Energy Efficiency and Throughput for TCP Traffic in Multi-Hop Wireless Networks

Sorav Bansal, Rajeev Gupta, Rajeev Shorey
Imran Ali†, Ashu Razdan*
IBM India Research Laboratory,
Block 1, Indian Institute of Technology,
Hauz Khas, New Delhi 110016, India

Archan Misra
IBM T J Watson Research Center,
19 Skyline Drive,
Hawthorne, NY 10532, USA

Abstract— **We study the performance metrics associated with TCP-regulated traffic in multi-hop, wireless networks that use a common physical channel (e.g., IEEE 802.11). In contrast to earlier analyses, we focus simultaneously on two key operating metrics– the energy efficiency and the session throughput. Using analysis and simulations, we show how these metrics are strongly influenced by the radio transmission range of individual nodes. Due to tradeoffs between the individual packet transmission energy and the likelihood of retransmissions, the total energy consumption is a convex function of the number of hops (and hence, of the transmission range). On the other hand, the TCP session throughput decreases supralinearly with a decrease in the transmission range. In certain scenarios, the overall network capacity can then be a concave function of the transmission range. Based on our analysis of the performance of an individual TCP session, we finally study how parameters such as the node density and the radio transmission range affect the overall network capacity under different operating conditions. Our analysis shows that capacity metrics at the TCP layer behave quite differently than corresponding idealized link-layer metrics.**

I. INTRODUCTION

Analyses of multi-hop, ad-hoc wireless networks typically concentrate on deriving bounds on the maximal achievable capacity, as a function of parameters such as the radio transmission range, node density, total number of nodes and the average distance traversed by traffic sessions. Thus, [1] showed that the end-to-end throughput available to each node is $O(\frac{1}{\sqrt{n}})$ (where n is the number of nodes) for random traffic patterns, and remains constant if the sessions exhibit appropriate localization properties. Similarly, [2] demonstrated the existence of a global scheduling algorithm that can provide a throughput of $\Omega(\frac{1}{\sqrt{n\log(n)}})$, when both the network layout and individual sessions end-points are distributed randomly. For networks where all nodes use the same physical radio channel (such as IEEE 802.11 [3] based ad-hoc LANs), the total network capacity is dependent on the transmission range of each node. This is, of course, expected, since a packet transmission by a node effectively precludes simultaneous transmissions by all nodes within this range (interference region).

The analysis of multi-hop wireless network performance presented in this paper differs from such prior analysis in two key aspects:

(i) Besides the network capacity, we also concentrate on another metric of interest: the *energy efficiency*, defined as the average total transmission energy required to *reliably* transmit a single packet (or byte) to its destination. Our metric includes the energy spent in potential retransmissions needed to overcome possible errors in the traffic path.

(ii) In contrast to greedy traffic sources used as the basis for maximal capacity analysis, we consider *TCP regulated flows*. It is well-known that the maximal achievable throughput of a TCP connection is a function of both its round-trip time (RTT) and the path loss rate–we shall show how both those parameters are affected by the underlying radio transmission range.

For the analysis in this paper, we assume that all nodes are identical in the sense that they all use the same transmission range R; we study the properties of TCP traffic as R is varied. Our focus is on treating R as a design parameter, and evaluating how changes in R affect the overall network performance in different operating conditions. For this discussion, we thus do not consider scenarios where individual nodes adaptively alter their transmission power levels based on various criteria, such as link distances or neighborhood node density. Further, we assume that the maximum capacity of the physical channel is C; for our studies with IEEE 802.11 LANs, we have used $C = 2$ Mbps.

We first demonstrate how the *energy-efficiency* metric is a function of the transmission range. In a variety of multi-hop wireless networks (such as battery-operated sensor networks), the energy efficiency is indeed the most critical metric, since it directly affects the network lifetime. Energy-aware ad-hoc routing algorithms typically choose a path that results in the minimum total transmission energy for a single packet; [4] shows why a more accurate objective should be the *minimum total effective transmission energy*, which focuses on reliable packet reception and includes the energy spent in one or more retransmissions.

We then study how the radio transmission range affects the *maximal achievable throughput* of a TCP session in such wireless networks. It is well known that the throughput of a TCP session (whose capacity is determined by the error rate and not network buffering constraints) varies as $O(\frac{1}{RTT*\sqrt{p}})$ [5], [6] if the path error rate p is small and as $O(\frac{1}{RTT*p})$ [7] if p is moderately high. We study how the range parameter, R, indirectly

† Department of Computer Science and Engineering, I.I.T, New Delhi
∗ Graduate Student, Computer Science Department, UCLA, USA

affects both p and RTT and hence, bounds the TCP session throughput. Additionally, we also consider the TCP throughput achieved over a chain of nodes using the 802.11 MAC layer, and observe how this throughput varies from the ideal maximum presented in [1].

Both studies mentioned above are compared with practical results obtained via simulations performed using IEEE 802.11. We subsequently use the analytical results to derive the *total network capacity* with TCP traffic for such ad-hoc networks. Since the capacity definition for TCP traffic is not immediately apparent, we define the network's TCP-centric capacity as the total (cumulative) goodput achieved by all TCP sessions. We then consider the impact of the transmission range R on this capacity in *two* different scenarios. In the first scenario, we assume that the number of TCP sessions, as well as the number of nodes are fixed. We then vary the total area A of the wireless network (implicitly varying the node density) and then observe how the cumulative goodput varies with changes in the transmission range of individual nodes. In the *second* scenario, we assume that the network is dispersed over a fixed area A and that the number of TCP sessions is proportional to the total number of network nodes. Our analysis shows, that in contrast to earlier studies based on maximal link-layer throughput, the throughput of the individual TCP is $O(\frac{1}{n^{\frac{3}{2}}})$ and the total network goodput is $O(n^{\frac{1}{4}})$ for moderate link error rates. We also use simulation studies with 802.11-based multi-hop wireless networks to quantitatively explore the validity of our analysis.

II. Related Work

It is widely recognized that network capacity is a major constraint in the effective deployment of multi-hop wireless networks. In networks where nodes use the same physical channel, the transmission range of individual nodes is a key determinant of capacity, since it effectively determines the extent of spatial reuse possible. [2] demonstrated that, under random session paths, the capacity of each individual session would degrade as $\Omega\left(\frac{1}{\sqrt{n}}\right)$ with an increase in n the number of nodes, while the total network capacity would grow as $\Omega(\sqrt{n})$. Moreover, [2] also showed how an ideal MAC protocol could be designed to provide each node at least $\Omega(\frac{1}{n*\log(n)})$ of the maximal channel capacity. [9] considered the design of an optimal MAC layer to maximize the total utilization of the shared channel over all the nodes in a multi-hop network. [1] considered how the IEEE 802.11 MAC algorithm performed relative to these bounds, and also showed that if the traffic patterns showed appropriate stochastic locality (more accurately, if the probability of the session distance decayed faster than D^{-2}), then the ideal throughput per session would remain a constant. These studies, however, consider idealized sources that are capable of injecting packets whenever permitted by the MAC layer. In particular, they do not consider the use of TCP traffic and the impact of transmission errors in the link layer on the maximal link utilization by such TCP sources.

Studies on energy-efficient communication for wireless networks typically focus on the routing problem alone: they are concerned solely with maximizing some measure of the total transmission energy or minimizing some function of the battery drainage. For example, [8] adapts Dijkstra's minimum cost path selection algorithm to find minimum total energy paths, by setting the link cost to the associated transmission energy. Such energy-efficient routing protocols assume that, when the physical distance of a hop is smaller, the wireless nodes are able to appropriately reduce their transmission power. Similarly, newer routing algorithms (e.g, [10]) seek to reduce a long-distance hop into a series of short-distance ones, thereby minimizing the total power usage. Battery-aware routing protocols ([11], [12]) often consider the residual energy level of the node's battery as a metric, and hence attempt to form routes using potentially less-drained nodes. Such studies do not however analyze how the selection of such energy-efficient paths impact other metrics such as session throughput: since modification of the transmission range implies modification of the session throughput, such power-conscious routing algorithms implicitly affect the network capacity.

The performance of TCP congestion avoidance under varying loss rates and RTT has been extensively analyzed in literature (e.g. [5], [6], [7]), especially for point-to-point links. For moderate to low loss rates, the TCP throughput varies inversely as the square-root of the loss probability. The interaction of TCP performance with the contention-based MAC scheduling in multi-access media is less clearly understood.

III. Energy Efficiency and Transmission Range

We consider a scenario where the transmitter radios are capable of dynamically altering their transmission power, based on the transmission distance. We first focus solely on the communication cost, and then show how the energy budget is changed substantially if we additionally consider the computing cost.

We also assume the use of omni-directional antenna; accordingly, if the maximum distance for acceptable reception is R, it follows that the coverage area for reliable reception is $\propto R^2$. Since the power attenuation with distance D is usually proportional to $D^K : K \geq 2$, it follows that the optimal transmission power needed to communicate over a radial distance R is proportional to R^K. Accordingly, an energy-efficient transmission scheme will ensure that the transmission energy over a single hop (or link), $E(R)$, of distance R is:

$$E(R) \propto R^K \qquad (1)$$

Given the above relationship between the optimal transmission energy and the total transmission distance, it is easy to see that the total energy associated with a single transmission event actually decreases if a hop is sub-divided into multiple smaller ones: clearly, if $D_1 + D_2 = D$, then $D_1^K + D_2^K < D^K$ if K > 2. Energy-efficient routing protocols thus usually seek to transmit a packet between a source S and a destination D using multiple short-distance hops, as opposed to a smaller number of long-distance hops. Indeed, minimum total-energy routing algorithms, such as [8], result in the formation of routes with a large number of short-range hops. This intuition is, however, misleading: the formulation neglects the fact that an increase in the hop-count leads to an increase in the packet error rate over the entire path, and thereby increases the likelihood of retransmissions and thus decreasing the session throughput.

Analysis in [4] shows that, in the absence of reliable link layers (or what is called the end-to-end retransmission or EER model), the actual *effective* energy per reliably transmitted packet over a $N-hop$ path (with nodes indexed as $(1,\ldots,N+1)$) is given by:

$$E^{EER}_{total} \propto \frac{\sum_{i=1}^{N} D^{K}_{i,i+1}}{\prod_{i=1}^{N}(1-p_{i,i+1})}, \tag{2}$$

where $p_{i,i+1}$ indicates the packet error rate of the i^{th} hop (between nodes i and $i+1$). On the other hand, if the number of permitted retransmissions on each link is unbounded (hence, each link ensures accurate delivery to the next hop), the total effective energy per packet (in the so called hop-by-hop or HHR model) is given by:

$$E^{HHR}_{total} \propto \sum_{i=1}^{N} \frac{D^{K}_{i,i+1}}{1-p_{i,i+1}}. \tag{3}$$

Analysis of the expression for the EER mode shows that, even if all the links have identical error rates, there is an optimal value for the number of hops associated with a specific transmission path. If the number of hops is smaller, the energy budget is dominated by the larger transmission energies needed to transmit over larger distances; if the number of hops is larger, it is the overhead associated with retransmissions that negates the energy gains associated with smaller individual hops. In contrast, if each link is allowed potentially unlimited number of retransmission attempts, the total effective energy always decreases with increasing N.

A. Transmission Energy Efficiency and Transmission Range

Before proceeding further, it is necessary to extend the analysis of *effective transmission energy* mentioned above. To apply our insights quantitatively to technologies, such as IEEE 802.11, we need to analyze the case where each link has an *upper bound on the maximum number of retransmission attempts*. This bound is a practical necessity to avoid abnormally large latencies and buffer overflows at the link layer. We assume that each link layer is permitted a total of max transmissions; clearly, such a restriction resurrects the possibility of end-to-end retransmissions in the case of forwarding failure at an intermediate link. Also, for analytical ease, we assume that all links have the same packet error rate p and the same transmission energy E.

Due to the considerably more involved nature of the calculations for effective energy in this case, we relegate the complete mathematical analysis to the Appendix, mentioning only the relevant features here.

Result 1: If each link has a transmission packet error rate p, then the conditional expected number of distinct transmissions, *given the successful forwarding over the link*, is given by:

$$T_{good} = \frac{1}{1-p} - \frac{max * p^{max}}{1-p^{max}},$$

and the expected number of distinct transmissions, *given the failure of the link forwarding process* is given by:

$$T_{bad} = max.$$

Result 2: In case of an end-to-end failure in reliable packet delivery (one of the N intermediate links failed to reliably forward the packet), the total number of expected distinct transmissions is given by:

$$T^{total}_{bad} = T_{bad} + T_{good} * (1-q) * \{ \frac{1 - N*(1-q)^{N-1} + (N-1)*(1-q)^{N}}{q * \{1-(1-q)^{N}\}} \}, \tag{4}$$

where $q = p^{max}$. Similarly, if the packet was indeed successfully forwarded to the destination node, the total number of expected distinct transmissions is:

$$T^{total}_{good} = N * T_{good} \tag{5}$$

By combining the above two results with the fact that the probability of successful packet end-to-end delivery is given by $(1-q)^N$ (where $q = p^{max}$), we can finally derive the following result:

Result 3: The total effective number of distinct packet transmissions needed for reliable packet delivery is given by:

$$T = T^{total}_{bad} * \frac{P_{fail}}{1-P_{fail}} + T^{total}_{good}, \tag{6}$$

where $P_{fail} = 1-(1-q)^N$.

Since T is really a function of N, p and max, we represent this result generically as $T(N,p,max)$. We defer the quantitative comparisons of our analytical expression with simulation results to the next sub-section and, instead, focus on the expected qualitative behavior. Clearly, in the limited-retransmission case, there is an optimal value for N, the number of hops: if N becomes too large, then the probability of an end-to-end error becomes non-negligible and the consequent effects of end-to-end retransmissions begin to dominate the energy budget. In fact, the approximate value of this optimal value can be obtained by realizing that, from the standpoint of energy consumption alone, a link with a packet error rate of p and a transmission bound of max is essentially equivalent to a link with no retransmissions but a link packet error rate of p^{max}. (This is not completely accurate when we consider the effects on protocols at higher layers; for example, link-layer retransmissions are likely to result in greater variation in the forwarding latency and hence, the possibility of spurious TCP-layer timeouts.) Accordingly, using the analysis in [4], the optimal value of N is, to a good approximation, given by $\frac{-1}{\log(1-p^{max})}$.

For a generalized ad-hoc network, it is now easy to see the connection between the transmission radius and effective energy. If we assume that the average distance between the endpoints of a session is $\bar{L}$, then a transmission range of R implies

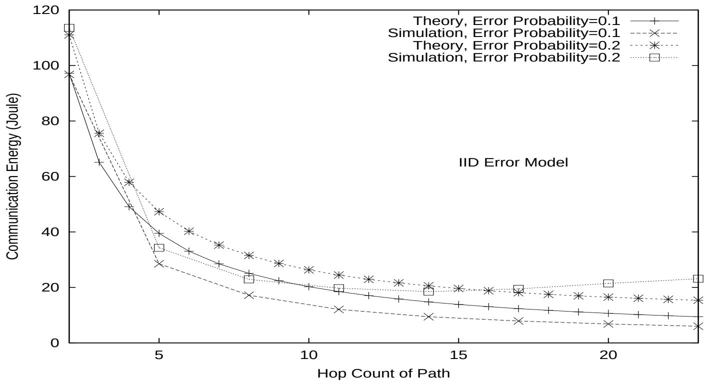

Fig. 1. Communication Energy versus Number of Hops (TxThresh = 1)

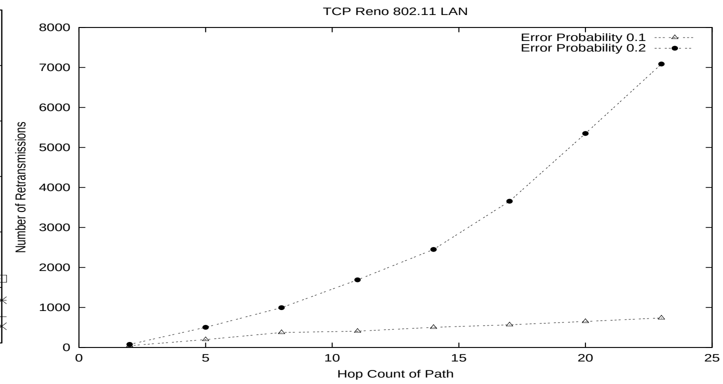

Fig. 2. Number of TCP Retransmissions versus Number of Hops (TxThresh = 1)

that the average number of hops, N is given by $\lceil \frac{\bar{L}}{R} \rceil$, or to a good approximation by $\frac{\bar{L}}{R}$. Accordingly, with a link layer bound of max on the number of retransmissions, equation (6) shows that the effective energy efficiency of the ad-hoc network is given by (ignoring proportionality constants):

$$E_{total} = R^K * T\left(\frac{\bar{L}}{R}, p, max\right) \tag{7}$$

Clearly, as long as the decrease in R in the expression (7) dominates over the corresponding increase in $T(.)$, the energy consumption per byte decreases. Beyond the optimal value for R, the decrease in the energy spent in any single transmission activity is negated by the larger increase in $T(.)$. From an energy efficiency perspective, there is an optimal value to the radius of acceptable reception quality R in an ad-hoc network; decreasing the transmission range below this optimal value does not lead to greater energy savings.

A.1 Applicability to the 802.11 Environment

We applied this analytical model to the 802.11-specific environment, using the 802.11 implementation in the ns-2 simulator. For our simulations, the distance between the source and destination was kept at 750 meters, while the transmission range was varied between $(30, 700)$ meters; N was thus varied from 2 to 24. The energy associated with each transmission was assumed to be (ignoring proportionality constants) given by $E \propto R^2$; the simulations were run for both uncorrelated (i.e., i.i.d) and correlated error models. For the results plotted here, we set the transmission power for a distance of 250 meters to 0.03346 W, and then computed the corresponding power for other transmission distances by appropriate scaling (proportional to the square of the distance).

The *communication* energy efficiency (the effective transmission energy per packet) was computed by determining the total transmission energy spent in transferring a 10 MB-sized file using a TCP flow from the source to the destination. Since the number of packets transferred reliably by TCP is the same for all simulations, the total communication energy consumption is a direct indicator of the transmission energy efficiency. The number of hops N was varied by simply inserting the corresponding

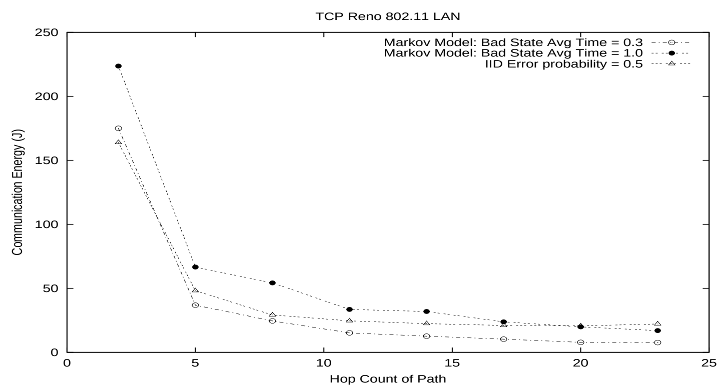

Fig. 3. Communication Energy versus Number of Hops (TxThresh = 4)

number of intermediate nodes between the source and destination. The total energy consumption is clearly a function of the maximum number of retransmissions supported at each link (the TxThresh parameter in ns-2). We present results here for TxThresh equal to 1 and 4; the corresponding value of max (see Equation (7)) was thus 2 and 5 respectively. We simulate the energy efficiency for TCP file transfer using two standard models for the link error: a) the two-state Markov-modulated channel model with correlated errors and b) the independent identically distributed (i.i.d) model with independent and identically distributed bit error rates.

Figure 1 plots the simulated total transmission energy consumption, under the i.i.d model, as N varies between 2 and 23 for two different values of p, 0.1 and 0.2, and TxThresh equal to 1. The figure also includes the energy efficiency values (with appropriate scaling) predicted by Equation (7). We can see that the theoretical model, while an accurate reflector of the overall trend, underestimates the energy consumption, especially for larger values of N. This is to be expected, since our analytical formulation does not include the energy spent in the 802.11 signaling (such as RTS/CTS/ACK packets), as well as the energy wastage in potential MAC layer collisions (which can be expected to occur more often for higher values of N). It is easy to see that, when the link layer permits only one retransmission, the optimal value of N (from simulations) is larger than 23 for

$p = 0.1$; even when the error rate is fairly large ($p = 0.2$), the optimal number of hops is approximately 15. The number of TCP level retransmissions for the two cases have also been plotted in Figure 2; as expected, the number of source-initiated retransmissions needed increases with increasing N.

To further study the impact of link-layer retransmissions, we also studied the total energy consumption with TxThresh equal to 4 and three link error rates:

a) The two-state Markov model where the average sojourn times in the Good and Bad states were 1.0 and 0.3 ms respectively.

b) The two-state Markov model where the average sojourn times in the Good and Bad state were identical and equal to 1.0 ms.

c) The i.i.d model with p set to 0.5 (a very high value).

Figure 3 plots these simulation results for the total transmission energy with TxThresh $= 4$; it is again seen that under all these operating conditions, the transmission energy consumption decreases as long as N is increased over any realistic range.

B. Total Energy Efficiency

The discussion and results of the previous section show that a larger number of hops, or equivalently a smaller transmission range, typically always increases the energy efficiency. This argument is, however, misleading, since this formulation ignores the *computing energy*– any node engaged in packet transmissions also expends ambient energy in addition to that consumed by the radio interface. In particular, we shall see in the next section that an increase in N typically leads to a corresponding drop in the TCP goodput, even if the physical distance between the source and destination nodes is unchanged. Hence, while the transmission energy efficiency may indeed increase with N, the resultant loss in throughput implies that the transfer of a fixed number of bytes will take a longer time. Since the total *computing energy* can be assumed to be proportional to the total activity duration, it should be clear that this cost will only increase with N.

To formally explore this concept, we repeated the energy-related simulations, taking care to measure the total time taken by TCP to reliably transfer the entire 10 MB file. If we then assume then P_a is the ambient or standby power spent by each node during the lifetime of the session, the computing energy expenditure over all the N nodes is equal to $P_a * N *$ simulation duration. Accordingly, the total energy consumption is now given by:

$$E_{total} = E(transmission) + N * P_a * \text{simulation duration.}$$

Figure 4 plots the variation in this total energy with changing N for the experiments using the two-state error model with Good and Bad sojourn times of 1.0 msec and 0.3 msec respectively (TxThresh=4). Similarly, Figure 5 plots the total energy consumption versus the number of hops for the two-state error model with Good and Bad sojourn times of 1.0 msec and 0.3 msec respectively (TxThresh = 1), and the i.i.d error model with $p = 0.1$ (TxThresh = 1). These results correspond to a choice of $P_a = 0.004$ W.

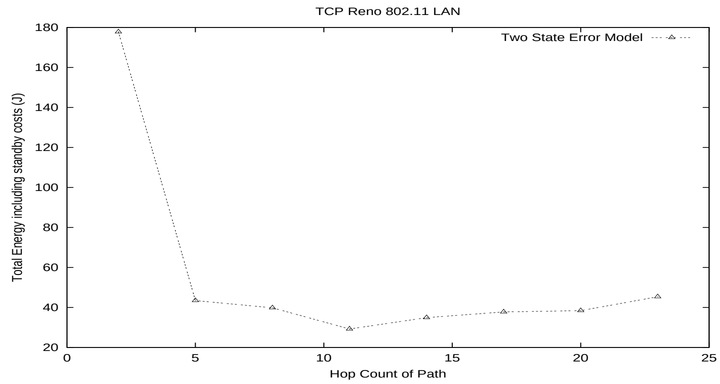

Fig. 4. Total Energy versus Number of Hops (TxThresh = 4)

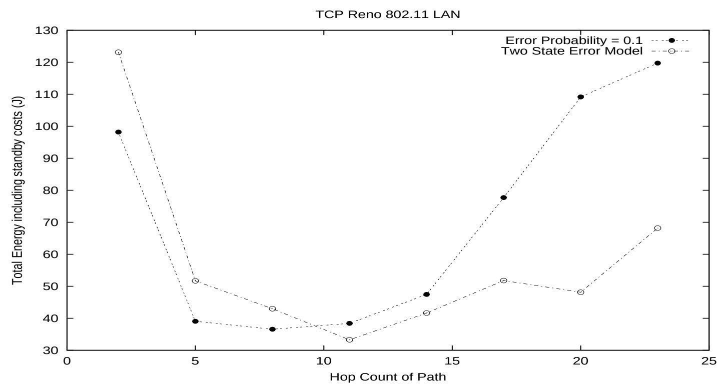

Fig. 5. Total Energy versus Number of Hops (TxThresh = 1)

It is easy to see that, when the total energy is considered, both graphs show minimum energy consumption for realistically small values of N. For example, if we consider only the transmission energy, the optimal value of N was certainly greater than 23 for the i.i.d channel with an error rate of 0.1. However, when the total energy consumption is considered, it is clear that increasing the number of hops beyond $\sim 10 - 12$ hops will prove to be disadvantageous. (In our simulated environment, an optimal hop count of 12 corresponds to a transmission range of ~ 65 meters.) Our studies thus clearly show that any adjustments to the transmission range to improve the network capacity (which we shall define appropriately in the section V) must consider the potential effect on the energy efficiency of the resulting network. If the transmission range is decreased such that the average number of hops traversed by a session increases beyond $\sim 10 - 15$, then any increase in network capacity comes only at the expense of lower energy efficiency.

IV. Maximum Throughput of a Single TCP Session

After analyzing the energy-related metrics of an ad-hoc network, we now consider the impact of the transmission range on the throughput achieved by an idealized TCP session. In this section, we assume the absence of any cross-traffic from other sessions; the path for the session of interest is thus simply a node-chain. Analysis in [1] showed that, for such a chain topology (where the nodes could interfere with their one and two-hop neighbors), the maximum ideal capacity is $\frac{C}{4}$; with 802.11 MAC-based scheduling, the maximum obtained throughput is usually around $\frac{C}{7}$. To achieve such an ideal throughput, the MAC layer must be the only bottleneck; in contrast to these analyses, we consider a persistent flow subject to the dynamics of TCP flow control. The throughput of a persistent TCP flow depends on the range of the magnitude of the error rates and the buffer capacity available at intermediate nodes.

If the TCP losses occur primarily due to link errors, and if buffer overflow is a fairly rare event, then the throughput of a TCP connection as a function of p and RTT is given by the well-known square-root formula:

$$\rho(RTT, p) \sim \frac{\kappa * MSS * 8}{RTT * \sqrt{p}}, \quad (8)$$

where RTT equals the round-trip delay, p equals the effective error rate, MSS indicates the packet size (in bytes) and where κ is an implementation-specific constant. (For example, κ is $\sim \sqrt{2}$ for TCP without delayed acknowledgments and ~ 1 with delayed acknowledgments.) The above equation holds as long as p does not become very greater than $\sim 15 - 20\%$ for most TCP versions; larger values of p lead to undesirable transients such as retransmission timeouts and a sharper drop in the TCP throughput.

On the other hand, if TCP losses occur primarily due to buffer overflows, the dynamics of the connection becomes much harder to analyze in the presence of multiple hops. In such a situation, the RTT is dominated by the various queuing delays; however, in general, the throughput of the TCP flow decreases with an increase in the RTT.

For practical ad-hoc topologies, the propagation delays are usually small–consequently, the RTT *is dominated by the queuing and transmission delays.* Assuming that nodes are homogeneous, the RTT is thus directly proportional to N, the number of hops, since each additional hop introduces queuing and transmission delays. If the error probability of each link is a constant p, the end-to-end error probability is given exactly by $1-(1-p)^N$; if $N * p << 1$, the end-to-end packet error rate is then approximately $N * p$. Accordingly, for ad-hoc networks operating under relatively small end-to-end packet error rates (say, less than $\sim 10\%$), the maximal throughput of a TCP connection should behave as the following function of N:

$$\rho \propto \frac{1}{N * \sqrt{N}} \propto \frac{1}{N^{\frac{3}{2}}}. \quad (9)$$

However, if the error rates are so low that the TCP flow almost never halves its window in response to a link loss, it should be clear that the throughput becomes independent of the link error probabilities. In such a case, since $RTT \propto N$, the TCP throughput will vary as:

$$\rho \propto \frac{1}{N} \quad (10)$$

For a fixed mean distance $\bar{L}$ (in absolute units) between the end-points of an ad-hoc session, the average number of hops, N, as a function of the transmission range R is given by $N = \frac{\bar{L}}{R}$. Accordingly, the maximum throughput of a persistent TCP flow will vary $\propto R^{\frac{3}{2}}$ if the flow is link-loss controlled, and $\propto R$ if the flow is buffer-loss controlled. Of course, the above equations hold good only when ρ is less than the theoretical goodput of the chain topology. For example, in a linear topology with ideal MAC scheduling and interference radius equal to the acceptable reception radius, the dynamics of TCP flow control act as the primary flow capacity constraint as long as $\rho \leq \frac{C}{3}$. If the inequality does not hold, then the session throughput is constrained, not by TCP dynamics, *but by the interference at the MAC layer* among simultaneous transmissions by neighboring nodes.

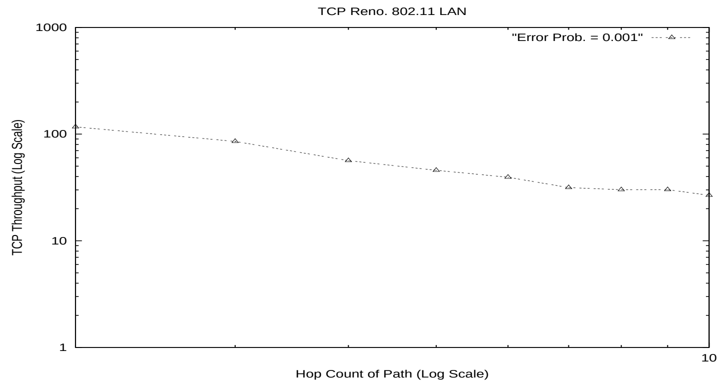

Fig. 6. Throughput versus Number of Hops (TxThresh = 1)

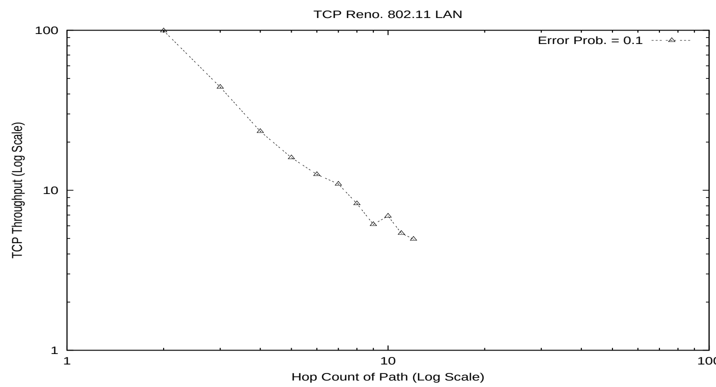

Fig. 7. Throughput versus Number of Hops (TxThresh = 1)

A. *Applicability to the 802.11 Environment*

To study the variation of TCP session throughput with the number of hops in the 802.11 environment, we performed simulations with our chain topology. As before, the distance between the session end-points was kept constant– the number of intermediate hops was varied by varying the transmission range. Moreover, we plotted $\log(\rho)$ against $\log(N)$; in this case, the slope of the resultant curve determines the exponent in the relationship between ρ and N.

Figure 6 plots the TCP throughput (in terms of packets/sec for an MSS of 512 bytes) against N on a logarithmic scale when the link error rate is very small (0.001) and TxThresh=1;

in this case, the resultant end-to-end loss rate is negligible and TCP is primarily buffer controlled. The slope of the curve is ~ -1, indicating fairly good agreement with our analysis. On the other hand, Figure 7 plots the TCP throughput (again in units of packet/sec for 512 byte packets) against N for $p = 0.1$ and TxThresh=1. In this case, the resultant error rate is moderately high; the slope of the curve is around -1.7 in this case, which indicates fairly close agreement with our theoretical analysis.

The results on the TCP throughput in such multi-hop networks are important from the capacity analysis standpoint. The results show that for TCP-controlled traffic, decreasing the transmission range actually penalizes the maximum session throughput, since the consequent increase in the number of hops increases both the RTT and the end-to-end loss rate. As we shall see in the next section, this phenomenon impacts the amount of TCP traffic that such a multi-hop, wireless network may be expected to carry.

V. TCP-based Ad-Hoc Network Capacity

Having studied both the energy-efficiency and the individual TCP session behavior with varying R, we now focus on the total capacity of the ad-hoc network. Most literature defines the network capacity Cap as the the total "one-hop throughput" or the "bit-distance product"– fundamentally speaking, this is a weighted sum of all the session throughputs, with the weight of each session equal to the distance (or the number of hops) over which it passes.

From a theoretical perspective, if the transmission (and interference) range of the ad-hoc nodes are R, then a node transmitting packets at the channel capacity C effectively prohibits any transmission activity for all nodes within the coverage area, which is $\propto R^2$. Accordingly, if the area of the ad-hoc network is A, and the transmission and interference radii are both R, the maximal ideal capacity of the ad-hoc network is $\frac{C*A}{\pi * R^2}$. In a more generic context, where reception and interference radii are not necessarily identical, the maximal network capacity Cap is $\propto \frac{A}{R^2}$. In general, we would thus expect the maximal ideal throughput to increase quadratically with a reduction in the transmission radius.

Since a greedy TCP flow (where $cwnd$ is the only constraint for packet generation at the transport layer) cannot avail of the maximal capacity, the concept of maximal TCP throughput and network capacity becomes trickier. It is also apparent that attempting to attain $\sim 100\%$ link utilization by pumping up the number of parallel TCP sessions is also not feasible, especially in wireless networks where the buffer capacity on individual nodes is fairly limited. We thus study the expected throughput dynamics for two different, but interesting, operational scenarios.

A. The Fixed Session, Variable Area Framework

We now attempt to formalize the notion of capacity in this scenario (fixed session, variable area). We recall from the previous section on capacity of a single TCP session that for a fixed mean distance $\bar{L}$ (in absolute units) between the end-points of an ad-hoc session, the average number of hops, N, as a function of the transmission range R is given by $N = \frac{\bar{L}}{R}$. For the link layer, the number of simultaneous active sessions decrease with increasing range R. On an average,

$$Cap \propto \frac{A}{R^2} \tag{11}$$

Now, for a fixed number of TCP sessions, capacity is proportional to TCP throughput (as long as the MAC layer bounds are not violated), i.e.,

$$Cap \propto \frac{1}{\left(\frac{\bar{L}}{R}\right)^{\frac{3}{2}}} \tag{12}$$

If R is very small, the average degree of connectivity of the graph is fairly small. The resultant sub-optimal paths imply that each packet has to travel a large number of hops (N) to reach to the destination. Accordingly, the TCP session throughput decreases with decreasing R, if R is below a certain value. Therefore the sum of the throughputs (over the fixed number of sessions) becomes smaller. On the other hand, if R is larger than a certain value, then the resultant MAC-layer channel interference and collisions limit the capacity of the TCP sessions. In this range of R, the TCP sessions are prevented from better exploiting the network by the larger delays caused due to collisions and backoffs at the MAC layer, hence, the TCP throughput (that decreases with increasing N) for each session is small. We can thus expect an optimal value of R. *To the right of this value (larger R), the network is MAC-layer constrained, with the channel interference dominating the throughput; to the left of this value (smaller R), the network is TCP-layer constrained (Equation 12), with the TCP sessions unable to pump enough packets into the network.*

Accordingly, it follows that for R smaller than this optimal value, the network capacity will degrade in proportion to the TCP throughput degradation ($\propto R^{\frac{3}{2}}$ from Equation 9), if p lies within a sensible operating range. To the right of this optimal value, the resultant throughput is determined by the competing effects of higher TCP-layer throughput (lower loss rates due to smaller N) and greater MAC contention. Thus, from Equation 12 and Equation 11, we would expect the 'capacity' in this range to vary as:

$$Cap \propto \frac{A}{R^2} \frac{1}{\left(\frac{\bar{L}}{R}\right)^{\frac{3}{2}}} \propto R^{\frac{-1}{2}}. \tag{13}$$

Figure 8, Figure 9 and Figure 10 show results for capacity as transmission range R is varied. In these simulations, 50 nodes were randomly distributed in a square grid area. 25 TCP connections were chosen randomly and every node was either a TCP source or a TCP destination, but not both. All our simulations with random topologies use DSR for computing the session paths; in the absence of mobility, the choice of paths (and consequent network performance) is expected to be independent of the choice of a specific ad-hoc routing protocol.

In Figure 8 we plot the capacity versus R for an error-free channel model and a square grid of 500m $\times$ 500m. We see that

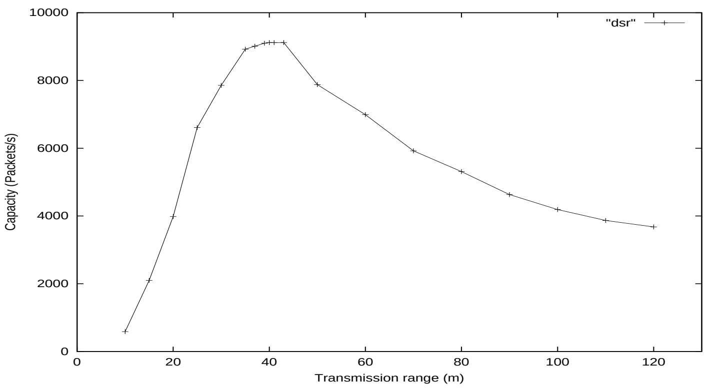

Fig. 8. Capacity versus Transmission Range (A = 500m × 500m)

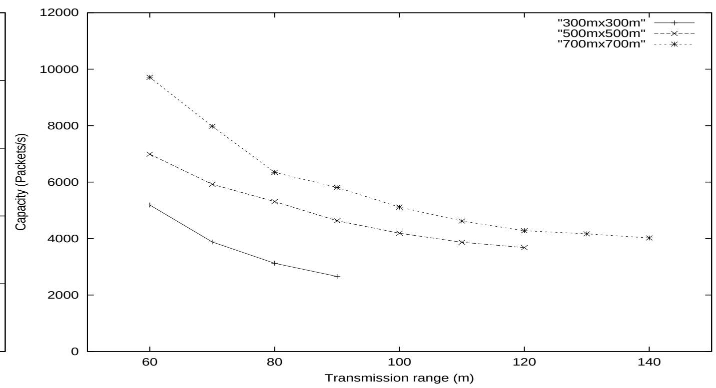

Fig. 10. Capacity versus Transmission Range for varying Density

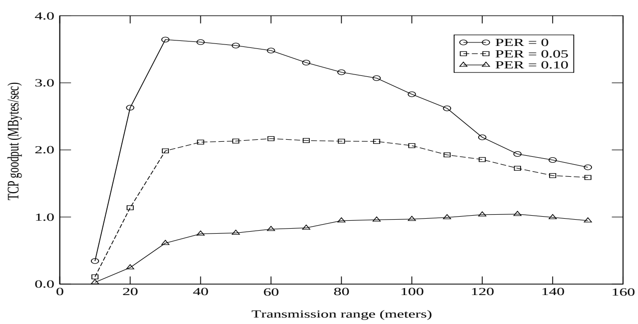

Fig. 9. TCP Throughput versus Transmission Range for varying PER (A = 500m × 500m)

the optimal value of R (from a capacity standpoint) is $\sim 35-40$ meters.

In Figure 9, we have plotted TCP goodput versus the transmission range for various link Packet Error Rates (PER) (for a constant 500m × 500m grid) under the IID error model. We see that as PER increases, the TCP goodput decreases and the the optimal transmission range (i.e., the range corresponding to maximum TCP goodput) increases. This can be explained by observing that a larger packet error rate implies a faster degradation in TCP throughput with the number of hops in a path. Thus, a value of R that is optimal for smaller p will prove sub-optimal for larger p. As R is increased, the average value of N, and hence $N * p$, the end-to-end error rate, decreases leading to more aggressive behavior. Of course, the resultant increase in the optimal value of R cannot be very large, since a larger R also implies greater delays and interference at the MAC layer.

It is also interesting to see what happens if the total area A of the wireless, multi-hop network is increased without varying the total number of nodes n or the transmission range R. If η is the node density, then clearly $A = \frac{n}{\eta}$. Further, for networks where the source and destination are chosen at random, the average distance of a session, $\bar{L}$ is clearly $\propto A^{\frac{1}{2}}$. If the transmission radius R is chosen to be greater than the optimal value, then Equation 13 shows that the total 'TCP capacity' is given by:

$$Cap \propto \frac{n^{\frac{1}{4}}}{R^{\frac{1}{2}}\eta^{\frac{1}{4}}} \tag{14}$$

Thus, in the fixed session, variable area and constant range framework, the capacity of the system is inverse in proportion to $\eta^{\frac{1}{4}}$*, or proportional to* $A^{\frac{1}{4}}$. In networks where the radio ranges cannot be adjusted, one must thus guard against packing too many nodes into too small an area.

In Figure 10, we plot the system capacity versus the transmission range for varying node densities by changing the area (300m×300m, 500m×500m, 700m×700m). The simulation is done for an error-free channel (i.e, PER = 0). It is seen from the plot that for a fixed transmission range, the capacity decreases with an increase in the density.

B. *The Variable Sessions, Fixed Area Framework*

In contrast to the assumptions of the previous section, now consider an operational mode where the coverage area, A, of the ad-hoc network is fixed. Further, the number of simultaneously active TCP sessions in the network, denoted by T, is directly proportional to n, the total number of ad-hoc nodes. Thus. mathematically

$$T = \gamma * n, \tag{15}$$

where γ indicates the probability that any given node is engaged in a TCP-based transfer at any instant.

This formulation is a useful model for understanding network dynamics under certain very practical situations. Consider, for example, the problem of covering a geographic area with a certain number of sensor (say thermal sensor) nodes. Each node is autonomously programmed to periodically activate itself, monitor the temperature and communicate it to a central authority. Thus, if the communication process happens for 15 minutes every hour, we have a model where the number of active sessions is $\frac{1}{4}^{th}$ of the total number of nodes n. The network designer would clearly be interested in evaluating how his choice of the nodal density (how closely to place the wireless nodes), denoted by η, affects the achievable network capacity.

To study the dependence of total capacity on η, we make the *fundamental assumption* that a larger η leads to a smaller transmission range R. In well-designed networks, the choice of R is actually based on the need to keep the average degree of each node, defined as the number of one-hop neighbors, moderately high; in fact, classical results [13] state that the optimal number of one-hop neighbors is ~ 6. As η increases, a node is able to find one-hop neighbors within a smaller radial distance, and consequently, can lower its transmission radius.

Then, since each TCP session, by our previous section, has $\rho \propto$ either $\left(\frac{R}{L}\right)^{\frac{3}{2}}$ (for moderate values of p) or $\propto \left(\frac{R}{L}\right)$ (for low values of p), it follows that the total capacity utilized by the ad-hoc network is then:

$$\begin{aligned} Cap &\propto \gamma * \eta * A * \left(\frac{R}{\bar{L}}\right)^{\frac{3}{2}} \text{ for moderate } p \\ &\propto \gamma * \eta * A * \left(\frac{R}{\bar{L}}\right) \text{ for very low } p \end{aligned} \quad (16)$$

We thus consider a fixed area A and progressively increase ad-hoc node density η. Since the optimal transmission radius (that needed to maintain a constant nodal degree) decreases as the square-root of the number of nodes, it is easy to see that node density and the transmission radius are related as

$$R \propto \frac{1}{\sqrt{\eta}}$$

By substituting this into equation (16), we finally get the ‘capacity’ of the TCP-based ad-hoc network as

$$Cap \propto \frac{\gamma * \eta * A}{\eta^{\frac{3}{4}} * \bar{L}^{1.5}} \propto \frac{\gamma * A * \eta^{\frac{1}{4}}}{\bar{L}^{\frac{3}{2}}}, \quad (17)$$

or,

$$Cap \propto \frac{\gamma * \eta * A}{\eta^{\frac{1}{2}} * \bar{L}^{0.5}} \propto \frac{\gamma * A * \eta^{\frac{1}{2}}}{\bar{L}^{\frac{1}{2}}}, \quad (18)$$

where Equation 17 holds for moderately low values of link error rates, and Equation 18 holds for very low values of link error rates.

To illustrate the validity of our conclusions, we ran simulations where the area was kept constant and the number of nodes was progressively increased. Figure 11 plots the TCP throughput against the logarithm of the node density, for an operating environment where the link packet error rate (i.i.d.) was only 0.001 and TxThresh=1. The slope of the graph in this case is ~ 0.6, showing the applicability of Equation 18 to this case (since the effective end-to-end error rate was very low).

It is interesting to contrast these results with those on the idealized link capacity in [1], which showed that, under similar operating conditions, the idealized link-layer network capacity would increase as $O(\sqrt{\eta})$. Clearly, the bursty nature of TCP traffic (which prevents us from indiscriminately increasing the total number of sessions), and the dependence of TCP session throughput on the link error rate can prevent TCP-based data traffic from achieving this ideal value.

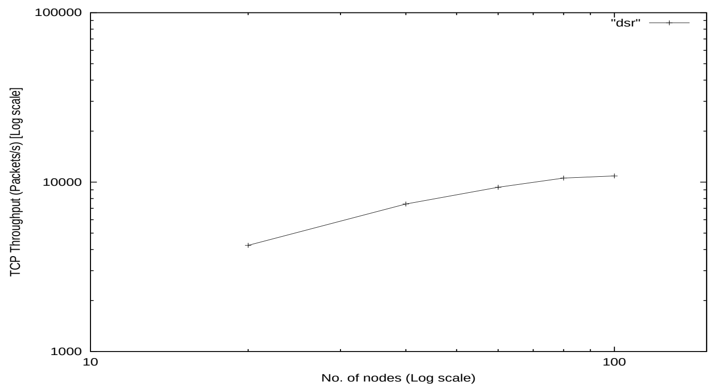

Fig. 11. TCP Throughput versus Node Density (Log Scale) for Optimum Range

VI. Conclusion

In this paper, we focus on the theoretical performance of TCP traffic over a multi-hop, wireless network where all links share the same physical channel. Our analysis shows that it is often difficult to simultaneously improve both the energy efficiency and the session throughput. We studied how the transmission energy efficiency decreases with an increase in the average number of hops for a session, and is a combined effect of smaller energy for individual packet transmissions and larger retransmission probabilities.

This improvement in energy efficiency, however, comes with a cost: the TCP session goodput. We showed how the goodput of an individual TCP flow decreases as either $O(\frac{1}{N^{\frac{3}{2}}})$, or as $O(\frac{1}{N})$, depending on the error rate of the link layer. Accordingly, a decrease in the radio transmission range R implies that a session takes longer to transfer a fixed quantity of data. When the ambient computing energy is considered in tandem with the communication costs, we show that a smaller R is beneficial from an energy perspective only up to a certain limit. A designer of such a multi-hop wireless network needs to consider these energy-bandwidth tradeoffs while deciding on the node density and radio transmission range.

Perhaps, most importantly, we show how the transport layer (TCP) capacity of the network differs from the idealized link-layer capacity. We consider two distinct operational modes and show that, in either case, the capacity is typically decided by the tradeoff between the maximum attainable throughput of a TCP session and the interference effects at the MAC layer. When the area of the network and the number of active sessions is fixed, the capacity is a concave function of the transmission range. On the other hand, if the number of active sessions scales linearly with the number of nodes, the effective TCP-layer capacity varies between $O(n^{\frac{1}{4}})$ and $O(n^{\frac{1}{2}})$, in contrast to the idealized bound of $O(n^{\frac{1}{2}})$.

In future, we need to extend our analysis and simulation studies to mobile ad-hoc environments, since node mobility will clearly impact both the energy consumption and the individual session goodput. Clearly, varying the transmission range R directly affects the frequency of link breakages in such mobile environments– since such breakages lead to both additional routing overheads and packet losses, they affect both the energy consumption and the overall goodput in non-trivial ways. Moreover, we also need to study the comparative performance of various ad-hoc routing protocols such as AODV, DSR and TORA, which are expected to differ in terms of important performance metrics such as the packet delivery ratio, the average forwarding latency and the energy consumption.

APPENDIX

In this appendix, we derive the expression for the total number of packet transmissions necessary for reliable delivery of a packet over an N hop path. The packet error rate for each hop is p and the maximum number of retransmissions at the link layer is max.

Since reliable link forwarding fails only when all max transmissions fail, the unconditional probability of link packet transmission failure, which we call q, is given by $q = p^{max}$; the corresponding probability of reliable link delivery (potentially using between $(1, \ldots, max)$ transmissions) is then $1 - q$. Since the total number of link transmissions, given that the link has reliably forwarded the packet, is a truncated geometric distribution with parameter p, the *conditional* expected number of transmissions, T_{good}, over a single link, is given by:

$$T_{good} = \sum_{i=1}^{max} i * p^i * (1-p) \; = \; \frac{1}{1-p} - \frac{max * p^{max}}{1-p^{max}}. \quad (19)$$

Since link packet delivery fails only after exactly max transmissions, the corresponding *conditional* number of transmissions, given forwarding failure is:

$$T_{bad} = max.$$

Now since each link fails to forward the packet independently with q, the unconditional probability of successful end-to-end delivery (without another source retransmission) is given by $P_{succ} = (1-q)^N$, and the unconditional probability of unsuccessful end-to-end delivery is given by

$$P_{fail} = 1 - (1-q)^N. \quad (20)$$

Next, we determine the expected number of total packet transmissions (over all the links that attempted to transmit a packet), T_{bad}^{total}, *given that the end-to-end forwarding attempt was unsuccessful*. Since a downstream node forwards packets only when all the upstream nodes successfully transmitted the packet, it is easy to see that the conditional probability that failure occurs at the i^{th} link is given by:

$$\begin{aligned} Prob_{fail}(i|\text{end-to-end failure}) &= \frac{Prob_{fail}(i)}{P_{fail}} \\ &= \frac{(1-q)^{i-1} * q}{P_{fail}}. \end{aligned}$$

If failure occurs at the i^{th} link, the expected number of total link-layer transmissions (over all the upstream nodes) is $(i-1) * T_{good} + T_{bad}$. Accordingly, the conditional mean number of total link-layer transmissions during link failure is:

$$\begin{aligned} T_{bad}^{total} &= T_{bad} + T_{good} * (1-q) * \{ \quad (21) \\ & \frac{1 - N*(1-q)^{N-1} + (N-1)*(1-q)^N}{q * \{1-(1-q)^N\}} \}. \end{aligned}$$

On the other hand, if the packet has been successfully received at the end-destination, it is clear that the total expected transmission energy is

$$T_{good}^{total} = N * T_{good}. \quad (22)$$

Since each end-to-end transmission attempt (initiated at the transport layer by the source) is independent of prior end-to-end retransmissions, the total number of end-to-end transmissions for reliable delivery is geometrically distributed with a mean of $\frac{1}{1-P_{fail}}$; hence, on average, the successful transmission of a packet involves $\frac{1}{1-P_{fail}} - 1$ failed end-to-end transmissions, followed by the final successful one. Accordingly, the total effective number of distinct packet transmissions is

$$T = T_{bad}^{total} * \frac{P_{fail}}{1-P_{fail}} + T_{good}^{total}, \quad (23)$$

where T_{bad}^{total}, T_{good}^{total} and P_{fail} are given by equations (22), (22) and (20) respectively.

REFERENCES

[1] J. Li, et al, "Capacity of Ad Hoc Wireless Networks", *Proceedings of ACM MOBICOM '01*, July 2001.

[2] P. Gupta and P.R. Kumar, "The Capacity of Wireless Networks", *IEEE Transactions on Information Theory,* March 2000.

[3] IEEE Computer Society LAN MAN Standards Committee, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", *IEEE Std. 802.11-1197.*

[4] S. Banerjee and A. Misra, "Minimum Energy Paths for Reliable Communication in Multi-Hop Wireless Networks", *Technical Report CS-TR 4312, University of Maryland at College Park*, December 2001.

[5] S. Floyd, "Connections with Multiple Congested Gateways in Packet-Switched Networks Part 1: One-way Traffic", *Computer Communication Review* October 1991.

[6] M. Matthis, J. Semke, J. Mahdavi and T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", *Computer Communications Review*, July 1997.

[7] J. Padhye, V. Firiou, J. Kurose and D. Towsley, "Modeling TCP Throughput: A Simple Model and its Empirical Validation", *Proceedings of ACM SIGCOMM'98*, September 1998.

[8] S. Singh and C.S. Raghavendra, "PAMAS-PowerAware Multi-Access Protocol with Signaling for Ad Hoc Networks", *ACM Communication Review*, July 1998.

[9] T. Shepard, "A Channel Access Scheme for Large Dense Packet Radio Networks", *Proceedings of ACM SIGCOMM'96*, August 1996.

[10] J. Gomez, A. Campbell, M. Naghshineh and C. Bisdikian, "PARO: A Power-Aware Routing Optimization Scheme for Mobile Ad hoc Networks", *draft-gomez-paro-manet-00.txt*, IETF, Work in Progress, February 2001.

[11] S. Singh, M. Woo and C.S. Raghavendra, "Power-Aware Routing in Mobile Ad Hoc Networks", *Proceedings of ACM MOBICOM '98*, October 1998.

[12] C.K. Toh, H. Cobb and D. Scott, "Performance Evaluation of Battery-Life-Aware Routing Schemes for Wireless Ad Hoc Networks", *Proceedings of IEEE ICC'2001*, June 2001.

[13] L. Kleinrock and J. Silvester, "Optimum Transmission Radii for packet Radio Networks or Why Six is a Magic Number", *Proceedings of the IEEE National Telecommunications Conference*, December 1978.