Final version of this manuscript is publisehd in Computer Networks, Vol. 54, Issue 4, 19 March 2010, Pages 573-588.

A Unified Model for Joint Throughput-Overhead Analysis of Mobile Ad Hoc Networks

Zhenzhen Ye and Alhussein A. Abouzeid
Dept. of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, USA
yez2@rpi.edu, abouzeid@ecse.rpi.edu

ABSTRACT

We develop an analytical framework to study the throughput and routing overhead for proactive and reactive routing strategies in random access mobile ad hoc networks. To characterize the coexistence of the routing control traffic and data traffic, we model the interaction as a multi-class queue model at each node, where the aggregate control traffic and data traffic are two different classes of customers of the queue. We investigate the scaling property of the throughput, maximum mobility degree supported by the network and mobility-induced throughput deficiencies, under both classes of routing strategies. The proposed analytical model can be extended to incorporate various optimization techniques in routing. We present one exemplary technique and discuss its impacts on the scaling properties of throughput and routing overhead. The connection between the derived throughput result and some well-known network throughput capacity results in the literature is also addressed.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Modeling techniques; C.2.2 [**Network Protocols**]: routing protocols

General Terms

Performance, Theory

Keywords

Routing, throughput, overhead, random access, mobile ad hoc networks

1. INTRODUCTION

Mobile ad hoc networks (MANETs) support a variety of new applications in many military and civilian settings due to the availability of portable wireless communication devices and the flexibility offered when networking them. A fundamental research question is the capability of supporting data traffic in large-scale MANETs, i.e., how feasible a large-scale MANET is. Several theoretical results have been recently discovered for the extreme performance points of a MANET under certain assumptions e.g. [1, 2, 3, 5, 4]. All of these works focus on characterizing the fundamental scaling properties of the performance of a MANET without the restriction of the choice of routing/relaying scheme used in the network – typically an idealized scheme is assumed. In parallel, various practical routing protocols have been proposed for MANETs (e.g. see [6, 7] and references therein). The performance of these routing protocols has been primarily evaluated by simulations. Simulation results reported in the literature show that the performance of a routing protocol heavily depends on several key parameters such as node density, mobility degree, traffic pattern, etc., to name a few. However, simulation itself as a tool is limited in that it provides no analytical expressions of the impacts of one or more of these parameters on the performance. Simulations also do not scale well. This work fills in this gap.

Thus, the key question that this work addresses is: *Given a particular routing strategy, how does the achievable performance scale with the network parameters?*

In this paper, we present a framework for addressing this question and provide answers for particular classes of routing protocols of MANETs: *proactive routing* and *reactive routing*. The number of nodes and the mobility degree in the network are two of the most important network parameters in evaluating the feasibility of a routing strategy in a large-scale dynamic wireless network. For the mobility metric, we use the average relative speed in the network, which has been empirically shown to be a good metric to quantify the dynamics of a MANET [8].

Unlike some of the recent theoretical works in studying the achievable throughput in a MANET that assume that mobility can be used for improving throughput, such as [2, 3, 4, 5], mobility actually plays an *opposite* role in many practical routing protocols, such as in situations with localized mobility, or mobility with variance that is much lower than the tolerable application delay. It is well known that under these conditions, mobility does not improve throughput. For example, in proactive routing, mobility-induced link breakages initiate routing-layer actions to update and propagate the topology information. In reactive routing, mobility-induced path breakages initiate routing-layer actions for repairing or re-discovering path(s) between the corresponding source-destination pair(s). In either case, the control overhead is generated and propagated over the network. (In this paper, the terms of *routing overhead*, *control overhead* and *control traffic* are exchangeable.) Such control traffic consumes a portion of network resources and thus affects the achievable throughput of the network. This issue has not been sufficiently considered in the theoretical literature.

Thus, in this paper, the coexistence of such mobility-induced control traffic with the user-generated (i.e. applications) data traffic in the network motivates the development of a generic *multi-class* queue model at a node, where the *aggregate* control traffic and *aggregate* data traffic are two different classes of customers of the queue. From the stability requirement of a queue, we first analyze the throughput of the network under both proactive and reactive routing strategies and characterize its scaling properties. Then we study the impact of mobility on the throughput performance of the network from two metrics: (i) the *critical* degree of mobility beyond which all the capacity of the network is consumed by control traffic; and (ii) the *mobility-induced throughput deficiency* which quantifies the negative effect of mobility on the throughput of proactive/reactive routing. The proposed analytical model can be readily extended to analyze the routing strategies incorporated with various optimization techniques, such as expanding ring search [9], route caching [10] and multi-point relaying [11]. We present one such technique - node selection in flooding operations, as an example, and investigate its impacts on the scaling properties of throughput and routing overhead. We also discuss the connection between the derived throughput result and well-known network throughput capacity results in [1, 2].

2. NETWORK MODEL

2.1 Topology and Traffic Setting

We consider $n + 1$ mobile nodes randomly distributed in a torus of unit area. The position of any node is assumed to be a stationary random process with stationary distribution uniform on the area of interest and the trajectories of different nodes are independently and identically distributed; each node is equipped with an omnidirectional antenna and the transmission range is given by r; the number of neighbors of any node is also a stationary random process with expected value $n\pi r^2$.

The external data traffic is uniformly distributed on each node within the network; each node is the source for one session, and the destination of another session[1]; the source-destination association does not change with time, although the nodes themselves move; sessions are independent and each session has the same average[2] data packet rate λ_d^E (i.e., throughput per session).

The (active) path length of any session[3], measured in number of hops, is assumed to be a stationary random process with the expected path length $\overline{H}$.

The size of the transmission buffer at a node is assumed to be infinite; the transmission of a packet is controlled by a random access MAC protocol.

2.2 Routing Strategies

The routing strategies under investigation are generic but have preserved basic properties in practical proactive and reactive routing.

2.2.1 *Proactive Routing*

The basic operation in proactive routing is to *periodically* update the topology information by each node. To make the topology information consistent in the network, this information is usually propagated over the whole network. In our proactive routing protocol model, each node has a clock that is not required to be synchronized. Each node notifies its existence via periodic transmission of "HELLO" messages to its neighboring nodes, and detects the link changes with its neighboring nodes by listening to the transmission of "HELLO" messages from its neighbors. All nodes follow the same time period for transmitting "HELLO" messages. In addition, each node periodically sends the link status changes, either detected by itself or received from another node, to its neighboring node. From the received "HELLO" messages and link change messages (if any), each node updates the stored network topology information on it and if necessary, the routes to destination nodes are also updated. The route between any source-destination node pair found by the protocol follows a shortest path [6]. The "HELLO" messages are used to maintain the local connectivity of a node, which has been widely adopted in wireless communication protocols to detect the PHY/MAC layer link status changes [12].

The control overhead in proactive routing consists of link change messages and "HELLO" messages. We denote the link change rate per node as $\lambda_{c,l}$ and the rate of transmitting "HELLO" messages at any node is $\lambda_{c,h}$. We assume that the control packet length for reporting a link change is the same as that of a "HELLO" message.

2.2.2 *Reactive Routing*

Unlike proactive routing, the routes in reactive routing are discovered and maintained in an *on-demand* fashion. In our reactive routing protocol model, each node also uses periodic "HELLO" messages to maintain its local connectivity. If there is mobility-induced path breakage between a source-destination node pair[4], a route-error (RERR) message will be sent to the source node along the path, by the node who detects the breakage. Once the RERR message arrives at the source node, the source node will initiate a route discovery process by flooding the route-request (RREQ) message over the network. Once the RREQ message reaches the destination node, the destination node will reply to this route request by unicasting a route-reply (RREP) message back to the source node, along the forwarding path of the first arrival RREQ[5]. In this basic version of reactive routing protocol model, we do not consider optimization techniques such as expanding ring search [9] which reduces the flooding overhead or route caching [10] which reduces the delay in route discovery. We refer interested readers to [22] for a detailed discussion on extensions of our analytical model to incorporate the impacts of these techniques.

The control overhead in reactive routing consists of RREQ, RREP and RERR messages and "HELLO" messages. The rates of RREQ, RREP and RERR messages are closely related to the breakage rate of a path. The path breakage

[1] The same setting has been used in [2, 3].

[2] All rates are average unless otherwise noted, and thus we will drop the word "average" in most of the discussions since it is obvious.

[3] In the case that multiple paths are available for a session, only one path is active and others are for backup.

[4] In this paper, we use "path" and "route" inter-changeably.

[5] We assume that links are bi-directional.

rate of any session is denoted by $\lambda_{c,p}$. The rate of transmitting "HELLO" messages at any node is the same as that in proactive routing (i.e., $\lambda_{c,h}$) since it only depends on local connectivity of a node. We also assume that the control packet lengths of RREQ, RREP and RERR are the same as that of a "HELLO" message.

2.3 A Multi-Class Queue Model for Data and Control Traffics

When control overhead is taken into account, each individual node can be modeled as a multi-class queueing system with two types of traffic, control packets and data packets. With the given traffic setting in Section 2, we observe that a large portion of the traffic going through a node is the relayed traffic for other source-destination pairs. This is especially true for large-scale networks [1]. Therefore the *aggregate* control traffic and *aggregate* data traffic on a node are approximately independent, though the data and control traffic from the same session are dependent. We allow any service discipline used at each node, as long as the service discipline is independent of sessions. The arrival rate of the aggregate data (control) traffic at the queue of a node is denoted as λ_d (λ_c), and the service rate for the aggregate data (control) traffic at the queue of a node is denoted as μ_d (μ_c). The utilization of a node for data (control) traffic is thus given by $\rho_d \triangleq \lambda_d/\mu_d$ ($\rho_c \triangleq \lambda_c/\mu_c$) and the total utilization of a node is $\rho \triangleq \rho_d + \rho_c$. The main notations used in this paper are summarized in Table 1.

One key issue in the proposed multi-class queue model is to determine the arrival rates and service rates of aggregate data and control traffic at a node. The arrival rate of the aggregate control traffic λ_c depends on the specific routing strategy used in the network and thus the proactive routing and reactive routing have different values of this arrival rate, which will be derived in Section 3. The service rates μ_d and μ_c will be given in Section 4. For the arrival rate of the aggregate data traffic λ_d, since routes found by both proactive routing and reactive routing strategies follow the shortest path fashion [6], it can be derived as follows.

For any session, given the length of the route between the source and the destination to be $h(\geq 1)$ hops, the data packet rate generated by the session (i.e., λ_d^E) will contribute to the arrival rates of data traffic at the queues of the source and $(h-1)$ relaying nodes. Thus the total data traffic rate contributed to the network by this session is $h\lambda_d^E$. There are $(n+1)$ independent sessions with the same data packet rate. Given that the path lengths of these sessions as $\{h_1, ..., h_{n+1}\}$, the total data traffic rate in the network is $\sum_i h_i\lambda_d^E$. Then, unconditioned on $\{h_1, ..., h_{n+1}\}$, the total data traffic rate in the network is $(n+1)\overline{H}\lambda_d^E$. On the other hand, for any session, given the positions of the source and the destination and the corresponding length of the route to be $h > 1$, any other node except the source and the destination has an equal probability to serve as a relay of the session since this probability only depends on the positions of nodes and nodes have an independent and identical stationary distribution on position. This indicates that nodes have the same traffic rate for relaying data. As every node also generates data traffic at the same rate (i.e., λ_d^E), the arrival rate of the aggregate data traffic at the queue of any node is thus given by

$$\lambda_d = \frac{1}{n+1}\left[(n+1)\overline{H}\lambda_d^E\right] = \overline{H}\lambda_d^E. \quad (1)$$

Table 1: List of Notations

Notation	Description
r	The transmission range of a node
$\overline{H}$	Average path length of a session (in hops)
$\bar{v}$	Average relative speed of nodes in the network
$\bar{u}$	Average absolute speed of a node
W	The transmission rate of a node (bit per sec)
L	A data packet length (in bit)
βL	A control packet length (in bit)
$1/\xi$	The mean duration of the back-off timer
Γ	Mobility-induced relative throughput deficiency
λ_d^E	Data packet rate (i.e., throughput) per session
$\lambda_{c,l}$	Link change rate per node
$\lambda_{c,h}$	"HELLO" message rate per node
$\lambda_{c,p}$	The breakage rate of a path
λ_c	The aggregate control traffic arrival rate at a node
λ_d	The aggregate data traffic arrival rate at a node
μ_c	The service rate of control traffic at a node
μ_d	The service rate of data traffic at a node
ρ_c	The utilization of a node for control traffic
ρ_d	The utilization of a node for data traffic
ρ	The utilization of a node

3. CONTROL TRAFFIC

3.1 The Mobility-induced Control Traffic in Proactive Routing

In proactive routing, the amount of mobility-induced control traffic at any node is determined by the link change rate $\lambda_{c,l}$ and the rate of transmitting "HELLO" messages $\lambda_{c,h}$. We determine these two quantities as follows.

To derive the link change rate $\lambda_{c,l}$ at an arbitrary node A, we consider the (possible) link between node A and another arbitrarily chosen node B. Since both node A and B are mobile, we consider the movement of node B relative to node A. We assume that the relative movement of node B at any time satisfies two conditions:

1. the relative speed $v \in [0, v_{max}]$ follows an arbitrary distribution $f_V(v)$, where $v_{max} > 0$ is the finite maximum relative speed of a node;

2. the relative movement direction $\theta \in [0, 2\pi)$ follows a *uniform* distribution.

These two conditions are compatible to our assumptions on the node (absolute) mobility in Section 2.1. One well-known mobility model satisfying these two conditions is the random-walk mobility model [14].

At any time t, consider a small time interval $[t, t+\Delta t]$ where $\Delta t > 0$. When the time interval is sufficiently small, the relative movement of node B can be seen as invariant, i.e., the relative speed v and direction θ are constants in this small interval. Denote the distance between node A and node B at time t as $d_{AB}(t)$. To find out the link change rate between node A and B, we consider two cases.

In the first case, node B is outside of the transmission range of node A at time t, i.e., $d_{AB}(t) > r$. There is a link change at time $t+\Delta t$ if $d_{AB}(t+\Delta t) \leq r$. We note that

$$\begin{aligned} & P[d_{AB}(t+\Delta t) \leq r, d_{AB}(t) > r|v] \\ = & \int_r^{r+v\Delta t} P[d_{AB}(t+\Delta t) \leq r|v,x] f_{d_{AB}}(x|v)dx \\ = & \int_r^{r+v\Delta t} P[d_{AB}(t+\Delta t) \leq r|v,x] 2\pi x dx, \end{aligned}$$

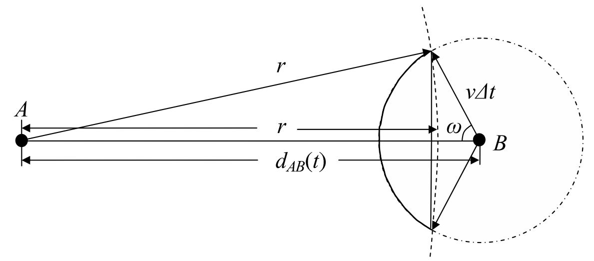

Figure 1: The illustration of the approximation (3) when node B is outside of the transmission range of node A at time t.

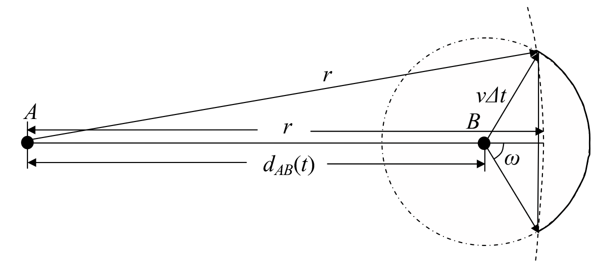

Figure 2: The illustration of the approximation (6) when node B is within the transmission range of node A at time t.

where $f_{d_{AB}}(x|v) = f_{d_{AB}}(x)$ is the probability density function (PDF) of the (random) distance between node A and B, which is independent of v; the upper-limit $r + v\Delta t$ of the integration is due to that $P[d_{AB}(t + \Delta t) \le r|v, d_{AB}(t) > r + v\Delta t] = 0$. For a given distance $d_{AB}(t) \in (r, r + v\Delta t]$,

$$P[d_{AB}(t + \Delta t) \le r|v, d_{AB}(t)] = \frac{2\omega}{2\pi}, \tag{2}$$

which is the probability that node B reaches the (solid) arc in Fig. 1 at time $t + \Delta t$, since node B has equal probability to reach any point on the circle centered at the position of node B at time t with a radius $v\Delta t$. For a sufficiently small Δt,

$$\omega \approx \arccos\left(\frac{d_{AB}(t) - r}{v\Delta t}\right). \tag{3}$$

With (2) and(3), we have

$$\begin{aligned} & P[d_{AB}(t + \Delta t) \le r, d_{AB}(t) > r|v] \\ \approx & \int_r^{r+v\Delta t} \frac{1}{\pi} \arccos\left(\frac{x - r}{v\Delta t}\right) 2\pi x dx = 2v\Delta t\left(r + \frac{\pi v\Delta t}{8}\right) \end{aligned} \tag{4}$$

In the second case, when node B is inside of the transmission range of node A at time t, i.e., $d_{AB}(t) \le r$, there is a link change at time $t + \Delta t$ if $d_{AB}(t + \Delta t) > r$. Similar to the first case, we have

$$\begin{aligned} & P[d_{AB}(t + \Delta t) > r, d_{AB}(t) \le r|v] \\ = & \int_{r-v\Delta t}^{r} P[d_{AB}(t + \Delta t) > r|v, x] 2\pi x dx, \end{aligned}$$

where the lower-limit $r - v\Delta t$ of the integration is due to that $P[d_{AB}(t + \Delta t) > r|v, d_{AB}(t) < r - v\Delta t] = 0$. For a given distance $d_{AB}(t) \in (r - v\Delta t, r]$,

$$P[d_{AB}(t + \Delta t) > r|v, d_{AB}(t)] = \frac{2\omega}{2\pi}, \tag{5}$$

which is the probability that node B reaches the (solid) arc in Fig. 2 at time $t + \Delta t$. For a sufficiently small Δt,

$$\omega \approx \arccos\left(\frac{r - d_{AB}(t)}{v\Delta t}\right). \tag{6}$$

With (5) and (6), we have

$$\begin{aligned} & P[d_{AB}(t + \Delta t) > r, d_{AB}(t) \le r|v] \\ \approx & \int_{r-v\Delta t}^{r} \frac{1}{\pi} \arccos\left(\frac{r - x}{v\Delta t}\right) 2\pi x dx = 2v\Delta t\left(r - \frac{\pi v\Delta t}{8}\right) \end{aligned} \tag{7}$$

Let $e_{AB}(\Delta t) \in \{0, 1\}$ be an indicator that there is a link change between node A and node B in a small interval Δt, where 1(0) denotes there is (not) a link change. Combining the above two cases, the probability that $e_{AB}(\Delta t) = 1$ is given by

$$\begin{aligned} & P[e_{AB}(\Delta t) = 1] \\ = & \int_0^{v_{max}} P[d_{AB}(t + \Delta t) > r, d_{AB}(t) \le r|v] f_V(v) dv \\ + & \int_0^{v_{max}} P[d_{AB}(t + \Delta t) \le r, d_{AB}(t) > r|v] f_V(v) dv \\ = & \int_0^{v_{max}} 4vr\Delta t f_v(v) dv \\ = & 4\bar{v}r\Delta t, \end{aligned} \tag{8}$$

where $\bar{v}$ is the average relative speed of a node.

Therefore the number of link changes associated with node A in a small interval Δt is $\sum_{B \ne A} e_{AB}(\Delta t)$ and its expected value is given by

$$\mathbb{E}\left[\sum_{B \ne A} e_{AB}(\Delta t)\right] = \sum_{B \ne A} \mathbb{E}[e_{AB}(\Delta t)] = 4n\bar{v}r\Delta t.$$

Since each link change event is observed by two nodes associated with the link, the link change rate *per node* is

$$\lambda_{c,l} = \lim_{\Delta t \to 0} \frac{1}{2} \frac{\mathbb{E}\left[\sum_{B \ne A} e_{AB}(\Delta t)\right]}{\Delta t} = 2n\bar{v}r. \tag{9}$$

On the other hand, as the periodic "HELLO" messages are used to maintain the local connectivity of a node and the status of the local connectivity changes at the same rate as the link change rate per node, the rate of transmitting "HELLO" messages at a node should be proportional to $2n\bar{v}r$. Therefore,

$$\lambda_{c,h} = 2c_1 n\bar{v}r, \tag{10}$$

where $c_1 > 0$ is a constant.

In proactive routing, the link change message is flooded over the network. Without any optimization in the flooding operation, each node in the network will re-broadcast the link change message once. Thus, the link change rate per node $\lambda_{c,l}$ contributes to the arrival rate of the control traffic at the queue of each node in the network. As there are $(n+1)$ nodes, the control traffic rate arriving at the queue of each node, due to link changes, is $(n + 1)\lambda_{c,l}$. On the other hand, since a "HELLO" message is only broadcasted to one-hop neighboring nodes, it only contributes to the arrival control traffic at the queue of the transmitting node. Therefore, the arrival rate of the aggregate control traffic at the queue of any node is

$$\lambda_c = (n + 1)\lambda_{c,l} + \lambda_{c,h} = (n + 1 + c_1)2n\bar{v}r. \tag{11}$$

3.2 The Mobility-induced Control Traffic in Reactive Routing

To determine the mobility-induced control traffic in reactive routing, the key is to find the path breakage rate $\lambda_{c,p}$. Given that there is a link between an arbitrary node pair A and B at time t, i.e., $d_{AB}(t) \leq t$, and the relative speed between A and B is v, the probability that the link is broken during a small time interval Δt is given by

$$\begin{aligned} & P[d_{AB}(t+\Delta t) > r | v, d_{AB}(t) \leq r] \\ = & \frac{P[d_{AB}(t+\Delta t) > r, d_{AB}(t) \leq r | v]}{P[d_{AB}(t) \leq r|v]} \\ = & \frac{\int_{r-v\Delta t}^{r} P[d_{AB}(t+\Delta t) > r | v, x] 2\pi x dx}{\pi r^2} \\ \approx & \frac{1}{\pi r^2} 2v\Delta t \left(r - \frac{\pi v \Delta t}{8} \right). \end{aligned} \quad (12)$$

Unconditioned on v, the link breakage probability of the link between node A and node B is

$$p_l(\Delta t) \triangleq P[d_{AB}(t+\Delta t) > r | d_{AB}(t) \leq r] = \frac{2\bar{v}\Delta t}{\pi r} - \frac{\overline{v^2}(\Delta t)^2}{4r^2}, \quad (13)$$

where $\overline{v^2} \triangleq \mathbb{E}[v^2]$.

Consider a path consisting of $h(\geq 1)$ links between a source-destination pair of any given session. Although the consecutive links along the path are generally dependent, the dependence between links which do not share a common node is negligible [14, 13]. Indexing the links as $l = 1, 2, ..., h$, according to their distances to the source node. We select a set of links with odd indices (i.e., $S_l \triangleq \{1, 3, 5, ...\}$). Since the links in S_l are separated at least one hop, they are approximately independent of each other. Let $P_p(\Delta t|h)$ be the probability that the path is broken in the interval Δt, we have

$$P_p(\Delta t|h) \geq 1 - (1 - p_l(\Delta t))^{h/2}, \quad (14)$$

where $p_l(\Delta t)$ is given in (13). On the other hand, let $l_i,\ i = 1, ..., h$ be the indicator of the link breakage of ith link in the path, with the help of union bound, we have

$$P_p(\Delta t|h) = P\left[\cup_i (l_i = 1)\right] \leq \sum_i P[l_i = 1] = h p_l(\Delta t). \quad (15)$$

From (14) and (15), the path breakage rate for a path with $h(\geq 1)$ links is bounded by

$$\lim_{\Delta t \to 0} \frac{1 - (1 - p_l(\Delta t))^{h/2}}{\Delta t} \leq \lim_{\Delta t \to 0} \frac{P_p(\Delta t|h)}{\Delta t} \leq \lim_{\Delta t \to 0} \frac{h p_l(\Delta t)}{\Delta t}$$

$$\Rightarrow \frac{\bar{v}h}{\pi r} \leq \lim_{\Delta t \to 0} \frac{P_p(\Delta t|h)}{\Delta t} \leq \frac{2\bar{v}h}{\pi r}. \quad (16)$$

The result in (16) is almost the same as the empirical formula given in [8] where the empirical formula is obtained from extensive simulation on different reactive routing protocols. Unconditioned on the path length h, the path breakage rate $\lambda_{c,p}$ is bounded by

$$\frac{\bar{v}\overline{H}}{\pi r} \leq \lambda_{c,p} \leq \frac{2\bar{v}\overline{H}}{\pi r}. \quad (17)$$

As mentioned in Section 2.2, the control traffic consists of RREQ, RREP and RERR messages and "HELLO" messages. The rates of RREQ, RREP and RERR messages for discovering and maintaining routes are closely related to $\lambda_{c,p}$. We specify the control traffic rates induced by RREQ, RREP and RERR as follows.

- RREQ traffic rate per node: the RREQ message is flooded over the network. Without any optimization in this flooding operation, all nodes except the destination node of the RREQ message will re-broadcast the RREQ message. For any session, the rate of this flooding operation is the same as the path breakage rate $\lambda_{c,p}$. Since any node participates the flooding operations of n sessions (except the session that it is the destination), the RREQ traffic rate at any node is $n\lambda_{c,p}$.

- RREP traffic rate per node: in any session, the RREP message is unicasted back from the destination node to the source node along the path discovered by the flooding operation. The rate of this RREP operation is the same as the path breakage rate $\lambda_{c,p}$. If the path found by the flooding operation has $h(\geq 1)$ hops, the RREP traffic will contribute to the control traffic at the queues of the destination node and $(h-1)$ relaying nodes. Since there are $(n+1)$ independent sessions with the identical path length distribution, the total RREP traffic rate in the network is given by $(n+1)\overline{H}\lambda_{c,p}$. Thus the RREP traffic rate per node is $\overline{H}\lambda_{c,p}$.

- RERR traffic rate per node: in any session, when there is a path breakage, the RERR message is unicasted back to the source node along the path, by one of the end nodes of the broken link which is closer to the source node. The rate of this RERR operation is the same as the path breakage rate $\lambda_{c,p}$. If the path breakage (statistically) equally happens at each link, then for an h-hop path, the RERR traffic will contribute to the control traffic at the queues of $(h-1)/2$ nodes along the path, in average. Since there are $(n+1)$ independent sessions with the identical path length distribution, the total RERR traffic rate in the network is $\frac{1}{2}(n+1)(\overline{H}-1)\lambda_{c,p}$. Thus the RERR traffic rate per node is $(\overline{H}-1)\lambda_{c,p}/2$.

The control traffic induced by "HELLO" messages in reactive routing is the same as that in proactive routing, since it only depends on the rate of local topology changes. We thus obtain the arrival rate of the aggregate control traffic at the queue of any node as

$$\lambda_c = \left(n + \frac{3}{2}\overline{H} - \frac{1}{2}\right)\lambda_{c,p} + \lambda_{c,h}. \quad (18)$$

4. A SERVICE MODEL

The (transmission) service for a data/control packet at a node is under the control of the MAC layer protocol. Random access is one of the most popular MAC strategies in large-scale MANETs, due to its simplicity and robustness in dynamic environments. We adopt a random access MAC protocol model, similar to that in [15]. The basic operation of this MAC model is:

1. before transmitting each packet (either a control packet or a data packet), a node counts down a random backoff timer; the duration of the timer is exponentially distributed with a mean $1/\xi$, where ξ is a positive constant;

2. the timer of a node freezes each time when an interfering node starts transmitting a packet, where an *interfering node* is defined as the neighboring nodes which lie within a distance of $2r$ of each other; the transmission duration for a data packet is set to be L/W, where L is the length of a data packet in bit and W is the transmission rate in bit per second; the transmission duration for a control packet is $\beta L/W$, where $\beta > 0$ and far less than one in practice;

3. when the timer expires, the node starts transmitting the packet and the back-off timers of all its interfering neighbors are immediately frozen; the timers will resume once the transmission is completed.

This random access MAC model captures the essential behavior of IEEE 802.11 MAC protocol and also provides sufficient information for determining the service rates μ_d and μ_c in the proposed queue model, though somewhat ideal in avoiding collisions.

Consider an arbitrary node in the network, we define

- the (random) number of interfering nodes around the node as U;
- the (random) number of interfering nodes which have transmission requirement for control (data) packets during the back-off time of the node as M_c (M_d)[6];
- the (random) number of times that the timer of the node is frozen due to the transmissions of control (data) packets of interfering nodes as Z_c (Z_d).

As nodes are uniformly distributed over a unit area, it is straightforward to see that U is binomial distributed. Thus

$$\mathbb{E}[U] = 4\pi r^2 n, . \tag{19}$$

For M_c and M_d, we have

$$\mathbb{E}[M_c] = \sum_u \mathbb{E}[M_c|U=u]P[U=u] = \sum_u \rho_c u P[U=u] = \rho_c \mathbb{E}[U],$$
$$\mathbb{E}[M_d] = \sum_u \mathbb{E}[M_d|U=u]P[U=u] = \sum_u \rho_d u P[U=u] = \rho_d \mathbb{E}[U],$$

where ρ_c (ρ_d) is the utilization of a node for control (data) traffic.

Let T denote the random back-off time of the node. We first consider $\mathbb{E}[Z_c]$. Given $T = t$, $M_c = m_c$ and assuming that m_c is constant during the back-off time of the node, the conditional distribution of Z_c is Poisson with associated parameter $m_c \xi t$ [15], i.e.,

$$P[Z_c = z_c | T = t, M_c = m_c] = \frac{e^{-m_c \xi t}(m_c \xi t)^{z_c}}{z_c!}.$$

It is straightforward to see that $\mathbb{E}[Z_c|T = t, M_c = m_c] = m_c \xi t$ and thus

$$\mathbb{E}[Z_c] = \mathbb{E}[M_c] = \rho_c \mathbb{E}[U]. \tag{20}$$

Similarly, we have

$$\mathbb{E}[Z_d] = \rho_d \mathbb{E}[U]. \tag{21}$$

For a data packet, the (random) service time X_d at the node includes the random back-off time t, total frozen time of the timer due to the transmissions of interfering nodes $Z_c \frac{\beta L}{W} + Z_d \frac{L}{W}$, and the transmission time of the data packet $\frac{L}{W}$, i.e.,

$$X_d = t + Z_c \frac{\beta L}{W} + Z_d \frac{L}{W} + \frac{L}{W}.$$

The mean service time of a data packet is given by

$$\mathbb{E}[X_d] = \frac{1}{\xi} + \frac{L}{W} + (\beta\rho_c + \rho_d)\mathbb{E}[U]\frac{L}{W}. \tag{22}$$

The corresponding service rate of a data packet is

$$\mu_d = \frac{1}{\mathbb{E}[X_d]}. \tag{23}$$

Similarly, for a control packet, the (random) service time X_c at the node includes the random back-off time t, total frozen time of the timer due to the transmissions of interfering nodes $Z_c \frac{\beta L}{W} + Z_d \frac{L}{W}$, and the transmission time of the data packet $\frac{\beta L}{W}$, i.e.,

$$X_c = t + Z_c \frac{\beta L}{W} + Z_d \frac{L}{W} + \frac{\beta L}{W}.$$

The mean service time of a control packet is given by

$$\mathbb{E}[X_c] = \frac{1}{\xi} + \frac{\beta L}{W} + (\beta\rho_c + \rho_d)\mathbb{E}[U]\frac{L}{W}. \tag{24}$$

The corresponding service rate of a control packet is

$$\mu_c = \frac{1}{\mathbb{E}[X_c]}. \tag{25}$$

5. THROUGHPUT ANALYSIS

In this section, we use the proposed queue model to characterize the scaling properties of throughput and overhead in proactive routing and reactive routing, respectively. Specifically, the scaling result of the throughput per source-destination pair (i.e., per session), the maximum mobility degree supported by the network and the mobility-induced throughput deficiencies, under both classes of routing strategies, are investigated.

5.1 The Throughput of Routing Protocols

First, we show the throughput per source-destination pair in both proactive routing and reactive routing in the following theorem.

THEOREM 5.1. *The throughput per source-destination pair (i.e., per session) of a MANET with either proactive routing or reactive routing is*

$$\lambda_d^E < \left[\frac{1 - \left(\frac{1}{\xi} + \frac{\beta L}{W} + 4\pi n r^2 \frac{\beta L}{W}\right)\lambda_c}{1 + \frac{C(n)}{B(n)}\lambda_c} \right] \frac{1}{B(n)\overline{\overline{H}}}, \tag{26}$$

where $C(n) \triangleq (1-\beta)^2 4\pi r^2 n \left(\frac{L}{W}\right)^2$ *and* $B(n) \triangleq \frac{1}{\xi} + \frac{L}{W} + 4\pi r^2 n \frac{L}{W}$.

PROOF. The maximum throughput, i.e., the maximum data packet rate λ_d^E that a node can generate, in a MANET with either proactive routing or reactive routing is limited by the stability of the queue, i.e.,

$$\begin{aligned}
\rho &= \rho_d + \rho_d = \frac{\lambda_d}{\mu_d} + \frac{\lambda_c}{\mu_c} < 1 \\
&\Rightarrow \lambda_d \left(\frac{1}{\xi} + \frac{L}{W}\right) + \lambda_c \left(\frac{1}{\xi} + \frac{\beta L}{W}\right) \\
&\quad + (\lambda_d + \lambda_c)(\beta\rho_c + \rho_d) 4\pi r^2 n \frac{L}{W} < 1. \qquad (27)
\end{aligned}$$

[6] Note that it is possible that a particular node has transmission requirements for both data and control packets.

From (22)-(25), since

$$\rho_d = \lambda_d \left(\frac{1}{\xi} + \frac{L}{W}\right) + \lambda_d \left(\beta\rho_c + \rho_d\right) 4\pi r^2 n \frac{L}{W}$$
$$\rho_c = \lambda_c \left(\frac{1}{\xi} + \frac{\beta L}{W}\right) + \lambda_c \left(\beta\rho_c + \rho_d\right) 4\pi r^2 n \frac{L}{W},$$

we have

$$\rho_d + \beta\rho_c = \frac{\lambda_d \left(\frac{1}{\xi} + \frac{L}{W}\right) + \beta\lambda_c \left(\frac{1}{\xi} + \frac{\beta L}{W}\right)}{1 - (\lambda_d + \beta\lambda_c) 4\pi r^2 n \frac{L}{W}}. \quad (28)$$

Taking (28) into (27), we obtain

$$\lambda_d < \frac{1 - \left(\frac{1}{\xi} + \frac{\beta L}{W} + 4\pi n r^2 \frac{\beta L}{W}\right)\lambda_c}{\left(\frac{1}{\xi} + \frac{L}{W} + 4\pi r^2 n \frac{L}{W}\right) + (1-\beta)^2 4\pi r^2 n \frac{L^2}{W^2}\lambda_c}. \quad (29)$$

From (1), (29) and $B(n), C(n)$ defined above, we obtain the result in (26). □

In (26), we note that $\lambda_d^E > 0$, i.e., the node can generate data traffic only if

$$\lambda_c < \left(\frac{1}{\xi} + \frac{\beta L}{W} + 4\pi n r^2 \frac{\beta L}{W}\right)^{-1}. \quad (30)$$

Recall that $1/\xi$ is the average back-off time for transmission before a node starts to transmit a packet in the random access MAC model and $(1 + 4\pi n r^2)\frac{\beta L}{W}$ is the (average) total transmission duration of control packets within a node's neighborhood when each node has a control packet to send. As each packet is randomly backed-off to avoid collisions in transmission and each node within the neighborhood has an equal opportunity to access the shared wireless channel, the arrived rate of control packets at the queue of a node, i.e., λ_c, should not exceed $\left[\frac{1}{\xi} + (1 + 4\pi n r^2)\frac{\beta L}{W}\right]^{-1}$, otherwise the node is overwhelmed by control traffic and the queue becomes unstable. Therefore no nonzero data rate can be supported.

Consider the asymptotic scenario that the number of nodes n is large (i.e., the density of nodes goes to infinity as the underlying area of the network is fixed as unit). We have the following scaling property of the throughput per source-destination.

Corollary 5.2. *The throughput of a MANET with either proactive routing or reactive routing is*

$$\lambda_d^E = O\left(\frac{W}{\sqrt{n \log n}}\right). \quad (31)$$

Proof. For a mobile network which is asymptotically connected, the critical transmission range (CTR) is given by $r = c_2\sqrt{\frac{\log n}{\pi n}}$, where $c_2 (\geq 1)$ is a constant [16]. On the other hand, as the stationary distribution of nodes is uniform in the area, in a dense network with a large value of n, the shortest path between any source-destination node pair is over nearly straight-line paths [1]. Since the expected distance between any source-destination pair is $\Theta(1)$, we have $\overline{H} r = \Theta(1)$, i.e., $\overline{H} = \Theta(1/r) = \Theta(\sqrt{n/\log n})$. Furthermore, by observing that the term in the bracket in (26) is no more than one and nonnegative when (30) holds, we have

$$\lambda_d^E < \frac{1}{B(n)\overline{H}} < \frac{W}{4\pi L n r^2 \overline{H}}. \quad (32)$$

The result of (31) then follows by observing that the transmission rate W and packet length L are constants. □

As the utilization ρ of a queue can be arbitrarily close to one, the right-hand side of (26) can be seen as the *maximum achievable throughput* in a *mobile* ad-hoc network, we denote it as τ_{max}. When the average relative speed of nodes in the network goes to zero, $\lambda_c \to 0$ and the maximum achievable throughput approaches $\frac{1}{B(n)\overline{H}}$. Thus, let

$$\eta_{max} \triangleq \frac{1}{B(n)\overline{H}} \quad (33)$$

be the maximum achievable throughput in a *static* ad hoc network. From the proof of Corollary 5.2, we see that $\eta_{max} = O(\frac{W}{\sqrt{n \log n}})$, which is comparable to Gupta-Kumar's throughput capacity result for random static networks in [1]. If a perfect deterministic scheduling as in [1] is used, instead of the random access MAC model in our analysis, the order of the throughput is expected to be achievable.

5.2 The Maximum Mobility Degree

From (30), we can characterize the maximum mobility degree that the network can support in proactive routing and reactive routing, respectively.

Theorem 5.3. *With the proactive routing strategy, the critical degree of the average relative speed in the network is*

$$\bar{v} = O\left(\frac{1}{(n \log n)^{3/2}}\right). \quad (34)$$

Proof. Obtained immediately from (11) and (30). □

Theorem 5.4. *With the reactive routing strategy, the critical degree of the average relative speed in the network is*

$$\bar{v} = O\left(\frac{1}{n^2}\right). \quad (35)$$

Proof. Obtained immediately from (17), (18) and (30). □

From Theorem 5.3 and 5.4, we note that, asymptotically, proactive routing is able to support a higher mobility degree than reactive routing. This is because that proactive routing asymptotically has a smaller (mobility-induced) control overhead than reactive routing, under the traffic setting in Section 2.1. Under the given traffic setting, each node is the source of a data session and thus the number of data sessions in the network linearly increases with the number of nodes (i.e., n). Consider the asymptotic scenario such that $n \to \infty$, the arrival rate of the aggregate control traffic at any node in proactive routing (in (11)) is dominated by the rate of broadcasting link change messages, i.e., $(n+1)\lambda_{c,l}$, where $\lambda_{c,l}$ increases with $\sqrt{n \log n}$. On the other hand, as $n \to \infty$, the arrival rate of the aggregate control traffic at any node in reactive routing (in (18)) is dominated by the rate of broadcasting RREQ messages, i.e., $n\lambda_{c,p}$, where $\lambda_{c,p}$ increases with $n/\log n$. Clearly $\lambda_{c,p}$ is asymptotically greater than $\lambda_{c,l}$, and consequently, reactive routing (asymptotically) induces a larger amount of control overhead than proactive routing. Our observation is somewhat surprising as it is usually believed that the *on-demand* type of routing reduces overhead by maintaining information for active routes only [6, 7]. However, one should notice that there are three major differences in network setting between the simulations in the literature (e.g., [17, 18, 19, 20], etc.) and ours. First, all simulations in these works were carried out in networks with a small or moderate size (usually less than 200 nodes in the network) while our result is obtained in an asymptotic scenario such that $n \to \infty$. There is no evidence to show that the simulation results obtained from these relatively small networks are a reliable indicator to the behavior of routing strategies in a large-scale network. Second, a fixed radio transmission range, independent of the network size, is often set in simulations, while our result is derived from a throughput optimal CTR setting [16] such that the transmission range of a node decreases with $\sqrt{\frac{\log n}{n}}$. The fixed transmission range setting is convenient to use in simulation but it can bias the explanations of the simulation results. For example, under the fixed transmission range setting, the increase of network size (i.e., node density in a fixed network region) would have little impact on the length of a path and thus has insignificant impact on the path breakage rate, which is in contrast to the well-known fact that the expected path length of a multihop transmission would increase with the network size to maximize the network throughput capacity. Third, in almost all referenced simulations, the effect of the number of sessions on the network is studied independent of the network size setting, while in this paper the number of sessions in the network linearly increases with the network size. Since it is well-known that the control overhead in reactive routing increases dramatically with the number of sessions in the network while proactive routing is rather insensitive to the increase of the number of sessions [17, 19, 20], our heavy traffic setting is expected to be unfavorable to reactive routing.

5.3 Mobility-Induced Throughput Deficiency

Since the maximum achievable throughput in a *mobile* ad hoc network with either proactive routing or reactive routing, i.e., τ_{max}, is strictly less than η_{max} when the average relative speed in the network is nonzero, *mobility actually decreases the capacity of the network with either proactive routing or reactive routing*. Intuitively, the reason for the decrease of the throughput is that both proactive routing and reactive routing assume the network is *quasi-static*, thus it tries to establish a *stable* route between any source-destination pair and maintain it either through periodically updating the network topology information (in proactive routing) or by detecting path breakage and re-establishing it via flooding route request packets (in reactive routing). In the underlying assumption of both proactive routing and reactive routing, the dynamics of network topology is far slower than the packet forwarding speed (i.e., network flow speed) and thus the objective of maintaining a stable path between any source-destination pair is feasible. Once the mobility is too high, these two routing strategies will not be able to work as the routing control message overheads dominate the network traffic and no data throughput can be supported.

To further quantify the *negative* effect of mobility on network throughput in both proactive routing and reactive routing, we introduce the metric of *mobility-induced relative throughput deficiency*, which is defined as $\Gamma \triangleq (\eta_{max} - \tau_{max})/\eta_{max}$. From (26) and (33), for a network with a fixed size n, we

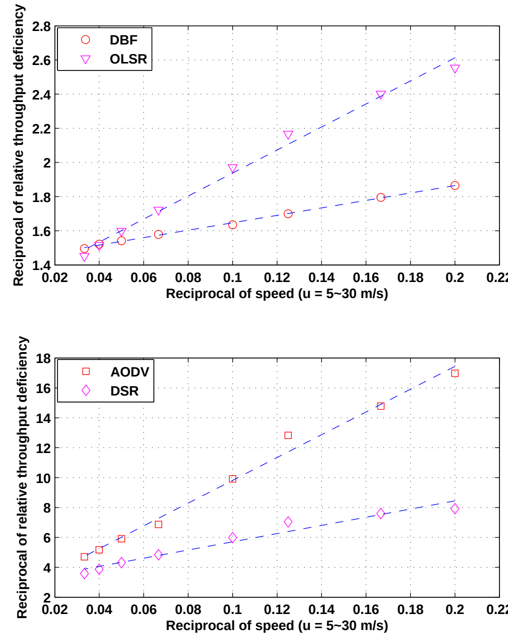

Figure 3: Results from Qualnet simulation indicate a linear relationship between the reciprocal of speed and the reciprocal of average relative throughput deficiency in routing protocols. The fitting curves are shown in dashed lines and the (relative) fitting error is $0.29\%, 1.47\%, 2.87\%$ **and** 4.51% **for DBF, OLSR, AODV and DSR, respectively.**

have

$$\Gamma^{-1} = D(n)\left(\lambda_c^{-1} + \frac{C(n)}{B(n)}\right), \tag{36}$$

where $D(n) \triangleq \left[\frac{C(n)}{B(n)} + \left(\frac{1}{\xi} + \frac{\beta L}{W} + 4\pi r^2 n \frac{\beta L}{W}\right)\right]^{-1}$. Furthermore, from (17), we can set $\lambda_{c,p} = \frac{c_3 \bar{v} \overline{H}}{\pi r}$ for a fixed n, where $c_3 \in [1, 2]$ and might be n dependent. Thus, from (11), (18) and (36), we obtain the following result.

Theorem 5.5. *For both proactive routing and reactive routing, the reciprocal of the mobility-induced relative throughput deficiency (i.e.,* Γ^{-1}*) is linearly increasing with* $\bar{v}^{-1}$*, i.e.,*

$$\Gamma^{-1} = a\bar{v}^{-1} + b, \tag{37}$$

where

$$a \triangleq \begin{cases} [2(n+1+c_1)nr]^{-1} D(n), & \textit{proactive routing} \\ \left[\left(n + \frac{3}{2}\overline{H} - \frac{1}{2}\right) \frac{c_3 \overline{H}}{\pi r} + c_1 2nr\right]^{-1} D(n), & \textit{reactive routing} \end{cases}$$

and $b \triangleq C(n)D(n)/B(n)$.

We demonstrate this linear relationship in a MANET with 40 nodes, using *QualNet* (ver. 3.9) network simulator [21]. Four different routing protocols provided by *QualNet*, i.e., dynamic Bellman-Ford (DBF), optimized link state routing (OLSR), ad hoc on-demand distance vector (AODV) routing and dynamic source routing (DSR), are tested, where the first two and the last two are the examples of proactive routing and reactive routing, respectively. A constant

Table 2: Scalability of Throughput and Overhead in Routing Strategies

w/o NS[1]	Proactive	Reactive
Max Mobility Degree $\bar{v}$	$O\left(\frac{1}{n^{3/2}(\log n)^{3/2}}\right)$	$O\left(\frac{1}{n^2}\right)$
Control Overhead Rate λ_c	$\Theta\left(n\sqrt{n\log n}\bar{v}\right)$	$\Theta\left(\frac{n^2}{\log n}\bar{v}\right)$
Throughput λ_d^E	$O\left(\frac{W}{\sqrt{n\log n}}\right)$	

w/ NS	Proactive	Reactive
Max Mobility Degree $\bar{v}$	$O\left(\frac{1}{n\sqrt{n\log n}}\right)$	$O\left(\frac{\log n}{n^2}\right)$
Control Overhead Rate λ_c	$\Theta\left(\frac{n^{3/2}}{\sqrt{\log n}}\bar{v}\right)$	$\Theta\left(\frac{n^2}{(\log n)^2}\bar{v}\right)$
Throughput λ_d^E	$O\left(\frac{W}{\sqrt{n\log n}}\right)$	

[1] NS = node selection in flooding (see Section 6.1).

(absolute) speed setting u is used in the simulation. Since the average relative speed $\bar{v}$ is linearly increasing with u in a constant speed setting, the mobility degree of the network is controlled by adjusting the value of u. From simulation, we find that the Pearson coefficient of correlation between u^{-1} and the measured relative throughput deficiency is -0.9980, -0.9745, -0.9426 and -0.9355 for DBF, OLSR, AODV and DSR, respectively, which indicates a strong correlation between $\bar{v}^{-1}$ and the relative throughput deficiency in any protocol under test. In Figure 3, for every protocol, the average relative throughput deficiency is obtained from simulation data, and its reciprocal is (approximately) linearly increasing with u^{-1} (or $\bar{v}^{-1}$).

6. EXTENSIONS OF THE MODEL

In this section, we discuss two extensions of the proposed analytical model. One is to analyze the routing strategies with various optimization techniques. As an illustration, we discuss the technique of node selection in flooding operations and its impacts on the scaling properties of throughput and routing overhead. The other is how the model is extended to study a routing strategy other than proactive or reactive routing. By using the well-known Grossglauser-Tse's relaying scheme as an example [2], we show that the technique in Section 5.1 can be extended to analyze this two-hop relaying scheme and the achievable throughput result obtained from (26) is comparable to the result in [2]. Due to space limit, we refer interested readers to [22] for more extensions and discussions of the proposed analytical model.

6.1 Node Selection in Flooding

In the basic version of the flooding operations, each node re-broadcasts the control message once. This is a common case in routing protocols when there is no cooperation among nodes. When the cooperation among nodes is incorporated into the protocol design, the number of nodes participating in the re-broadcasting can be significantly reduced. We show here that an optimized node selection for participating in flooding can improve the maximum mobility degree supported by the network, though there is no change in the scaling property of throughput in (31). The summary of the scaling results is given in Table 2. First, we characterize the minimum number of nodes necessary to participate in a flooding operation. Due to space limit, we skip the proof of the following lemma and refer interested readers to [22].

LEMMA 6.1. *The minimum number of nodes necessary to participate in a flooding operation scales as* $\Theta(n/\log n)$.

From Lemma 6.1, (10) and (17), it is straightforward to see that the arrival rate of the aggregate control traffic at the queue of any node has been reduced to

$$\lambda_c = \begin{cases} \Theta\left(\frac{n^{3/2}}{\sqrt{\log n}}\bar{v}\right) & \text{proactive routing} \\ \Theta\left(\frac{n^2}{(\log n)^2}\bar{v}\right) & \text{reactive routing} \end{cases}.$$

Thus, from (30), we obtain the following results for the maximum relative speed supported by the network.

COROLLARY 6.2. *1. With proactive routing and the optimized node selection in flooding, the critical degree of the average relative speed in the network is*

$$\bar{v} = O\left(\frac{1}{n\sqrt{n\log n}}\right). \tag{38}$$

2. With reactive routing and the optimized node selection in flooding, the critical degree of the average relative speed in the network is

$$\bar{v} = O\left(\frac{\log n}{n^2}\right). \tag{39}$$

In practice, the idea of node selection in flooding has been used in the OLSR protocol, where a multi-point relaying (MPR) strategy is applied to select a set of neighboring nodes to re-broadcast its control message and a significant reduction of overhead in flooding has been observed [11].

6.2 Grossglauser-Tse's Relaying Scheme

One well-known extreme throughput performance point of MANETs is Grossglauser-Tse's throughput capacity result [2]. A two-hop relay scheme has been proposed in [2] to take advantage of a high mobility degree in the network. In the two-hop relaying scheme, each packet is relayed at most once. As there is no attempt to maintain any topology information or a stable route between any source-destination pair in the two-hop relaying scheme, $\lambda_c = 0$ in (26). In addition, similar to the queue model in Fig. 3 in [2], the service discipline of the queue model at any relay node is also unrestricted in our setting. Thus the result in (26) can be directly applied to the two-hop relaying scheme. By observing that $\overline{H} \in [1,2]$ in the two-hop relaying scheme, from (32), we have

$$\lambda_d^E < \frac{1}{B(n)} < \frac{W}{4\pi L n r^2}.$$

Note that our analysis for the random access MAC model is based on the *protocol channel model* [1], not the *physical channel model* used in [2]. In order to find the connection between Grossglauser-Tse's result and our result, we need a comparable transmission range r to that in the physical channel model. In the protocol in [2], any transmission is intended to the nearest neighboring node. They observe that *the received signal power at the nearest neighboring node is of the same order as the total interference from* $\Theta(n)$ *number of interferers.* In other words, the probability of successful transmission of a packet to the nearest neighbor will not vanish with the increase of node density, i.e., $P_r\{SIR > \gamma\} \to \epsilon > 0$, as $n \to \infty$, where γ is the Signal-to-Interference-Ratio (SIR) threshold and ϵ is a positive constant. Equivalently, in our channel model, as the interference area is set as $4\pi r^2$, the transmission will be successful if the nearest neighboring node is within the disk centered at the transmitting node with a radius r. In a steady state, neighboring nodes are randomly uniformly distributed in the unit area with the density n, the PDF of the distance to the nearest neighboring node is given by $f(d) = 2\pi n d e^{-n\pi d^2}$ [23] and the probability that the nearest neighbor is within r is $1 - e^{-n\pi r^2}$. Let $Pr\{SIR > \beta_{th}\} = 1 - e^{-n\pi r^2} = \epsilon$, we obtain the equivalent transmission range r in our channel model as

$$r = \sqrt{\frac{1}{n\pi}\log\left(\frac{1}{1-\epsilon}\right)}.$$

Thus the throughput per source-destination pair is given by

$$\lambda_d^E < \frac{W}{4L\log\frac{1}{1-\epsilon}} \Rightarrow \lambda_d^E = O(1), \qquad (40)$$

which is comparable to the result in [2].

7. CONCLUSIONS

We have proposed an analytical framework to study the throughput and routing overhead for practical routing strategies in random access MANETs. By considering the coexistence of mobility-induced control traffic and data traffic at any node in the network, we have modeled the individual node as a generic multi-class queue. The analysis has shown that (i) the throughput per source-destination pair for both proactive routing and reactive routing is $O(\frac{W}{\sqrt{n\log n}})$; (ii) there is a strong *linear* relationship that characterizes how mobility reduces the throughput per session, for both proactive routing and reactive routing; (iii) the existence of the critical mobility degree under any given routing strategy indicates that the data traffic can be effectively supported only if the maximum mobility degree does not exceed the critical value, otherwise the total capacity of the network is consumed by routing control traffic.

The proposed analytical model can be extended to incorporate various routing design options and has a close connection to well-known throughput capacity results in the literature. This shows that the model serves as an effective tool to characterize the throughput performance of routing strategies in MANETs.

8. REFERENCES

[1] P. Gupta and P. R. Kumar, “The capacity of wireless networks”, *IEEE Trans. Information Theory*, vol. 46, no. 2, pp. 388-404, Mar. 2000.

[2] M. Grossglauser and D. Tse, “Mobility inceases the capacity of ad hoc wireless networks”, *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477-486, Aug. 2002.

[3] G. Sharma, R. Mazumdar and N. Shroff, “Delay and Capacity Trade-offs in Mobile Ad Hoc Networks: A Global Perspective”, in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[4] N. Bansal and Z. Liu, “Capacity, Delay and Mobility in Wireless Ad-Hoc networks”, in *Proc. IEEE INFOCOM*, San Francisco, CA, Apr. 2003.

[5] J. Mammen and D. Shah, “Throughput and delay in Random Wireless networks with Restricted Mobility”, *IEEE Trans. Information Theory*, vol. 53, no. 3, pp. 1108-1116, Mar. 2007.

[6] E. Royer and C.-K. Toh, “A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks”, *IEEE Personal Communications*, vol. 6, no. 2, pp.46-55, Apr. 1999.

[7] M. Abolhasan, T. Wysocki and E. Dutkiewicz, “A review of routing protocols for mobile ad hoc networks”, *Ad Hoc Networks (Elsevier)*, no. 2, pp. 1-22, 2004.

[8] F. Bai, N. Sadagopan, B. Krishnamachari and A. Helmy, “Modeling path duration distributions in MANETs and their impact on reactive routing protocols”, *IEEE Journal of Selected Areas on Communications*, vol. 22, no. 7, pp.1357-1373, Sept. 2004.

[9] C. Perkins, E. Royer and S. R. Das, “Ad hoc On-Demand Distance Vector (AODV) Routing draft-ietf-manet-aodv-13”, *Mobile Ad Hoc Networking Working Group, INTERNET DRAFT*, Feb. 2003.

[10] David B. Johnson, David A. Maltz, and Josh Broch, “DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks”, in *Ad Hoc Networking*, edited by Charles E. Perkins, Chapter 5, pp. 139-172, Addison-Wesley, 2001.

[11] A. Laouiti, A. Qayyum and L. Viennot, “Multipoint Relaying: An Efficient Technique for Flooding in Mobile Wireless Networks”, in *Proc. IEEE HICSS*, Hawaii, 2001.

[12] J. Sucec and I. Marsic, “Clustering overhead for hierarchical routing in mobile ad hoc networks”, in *Proc. IEEE INFOCOM*, New York, USA, Jun. 2002.

[13] Y. Han, R. J. La and A. M. Makowski, “Distribution of path durations in mobile ad-hoc networks - Palm’s theorem at work”, in *Proc. 16th ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, 2004.

[14] S. Xu, K. L. Blackmore and H. M. Jones, “An analysis framework for mobility metrics in mobile ad hoc networks”, *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, Jan. 2007.

[15] N. Bisnik and A. Abouzeid, “Queueing network models for delay analysis of multihop wireless ad hoc networks”, in *Proc. IEEE IWCMC*, pp.773-778, 2006.

[16] P. Santi, “The critical transmitting range for connectivity in mobile ad hoc networks”, *IEEE Trans. Mobile Computing*, vol. 4, no. 3, pp. 310-317, May/Jun. 2005.

[17] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu and J. Jetcheva, “A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols”, in *Proc. ACM MobiCom*, Dallas, TX, Oct. 1998.

[18] S. R. Das, R. Castaneda, J. Yan and R. Sengupta, “Comparative performance evaluation of routing protocols for mobile, ad hoc networks”, in *Proc. IEEE ICCCN*, pp. 153-161, 1998.

[19] T. Clausen, P. Jacket and L. Viennot, “Comparative study of Routing Protocols for Mobile Ad Hoc Networks,”, in *Proc. IFIP Med-Hoc-Net*, Sept. 2002.

[20] C. Mbarushimana and A. Shahrabi, “Comparative Study of Reactive and Proactive Routing Protocols Performance in Mobile Ad Hoc Networks”, in *Proc. IEEE AINAW*, May 2007.

[21] “QualNet 3.9 Programmer’s Guide”, *Scalable Network Technologies, Inc.*, Dec. 2005.

[22] Z. Ye and A. A. Abouzeid, “A Unified Model for Joint Throughput-Overhead Analysis of Mobile Ad Hoc Networks”, *Tech. Report, ECSE department, RPI*, Troy, NY, May 2008.

[23] C. Bettstetter, “On the minimum node degree and connectivity of a wireless multihop network”, in *Proc. ACM MobiHoc*, pp. 80-91, 2002.