

# The Effect of Introducing Redundancy in a Probabilistic Forwarding Protocol

Vinay Kumar B.R.<sup>†</sup>Roshan Antony<sup>‡</sup>Navin Kashyap<sup>†</sup>

**Abstract**—This paper is concerned with the problem of broadcasting information from a source node to every node in an ad-hoc network. Flooding, as a broadcast mechanism, involves each node forwarding any packet it receives to all its neighbours. This results in excessive transmissions and thus a high energy expenditure overall. Probabilistic forwarding or gossiping involves each node forwarding a received packet to all its neighbours only with a certain probability  $p$ . In this paper, we study the effect of introducing redundancy, in the form of coded packets, into a probabilistic forwarding protocol. Specifically, we assume that the source node has  $k$  data packets to broadcast, which are encoded into  $n \geq k$  coded packets, such that any  $k$  of these coded packets are sufficient to recover the original  $k$  data packets. Our interest is in determining the minimum forwarding probability  $p$  for a “successful broadcast”, which we take to be the event that the expected fraction of network nodes that receive at least  $k$  of the  $n$  coded packets is close to 1. We examine, via simulations and analysis of a number of different network topologies (e.g., trees, grids, random geometric graphs), how this minimum forwarding probability, and correspondingly, the expected total number of packet transmissions varies with the amount of redundancy added. Our simulation results indicate that over network topologies that are highly connected, the introduction of redundancy into the probabilistic forwarding protocol is useful, as it can significantly reduce the expected total number of transmissions needed for a successful broadcast. On the other hand, for trees, our analysis shows that the expected total number of transmissions needed increases with redundancy.

**Index Terms**—ad-hoc networks, broadcast, gossip, probabilistic forwarding, grid, tree, random geometric graph

## I. INTRODUCTION

An ad-hoc network is a network of nodes which communicate with each other without relying on any centralized infrastructure. They are an integral part of defence operations and rescue missions. Instances include reconnaissance by soldiers, rescue during earthquakes, surveillance using drones, Geographical Information Systems etc.

Many such applications necessitate certain information to be broadcast from a source node to all the other nodes in the network. Flooding is a common strategy for broadcasting content to all nodes. In this strategy, each node, upon receiving a new message packet, forwards it to all its one-hop neighbours. While this strategy is simple and easy to implement, it is wasteful in terms of overall power consumption as the

total number of transmissions across all nodes in the network can be quite high. It is also known to result in the ‘broadcast storm’ problem [1].

An attractive alternative that has been considered in the literature is probabilistic retransmission [2] or probabilistic forwarding, in which each node in the network, upon receiving a new packet, decides to broadcast it to its one-hop neighbours with probability  $p$ , and takes no action with probability  $1 - p$ .

The probabilistic forwarding algorithm has been well studied with reference to ad-hoc networks. It is also referred to as the gossip protocol in some literature; for instance, in [3], gossiping is compared with flooding and is found to save up to 35% message overhead. It is also used alongside routing protocols to improve network performance in terms of end-to-end latency and throughput.

In our scenario, a source node needs to transmit  $k$  message packets to a large fraction of nodes in the network. These  $k$  message packets are encoded into  $n$  coded packets and are transmitted by the source. We assume that, any node receiving at least  $k$  out of these  $n$  coded packets can decode the original  $k$  message packets. The source transmits all  $n$  coded packets with probability 1, whereas other nodes in the network employ the probabilistic forwarding algorithm.

Our goal is to analyze the performance of the above algorithm. In particular, we wish to find the minimum retransmission probability  $p$  for which the expected fraction of nodes receiving at least  $k$  out of the  $n$  coded packets is close to 1, which we deem a “successful broadcast”. In other words, we wish to determine the probability with which every node needs to retransmit a received packet, so that (with high probability) almost all the nodes in the network can decode the  $k$  message packets which the source intended to communicate. This probability yields the minimum value for the expected total number of transmissions across all the network nodes for a successful broadcast. We study the variation of the expected total number of transmissions with redundancy.

Our simulation results show that, over a variety of network topologies that are highly connected — for example, grids, and random geometric graphs above the connectivity threshold — the expected total number of transmissions by all the nodes of the network decreases initially to a minimum and then increases as the redundancy  $\rho = \frac{n-k}{k}$  is increased. This means that there is some value of redundancy which is optimal, in the sense that it minimizes the number of transmissions. Consequently, a network in the grid or the random geometric topology performs best when operated at this value of redun-

<sup>†</sup>Vinay Kumar B.R. and N. Kashyap are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. Email: {vinaykb, nkashyap}@iisc.ac.in

<sup>‡</sup>Roshan Antony is currently with Qualcomm India, Bangalore. Email: roshanantony@outlook.com

dancy and the corresponding minimum forwarding probability. On the other hand, over trees, our simulations and analysis indicate that there is no benefit to introducing redundancy in the probabilistic forwarding protocol: the expected total number of transmissions increases with redundancy.

The rest of the paper is organized as follows. In Section II, we provide a theoretical framework for the problem we are trying to address. Section III has simulation results for the probabilistic forwarding algorithm on different types of network topologies, such as random geometric graphs, grids and trees. In Section IV, we present a mathematical analysis of the protocol on trees, which explains the simulation results obtained for that topology. Section V discusses and gives some heuristic insight into our results.

## II. PROBLEM FORMULATION

Consider a graph  $G = (V, E)$  where  $V$  is the vertex set with  $N$  vertices (nodes) and  $E$  is the set of edges (communication links). A source node  $s \in V$  has  $k$  message packets which need to be broadcast in the network. The source  $s$  encodes the  $k$  messages into  $n$  coded packets using a Maximum Distance Separable (MDS) code (see e.g., [4, Ch. 11]). Such a code is able to correct up to  $n - k$  erasures. Thus, on receiving any  $k$  of these  $n$  coded packets, a node can retrieve the original  $k$  message packets by treating the unreceived packets as erasures. We assume that all the required encoding/decoding operations are carried out over a sufficiently large field, so that an MDS code with the necessary parameters exists. The redundancy introduced in this scheme is  $\rho = \frac{n-k}{k}$ .

The source node broadcasts all  $n$  coded packets to its one-hop neighbours, after which the probabilistic forwarding protocol takes over. A node receiving a particular packet for the first time, transmits it to all its one-hop neighbours with probability  $p$  and takes no action with probability  $1 - p$ . Each packet is transmitted independently of other packets and other nodes. This probabilistic retransmission continues until the time there are no further transmissions in the system. This maximum time is finite since each node in the network decides to transmit a particular packet only the first time it is received. Subsequent receptions of that packet are ignored.

We are interested in the following scenario. Let  $R_{k,\rho}$  be the number of nodes, including the source node, that receive at least  $k$  out of the  $n$  packets. Given a  $\delta \in (0, 1)$ , let  $p_{k,\rho,\delta}$  be the minimum forwarding probability  $p$  such that  $\mathbb{E} \left[ \frac{R_{k,\rho}}{N} \right] \geq 1 - \delta$ . The performance measure of interest, denoted by  $\tau_{k,\rho,\delta}$ , is the expected total number of transmissions across all nodes when the forwarding probability is set to  $p_{k,\rho,\delta}$ . Here, it should be clarified that each network node transmits a given coded packet at most once; a single (broadcast) transmission of a packet by a node is received by all its one-hop neighbours. Since each network node transmits each of the  $n$  coded packets independently with probability  $p_{k,\rho,\delta}$ , the expected number of transmissions by a particular node, assuming it receives all  $n$  packets, is  $np_{k,\rho,\delta}$ . However, a node may not receive all  $n$  packets owing to the structure of the graph. Hence, by independence of packet transmissions across nodes, and the

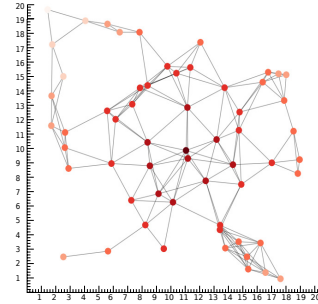


Fig. 1. Random geometric graph.

fact that the source node broadcasts all  $n$  packets, the expected total number of transmissions,  $\tau_{k,\rho,\delta}$ , is bounded above by  $n + (N - 1)np_{k,\rho,\delta}$ . Our aim is to determine, for a given  $k$  and  $\delta$ , how  $\tau_{k,\rho,\delta}$  varies with  $\rho$ , and the value of  $\rho$  at which it is minimized.

## III. SIMULATION RESULTS

Simulations were performed on random geometric graphs, grids and binary trees. For each of these graphs, the value of  $k$  and  $\delta$  was fixed initially and transmission of  $n$  packets was carried out from a source node as explained in the previous section. The value of  $p$  was decreased from 1 and the average number of nodes that receive  $k$  out of the  $n$  packets was computed over 500 simulation trials. The minimum forwarding probability for which this average value exceeded  $1 - \delta$  was recorded as  $p_{k,\rho,\delta}$  along with the corresponding value of  $\tau_{k,\rho,\delta}$ . Both these quantities were plotted as a function of the redundancy  $\rho$ . The results are summarized below.

### A. Random geometric graph (RGG)

RGGs have been widely used to model ad-hoc networks [5]. In an RGG, nodes are distributed uniformly in some region, and two nodes are connected by an edge iff they are at most at a prescribed distance  $r$  from each other. In our simulations, 60 nodes are deployed uniformly in a rectangular region of  $20 \times 20$  units, as shown in Fig. 1. These form the vertices of the graph. The distance  $r$  is chosen so that the RGG operates well above the connectivity threshold<sup>1</sup>. A node is chosen randomly to be the source  $s$ , and the probabilistic forwarding algorithm is simulated. The simulation is done for  $k = 100$ , and  $\delta = 0.1$  and  $0.05$ , and the results are shown in Fig. 2.

It is observed that, for fixed  $k$  and  $\delta$ , the minimum retransmission probability  $p_{k,\rho,\delta}$  decreases as the redundancy  $\rho$  is increased. This is expected since with a larger number of coded packets  $n$ , we can afford to retransmit each packet with a smaller probability while still being able to deliver  $k$  out of the  $n$  packets to a  $1 - \delta$  fraction of the nodes in the network.

A much more interesting trend is that of the expected total number of transmissions  $\tau_{k,\rho,\delta}$ , which initially decreases and then grows gradually as the redundancy  $\rho$  is increased. There is thus an optimal value of  $\rho$  that minimizes  $\tau_{k,\rho,\delta}$ . This happens due to an interplay between two opposing factors: an increase

<sup>1</sup>The connectivity threshold is the least distance  $r$  for which the RGG is connected, with high probability [6, Ch. 7, pp. 158–159].

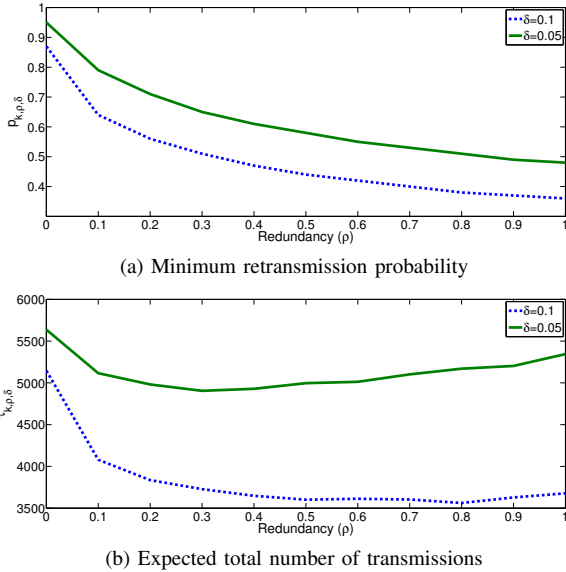


Fig. 2. Probabilistic forwarding on a RGG of 60 nodes in a  $20 \times 20$  area with radius  $r = 5.5$  units.

in  $\rho$  leads to a decrease in  $p_{k,\rho,\delta}$ , which contributes towards a decrease in  $\tau_{k,\rho,\delta}$ . But this is opposed by the fact that a higher redundancy tends to increase the number of transmissions, since there are a larger number of packets to be transmitted in the network. The initial decrease in  $\tau_{k,\rho,\delta}$  can be attributed to the dominant effect of the initial steep decrease in  $p_{k,\rho,\delta}$ . However, as the redundancy is further increased, the decrease in  $p_{k,\rho,\delta}$  becomes more gradual. In this regime, as the number of coded packets  $n$  increases, the gain obtained via the slight decrease in  $p_{k,\rho,\delta}$  is more than offset by the fact that there are more packets to be transmitted in the network. This trend can be seen much more clearly in a grid topology.

### B. Two-Dimensional (2-D) Grid

Consider a  $31 \times 31$  square grid as shown in Fig. 7(a). The source node is assumed to be at the center of the grid and the probabilistic forwarding algorithm is implemented. The simulation results for  $k = 100$  and  $\delta = 0.1, 0.05$  are shown in Fig. 3. The decrease in  $p_{k,\rho,\delta}$  as a function of  $\rho$  is similar to that seen in the RGG setting above. The variation of  $\tau_{k,\rho,\delta}$  with  $\rho$  is also similar to that observed in the RGG, but the trend is more pronounced here. Note that even when  $\rho = 0$ ,  $p_{k,\rho,\delta} \neq 1$ , since we only ask for a  $1 - \delta$  fraction of the nodes to receive at least  $k$  packets.

### C. Binary tree

A rooted binary tree of height  $H$  is the graph depicted in Fig. 5. The probabilistic forwarding strategy with coded packets is simulated over a rooted binary tree of height 10. The root of the tree is the source node. The simulation results are shown in Fig. 4, again for  $k = 100$  and  $\delta = 0.1, 0.05$ . Here, we find that the minimum probability of retransmission decreases with increase in redundancy, but the rate of decrease is much lower than that in the previous two topologies. Consequently, the expected total number of transmissions  $\tau_{k,\rho,\delta}$  always

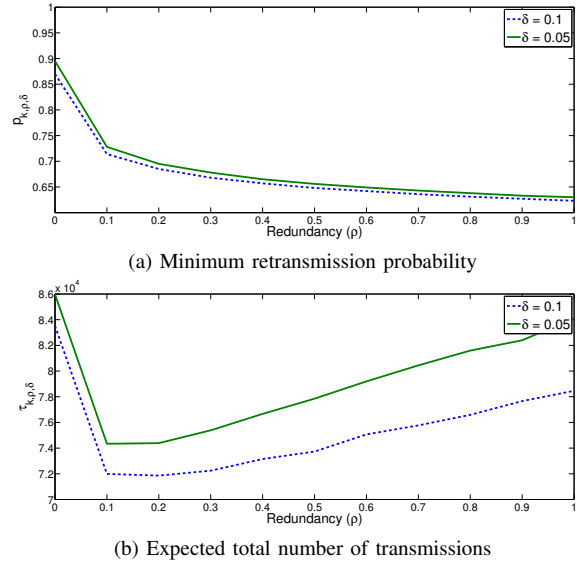


Fig. 3. Probabilistic forwarding on a  $31 \times 31$  grid.

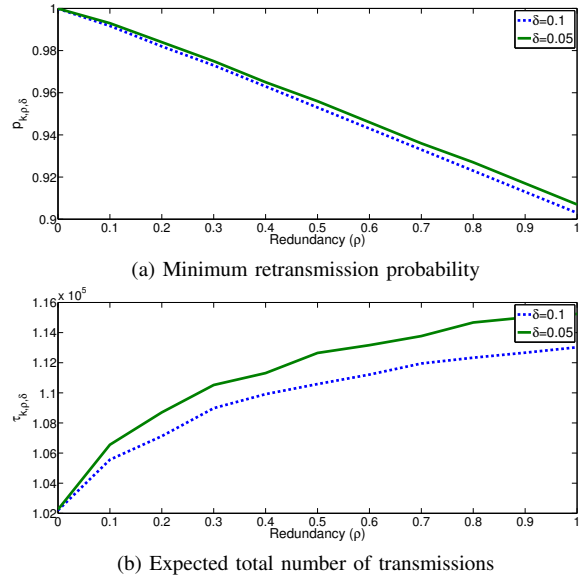


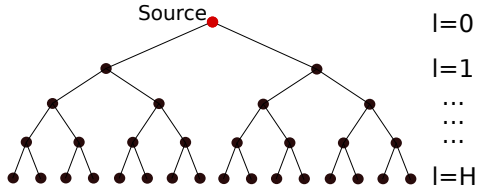
Fig. 4. Probabilistic forwarding on a binary tree of height  $H = 10$ .

increases with redundancy. A theoretical analysis of these observations is presented in the following section.

## IV. ANALYSIS OF PROBABILISTIC FORWARDING ON TREES

In this section, we analyze the probabilistic forwarding mechanism on trees. In particular, we study rooted binary trees. However, the results in this section can be easily extended to  $d$ -ary trees.

Consider again a rooted binary tree of height  $H \geq 1$  as shown in Fig. 5. The tree consists of  $H$  levels, with the root node at level  $l = 0$ , and for  $l = 1, 2, \dots, H - 1$ , each node at level  $l$  having two children at level  $l + 1$ . Thus, there are  $2^l$  nodes at level  $l$ , for  $l = 0, 1, 2, \dots, H$ , so that the total number of nodes in the tree is  $N = \sum_{l=0}^H 2^l = 2^{H+1} - 1$ . The root node is taken to be the source node. It initially has  $k$  message packets which it encodes into  $n = k(1 + \rho)$  coded packets using an MDS code and transmits all of them (with probability


 Fig. 5. A rooted binary tree of height  $H$ .

1) to each of its children. Subsequent transmissions of the  $n$  coded packets follow the probabilistic forwarding mechanism for some fixed value of the retransmission probability  $p \in [0, 1]$ . Note that nodes that share a common parent receive the same packets and hence will possess the same number of packets at the end of the probabilistic forwarding mechanism. We will assume that the nodes at level  $H$  (i.e., the leaf nodes) do not transmit, as there is nothing to be gained in allowing them to do so.

As explained in Section II, the reception of any  $k$  out of the  $n$  coded packets suffices to recover the  $k$  message packets. The notation  $R_{k,\rho}$  was introduced there to denote the random number of nodes (including the source node) that possess at least  $k$  out of the  $n$  coded packets at the end of the protocol. We further let  $T_{k,\rho}$  be the total number of transmissions that take place across all  $N$  nodes during the course of the protocol. Our primary interest is in determining, for a given  $\delta \in (0, 1)$ , the quantity  $\tau_{k,\rho,\delta}$ , which denotes the minimum expected value of  $T_{k,\rho}$ , minimized over all  $p \in [0, 1]$  such that  $\mathbb{E}\left[\frac{R_{k,\rho}}{N}\right] \geq 1 - \delta$ . As we will see below,  $\mathbb{E}[T_{k,\rho}]$  is a monotonically increasing function of  $p$ , so that it is also of interest to determine  $p_{k,\rho,\delta}$ , which is the minimum probability  $p \in [0, 1]$  such that  $\mathbb{E}\left[\frac{R_{k,\rho}}{N}\right] \geq 1 - \delta$ .

We can express  $R_{k,\rho}$  as  $\sum_{l=0}^H R_l$ , where  $R_l$  is the number of nodes at level  $l$  that hold at least  $k$  of the  $n$  packets. Note that  $R_0 = 1$ . Similarly,  $T_{k,\rho} = \sum_{l=0}^{H-1} T_l$ , where  $T_l$  is the number of transmissions by nodes at level  $l$ . Note that  $T_0 = n$ , since the source node always transmits all  $n$  coded packets. Also,  $T_H = 0$ , since leaf nodes are assumed not to transmit. We compute the expected values of  $R_l$  and  $T_l$  next.

Note that there is only a single path from the root to any node in the tree. Thus, for a node  $\mathbf{v}$  at level  $l$  to receive the  $j$ th coded packet from the root, all the intermediate nodes on the unique path from the root to  $\mathbf{v}$  need to transmit the  $j$ th packet. Hence, for  $l \geq 1$ ,

$$\mathbb{P}(\text{node } \mathbf{v} \text{ at level } l \text{ receives the } j\text{th packet}) = p^{l-1}. \quad (1)$$

Since distinct packets are transmitted independently of each other, we have

$$\begin{aligned} & \mathbb{P}(\text{node } \mathbf{v} \text{ at level } l \text{ receives at least } k \text{ out of } n \text{ packets}) \\ &= \sum_{k'=k}^n \binom{n}{k'} p^{(l-1)k'} (1 - p^{l-1})^{n-k'} \\ &= \mathbb{P}(Z_{l-1} \geq k) \end{aligned}$$

where  $Z_{l-1} \sim \text{Bin}(n, p^{l-1})$  is a binomial random variable with parameters  $n$  and  $p^{l-1}$ . Summing the above over all nodes  $\mathbf{v}$

at level  $l$ , we obtain  $\mathbb{E}[R_l] = 2^l \mathbb{P}(Z_{l-1} \geq k)$ , and hence,

$$\mathbb{E}[R_{k,\rho}] = 1 + \mathbb{E}\left[\sum_{l=1}^H R_l\right] = 1 + \sum_{l=1}^H 2^l \mathbb{P}(Z_{l-1} \geq k). \quad (2)$$

Following (1), we also see that the probability that a node  $\mathbf{v}$  at level  $l$  receives and retransmits the  $j$ th packet equals  $p^l$ , for  $l = 1, 2, \dots, H-1$ . Hence, the expected number of transmissions of the  $j$ th packet by nodes at level  $l$  is equal to  $2^l p^l$ . Summing over  $j = 1, 2, \dots, n$ , we obtain that  $\mathbb{E}[T_l] = n(2p)^l = k(1 + \rho)(2p)^l$ , and as a consequence,

$$\mathbb{E}[T_{k,\rho}] = \sum_{l=0}^{H-1} \mathbb{E}[T_l] = k(1 + \rho) \frac{(2p)^H - 1}{2p - 1}.$$

Thus,  $\mathbb{E}[T_{k,\rho}]$  is a monotonically increasing function of  $p$ , from which we infer that

$$\tau_{k,\rho,\delta} = k(1 + \rho) \frac{(2p_{k,\rho,\delta})^H - 1}{2p_{k,\rho,\delta} - 1}. \quad (3)$$

Consequently, to understand how  $\tau_{k,\rho,\delta}$  behaves as a function of the redundancy  $\rho$ , for fixed  $k$  and  $\delta$ , it is necessary to understand how  $p_{k,\rho,\delta}$  varies with  $\rho$ . In the remainder of this section, we derive a good approximation for  $p_{k,\rho,\delta}$  valid for all  $\delta$  sufficiently close to 0 and for all sufficiently large  $k$ . We will also assume that the height  $H$  of the tree is large enough that  $2^{H+1} \gg 1$ .

From (2), we see that  $p_{k,\rho,\delta}$  is the least value of  $p \in [0, 1]$  for which

$$\frac{1}{2^{H+1} - 1} + \frac{\sum_{l=0}^{H-1} 2^{l+1} \mathbb{P}(Z_l \geq k)}{2^{H+1} - 1} \geq 1 - \delta. \quad (4)$$

where  $Z_l \sim \text{Bin}(k(1 + \rho), p^l)$  for  $l = 0, 1, \dots, H-1$ . To proceed, we need the following lemma.

**Lemma IV.1.** For  $\zeta \sim \text{Bin}(n, p)$  and  $0 \leq k \leq n$ ,  $\mathbb{P}(\zeta \geq k)$  is a continuous, monotonically increasing function of  $p$ .

*Proof.* Note first that  $\mathbb{P}(\zeta \geq k) = \sum_{k' \geq k} \binom{n}{k'} p^{k'} (1 - p)^{n-k'}$ , which is a polynomial in  $p$ , and hence  $\mathbb{P}(\zeta \geq k)$  is continuous in  $p$ . Monotonicity is by a standard coupling argument: Let  $U_i, i = 1, 2, \dots, n$ , be i.i.d.  $\text{Unif}[0, 1]$  random variables. For  $p \leq p'$ , let  $X_i = \mathbb{I}_{\{U_i \leq p\}}$  and  $X'_i = \mathbb{I}_{\{U_i \leq p'\}}$ , so that the  $X_i$ s are i.i.d.  $\text{Ber}(p)$  and the  $X'_i$ s are i.i.d.  $\text{Ber}(p')$ . Then,  $\zeta = \sum_{i=1}^n X_i$  is  $\text{Bin}(n, p)$ , while  $\zeta' = \sum_{i=1}^n X'_i$  is  $\text{Bin}(n, p')$ . By construction,  $X_i(U_i) \leq X'_i(U_i)$ , and hence,  $\zeta \leq \zeta'$  almost surely. Thus,  $\mathbb{P}(\zeta \geq k) \leq \mathbb{P}(\zeta' \geq k)$ .  $\square$

The lemma above shows that the LHS of (4) monotonically increases from  $\frac{1}{2^{H+1}-1}$  to 1, as  $p$  goes from 0 to 1. Thus, if  $\delta > 0$  is such that  $1 - \delta$  lies between  $\frac{1}{2^{H+1}-1}$  and 1, then by continuity of the LHS,  $p_{k,\rho,\delta}$  is the least value of  $p$  for which the inequality in (4) holds with equality. To obtain an estimate of  $p_{k,\rho,\delta}$ , we re-write (4) as

$$\frac{\sum_{l=0}^{H-1} 2^{l+1} \mathbb{P}(Z_l \geq k)}{2^{H+1} - 1} \geq 1 - \delta - \frac{1}{2^{H+1} - 1}, \quad (5)$$

and obtain an approximation of the LHS above using the fact that  $Z_l$  concentrates about its mean,  $k(1 + \rho)p^l$ . We will in

fact neglect the term  $\frac{1}{2^{H+1}-1}$  on the right-hand-side (RHS), since we assume that  $2^{H+1} \gg 1$ .

We divide the terms in the summation on the LHS of (5) into two cases, as follows. Set  $l^* := \left\lfloor \frac{\log(\frac{1}{1+\rho})}{\log p} \right\rfloor$ . For simplicity, assume that  $p$  and  $\rho$  are such that  $\frac{\log(\frac{1}{1+\rho})}{\log p}$  is not an integer. (We will later deal with the situation when this assumption does not hold.) Then, we either have  $l \leq l^*$  (Case 1), in which case  $p^l > \frac{1}{1+\rho}$ , or  $l > l^*$  (Case 2), in which case  $p^l < \frac{1}{1+\rho}$ .

- **Case 1:**  $0 \leq l \leq l^*$ , i.e.,  $p^l > \frac{1}{1+\rho}$ . Using the Chernoff bound for the sum of  $n$  independent Bernoulli random variables (see e.g. [7]), we get

$$\begin{aligned} \mathbb{P}(Z_l \geq k) &\geq 1 - \mathbb{P}(Z_l \leq k) \\ &\geq 1 - \exp \left[ -k(1+\rho) \mathbb{D} \left( \frac{1}{1+\rho} \parallel p^l \right) \right], \end{aligned} \quad (6)$$

where  $\mathbb{D}(\cdot \parallel \cdot)$  is the relative entropy, defined as  $\mathbb{D}(r \parallel s) = r \log \frac{r}{s} + (1-r) \log \frac{1-r}{1-s}$ . Thus, for all sufficiently large  $k$ , we have  $P(Z_l \geq k) \approx 1$  for  $l = 0, 1, \dots, l^*$ .

- **Case 2:**  $l^* < l \leq H-1$ , i.e.,  $p^l < \frac{1}{1+\rho}$ . By virtue of Lemma IV.1 and the Chernoff bound for the sum of  $n$  independent Bernoulli random variables, we have

$$\begin{aligned} \mathbb{P}(Z_l \geq k) &\leq \mathbb{P}(Z_{l^*+1} \geq k) \\ &= \exp \left[ -k(1+\rho) \mathbb{D} \left( \frac{1}{1+\rho} \parallel p^{l^*+1} \right) \right]. \end{aligned} \quad (7)$$

Thus, for all sufficiently large  $k$ , we have  $P(Z_l \geq k) \approx 0$  for  $l = l^* + 1, \dots, H-1$ .

Using the above approximations, we obtain the following estimate for the sum in the LHS of (5): for all sufficiently large  $k$ ,

$$\begin{aligned} \sum_{l=0}^{H-1} 2^{l+1} \mathbb{P}(Z_l \geq k) &= \sum_{l=0}^{l^*} 2^{l+1} \mathbb{P}(Z_l \geq k) + \sum_{l=l^*+1}^{H-1} 2^{l+1} \mathbb{P}(Z_l \geq k) \\ &\approx \sum_{l=0}^{l^*} 2^{l+1} = 2^{l^*+2} - 2. \end{aligned}$$

With this, the LHS of (5) becomes

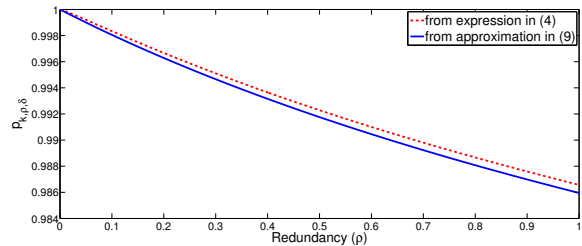
$$\frac{\sum_{l=0}^{H-1} 2^{l+1} \mathbb{P}(Z_l \geq k)}{2^{H+1}-1} \approx \frac{2^{l^*+2} - 2}{2^{H+1}-1} \approx 2^{l^*-H+1},$$

using our assumption that  $2^{H+1} \gg 1$ . Thus, for all  $k$  sufficiently large, the inequality (5) becomes

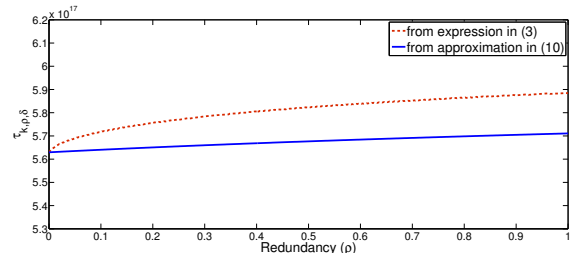
$$2^{l^*-H+1} \gtrsim 1 - \delta \quad (8)$$

The exponent in the LHS above is an integer, and for  $0 \leq \delta < \frac{1}{2}$ , the inequality is effectively the same as  $2^{l^*-H+1} \geq 1$ , which holds iff  $l^* \geq H-1$ . Replacing  $l^*$  by  $\left\lfloor \frac{\log(\frac{1}{1+\rho})}{\log p} \right\rfloor$ , we find that  $l^* \geq H-1$  iff  $p \geq \left(\frac{1}{1+\rho}\right)^{\frac{1}{H-1}}$ .

At this point, we would like to comment on what happens to the analysis above when  $\frac{\log(\frac{1}{1+\rho})}{\log p}$  is in fact an integer, for



(a) Minimum retransmission probability



(b) Expected total number of transmissions

Fig. 6. Comparison between the true values of  $p_{k, \rho, \delta}$  and  $\tau_{k, \rho, \delta}$  obtained from (4) and (3), and their corresponding expressions in (9) and (10), for  $k = 500$ ,  $\delta = 0.1$  and  $H = 50$ .

instance, when  $p = \left(\frac{1}{1+\rho}\right)^{\frac{1}{H-1}}$ . In this case,  $l^* = \frac{\log(\frac{1}{1+\rho})}{\log p}$ , and the bound obtained in Case 1 would hold for all  $l \leq l^* - 1$ , while the bound in Case 2 would hold for all  $l \geq l^* + 1$ . For  $l = l^* = \frac{\log(\frac{1}{1+\rho})}{\log p}$ , the mean of  $Z_{l^*}$  is  $k(1+\rho)p^{l^*} = k$ , and hence, for large  $k$ , we have  $\mathbb{P}(Z_{l^*} \geq k) \approx \frac{1}{2}$ . This additional term needs to be accounted for on the LHS of (5), but this does not result in a significant change in the ensuing analysis.

We can therefore conclude from the discussion following (8) that the least value of  $p$  for which the inequality in (5) holds is, effectively,

$$p_{k, \rho, \delta} = \left( \frac{1}{1+\rho} \right)^{\frac{1}{H-1}}. \quad (9)$$

Correspondingly, the expected total number of transmissions across all  $N$  nodes of the tree is given, via (3), by

$$\begin{aligned} \tau_{k, \rho, \delta} &= k(1+\rho) \frac{2^H \left(\frac{1}{1+\rho}\right)^{\frac{H}{H-1}} - 1}{2 \left(\frac{1}{1+\rho}\right)^{\frac{1}{H-1}} - 1} \\ &= k \frac{2^H \left(\frac{1}{1+\rho}\right)^{\frac{H}{H-1}} - (1+\rho)}{2 \left(\frac{1}{1+\rho}\right)^{\frac{1}{H-1}} - 1} \end{aligned} \quad (10)$$

The expressions in (9) and (10) have been plotted in Fig. 6 for  $k = 500$  message packets and a binary tree of height  $H = 50$ . Also, plotted are the true values of  $p_{k, \rho, \delta}$  obtained from the inequality (4) with  $\delta = 0.1$ , and the corresponding  $\tau_{k, \rho, \delta}$  from (3). The plots allow us to conclude that there is no benefit in introducing redundancy in the form of coding to the probabilistic retransmission protocol on a tree, since the expected total number of transmissions increases with redundancy.



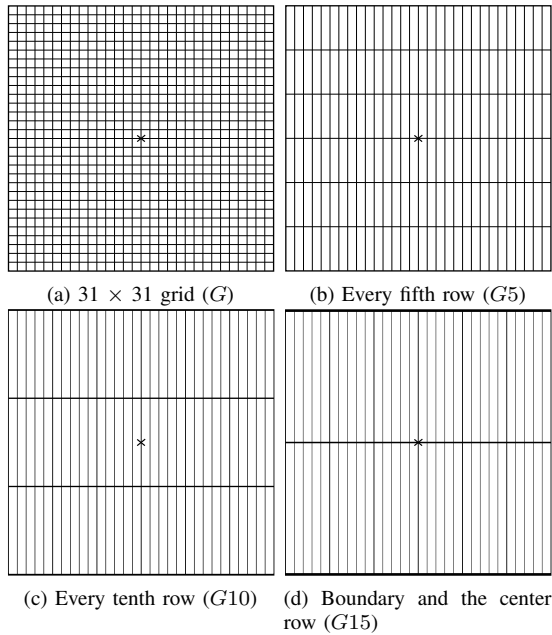
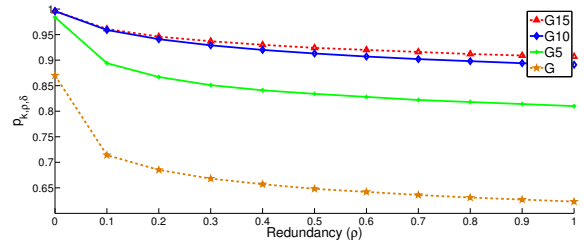


Fig. 7. Graphs to illustrate the importance of multiple paths.

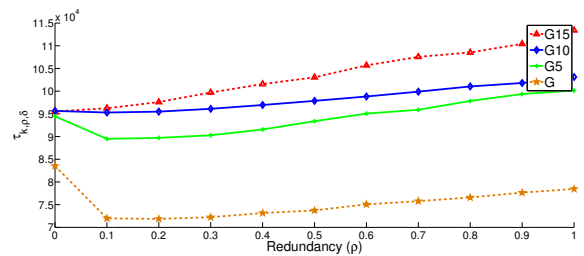
## V. DISCUSSION

The simulations and analysis of the last two sections reveal that there is a significant benefit to introducing coding-based redundancy into the probabilistic retransmission protocol when the underlying network topology is highly connected (as in a large grid), but not so when the underlying network is a tree. The benefit is in terms of a reduction in the overall number of transmissions needed for a successful broadcast. This phenomenon seems to arise from the availability of “multipath diversity” in the network, i.e., the existence of multiple paths between the source node and any other node in the network. Indeed, in a binary tree, there is only one path from the root to any other node, whereas in a large grid, there is abundant multipath diversity.

To test our multipath diversity hypothesis more systematically, we performed further simulations of the probabilistic forwarding protocol on graphs with different levels of multipath diversity. Starting with the  $31 \times 31$  grid  $G$  depicted in Fig. 7(a), we systematically deleted edges to obtain subgraphs  $G_5$ ,  $G_{10}$  and  $G_{15}$  with lower multipath diversity. Specifically, the graph  $G_q$  (for  $q = 5, 10, 15$ ) was obtained from the grid  $G$  as follows. The nodes of  $G$  form a  $31 \times 31$  array, whose rows can be indexed by the integers  $0, 1, 2, \dots, 30$ , with 0 denoting the index of the topmost row. Then,  $G_q$  is obtained from  $G$  by retaining the horizontal edges connecting adjacent nodes in row  $j$ , for every  $j$  that is a multiple of  $q$ , and deleting all other horizontal edges — see Figs. 7(b)–(d). The multipath diversity evidently decreases as  $q$  increases. The results of our simulations, for  $k = 100$  packets, with the expected fraction of nodes receiving at least  $k$  packets being  $1 - \delta = 0.9$ , are shown in Fig. 8. In these simulations, the source node is the node at the centre of the grid, depicted by a ‘x’ in each of the graphs in Fig. 7.



(a) Minimum forwarding probability



(b) Expected total number of transmissions

Fig. 8. Simulation results on  $G$ ,  $G_5$ ,  $G_{10}$  and  $G_{15}$ .

Clearly, the results plotted in Fig. 8 support our hypothesis that multipath diversity in a network is an important factor in determining whether or not the probabilistic forwarding protocol over the network would benefit from coding-based redundancy. In future work, we hope to be able to precisely characterize how multiple paths lead to a reduction in the overall number of transmissions needed for a successful broadcast.

## ACKNOWLEDGEMENTS

The research presented in this paper was supported in part by the DRDO-IISc “Frontiers” Research Programme, and by a Cisco PhD Fellowship awarded to the first author.

## REFERENCES

- [1] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, “The broadcast storm problem in a mobile ad hoc network,” *Wireless Networks*, vol. 8, no. 2/3, pp. 153–167, 2002.
- [2] Y. Sasson, D. Cavin, and A. Schiper, “Probabilistic broadcast for flooding in wireless mobile ad hoc networks,” in *Wireless Communications and Networking, 2003*, vol. 2. IEEE, 2003, pp. 1124–1130.
- [3] Z. J. Haas, J. Y. Halpern, and L. Li, “Gossip-based ad hoc routing,” *IEEE/ACM Transactions on Networking (ToN)*, vol. 14, no. 3, pp. 479–491, 2006.
- [4] R. M. Roth, *Introduction to Coding Theory*. Cambridge Univ. Press, 2006.
- [5] A. Sinclair and A. Stauffer, “Mobile geometric graphs, and detection and communication problems in mobile wireless networks,” *arXiv preprint arXiv:1005.1117*, 2010.
- [6] R. Vaze, *Random Wireless Networks*. Cambridge Univ. Press, 2015.
- [7] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, 2013.