figure 29: Average number of packets x_1^5 and end-to-end delay of D_{1-5}.

experiments on a series of sample networks. In the experiments, each node generates traffic to all the other nodes (i.e. full mesh of traffic demands). All links switch between on/off randomly following the two-state Markov process. Table 8 shows the computation time of the average number of packets at each node in the sample networks with Poisson and CBR traffic loads over the time $[0, 600]$ sec by both the fluid flow model and simulation. All the computations run on a PC with Intel i5-450M 2.4GHz processor and 4GB memory. The reported computation time by hybrid model excludes the short time spent during the off-line stage (i.e. setting up network parameters and then calculating data pair (ρ, x) and utilization function $G_i(\cdot)$). For the purpose of fair comparison, the simulation launch latency is not included in the computation time results. 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 does not count in the table. In all sample networks, we set the average packet arrival rate at all nodes as 0.02 pkt/s for both Poisson and CBR traffic loads, and the service rate is assigned to be 1 pkt/s. For the network with CBR traffic, the utilization function is obtained as $0.0023x^5 - 0.0329x^4 + 0.1898x^3 - 0.5756x^2 + 1.025x$. Following the nonstationary simulation procedure, we execute 5000 independent runs and collect the ensemble average number of packets at each node in OPNET. All the results from hybrid model are within 98% confidence interval of the simulation results. As seen from the

table, the simulation time seems to grow dramatically, which is generally a complex function of the number of nodes, traffic load, topology change, accuracy desired, etc. For the hybrid model implemented in MATLAB, we use the quadratic polynomial to fit the computation time data versus node number M ranging from 3 to 200 and obtain the growth rate as $\Theta(0.42M^2 - 6.1M + 8.3)$ in the case of Poisson traffic and $\Theta(0.91M^2 - 8.2M + 16.7)$ in the case of CBR traffic, which are both within our expected upper bound above. It is worthy to note that the differential-equation system based fluid flow model is solvable despite of its size (i.e. number of differential equations), as long as such a differential-equation system satisfies the condition of Picard-Lindelöf theorem to guarantee the existence and uniqueness of the solutions [140]. Moreover, we evaluate the accuracy of fluid flow model for multihop wireless networks by following the same approach in Section 3.7. As shown in Table 8, our proposed fluid flow model can offer reasonably accurate results with significant improvements in the computation time compared to standard simulation tools.

Table 8: Computation Time and Accuracy Comparison

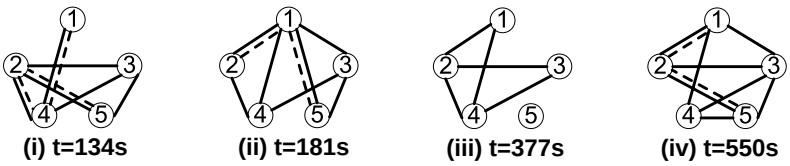
# of nodes	# of diff. equations	Simulation (sec)		Fluid flow model (sec)		Model accuracy (%)	
		Poisson	CBR	Poisson	CBR	Poisson	CBR
3	6	140.26	138.19	2.15	2.23	0.32	0.46
5	20	697.27	683.52	2.65	4.29	0.85	0.92
7	42	3418.65	3421.59	4.72	10.01	1.26	1.38
9	72	16761.26	17025.38	8.44	20.73	1.54	1.73
11	110	82178.61	81092.42	14.95	37.93	1.82	1.91
13	156	401246.82	392626.74	25.14	64.44	1.98	2.02

4.7 NETWORK PERFORMANCE ANALYSIS VIA HYBRID MODEL

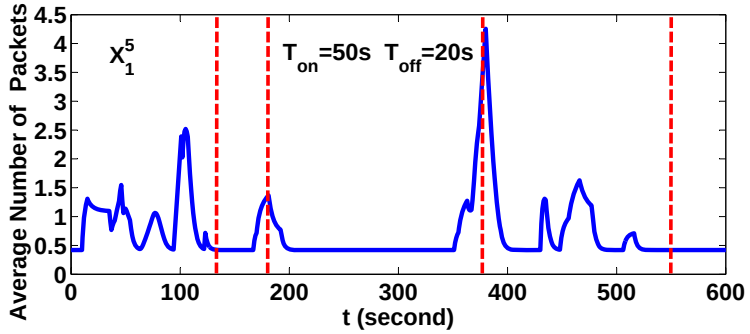
4.7.1 Node Mobility and Traffic Load Impact

We first consider the impact of nodes mobility on the performance of multihop wireless networks by using the random waypoint mobility (RWM) model for node movement. The stochastic properties of this mobility model were studied in [70], and it was observed that the RWM model can be characterized by the link connectivity as a function of time regardless of the detailed movement pattern. Actually, the 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, and both link-on and link-off durations follow an exponential distribution. The Markov process based link connectivity model can be used to greatly reduce the computation load in comparison of a detailed node mobility simulation.

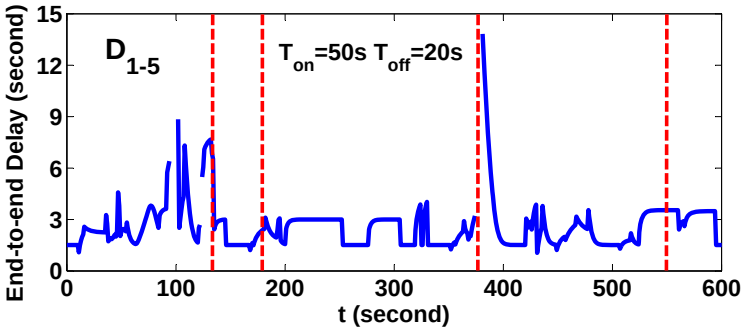
A five-node network with Poisson traffic was set up with mean link on lifetime $T_{on} = 50$ sec and mean link off lifetime $T_{off} = 20$ sec. All links are assumed to have the same T_{on} and T_{off}. In the experiments, the network parameters are the same as in the three node network case. The mean rate (pkt/s) of the Poisson traffic generated by each node is $\gamma_1^3 = 0.22, \gamma_1^5 = 0.28, \gamma_2^5 = 0.12, \gamma_3^5 = 0.21, \gamma_4^5 = 0.16$. We illustrate the typical dynamic network performance by plotting the traffic destined for node 5 at node 1, as seen in Figure 30. We show four snapshots of the network topology at different times in Figure 30(a) and mark all the routes of class 5 (destined for node 5) traffic from node 1 by dot lines in the topologies. Then, we associate the performance results with each topology by aligning them at the same time instant in Figure 30(b)(c). The results conform with the facts that when the direct link breaks as topologies (i) and (iv) of Figure 30(a), the traffic has to go through multiple hops to the destination resulting in longer end-to-end delay, while the average number of packets at the source node remains the same as when the direct link exists. Once destination node 5 is isolated from the network as in topology (iii), the instantaneous end-to-end delay D_{1-5} becomes infinitely large, and the packets have to be queued up in the buffer of source node 1. When node 1 helps forward packets of other nodes to destination node 5 at topology (ii), more packets in the buffer cause longer queueing delay in node 1.



(a) Sample topologies



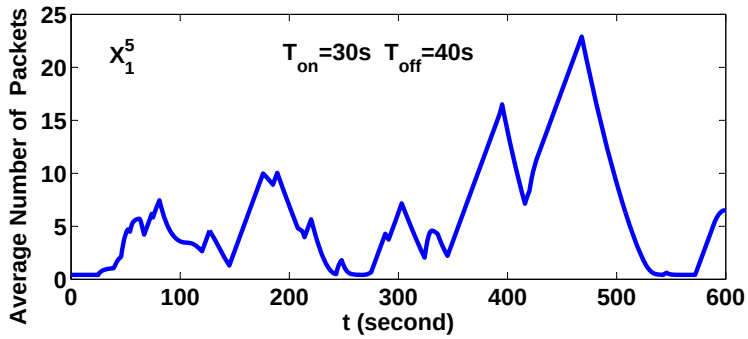
(b) Average number of packets destined for node 5 at node 1 buffer



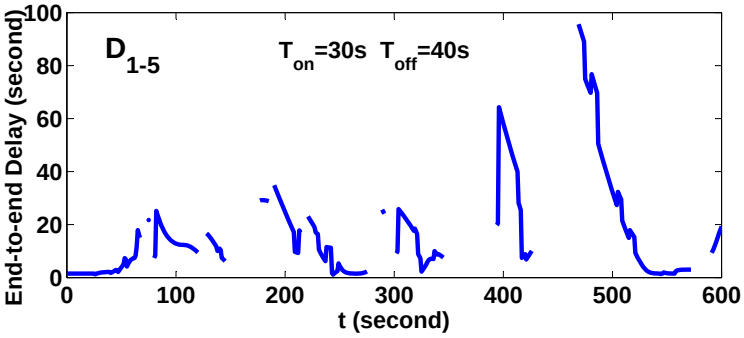
(c) Average end-to-end delay from node 1 to node 5

Figure 30: Dynamic behavior of the traffic destined for node 5 at node 1 buffer with node mobility model $T_{on} = 50$s, $T_{off} = 20$s.

Next we increased the average link off lifetime $T_{off} = 40s$ and decreased the link on lifetime $T_{on} = 30s$ for each pair of nodes to observe a higher level of nonstationarity, since each node will have less overall connectivity and is more likely to redirect traffic. All the other network parameters remain the same as before. Typical results are given in Figure 31, for traffic at node 1 destined for node 5. When the effect of long average link off lifetime starts to set in, it becomes more difficult for each node to find any intermediate node to relay the traffic. When the source node cannot find any alternative path to reroute the traffic, the instantaneous end-to-end delay becomes infinitely large and appears to be disconnected, and meanwhile the packets are accumulated linearly.



(a) Average number of packets destined for node 5 at node 1 buffer



(b) Average end-to-end delay from node 1 to node 5

Figure 31: Dynamic behavior of the traffic destined for node 5 at node 1 buffer with node mobility model $T_{on} = 30$s, $T_{off} = 40$s.

The concept of two-state Markov model can represent various mobility scenarios by using different combinations of T_{on} and T_{off}. For example, the random waypoint mobility of the

wireless nodes with a smaller radio range can be implicitly represented by relatively larger T_{off}. Alternately, in a random waypoint group mobility, if two nodes are from the same group, the average T_{on} of the link between these two nodes should be longer, since they tend to be moving with a comparable speed in a similar direction.

We then set up a 30-node network with full mesh contention-free CBR traffic, and and 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 4.3.1 and 4.3.2. 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 4.2.2.3. 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 will change to the polynomial $G(x)$, which is determined by using the curve fitting to Equation (.7). 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 Figure 32(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 Figure 32(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 Figure 32(c)-(e). Figure 32(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 Figure 32(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 Figure 32(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). Figure 32(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 Figure 32(d). Meanwhile, the average end-to-end delay per traffic stream climbs up in Figure 32(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, Figure 32(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 with 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 $d_{ij}(t) \leq R$, 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 Figure 32(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 Figure 32(d). Also, the increased queuing delay at each node prolongs the end-to-end delay of the traffic in Figure 32(e). Due to the link-level error, Figure 32(c) shows that the average utilization per link in the scenario (35s, 35s, 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 Figure 32(f).

4.7.2 Comparison with Steady State Performance Modeling Technique

To further illustrate the exclusive capability of capturing time varying behavior of multihop wireless networks by fluid flow model, we setup a 30-node network with full mesh traffic and numerically compare the results from pointwise stationary model, given in Figure 33. The network configuration is as follows: the exponentially distributed packet size with the mean $1/\mu$ of 1250 bytes, the link capacity C of 10^4 bps, the forwarding delay δ of 0.1 second, the link propagation delay ε of 0.1 microseconds and minimum hop routing. The link connectivity $a_{ij}(t)$ is assigned a random number between 0.95 and 1 if this link exists, otherwise $a_{ij}(t)$ is 0. We use the random waypoint mobility (RWM) model to represent node mobility.

In Figure 33(a)-(b), we change the mobility model (T_{up}, T_{down}) from (50s, 5s) to (50s, 20s) to represent different cases of network connectivity and keep the averaged rate of offered traffic γ_i^j as 0.0125 pkt/s. The number of packets at node 3's buffer and destined for node 30, i.e. x_3^{30}, is selected as an example of performance metric in this study. We observe that shorter link up lifetime with longer link down lifetime for each pair of nodes results in higher level of nonstationarity, since nodes will get less overall connectivity time and be more likely to redirect the traffic to others. The idea of pointwise stationary model is to sample the time varying network conditions and apply them into steady-state queuing formula to approximate the nonstationary system performance. Although this model offers an easy technique to calculate quantities of interest, it is actually in error for instantaneous values and return even poor results for highly-dynamic network, as seen in Figure 33(b). Then, we

remain the node mobility pattern and increase the full-mesh traffic load at each link in Figure 33(c). The growth of offered traffic load at a node inevitably results in packet accumulation in its buffer due to limited link capacity in such a less-connected network. However, since pointwise stationary model calculates the steady state results independently at each time instant, it does not consider the dependencies of buffer occupancy between adjacent time instants and thus cannot capture such a continuous-time queuing behavior. Instead, fluid flow model constructs continuous-time differential equations by following flow conservation principle. Hence, it can successfully demonstrate such a phenomena of network congestion.

4.7.3 Discussion on Steady State and Time Varying Behavior

It is increasingly noted that multihop wireless networks must not only perform well in steady state, but must also have acceptable performance under transient or nonstationary conditions. Here we study the network performance from both steady state and time varying perspectives. The nonstationarity of the network comes from topology change modeled by a two-state Markov process, where the average link durations T_{up} and T_{down} are being manipulated. The jitteriness of this two-state Markov process is defined as $\frac{1}{T_{up}+T_{down}}$, which represents the average times of going through this two-state cycle per unit time. In the numerical experiment, we fix the traffic load at 0.02 pkt/s and decrease T_{up} and T_{up} proportionally from (100s, 40s) to (25s, 10s) and then (5s, 2s), so that the corresponding jitteriness of the mobility model is increased accordingly. As a result, Figure 34 shows that both performance metrics (i.e. x_1^{30} and D_{1-30}) have more frequent variations and the nonstationary period approaches to dominate the network behavior. Note that the steady state network behavior is the same in all three cases as shown in Figure 35, even though the time varying behavior is quite different.

On the one hand, to study the steady state behavior, the network performance is measured over a fixed time period, and then averaged over that period. On the other hand, the time varying behavior describes the network performance at specific time instants, so that the network behavior can be shown as a function of time. To quantify the dynamic of network performance, we define instantaneous variation (IV) as the difference of a certain

performance metric (P) between two samples separated by a small time interval $[t_i, t_{i+1}]$ with length δ (i.e. $IV(t_i) = P(t_i) - P(t_{i+1})$, where $t_{i+1} - t_i = \delta$). The choice of δ depends on the change rate of network performance. According to [18], an order of magnitude bound on the change rate of any network performance metric can be determined by considering the event rate for the whole network under simplifying assumptions. If all of the time varying events taking place in the network are assumed to have exponentially distributed inter-event times with average rate r_i, then the overall average system event rate is $r = \sum_i r_i$, and no performance metric can change more rapidly than r on the average. In our experiment, the nonstationarity of the network comes from topology change, and the link connectivity is modeled as a two-state Markov process with the average transition rates $1/T_{up}$ and $1/T_{down}$. Thus, by setting the sample interval $\delta < \frac{1}{1/T_{up}+1/T_{down}}$, we will not miss any significant transient behavior within the interval. In Figure 35, both the time average and instantaneous variation values of different node mobility cases are shown. In the bar charts, the value of instantaneous variation is actually the average of the instantaneous variation between two successive samples over 600 seconds, and the sample interval δ is set to be 1 second. The time average value is obtained by simply averaging all samples over 600 seconds. The sample values of the mobility cases (100s, 40s), (25s, 10s), and (5s, 2s) come from the results in Figure 35. As illustrated from the bar charts, since the ratio between T_{up} and T_{down}) stays unchanged at various mobility cases, the total link up or down duration over a long time remains the same among cases on average. Thus, the time average values of x_1^{30} and D_{1-30} are shown to be similar. However, the instantaneous variation of x_1^{30} and D_{1-30} grows with the decrease of T_{up} and T_{down}). The reason is that the increase of jitteriness in node mobility model raises the possibility of link transition between up and down during the time interval δ, which inevitably leads to the variation of network behavior.

The results from Figure 30-35 confirm that our hybrid model is able to promptly respond to ongoing nonstationary condition of multihop wireless networks and allow us to analyze their time varying behavior, which is generally an important QoS consideration in assessment of network performance.

4.8 SUMMARY

In this chapter, we propose a performance modeling technique to study the time varying behavior of multihop wireless networks with CBR traffic, using numerical method based queueing analysis. Network queues are modeled using fluid flow based differential equations and solved using 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 this hybrid model to examine the effects of node mobility and traffic load dynamics on the performance of a moderate size network.

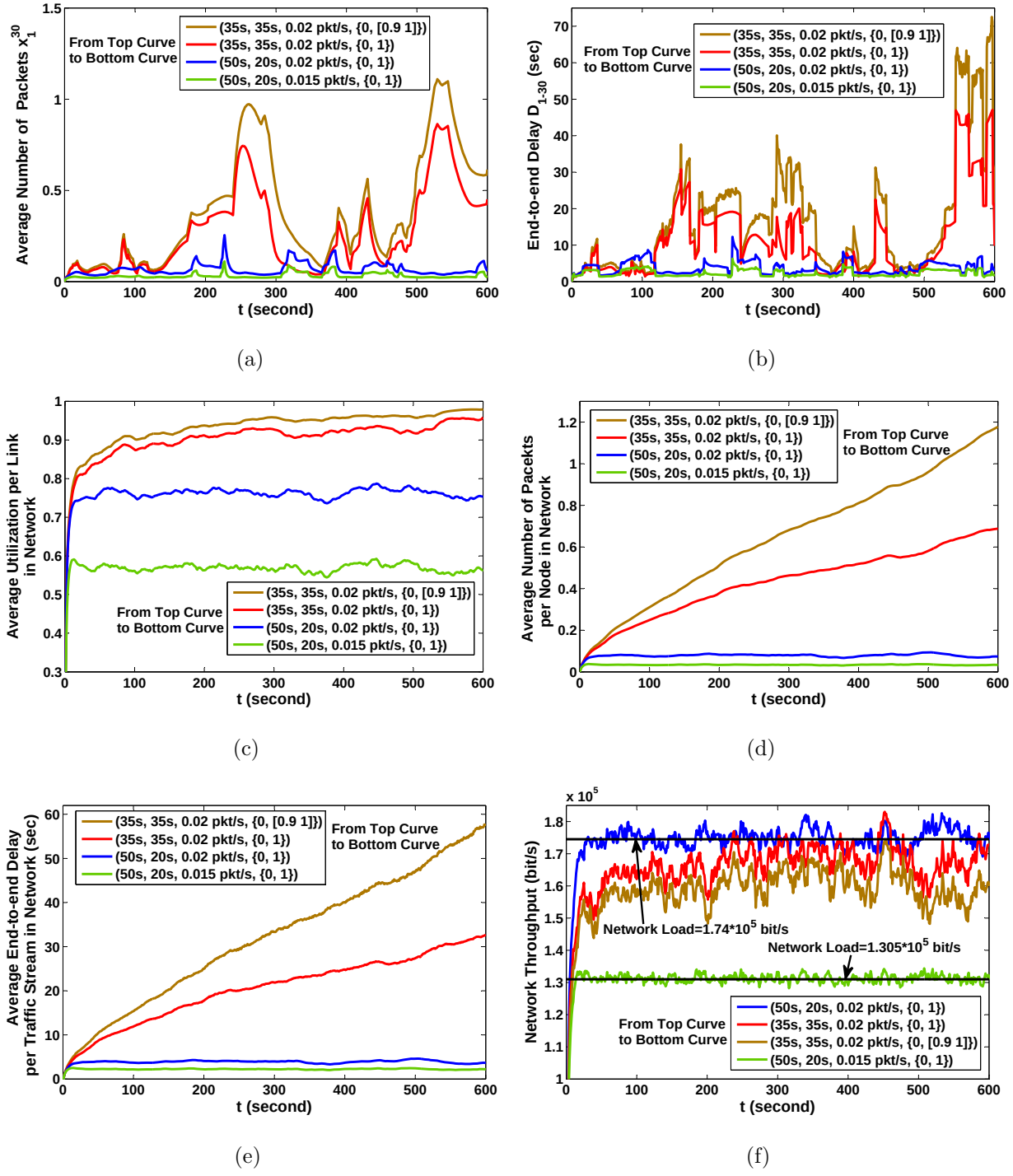
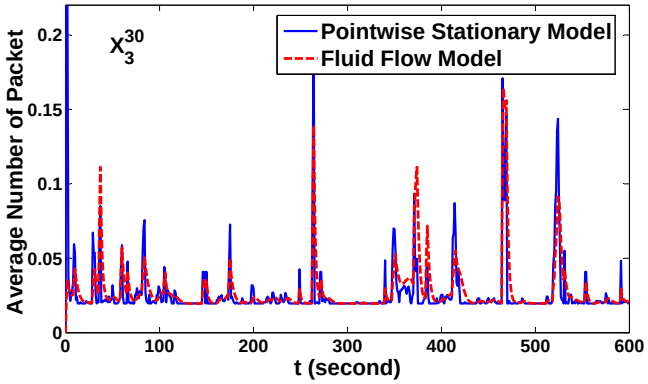
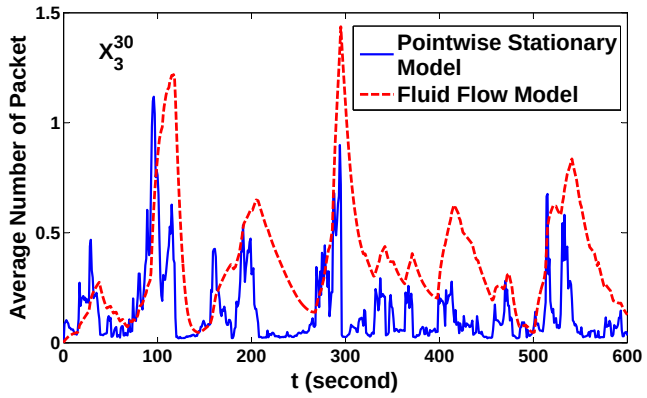


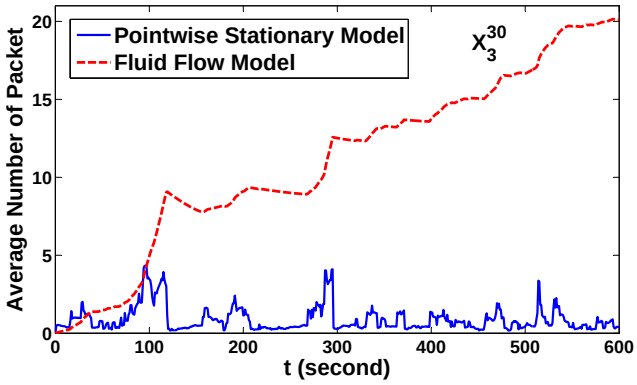
Figure 32: Various network performance measures impacted by traffic load, node mobility and link quality (i.e. $(T_{up}, T_{down}, \gamma, a_{ij})$).



(a) $(T_{up}, T_{down}, \gamma) = (50\text{s}, 5\text{s}, 0.0125\ \text{pkt/s})$

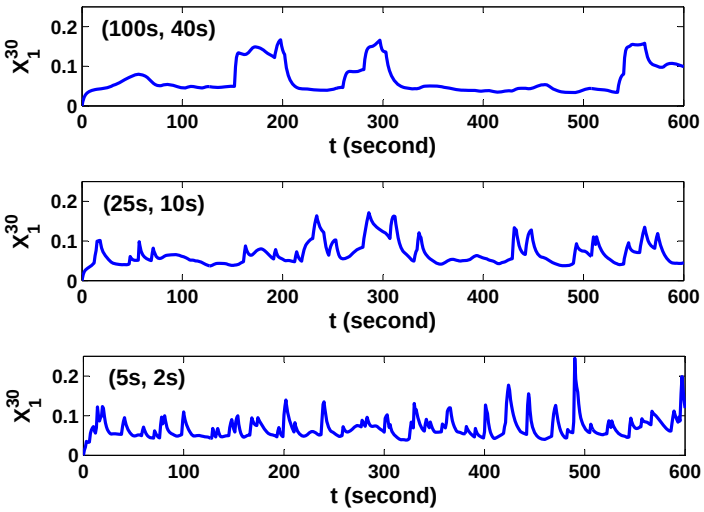


(b) $(T_{up}, T_{down}, \gamma) = (50\text{s}, 20\text{s}, 0.0125\ \text{pkt/s})$

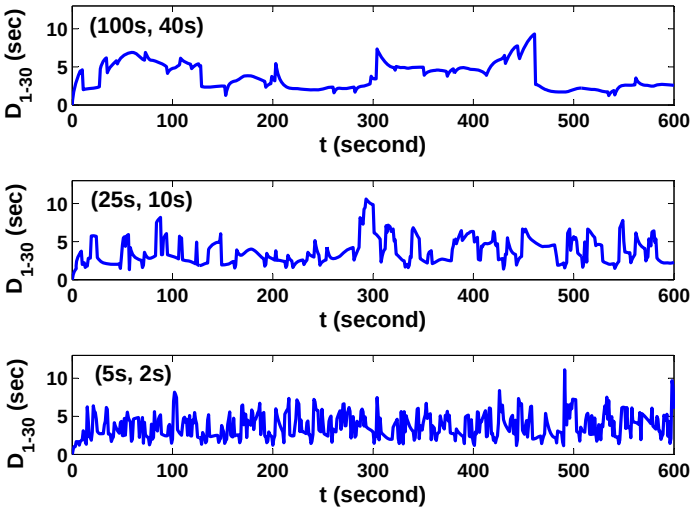


(c) $(T_{up}, T_{down}, \gamma) = (50\text{s}, 20\text{s}, 0.03\ \text{pkt/s})$

Figure 33: The dynamic behavior of x_3^{30} at various node mobility patterns and traffic loads.

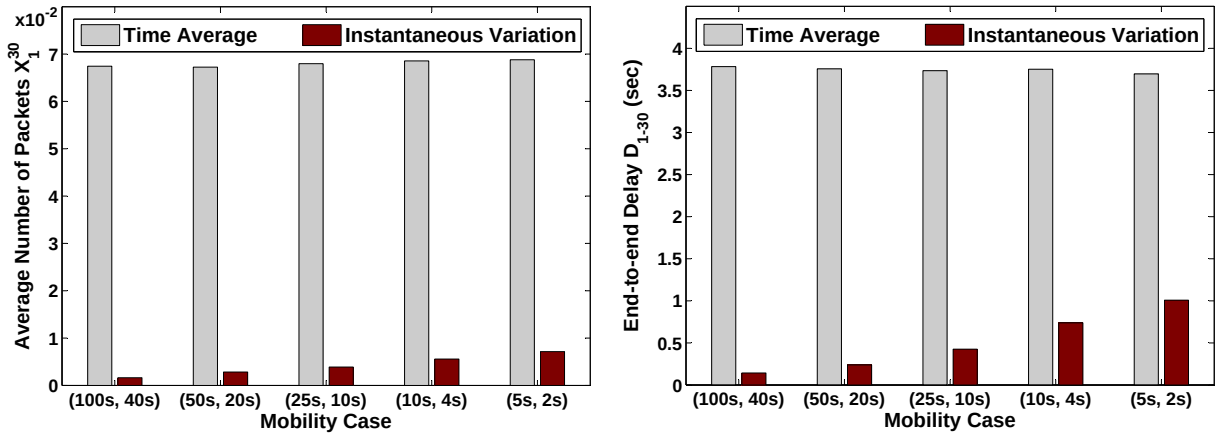


(a)



(b)

Figure 34: The time varying behavior of x_1^{30} and D_{1-30} in various mobility cases.



(a) (b)

Figure 35: The time average and the instantaneous variation of x_1^{30} and D_{1-30} in various mobility cases.

5.0 CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

Wireless networks are playing an important role in our society that rapidly evolves toward a pervasive computing age. There has been a great amount of research and development on wireless network communication and protocol issues. While there has been massive efforts on developing simulation models to estimate wireless network performance, most existing simulation approaches are known to be lacking fidelity and scalability. The performance of wireless networks is normally studied via simulation over a fixed time horizon using a steady-state type of statistical analysis procedure. However, due to dynamic nature of network topology, such an approach may be inappropriate in many cases as the network may spend most of the time in a transient or nonstationary state. Simulation studies of time varying behavior for such networks are possible, though computationally intensive. Specifically, one must perform a simulation study following the nonstationary simulation approach with the essence of ensemble average instead of time average. The basic approach is to observe the system behavior versus time over an ensemble of statistically identical but distinct independent runs. The quantities of interest (e.g. mean queue length at every node) are averaged across the ensemble of runs at a particular time instant and confidence intervals are calculated then from the ensemble. Many such points are obtained at different time instants and the system behavior is determined as a function of time. The main difficulty in conducting nonstationary simulation is the large number of runs (typically thousands) that must be generated conducted in order to get a representative ensemble from which a statistically accurate portrayal of the system behavior can be determined. Hence, significant amount of CPU time is required for even small sized networks and this approach is quite

difficult to scale.

The goal of this work is to develop an integrative framework incorporating a set of queuing and stochastic modeling technique to efficiently approximate the time varying performance of wireless networks. Network queues are modeled using fluid flow based differential equations which are solved using numerical methods, while traffic load, dynamic topology and network protocols are modeled using stochastic modeling techniques. Our performance modeling technique provides an insight into the joint effect by traffic, topology and protocols in wireless networks. Numerical and simulation experiments show that our fluid-flow based performance model can provide reasonably accurate results much more computation-efficiently than standard discrete event simulators.

In Chapter 3, we propose a fluid flow based performance model to evaluate the time varying behavior of single-hop vehicular networks. According to IEEE 802.11p standard designed for vehicular networks, the RF transmitter embedded in each vehicle accesses the wireless channel via contention based CSMA/CA mechanism and broadcasts the safety-related packets to all its sourounding vehicles. As a result, the packet service times at different vehicles in the same carrier sensing range are strongly coupled. The bandwidth shared by each vehicle is an order of magnitude less when 10 other vehicles are active than when only one single vehicle is active. Due to high-speed mobility, vehicles might move into/out of the carrier sensing range of other vehicles, so that the number of vehicles contending to access the channel varies over time. We first derive the packet service time distribution as a function of network hearing topology. Then, the fluid flow model is constructed for vehicular networks by considering the vehicle mobility and the resulting time varying packet service time and its variance. The fluid flow based model is shown to be accurate and scalable according to our numerical comparisons with simulation results. Moveover, we apply the proposed model to evaluate the nonstationarity of vehicle networks performance impacted by traffic load, vehicle velocity and vehicle density.

Following the same fluid flow modeling approach, we develop the time varying performance model to approximate the dynamic behavior of multihop wireless networks in Chapter 4. In such a network, all nodes have to collaborate with one another to dynamically route the traffic using wireless links. The traffic flow may go through multiple hops to reach des-

tination, with each node acting as a router. Besides the nonstationary features inherent in single-hop transmission discussed above, the routes in multihop wireless networks are prone to failures due to node mobility. Since network nodes may move arbitrarily, the network topology is expected to change frequently and unpredictably. Thus, one would expect that transient/nonstationary conditions to occur often. We first derive the single-node fluid flow model by deriving server utilization function for a variety of queueing systems. On the basis of the single-node model, we then construct the fluid flow model for multihop wireless networks by considering the time varying network connectivity and traffic routing. Numerical and simulation experiments show that the fluid flow model for multihop wireless networks is a fairly-accurate and scalable tool to approximate the dynamic network behavior. 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.

We believe that this performance modeling approach is a valuable tool for evaluating the time-varying behavior of wireless networks in an efficient manner. With the computation time saved by the fluid flow based modeling technique, it is a tremendous gain in flexibility for modeling complex network protocols or wireless channel characteristics in a specific environment, in order to add higher level of fidelity into the proposed model. Additionally, since many network controls are designed and implemented on the basis of steady state performance, they may not make optimum use of network resources under transient/nonstationary period. Therefore, our proposed models can serve as the basis for the application of control theory to develop dynamic network control algorithms.

5.2 FUTURE WORK

5.2.1 Model Extension

The fluid flow based performance model for wireless networks can be extended at different layers including physical, MAC and network layer.

To improve the fidelity of the model in the physical layer, it is possible to adjust the connectivity elements in the adjacency matrix as a function of the received signal strength (RSS) between two nodes using propagation channel models. Such a function can be determined by the modulation scheme used at the receiver side. The propagation channel models could be constructed by considering the effect of exponential path loss, multipath fading and shadowing. For a specific indoor/outdoor environment, we can determine all the possible parameters of the channel models by using the site-survey measurement data on RSS. By doing this, the connectivity between two nodes might be assigned by a decimal number and then incorporated into our model.

Besides the contention-based MAC discussed in this work, wireless network could also employ contention-free MAC, which can be described by the packet service time of our model. Since contention causes packet collisions in wireless network, a contention-free MAC can improve the reliability of wireless delivery. Possible contention-free MAC includes CDMA, TDMA, FDMA or other token based schemes, and they can be implemented in both centralized and distributed manners. For contention-free MAC, resource reservation delay has to be taken into account to determine the packet service time in the fluid flow model.

For network layer, more detailed routing schemes can be embedded into our model. Multipath routing is a routing technique of using multiple alternative paths through a network, which can yield a variety of benefits of fault tolerance and increased bandwidth. To model multipath routing, it is straightforward to assign the routing parameter in the model with a decimal number, which is calculated by the portion of traffic going through this path. Moreover, the practical routing algorithms, such as AODV and DSR, typically include the routing discovery phase, which could significantly impact the performance of a highly-dynamic network. In the fluid flow model, such a discovery phase can be modeled by adding a short query packets delivery process at the beginning of network connectivity change each time.

5.2.2 Hybrid Packet/fluid Simulation

Packet-level simulation models the detailed behavior of every packet in the network, and results in an accurate picture of overall network behavior. Alternatively, fluid-level

simulation abstracts the aggregated packets as a flow in a way that smooths over unessential details. Such a simulation approach only consider changes in rates of traffic flows, and allows one to sketch the network behavior using less computational effort.

The future work is to combine the standard packet-level simulators with fluid flow based performance modeling technique to leverage the strong points of both two methods. The hybrid method uses packet-level simulation to provide the behavior every packet for which more detail is required and fluid flow model to represent aggregations of flows for which less detail is needed. By treating data traffic as a flow and solving a set of differential equations to obtain statistical data, the fluid flow model can not only speed up simulation but also capture the dynamic behavior of that traffic flow.

5.2.3 Dynamic Network Control

The modeling techniques proposed in this dissertation can be applied to a variety of control mechanisms in wireless networks to better support a required quality of service in a real-time manner.

A dynamic control of media access in vehicular networks can be possibly developed on the basis of fluid flow model in Chapter 3. One weakness of media access control of vehicular networks is that static parameters are always used even though transient periods occur often in some scenarios. Due to high-speed mobility, transient periods might dominant the network behavior. Hence, the static MAC may not make optimum use of network resources after the network hearing topology is changed. Furthermore, numerical results confirm that as the transmission queue becomes heavily loaded, the settling time is in the unit of seconds. During such a time period, network hearing topology could update and the unsettled queue could be perturbed all the time and remain nonstationary. Consequently, the dynamic control of media access is underscored and settling time can provide some ideas as to how often dynamic control strategies should be updated based on the assumption of quasi-static conditions. The dynamic control of media access in vehicular networks include admission control, contention window sizing, and etc.

Based on the model for multihop wireless networks in Chapter 4, a dynamic optimal

control problem can be formulated to determine the minimum-delay route for the traffic flow from source node to destination node. The traditional routing algorithms designed for static networks ignore the transient behavior of the network. As a result, they may not obtain the optimal path or even feasible path for a traffic flow due to frequent changes of network connectivity. As we know, the queuing delay and the transmission delay of each node along the path are considered as the main factor of the end-to-end delay in a network. In a sense, the number of packets in the nodes along the path is a measure of the end-to-end delay. To seek the optimal route lasting for a certain time interval, it is first assumed that the network connectivity remains the same during such a period. Then, all the possible paths for traffic flow can be decided from network connectivity prior by graph theoretic algorithms. Finally, according to the time varying number of packets in each node obtained from the fluid flow model proposed in Chapter 4, we can determine the optimal route of the traffic flow by selecting the path with the minimum number of packets accumulated along.

APPENDIX

APPROXIMATION FOR $N*D/D/1$ QUEUE IN THE HEAVY TRAFFIC REGIME

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 $dr = 1$ could be considered as infinitesimal compared with N, and the values of $Q(r)$ at two adjacent steps becomes very close. Then, the $E[x]$ can be approximated as the integration of $Q(r)$, that is:

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

In the heavy traffic regime $\rho \to 1$, we have the following result from [49], based on the Brownian approximation.

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

By substituting Equation (.2) into (.1), we have

$$\begin{aligned} x &\cong \int_0^{N-1} e^{-2r\left(\frac{r}{N}+\frac{1-\rho}{\rho}\right)} dr \\ &= \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} \left(\operatorname{erf}\left(\frac{(\rho-1)\sqrt{2N}}{2\rho} \right) + \operatorname{erf}\left(\frac{(N\rho-2\rho+N)}{\sqrt{2N}\rho} \right) \right) \end{aligned} \tag{.3}$$

Since $N \gg 1$ and $\rho \to 1$, the argument of the second erf term in Equation (.3) 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 (.3) becomes:

$$x \cong \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} \left(\operatorname{erf}\left(\frac{(\rho-1)\sqrt{2N}}{2\rho} \right) + 1 \right) \tag{.4}$$

Since $\rho \leq 1$ and $\text{erfc}(x) = \text{erfc}(-x) + 1$, Equation (.4) can be rewritten as

$$x \cong \frac{\sqrt{2\pi N}}{4} e^{\frac{N(1-\rho)^2}{2\rho^2}} \text{erfc}\left(\frac{(1-\rho)\sqrt{2N}}{2\rho}\right) \tag{.5}$$

In [153], 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

$$\exp(x^2)\text{erfc}(x) \approx \frac{1}{Ax + \sqrt{1 + Bx^2}} \tag{.6}$$

where $A = 377/324$, and $B = 314/847$. Since the server utilization is defined as $\rho = N/D$, we substitute N by ρD in (.5). By using the approximation (.6) in (.5), 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

$$x \cong \frac{\sqrt{2\pi\rho D}}{\frac{2A(1-\rho)\sqrt{2\rho D}}{\rho} + 2\sqrt{4 + \frac{2B(1-\rho)^2 D}{\rho}}} \tag{.7}$$

BIBLIOGRAPHY

[1] *IEEE 802.11p Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 6: wireless access in vehicular environments*, Std., 2010.

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

[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] H. Zhang, Z. Zhang, and H. Dai, "Gossip-based information spreading in mobile networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5918–5928, 2013.

[10] X.-L. Huang, G. Wang, Q.-Q. Sun, and Z.-R. Cai, "Qos-adaptive routing protocol design for multi-hop cognitive radio networks considering multi-link interference," *International Journal of Sensors Wireless Communications and Control*, vol. 1, no. 2, pp. 88–92, 2011.

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

[12] *NS-2 Simulaiton Tool*, http://www.isi.edu/nsnam/ns/.

[13] *NS-3 Simulaiton Tool*, http://www.nsnam.org/.

[14] *QualNet Simulaiton Tool*, http://www.qualnet.com/.

[15] *Glomosim Simulaiton Tool*, http://pcl.cs.ucla.edu/projects/glomosim/.

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

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

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

[19] A. Boukerche and L. Bononi, "Simulation and modeling of wireless, mobile, and ad hoc networks," *Mobile ad hoc networking*, pp. 373–409, 2004.

[20] Z. Ji, J. Zhou, M. Takai, and R. Bagrodia, "Scalable simulation of large-scale wireless networks with bounded inaccuracies," in *Proc. of ACM international symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2004, pp. 62–69.

[21] W. Kasch, J. Ward, and J. Andrusenko, "Wireless network modeling and simulation tools for designers and developers," *IEEE Communications Magazine*, vol. 47, no. 3, pp. 120–127, 2009.

[22] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[23] Y. Xiao, "Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1506–1515, 2005.

[24] O. Tickoo and B. Sikdar, "Modeling queueing and channel access delay in unsaturated IEEE 802.11 random access mac based wireless networks," *IEEE/ACM Transaction on Networking*, vol. 16, no. 4, pp. 878–891, 2008.

[25] E. Ghadimi, A. Khonsari, A. Diyanat, M. Farmani, and N. Yazdani, "An analytical model of delay in multi-hop wireless ad-hoc networks," *Wireless Networks*, vol. 17, no. 7, pp. 1679–1697, 2011.

[26] J. Li, C. Blake, D. S. De Couto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *Proc. of International Conference on Mobile Computing and Networking*, 2001, pp. 61–69.

[27] J. W. G. Wang and Y. Zheng, "Optimum energy efficient communications for hybrid ARQ systems," in *Proc. of IEEE GLOBECOM*, 2013, pp. 1–6.

[28] J. Wu, G. Wang, and Y. Zheng, "Energy efficiency and spectral efficiency tradeoff in type-I ARQ systems," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 356–366, 2014.

[29] G. W. Jingxian Wu and Y. Zheng, "Energy and spectral efficient transmissions of coded ARQ systems," in *IEEE International Conference on Communications (ICC)*, 2013, pp. 5883–5887.

[30] 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.

[31] I. Broustis, K. Papagiannaki, S. Krishnamurthy, M. Faloutsos, and V. Mhatre, "Measurement-driven guidelines for 802.11 WLAN design," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 722–735, 2010.

[32] 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*, 2004, pp. 121–132.

[33] D. Giustiniano, D. Malone, D. Leith, and K. Papagiannaki, "Measuring transmission opportunities in 802.11 links," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1516–1529, 2010.

[34] 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*, 2011, pp. 775–783.

[35] D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," in *Proc. of ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2004, pp. 78–82.

[36] C. Newport, D. Kotz, Y. Yuan, R. S. Gray, J. Liu, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," *Simulation*, vol. 83, no. 9, pp. 643–661, 2007.

[37] C. Tala, L. Ahumada, D. Dujovne, S.-U. Rehman, T. Turletti, and W. Dabbous, "Guidelines for the accurate design of empirical studies in wireless networks," in *Testbeds and Research Infrastructure*. Springer, 2012, pp. 208–222.

[38] G. Bianchi, A. Di Stefano, C. Giaconia, L. Scalia, G. Terrazzino, and I. Tinnirello, "Experimental assessment of the backoff behavior of commercial IEEE 802.11b network cards," in *Proc. of IEEE INFOCOM*, 2007, pp. 1181–1189.

[39] D. Wu, D. Gupta, S. Liese, and P. Mohapatra, "QuRiNet: quail ridge natural reserve wireless mesh network," in *Proc. of International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2006, pp. 109–110.

[40] D. Wu, D. Gupta, and P. Mohapatra, "Quail ridge wireless mesh network: experiences, challenges and findings," in *Proc. of International Conference on Testbeds and Research Infrastructure for the Development of Networks and Communities*, 2007, pp. 1–6.

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

[42] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET simulation studies: the incredibles," *ACM Mobile Computing and Communications Review (SIGMOBILE)*, vol. 9, no. 4, pp. 50–61, 2005.

[43] N. Sarkar and J. Gutiérrez, "Revisiting the issue of the credibility of simulation studies in telecommunication networks: highlighting the results of a comprehensive survey of ieee publications," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 218–224, 2014.

[44] D. Cavin, Y. Sasson, and A. Schiper, "On the accuracy of MANET simulators," in *Proc. of ACM Workshop on Principle of Mobile Computing*, 2002, pp. 38–43.

[45] M. Segata and R. Lo Cigno, "Simulation of 802.11 PHY/MAC: The quest for accuracy and efficiency," in *Proc. of Conference on Wireless On-demand Network Systems and Services (WONS)*, 2012, pp. 99–106.

[46] A. Boukerche and L. Bononi, *Simulation and Modeling of Wireless Mobile Ad Hoc Networks*. IEEE Press and John Wiley and Sons, Inc., 2003.

[47] A. Vasan and A. Shankar, "Timestepped stochastic simulation of 802.11 WLANs," in *Proc. of International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, 2007, pp. 15–22.

[48] 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)*, 2003.

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

[50] J.-Y. L. Boudec, *Performance Evaluation of Computer and Communication Systems*. EFPL Press, 2011.

[51] C. Nuzman, I. Saniee, W. Sweldens, and A. Weiss, "A compound model for TCP connection arrivals for LAN and WAN applications," *Computer Networks*, vol. 40, no. 3, pp. 319–337, 2002.

[52] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. K. Shankaranarayanan, "Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots," in *Proc. of IEEE INFOCOM*, 2011, pp. 2921–2929.

[53] F. Hernández-Campos, M. Karaliopoulos, M. Papadopouli, and H. Shen, "Spatio-temporal modeling of traffic workload in a campus WLAN," in *Proc. of International Workshop on Wireless Internet*, 2006.

[54] G. He, J. C. Hou, W.-P. Chen, and T. Hamada, "Characterizing individual user behaviors in WLANs," in *Proc. of ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, 2007, pp. 132–137.

[55] B. L. M. Hisashi Kobayashi, *System Modeling and Analysis: Foundations of System Performance Evaluation*. Prentice Hall, 2008.

[56] L. Kleinrock, *Queueing Systems*. John Wiley and Sons, 1976.

[57] R. Jain and S. Routhier, "Packet trains–measurements and a new model for computer network traffic," *IEEE Journal on Selected Areas in Communications*, vol. 4, no. 6, pp. 986–995, 1986.

[58] H. G. Ferguson P., *Quality of Service: Delivering QoS on the Internet and in Corporate Networks*. John Wiley and Sons, 1988.

[59] M. Grossglauser and S. Keshav, "On CBR service," in *Proc. of IEEE INFOCOM*, vol. 1, 1996, pp. 129–137.

[60] K. L. Phuoc Tran-Gia, Dirk Staehle, "Source traffic modeling of wireless applications," *International Journal of Electronics and Communications*, vol. 55, pp. 27–36, 2001.

[61] W. Jiang and H. Schulzrinne, "Analysis of on-off patterns in VoIP and their effect on voice traffic aggregation," in *Proc. of International Conference on Computer Communications and Networks*, 2010, pp. 82–87.

[62] V. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, vol. 32, no. 3, pp. 70–81, 1994.

[63] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[64] 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.

[65] D. Pong and T. Moors, "The impact of random waypoint mobility on infrastructure wireless networks," in *Proc. of International Conference on Parallel and Distributed Systems*, vol. 2, 2005, pp. 140–144.

[66] E. Hyytiä and J. Virtamo, "Random waypoint mobility model in cellular networks," *Wireless Networks*, vol. 13, no. 2, pp. 177–188, 2007.

[67] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. of IEEE INFOCOM*, vol. 2, 2003, pp. 844–852.

[68] S. Khurana, A. Kahol, and A. Jayasumana, "Effect of hidden terminals on the performance of IEEE 802.11 MAC protocol," in *Proc. of IEEE Conference on Local Computer Networks*, 1998, pp. 12 –20.

[69] S. Khurana, A. Kahol, S. Gupta, and P. Srimani, "Performance evaluation of distributed co-ordination function for IEEE 802.11 wireless LAN protocol in presence of mobile and hidden terminals," in *Proc. of International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, 1999, pp. 40–47.

[70] 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.

[71] T. Lin and S. F. Midkiff, "Mobility versus link stability in the simulation of mobile ad hoc networks," in *Proc. of Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS)*, 2003, pp. 3–8.

[72] N. Abramson, "The ALOHA systemĺcanother alternative for computer communications," *Cluster Computing*, vol. 5, pp. 187–201, 1970.

[73] ——, "Packet switching with satellites," in *Proc. of National Computer Conference and Exposition*, 1973, pp. 695–702.

[74] L. G. Roberts, "ALOHA packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, 1975.

[75] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Transactions on Communications*, vol. 25, no. 1, pp. 117–128, 1977.

[76] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: part I-carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, pp. 1400–1416, 1975.

[77] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: part II–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution," *IEEE Transactions on Communications*, vol. 23, pp. 1417–1433, 1975.

[78] P. Karn, "MACA-a new channel access method for packet radio," in *Proc. of 9th ARRL/CRRL Amateur radio computer networking conference*, vol. 140, 1990, pp. 134–140.

[79] Y. Cai, K. Xu, Y. Mo, B. Wang, and M. Zhou, "Improving WLAN throughput via reactive jamming in the presence of hidden terminals," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 1085–1090.

[80] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, Std., 1997.

[81] P. Chatzimisios, A. Boucouvalas, and V. Vitsas, "IEEE 802.11 packet delay-a finite retry limit analysis," in *Proc. of IEEE GLOBECOM*, vol. 2, 2003, pp. 950–954.

[82] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement," in *Proc. of IEEE INFOCOM*, vol. 2, 2002, pp. 599–607.

[83] O. Y. G. W. Liang Zhang, Yantai Shu, "Study of medium access delay in IEEE 802.11 wireless networks," *IEICE Transaction in Communications*, vol. 89, pp. 1284–1293, 2006.

[84] Q. Ni, T. Li, T. Turletti, and Y. Xiao, "Saturation throughput analysis of error-prone 802.11 wireless," *Wireless Communications and Mobile Computing (WCMC)*, vol. 5, pp. 945–956, 2005.

[85] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 785–799, 2000.

[86] Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Networks*, vol. 7, pp. 159–171, 2001.

[87] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs: Research articles," *Wireless Commuications and Mobile Computing (WCMC)*, vol. 4, no. 8, pp. 917–931, 2004.

[88] C. Xu, K. Liu, G. Liu, and J. He, "Accurate queuing analysis of IEEE 802.11 MAC layer," in *Proc. of IEEE GLOBECOM*, 2008, pp. 1–5.

[89] P. Engelstad and O. Osterbo, "Analysis of the total delay of IEEE 802.11e EDCA and 802.11 DCF," in *Proc. of IEEE International Conference on Communications*, vol. 2, 2006, pp. 552–559.

[90] K. Huang, K. R. Duffy, and D. Malone, "On the validity of IEEE 802.11 MAC modeling hypotheses," *IEEE/ACM Transactions on Networking*, vol. 18, no. 6, pp. 1935–1948, 2010.

[91] M. Garetto and C. f. Chiasserini, "Performance analysis of the 802.11 distributed coordination function under sporadic traffic," *Springer-Networking*, 2005.

[92] J. Martin, *Communication Satellite Systems.* Prentice Hall, 1978.

[93] A. J. Viterbi, *Principles of Spread Spectrum Communication.* Addison-Wesley, 1995.

[94] C. Coutras, S. Gupta, and N. B. Shroff, "Scheduling of real-time traffic in IEEE 802.11 wireless LANs," *Wireless Networks*, vol. 6, no. 6, pp. 457–466, 2000.

[95] B. Sikdar, "Delay analysis of IEEE 802.11 PCF MAC based wireless networks," in *Proc. of GLOBECOM*, vol. 1, 2005, pp. 5–28.

[96] H. Li, D. Yu, and Y. Gao, "Spatial synchronous TDMA in multihop radio network," in *Proc. of IEEE International Conference on Vehicular Technology*, vol. 3, 2004, pp. 1334–1338.

[97] Z. Yang and J. Garcia-Luna-Aceves, "Hop-reservation multiple access (HRMA) for ad-hoc networks," in *Proc. of IEEE INFOCOM*, vol. 1, 1999, pp. 194–201.

[98] A. Muqattash and M. Krunz, "CDMA-based MAC protocol for wireless ad hoc networks," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking & Computing*, 2003, pp. 153–164.

[99] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proc. of International Conference on Mobile Computing and Networking*, 2001, pp. 70–84.

[100] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," in *Proc. of International Conference on Mobile Computing and Networking*, 2001, pp. 85–96.

[101] A. Cerpa and D. Estrin, "ASCENT: Adaptive self-configuring sensor networks topologies," in *Proc. of IEEE INFOCOM*, vol. 3, 2002, pp. 1278–1287.

[102] X. Chen, H. Refai, and X. Ma, "A quantitative approach to evaluate dsrc highway inter-vehicle safety communication," in *Proc. of IEEE GLOBECOM*, 2007, pp. 151–155.

[103] M. Xiaomin, C. Xianbo, R. Hazem H *et al.*, "Performance and reliability of DSRC vehicular safety communication: a formal analysis," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–13, 2009.

[104] M. Hassan, H. Vu, and T. Sakurai, "Performance analysis of the IEEE 802.11 MAC protocol for DSRC safety applications," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3882–3896, 2011.

[105] X. Yin, X. Ma, and K. S. Trivedi, "An interacting stochastic models approach for the performance evaluation of DSRC vehicular safety communication," *IEEE Transactions on Computers*, vol. 62, no. 5, pp. 873–885, 2013.

[106] R. de Haan, "Queueing models for MANETs," Ph.D. dissertation, 2009.

[107] D. Vassis and G. Kormentzas, "Performance analysis of ieee 802.11 ad hoc networks in the presence of exposed terminals," *Ad Hoc Networks*, vol. 6, no. 3, pp. 474–482, 2008.

[108] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: part IV-stability considerations and dynamic control in carrier sense multiple access," *IEEE Transactions on Communications*, vol. 25, no. 10, pp. 1103–1119, 1977.

[109] J. Sheng and K. Vastola, "Performance modeling of 802.11 ad hoc networks with time-varying carrier sense range and physical capture capability," in *Proc. of IEEE GLOBECOM*, 2008, pp. 1–5.

[110] P. C. Ng and S.-C. Liew, "Throughput analysis of IEEE 802.11 multi-hop ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 2, pp. 309–322, 2007.

[111] S. Ray, D. Starobinski, and J. B. Carruthers, "Performance of wireless networks with hidden nodes: A queuing-theoretic analysis," *Computer Communications*, vol. 28, no. 10, pp. 1179–1192, 2005.

[112] K. Medepalli and F. Tobagi, "Towards performance modeling of IEEE 802.11 based wireless networks: A unified framework and its applications," in *Proc. of IEEE INFOCOM*, 2006, pp. 1–12.

[113] N. Bisnik and A. A. Abouzeid, "Queuing network models for delay analysis of multihop wireless ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 1, pp. 79–97, 2009.

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

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

[116] W. K. Grassmann, "Computational methods in probability theory," *Handbooks in Operations Research and Management Science*, vol. 2, pp. 199–254, 1990.

[117] A. Rindos, S. Woolet, I. Viniotis, and K. Trivedi, "Exact methods for the transient analysis of nonhomogeneous continuous time markov chains," in *Computations with Markov Chains*, W. Stewart, Ed. Springer US, 1995, pp. 121–133.

[118] P. Glynn, *Operations Research and Management Science: Stochastic Models*. North Holland, New York, 1990, ch. Diffusion Approximations, pp. 145–198.

[119] D. Jagerman, "Approximating mean waiting times in transient GI/G/1 queues," *Bell System Technical Journal*, vol. 61, no. 8, 1982.

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

[121] 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, 2004.

[122] 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)*, 2008, pp. 77:1–77:9.

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

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

[125] 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)*, 2010, pp. 271–280.

[126] 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.

[127] ——, "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, pp. 1090–1096.

[128] K. Xu, D. Tipper, Y. Qian, P. Krishnamurthy, and S. Tipmongkonsilp, "Time-varying performance analysis of multihop wireless networks with CBR traffic," *accepted for publication in IEEE Transactions on Vehicular Technology*, 2014.

[129] B.-A. H and D. Tipper, "A call level adaptive bandwidth allocation scheme based on lyapunov control theory," in *Proc. of IEEE Symposium on Computers and Communications (ISCC)*, June 2006, pp. 655–661.

[130] P. Siripongwutikorn, "A survey of adaptive bandwidth control algorithms," *IEEE Communications Surveys*, vol. 5, no. 1, pp. 2–14, 2003.

[131] 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.

[132] Y. Yao, L. Rao, X. Liu, and X. Zhou, "Delay analysis and study of IEEE 802.11p based DSRC safety communication in a highway environment," in *Proc. IEEE INFOCOM*, 2013, pp. 1591–1599.

[133] W. Kramer and M. Langenbach-Belz, "Approximate formulae for the delay in the queuing system GI/G/1," in *Proc. of International Teletraffic Congress*, 1976, p. 235.

[134] T. Sobh, K. Elleithy, and A. Mahmood, *Novel Algorithms and Techniques in Telecommunications and Networking*. Springer, 2010.

[135] J. R. Dormand and P. J. Prince, "A family of embedded runge-kutta formulae," *Journal of computational and applied mathematics*, vol. 6, no. 1, pp. 19–26, 1980.

[136] P. Bogacki and L. F. Shampine, "A 3 (2) pair of Runge-Kutta formulas," *Applied Mathematics Letters*, vol. 2, no. 4, pp. 321–325, 1989.

[137] L. F. Shampine and M. K. Gordon, *Computer solution of ordinary differential equations: the initial value problem*. WH Freeman San Francisco, 1975.

[138] L. F. Shampine, M. W. Reichelt, and J. A. Kierzenka, "Solving index-1 DAEs in MATLAB and Simulink," *SIAM review*, vol. 41, no. 3, pp. 538–552, 1999.

[139] L. F. Shampine and M. W. Reichelt, "The matlab ode suite," *SIAM journal on scientific computing*, vol. 18, no. 1, pp. 1–22, 1997.

[140] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.

[141] M. Torrent-Moreno, "Inter-vehicle communications: achieving safety in a distributed wireless environment," Ph.D. dissertation, University Karlsruhe, 2007.

[142] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: a framework to systematically analyze the impact of mobility on performance of routing protocols for ad-hoc networks," in *Proc. of IEEE INFOCOM*, 2003, pp. 825–835.

[143] 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.

[144] A. R. Odoni and E. Roth, "An empirical investigation of the transient behavior of stationary queueing systems," *Operations Research*, vol. 31, no. 3, pp. 432–455, 1983.

[145] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghi, and E. Knightly, "Distributed multi-hop scheduling and medium access with delay and throughput constraints," in *Proc. of ACM MobiCom*, 2001, pp. 200–209.

[146] A. Veres, A. T. Campbell, M. Barry, and S. Li-Hsiang, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2081–2093, 2001.

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

[148] 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.

[149] W. Whitt, "The queueing network analyzer," *Bell System Technical Journal*, vol. 62, no. 9, pp. 2779–2815, 1983.

[150] L. M. Feeney, "An energy consumption model for performance analysis of routing protocols for mobile ad hoc networks," *Mobile Networks and Applications*, vol. 6, no. 3, pp. 239–249, 2001.

[151] K. Xu and I. Howitt, "Realistic energy model based energy balanced optimization for low rate WPAN network," in *Proc. of IEEE Southeastcon Conference*, 2009, pp. 261–266.

[152] K. Xu and M. Zhou, "Energy balanced chain in IEEE 802.15.4 low rate WPAN," in *Proc. of International Conference on Computing, Networking and Communications (ICNC)*, 2013, pp. 1010–1015.

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