Optimal Spatial Reuse in Poisson Multi-hop Networks

Kostas Stamatiou and Martin Haenggi
Department of Electrical Engineering
University of Notre Dame, Notre Dame, IN 46556
{kstamati,mhaenggi}@nd.edu

Abstract—We consider a wireless multi-hop network with sources that are Poisson distributed and relays which are placed on the source-destination line. Given a combined TDMA/ALOHA MAC protocol, we explore the following question of optimal spatial reuse: Increasing the number of nodes that are simultaneously scheduled to transmit in a route allows nodes to transmit more often. At the same time, it results in an increase of intra-route and inter-route interference, which has a negative impact on the end-to-end delay and throughput. In a regime of large source-destination distances \( R \), we find that it is delay-optimal for either only one node, or a number of nodes that increases linearly in \( R \), to be scheduled in each slot, depending on the ALOHA probability. If the transmission probability is also optimized, we find that maximum spatial reuse is delay-optimal. Scaling laws for the end-to-end delay and throughput are derived in all cases.

I. INTRODUCTION

The premise of multi-hop transmission in wireless networks is the deployment of intermediate nodes to relay packets from the source to the destination, in scenarios where direct communication is not possible due to power or interference limitations. This paper addresses the design issue pertinent to multi-hop networks, of jointly optimizing the number of hops and intra-route spatial reuse in order to minimize the end-to-end delay. The tradeoff involved is that allowing more nodes to simultaneously access the channel per route leads to a higher spatial reuse, but, at the same time, it increases interference both within the route and in the rest of the network.

Our framework, which is an extension of [1], [2], encompasses random node placement, a channel model with fading, path-loss and interference, and queueing delays associated with multi-hop transmission. In particular, we consider a network of Poisson distributed sources, each with its own destination, and relays placed on the source-destination line. The MAC protocol is a combination of ALOHA and TDMA: Within each route, a group of nodes with a given spatial separation, is given a TDMA token which allows them to transmit with a certain probability. In the next slot, the token is passed to the next group of nodes and so on, until all groups have been given their turn and the TDMA cycle starts again. This protocol is selected in light of its relative simplicity and the fact that it allows us to “tune” the intra-route spatial reuse in a straightforward manner. We find that the optimal reuse depends on the transmit probability. If the latter is also optimized, then maximum reuse is optimal. Scaling laws for the end-to-end delay and throughput are derived for large source-destination distances.

Previous work on Poisson networks has mostly studied the single-hop case, e.g., [3], [4] and multi-hop extensions have focused on throughput but not delay [5]. In [6], the issue of optimal spatial reuse was addressed for line networks in terms of the achievable end-to-end rates, but only a single route was considered.

II. SYSTEM MODEL

A. Network setting

The network consists of an infinite number of sources at locations \( \{x_i\} \), which form a homogeneous PPP \( \Psi = \{x_i\} \subset \mathbb{R}^2 \) of density \( \lambda \). Each source has a destination at distance \( R \) and a random orientation. Packets are relayed from the source to its destination by \( N - 1 \) equidistantly placed relays, \( N \in \mathbb{Z}^+ \) (if \( N = 1 \), we have single-hop transmission).

The sources are backlogged, i.e., they always have packets to transmit. Each relay has an infinite buffer, where packets that are received from the previous node in the route can be stored in a first-in, first-out fashion. Time is divided into packet slots. Within a route of \( N \) hops, a TDMA/ALOHA protocol is observed according to which, at any given time, nodes at a distance of \( d \) hops, \( d = 1, \ldots, N \), are allowed to transmit with a certain probability. If the node is a source, this probability is \( p \) and if it is a relay, it is \( p_r \). Let us label the relays with the numbers \( 1 \) to \( N - 1 \). The protocol operation can be described as follows: At slot 1, the source is allowed to transmit with probability \( p \) and the relays \( d, 2d, \ldots \) are each allowed to transmit\(^1\) independently with probability \( p_r \). At slot 2, the relays \( d + 1, 2d + 1, \ldots \) are each allowed to transmit with probability \( p_r \), and so on, until \( d \) groups of nodes have been given their turn and it is time for the first group to be scheduled again. Note that:

- The number of nodes per group in the typical route may vary, but it is at most \( \lceil \frac{\pi}{\theta} \rceil \) and, on average, \( \frac{\pi}{\theta} \).
- The case \( d = 1 \) corresponds to maximum spatial reuse and the case \( d = N \) to no intra-route spatial reuse.

If the signal-to-interference-and-noise ratio is above a target threshold \( \theta \), a packet is successfully received. If it is not, the transmitting node is informed via an ideal feedback channel\(^1\)Note that a relay which is allowed to transmit is a potential transmitter. It may not transmit as its queue might be empty.

\(^1\)Note that a relay which is allowed to transmit is a potential transmitter. It may not transmit as its queue might be empty.
and the packet remains at the head of its queue until the node gets another opportunity to transmit. We also assume that the nodes have access to a common clock (obtained, e.g., by GPS), i.e., the network is synchronized at the slot level. However, the TDMA schedules need not be aligned in any way, i.e., different groups of nodes across routes might be scheduled in the same slot.

Note that $N$, $d$, $p$ are design parameters, i.e., they are optimizable according to the desired metric(s) for given values of $R$, $\lambda$, $p_r$ and $\theta$.

**B. Physical layer**

The channel between two nodes at distance $r$ includes Rayleigh fading (with a coherence time of one slot) and path-loss according to the law $r^{-b}$, where $b > 2$ is the path-loss exponent. For ease of exposition, we consider an interference-limited setting, i.e., thermal noise is assumed to be negligible and all nodes have the same transmit power, normalized to one. (The analysis can be extended to include thermal noise.)

Consider the hop/slot in the typical route which is subject to the largest number of intra-route interferers, and let the corresponding receiving relay (RX) be located (without loss of generality) at the origin. The signal-to-interference-ratio (SIR) is a random variable (r.v.) defined as

$$\text{SIR} = \frac{A(R/N)^{-b}}{\sum_{z \in \Pi} e_z z B_z \|z\|^{-b} + \sum_{n=1,n \neq z} \left( \frac{\lambda}{\lambda_k} \right)^{r_n^{-b}} e_n B_n}$$

(1)

where

- $A$ is the fading coefficient between RX and its transmitting node, exponentially distributed with unit mean.
- $\Pi$ is the point process of *inter-route* nodes, scheduled at the given slot.
- $e_z = 1$, when the node at location $z$ transmits a packet. If the node is a source, then $\mathbb{P}(e_z = 1) = p$ (the respective probability for a relay follows in the next section).
- $B_z$ is the fading coefficient between the node located at $z$ and RX, and is exponentially distributed with unit mean.
- $n$ is the index of the intra-route node scheduled to transmit in the given slot. For $d = 1, \ldots, N - 1$, the distance of that node from RX is

$$r_n = \begin{cases} \left[ \frac{N}{2d} \right] d - nd + 1, & n = 1, \ldots, \left[ \frac{N}{2d} \right] - 1 \\ nd - \left[ \frac{N}{2d} \right] d - 1, & n = \left[ \frac{N}{2d} \right] + 1, \ldots, \left[ \frac{N}{d} \right] \end{cases}$$

(The index $\left[ \frac{N}{d} \right]$ corresponds to the desired transmitter.)
- $\{e_n, B_n\}$ are defined similarly to $\{e_z, B_z\}$.

We denote the total intra-route and inter-route interference as $I_i$ and $I_o$, respectively. The respective SIRs are denoted as SIR$_i$ and SIR$_o$.

**C. Definition of metrics**

The mean end-to-end delay $D$ corresponding to the typical route is defined as the mean total time (in slots) that it takes a packet to travel from the head of the source queue to its destination. $D$ is the sum of the mean service time at the source, and the service times and waiting times along the relays of the route. The service time is measured from the moment a packet reaches the head of the queue until it is successfully received by the next node. The waiting time is measured starting from the moment a packet arrives at a relay’s queue until it becomes the head-of-line packet, i.e., all packets in front of it have been successfully transmitted to the next node. The route throughput $T$ is defined as the expected number of packets successfully delivered to the destination per slot. By definition, $T > 1/D$, i.e., the inverse of the delay provides a lower bound on the throughput. As a result, by minimizing $D$, a lower bound on the throughput is also maximized. The next two sections focus on the evaluation and optimization of $D$.

**III. DELAY ANALYSIS**

In order to make the analysis tractable, assume that packet successes across all hops on all routes are independent events. The assumption is based on the observation that if the probability that a node is a transmitter is small, then, in combination with fading, a sufficient degree of randomization is achieved in the network\(^2\). Moreover, consider the worst-case scenario where packet success probabilities across hops of the typical route are all equal to the smallest one, corresponding to the receiver(s) subject to the largest number of intra-route interferers. Denote this probability by $p_s$. It is then understood that a necessary condition for the relay buffers to be stable is that $p < p_r$, as then the packet arrival probability to the first relay, $p_{ps}$, is smaller than the packet departure probability from the first and all subsequent relays, $p_r p_s$. As formally shown in [7], packet arrivals to all relays are iid geometric with parameter $p_{ps}$, and the probability that a relay located at $z$ transmits a packet is simply $\mathbb{P}(e_z = 1) = p$.

\(^2\)In the case $d = 1$, a relay is either allowed to transmit or receive at a given time, which creates correlation between the success probabilities. However, such a scenario arises when successive nodes actually have packets to transmit, which, as we will see in Section IV, is unlikely for sufficiently small $p$. 
If $d = N$, then, since $\Psi$ is a PPP, it follows from the displacement theorem [8] that the point processes of potential inter-route interferers $\Pi$, and actual inter-route interferers $\Pi_2 = \{z: e_z = 1, t_z = 1\}$, are PPPs with densities $\lambda$ and $\lambda p$ respectively. By Corollary 3.2 in [9], the hop success probability is therefore

$$p_s = \mathbb{P}(\text{SIR}_o \geq \theta) = e^{-\lambda pc(R/N)^2},$$

where $c = \Gamma(1+2/b)\Gamma(1-2/b)\pi^2/\theta$ is the spatial contention parameter [4] and $\Gamma(x), x > 0$, is the gamma function.

For $d < N$, $\Pi_2$ can be well approximated as a PPP of density $\lambda = \lambda kp$, where $k = N/d$ is the average number of scheduled nodes per route. This is shown in Fig. 1, where we have evaluated the success probability via simulation taking into account only inter-route interference, $\mathbb{P}(\text{SIR}_o \geq \theta)$, and compared it to the expression $e^{-\lambda kp(R/N)^2}$, for different $k$, and $N = k, 2k$ (i.e., $d = 1, 2$).

Since $\mathbb{P}(A \geq x) = e^{-x}$, $p_s = \mathbb{P}(\text{SIR} \geq \theta)$, where the SIR is defined in (1), can be written as

$$p_s = \Phi_{I_1}(\gamma)\Phi_{I_2}(\gamma),$$

where $\Phi_{I_1}(\gamma) = \mathbb{E}[e^{-X_1}]$, $s > 0$, denotes the Laplace transform of the pdf of the r.v. $X$ and $\gamma \triangleq \frac{R}{N} \theta$. Based on the observation of the previous paragraph, we can approximate $\Phi_{I_1}(\gamma)$ by [8]

$$\Phi_{I_1}(\gamma) \approx e^{-\lambda kp(R/N)^2} = e^{-\lambda c R^2 \frac{p}{\pi \sigma^2}}.$$ \hfill (3)

Since $\{e_n, B_n\}$ are independent, it is also straightforward to show that [4]

$$\Phi_{I_2}(\gamma) = \prod_n \left(1 - \frac{p}{1 + \frac{p}{n} / \gamma}\right),$$ \hfill (4)

where $n$ takes values as in (1). Note that, for $d = 1$, $\Phi_{I_1}(\gamma)$ includes the term $1 - p$, which is the probability that the receiver has no packets in its queue. Moreover, $\Phi_{I_1}(\gamma)$ does not depend on the hopping distance $R/N$.

A. Delay and throughput expressions

Following the analysis in [1], the service time for the head-of-line packet at the source is

$$H = \frac{d}{pp_s} - d + 1,$$

and, similarly, the service time for the head-of-line packet at a relay is

$$H_r = \frac{d}{p_r p_s} - d + 1.$$

Moreover, the waiting time at the queue of a relay is

$$Q_r = \frac{d p}{p_r (p_r - p) p_s}.$$

The end-to-end delay of the typical route is therefore

$$D = H + (N - 1)(H_r + Q_r) = \frac{d}{pp_s} + d(N - 1)\frac{1 - pp_s}{(p_r - p) p_s} - N(d - 1).$$ \hfill (5)

Since a packet is received by the destination every $d$ slots with probability $pp_s$, the first term is the inverse of the (stable) end-to-end throughput $T = pp_s/d$. The second term is dominated by the value of $(p_r p_s - pp_s)^{-1}$, i.e., the inverse of the difference between the packet service and packet arrival rate at the buffer of each relay. Therefore, in order to minimize $D$, the end-to-end throughput and the time spent in the relay queues must be optimally traded off. Given that for small values of $p$, $1 - pp_s \approx 1$, the simpler expression

$$D = \frac{d}{pp_s} + \frac{d(N - 1)}{(p_r - p) p_s} - N(d - 1).$$ \hfill (6)

provides a tight upper bound to $D$.

IV. DELAY OPTIMIZATION

In this section, we explore the dependence of $D$ on the parameters $N, d, p$. For convenience, we let $N \in [1, +\infty)$ and $d \in [1, N]$ and set $\Phi_{I_1}(\gamma) \approx 1$, i.e., we temporarily ignore intra-route interference. As seen in Fig. 2, $\Phi_{I_1}(\gamma)$ is insensitive to $N$ for $N \geq 10$ and a given $d$. While, for a given $N$, it quickly approaches unity as $d$ increases. In the first case, we thus expect that the approximation will yield a constant, albeit small, performance gap for small $p$, while, in the latter case, this gap will decrease with increasing $d$. These qualitative observations are confirmed by numerical examples.

First, we observe that the following scaling law holds.

**Proposition 1** As $R \to \infty$, $\bar{D} = O(R^2)$ only if $Nd/p = \Theta(R^2)$.

**Proof:** Due to the fact that $p_s \approx e^{-\lambda c R^2}$, we can see from (6) that $Nd/p = o(R^2)$ implies that $\bar{D} = \omega(e R^2)$. If $Nd/p = o(R^2)$, then, fixing $d$ and $p$ implies that $N = \omega(R^2)$, hence $\bar{D} = \omega(R^2)$. \hfill $\blacksquare$

According to Proposition 1, $Nd/p \neq \Theta(R^2)$ results in a scaling of $\bar{D}$ which is worse than quadratic. Hence, in optimizing $D$ over $N, d, p$, we constrain the parameter set to
satisfy \( N d/p = \Theta(R^2) \). Under this constraint, the success probability satisfies \( p_s = e^{-\Theta(1)} \).

A. Fixed source transmission probability \( p \)

Suppose that \( p \) is fixed. The following proposition characterizes the - jointly - delay-optimal \( N \) and \( d \).

**Proposition 2** Denote by \( N_o \) and \( d_o \) the values of \( N \) and \( d \) that jointly minimize (6) when \( R \to +\infty \) and define

\[
q = \frac{1}{p} - \frac{1}{p_r - p}.
\]

(7)

If \( q \leq 0 \), then \( N_o = d_o = \Theta(R) \). If \( q > 1 \), then \( d_o < N_o \) and \( N_o = \Theta(R) \), \( d_o = \Theta(R) \).

**Proof:** From Proposition 1 and (6) we have that

\[
\bar{D} = \frac{q e^{\Theta(1)} d + \frac{gp}{p_r - p} e^{\Theta(1)} - gp \left( 1 - \frac{1}{d} \right)}{1 - \frac{gp}{d}}.
\]

(8)

where \( q \) is defined in (7) and \( g \equiv \frac{N d}{p} \). The derivative with respect to \( d \) is

\[
\frac{\partial \bar{D}}{\partial d} = \frac{q e^{\Theta(1)} - gp}{(p_r - p)^2}.
\]

If \( q \leq 0 \), the derivative is always negative and the minimum \( \bar{D} \) is achieved for the maximum possible \( d \), i.e., \( d_o = N_o \). Since \( N_o d_o = \Theta(R^2) \), we have \( d_o = N_o = \Theta(R) \). If \( q > 1 \), then (8) is minimized for \( d_o = \sqrt{\frac{gp}{d}} < \sqrt{gp} < N_o \). Since \( g = \Theta(R^2) \), we have that \( N_o = \Theta(R) \) and \( d_o = \Theta(R) \).

**Remarks:**

1) The parameter \( q \) depends on the relation of \( p \) to \( p_r \).

- If \( p \geq p_r/2 \), i.e., \( q \leq 0 \), then allowing only one node to transmit per route minimizes the end-to-end delay. The intuition behind this rule is that, since the traffic is heavy in the relay queues (high-traffic regime), it is preferable to keep interference low at the expense of spatial reuse. On the other hand, if the system is operated in a low-traffic regime, i.e., \( q > 1 \), the delay is minimized if \( d_o \) and \( N_o \) are linear functions of \( R \), i.e., if a constant number of nodes \( k_o = N_o/d_o \) (as a function of \( R \)) is scheduled to transmit in each slot. As seen in Fig. 3, the transition of \( q \) from negative to positive values is very steep around \( p = p_r/2 \) (hence the range \( q \in (0, 1) \) is inconsequential). We may speak of a phase transition from no spatial reuse to spatial reuse, as \( p \) becomes smaller than \( p_r/2 \).

2) From (6), we can see that, in both high and low traffic regimes, \( \bar{D} = \Theta(R^2) \) and \( T = \Theta(R^{-1}) \). That no improvement in delay is achieved by allowing spatial reuse is explained by the fact that, on the one hand, nodes get an opportunity to transmit more often, on the other hand the interference in the network increases. The benefit of spatial reuse is manifested in terms of a throughput gain, which is of the order of \( \sqrt{qe^{\Theta(1)}} \).

**Example 1:** Consider a network with parameters \( \lambda = 10^{-4} \) routes/m\(^2 \), \( p_r = 0.1 \), \( p = 0.04 \), \( b = 4 \) and \( \theta = 6 \) dB.

We numerically optimize (5) over \( N \in \mathbb{Z}^+ \) and \( d = 1, \ldots, N \), and plot the results in Figs. 4-8. We observe that the results of Proposition 2 are confirmed and that taking into account intra-route interference results in a larger \( d \) (or smaller \( N \) and \( T \)) than the case where only inter-route interference is taken.
into account. The corresponding delay curves, however, are indistinguishable.

These results illustrate that, for sufficiently small \( p \), a throughput gain is achieved compared to the case of no spatial reuse. However, as seen in Fig. 8, this comes at a cost in terms of the required number of hops. For a fair comparison between the two cases in the sense of required resources, i.e., relays, we constrain the number of hops to be no larger than 50, which is the maximum value of \( N \) with no spatial reuse for \( R = 8000 \) m. The network throughput per unit area is plotted vs. \( R \) in Figs. 9. At \( R \approx 2000 \) m, the constraint on \( N \) takes effect, so, by Proposition 1, \( d \) starts to increase quadratically with \( R \), hence the throughput decreases as \( 1/R^2 \), until \( d = N \). The main message of Fig. 9 is that judicious (over no) spatial reuse results in a throughput gain which depends on \( R \), given the constraint placed on \( N \).

B. Variable source transmission probability \( p \)

We now consider the scenario where \( N, d \) and \( p \) are jointly optimized. We have the following result.

**Proposition 3** Denote by \( N_o, d_o \) and \( p_o \) the values of \( N, d \) and \( p \) that jointly minimize (6) when \( R \to \infty \). Then \( p_o \to 0 \), \( N_o p_o = \Theta(1) \) and \( d_o, N_o^2 = \Theta(R^2) \).

**Proof:** \( D \), given in (8), is strictly convex in \( p \in (0, p_r) \) and \( \lim_{p \to 0^+} D = \lim_{p \to p_r^-} D = +\infty \). As a result, the optimal \( p, p_o \), is obtained by setting \( \partial D/\partial p |_{p=p_o, d=d_o} = 0 \) or

\[
\frac{d_o}{p_o^2} + \frac{d_o}{(p_r - p_o)^2} = \frac{g p_r}{(p_r - p_o)^2} + \frac{g e^{\Theta(1)}}{d_o} \left( \frac{1}{d_o} - 1 \right). \tag{9}
\]

We have the following cases:

1) \( p_o = \Theta(1) \): The scaling with \( R \) on both sides is the same if \( d_o = \Theta(R^2) \). However, the constraint \( N_o d_o/p_o = \Theta(R^2) \) would then imply that \( N_o = \Theta(1) \), which violates the requirement \( d_o \leq N_o \).

2) \( p_o \to 0 \): Due to the constraint \( d \geq 1 \), it is necessary that \( d = \omega(1) \). Since \( g = \Theta(R^2) \), (9) implies that \( d_o/p_o^2 = \Theta(R^2) \).
Given the selected parameters, we obtained that $R_{\text{optimal}}$ for the whole range of scales as $R$ in Proposition 3. Moreover, we verified that $R_{\text{optimal}}$ was negligible. Figs. 4, 10-11 confirm the scaling laws derived with no constraint on $p$. The scaling is $p = \Theta(R^{-1})$; with no reuse, $p = \Theta(R^{-2/3})$ [2], $(\lambda = 10^{-4}$ routes/m², $p_r = 0.1$, $b = 4$ and $\theta = 6$ dB.)

Since $N_o d_o / p_o = \Theta(R^2)$, we have that $N_o p_o = \Theta(1)$ and $d_o N_o^2 = \Theta(R^2)$.

Remarks:
1) Eq. (6) can be rewritten as:

$$\bar{D} = \frac{d}{p_s} \left( \frac{1}{p} + \frac{N-1}{p_r - p} - N \right) + N.$$  

Since $p_0 \to 0$ and $N_0 p_0 = \Theta(1)$, $d = 1$, i.e., maximum spatial reuse minimizes $\bar{D}$. As a result, $N_o = \Theta(R)$ and $p_0 = \Theta(R^{-1})$. The scaling $p_o = \Theta(R^{-1})$ also implies that the probability of any two consecutive relays in the typical route transmitting a packet is of the order $N p_0^2 = \Theta(R^{-1})$, i.e., it goes to zero as $R$ grows large.

2) $\bar{D} = \Theta(R)$ and $T = \Theta(R^{-1})$. Note that these scaling laws are derived with no constraint on $N$. If $N$ (hence $d$) is constrained, then, by Proposition 1, $p = \Theta(R^{-2})$.

Example 2: Consider a network with the same parameters as in Example 1, with the only difference that $p$ is allowed to vary in $(0, 0.1)$. The numerically optimized $D$ is shown in Fig. 4 and the respective throughput and optimal $p$ in Figs. 10-11. Given the selected parameters, we obtained that $d = 1$ was optimal for the whole range of $R$. Moreover, we verified that the effect of intra-route compared to inter-route interference was negligible. Figs. 4, 10-11 confirm the scaling laws derived in Proposition 3.

V. CONCLUSIONS

We proposed a framework to characterize the delay-optimal number of hops and intra-route spatial reuse in Poisson multi-hop networks. The scaling of the delay and throughput as functions of $R$ were characterized. Our results have applications in routing algorithms for multi-hop networks where the relays are also randomly located, e.g., a routing protocol can select the relays which are found closest to the ideal locations determined by the analysis.

ACKNOWLEDGMENTS

The partial support of the DARPA/IPTO IT-MANET program (grant W911NF-07-1-0028) is gratefully acknowledged.

REFERENCES