Broadcasting on Trees

Alexander Mozeika[1]

[1]Status.im

E-mail: alexander.mozeika@status.im

Abstract. In this document we consider various mathematical and statistical aspects of broadcasting on trees.

Contents

1. Broadcasting on Linear Trees

1.1. Single-variable model of failures

- We assume that a node sends a message through K communication paths where each path contains exactly L nodes.
- We assumed that $L \times K$ nodes where sampled from the population of N nodes where M nodes are "faulty".
- If a path contains at least one faulty node then *communication* failure occurred.
- If all K paths have communication failure then *broadcast* failure occurred.
- If communication paths are sampled with *replacement* from N nodes with $M < N$ faulty then the probability that a node is faulty is $q = M/N$.
- Let us assume that we observe $\mathbf{s}(1), \ldots, \mathbf{s}(K)$, where $\mathbf{s}(\mu) \in \{0,1\}^L$, sampled from the probability distribution

$$\begin{aligned} \mathrm{P}(\mathbf{s}(\mu)) &= \prod_{\ell=1}^{L} q^{s_\ell(\mu)}(1-q)^{1-s_\ell(\mu)} \\ &= q^{\sum_{\ell=1}^{L} s_\ell(\mu)}(1-q)^{L-\sum_{\ell=1}^{L} s_\ell(\mu)} \end{aligned} \tag{1}$$

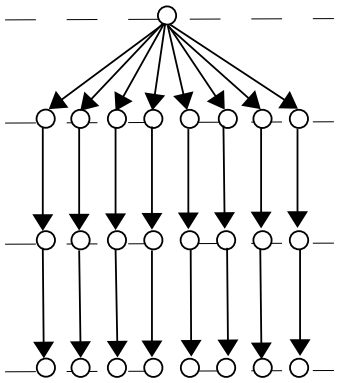

Figure 1. Broadcasting on linear trees. A message is sent through K linear trees with L layers.

- We are interested in probability of the event $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell=1}^{L} x_\ell(\mu) \geq 1] = K$, i.e. the broadcast failure event. To this end we compute moment generating function (MGF) of the random variable $n = \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu) \geq 1\right]$ as follows

$$\begin{aligned}
\Gamma[\lambda] &= \sum_{\mathbf{s}(1)} \cdots \sum_{\mathbf{s}(K)} \prod_{\mu=1}^{K} \mathrm{P}(\mathbf{s}(\mu)) \\
&\quad \times \mathrm{e}^{\lambda \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu) \geq 1\right]} \\
&= \prod_{\mu=1}^{K} \sum_{\mathbf{s}(\mu)} q^{\sum_{\ell=1}^{L} s_\ell(\mu)} (1-q)^{L - \sum_{\ell=1}^{L} s_\ell(\mu)} \mathrm{e}^{\lambda \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu) \geq 1\right]} \\
&= \left[\left[1-(1-q)^L\right] \mathrm{e}^{\lambda} + (1-q)^L\right]^K \\
&= \sum_{n=0}^{K} \binom{K}{n} \left[1-(1-q)^L\right]^n (1-q)^{L(K-n)} \mathrm{e}^{\lambda n} \qquad (2)
\end{aligned}$$

  From above follows that distribution of the random variable $n = \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu) \geq 1\right]$ is the binomial $\binom{K}{n} \left[1-(1-q)^L\right]^n (1-q)^{L(K-n)}$ and hence probability of the broadcast failure event $n = K$ is given by

$$\left[1-(1-q)^L\right]^K. \qquad (3)$$

- We note that in the limit $L \to \infty$, such that $K/L \to 0$, the probability of broadcast failure $\left[1-(1-q)^L\right]^K \to 1$ and in the limit $K \to \infty$, such that $L/K \to 0$, the probability $\left[1-(1-q)^L\right]^K \to 0$.

- Let us now assume that q is the probability that node is "curious". Then the event $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell=1}^{L} x_\ell(\mu) = L] \geq 1$ is the *anonymity* failure event, i.e. there is at least one path where *all* nodes are curious.
- We are interested in probability of the event $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell=1}^{L} x_\ell(\mu) = L] \geq 1$. To this end we compute MGF of the random variable $a = \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu) = L\right]$ as follows

$$\begin{aligned}\Gamma[\lambda] &= \sum_{\mathbf{s}(1)} \cdots \sum_{\mathbf{s}(K)} \prod_{\mu=1}^{K} \mathrm{P}(\mathbf{s}(\mu)) \\ &\quad \times \mathrm{e}^{\lambda \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu)=L\right]} \\ &= \prod_{\mu=1}^{K} \sum_{\mathbf{s}(\mu)} q^{\sum_{\ell=1}^{L} s_\ell(\mu)} (1-q)^{L-\sum_{\ell=1}^{L} s_\ell(\mu)} \mathrm{e}^{\lambda \mathbf{1}\left[\sum_{\ell=1}^{L} x_\ell(\mu)=L\right]} \\ &= \left[q^L \mathrm{e}^{\lambda} + 1 - q^L\right]^K = \sum_{a=0}^{K} \binom{K}{a} q^{La} (1-q^L)^{K-a} \mathrm{e}^{\lambda a} \qquad (4)\end{aligned}$$

  From above follows that distribution of the random variable a is the binomial $\binom{K}{a} q^{La}(1-q^L)^{K-a}$ and hence probability of the anonymity failure event $a \geq 1$ is given by

$$\sum_{a=1}^{K} \binom{K}{a} q^{La} (1-q^L)^{K-a} = 1 - (1-q^L)^K. \qquad (5)$$

- We note that in the limit $L \to \infty$, such that $K/L \to 0$, the probability of anonymity failure $1-(1-q^L)^K \to 0$ and in the limit $K \to \infty$, such that $L/K \to 0$, the probability $1-(1-q^L)^K \to 1$.

1.2. Two-variable model of failures

- We assume that a node sends a message through K communication paths where each path contains exactly L nodes (see Figure 1).
- We assumed that $L \times K$ nodes where sampled with *replacement* from the population of N nodes.
- We assume that M_F nodes in the population are "faulty"‡ and the probability that a node is faulty is $q_F = M_F/N$.
- If a path contains at least one faulty node then *communication* failure occurred.
- If *all* nodes in a communication path are *non-faulty* then this is a *functioning* communication path.
- If *all* K paths have communication failure then *broadcast* failure occurred.
- We assume that M_A nodes in the population are "adversarial"§ and the probability that a node is adversarial is $q_A = M_A/N$.
- Let us define the binary variable $\sigma_\ell(\mu) \in \{0, 1\}$ for a node ℓ in the communication path μ. A node is faulty/not-faulty when $\sigma_\ell(\mu) = 0/1$ with probability $q_F/1-q_F$.

‡ Faulty node is unable to relay a message.
§ Adversarial nodes are controlled by an adversary which can make nodes faulty, use them for traffic analysis, etc.

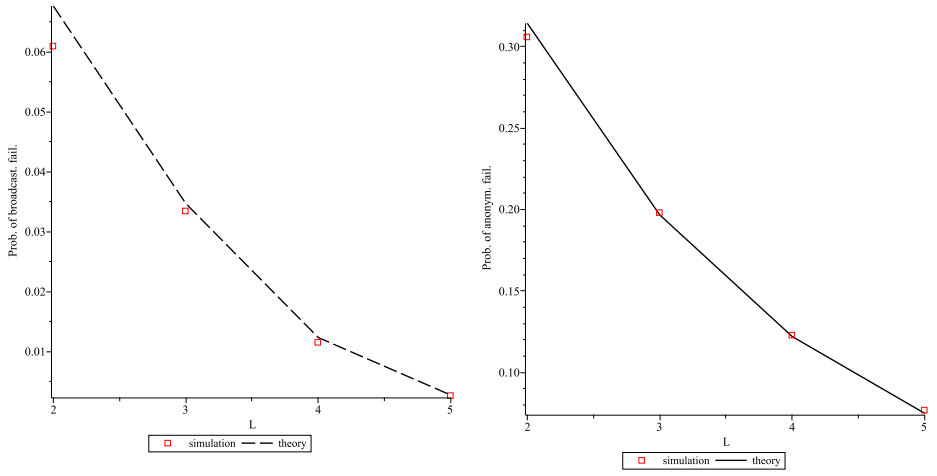

Figure 2. Analysis of failures in 2^L linear trees with L layers (see Figure 1). Left: The probability of broadcast failure plotted as a function of number of layers L for the fraction $q = 0.3$ of faulty nodes. Right: The probability of anonymity failure plotted as a function of the number of layers L for the fraction $q = 0.3$ of "curious" nodes. In simulation probabilities were computed from $M = 10^4$ samples.

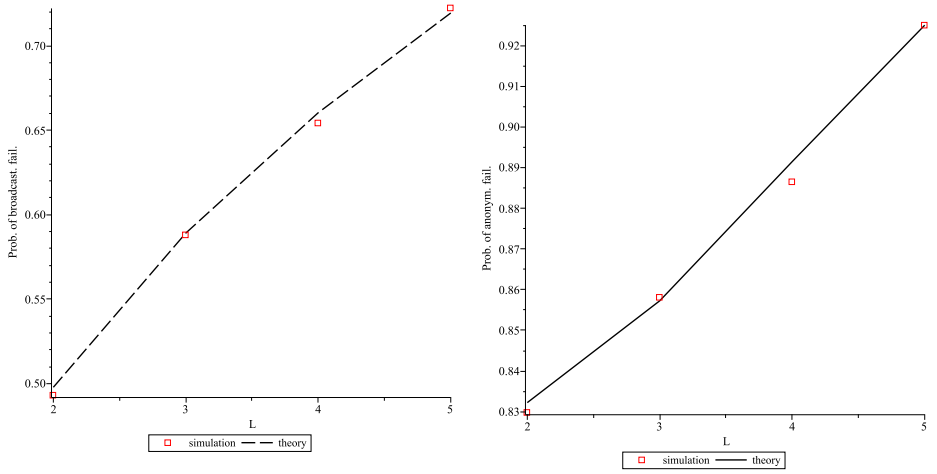

Figure 3. Analysis of failures in 2^L linear trees with L layers (see Figure 1). Left: The probability of broadcast failure plotted as a function of number of layers L for the fraction $q = 0.6$ of faulty nodes. Right: The probability of anonymity failure plotted as a function of the number of layers L for the fraction $q = 0.6$ of "curious" nodes. In simulation probabilities were computed from $M = 10^4$ samples.

- We assume that $\boldsymbol{\sigma}(1), \ldots, \boldsymbol{\sigma}(K)$, where $\boldsymbol{\sigma}(\mu) \in \{0,1\}^L$, sampled from the

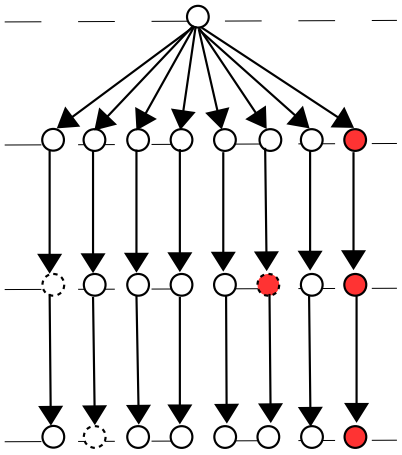

Figure 4. Broadcasting on linear trees. A message is sent through K linear trees with L layers.

probability distribution

$$\begin{aligned} \mathrm{P}(\boldsymbol{\sigma}(\mu)) &= \prod_{\ell=1}^{L} (1-q_F)^{\sigma_\ell(\mu)} q_F^{1-\sigma_\ell(\mu)} \\ &= (1-q_F)^{\sum_{\ell=1}^{L}\sigma_\ell(\mu)} q_F^{L-\sum_{\ell=1}^{L}\sigma_\ell(\mu)} \end{aligned} \tag{6}$$

- If path μ contains at least one faulty node, i.e. $\sum_{\ell=1}^{L}\sigma_\ell(\mu) < L$, then *communication* failure occurred.
- If all K paths have communication failure, i.e. $\sum_{\mu=1}^{K}\mathbf{1}[\sum_{\ell=1}^{L}\sigma_\ell(\mu) < L] = K$, then *broadcast* failure occurred.
- Also we define the binary variable $s_\ell(\mu) \in \{0,1\}$. A node is "honest"/"adversarial" when $s_\ell(\mu) = 0/1$ with probability $1-q_A/q_A$.
- We assume that $\mathbf{s}(1),\ldots,\mathbf{s}(K)$, where $\mathbf{s}(\mu) \in \{0,1\}^L$, are sampled from the probability distribution

$$\begin{aligned} \mathrm{P}(\mathbf{s}(\mu)) &= \prod_{\ell=1}^{L} q_A^{s_\ell(\mu)} (1-q_A)^{1-s_\ell(\mu)} \\ &= q_A^{\sum_{\ell=1}^{L} s_\ell(\mu)} (1-q_A)^{L-\sum_{\ell=1}^{L} s_\ell(\mu)} \end{aligned} \tag{7}$$

- If $\sum_{\mu=1}^{K}\mathbf{1}[\sum_{\ell_1=1}^{L}\sigma_{\ell_1}(\mu) = L]\,\mathbf{1}[\sum_{\ell_2=1}^{L}s_{\ell_2}(\mu) \geq 1] = \sum_{\mu=1}^{K}\mathbf{1}[\sum_{\ell_1=1}^{L}\sigma_{\ell_1}(\mu) = L]$ and $\sum_{\mu=1}^{K}\mathbf{1}[\sum_{\ell_1=1}^{L}\sigma_{\ell_1}(\mu) = L] \geq 1$, i.e. *all* functioning communication paths have at least *one* adversarial node, then adversary has *opportunity* to cause *broadcast* failure.
- If $\sum_{\mu=1}^{K}\mathbf{1}[\sum_{\ell_1=1}^{L}\sigma_{\ell_1}(\mu) = L]\,\mathbf{1}[\sum_{\ell_2=1}^{L}s_{\ell_2}(\mu) = L] \geq 1$, i.e. there is at least *one* functioning communication paths where *all* nodes are adversarial, then adversary has *opportunity* to cause *anonymity* failure.

- We are interested in probability of the above two events. The latter are functions of the random variables $n(\mu) = \sum_{\ell=1}^{L} \sigma_\ell(\mu)$ (number of non-faulty nodes in comm. path μ) and $n_A(\mu) = \sum_{\ell=1}^{L} s_\ell(\mu)$ (number of adversarial nodes in comm. path μ) which are *independent*. The prob. distribution of $n(\mu)$ is the binomial

$$\mathrm{P}(n(\mu)|q_F) = \binom{L}{n(\mu)} (1-q_F)^{n(\mu)} q_F^{L-n(\mu)} \tag{8}$$

and prob. distribution of $n_A(\mu)$ is the binomial

$$\mathrm{P}(n_A(\mu)|q_A) = \binom{L}{n_A(\mu)} q_A^{n_A(\mu)} (1-q_A)^{L-n_A(\mu)}. \tag{9}$$

- Above allows us to compute the probability of the event $\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]\, \mathbf{1}[n_A(\mu) \geq 1] = \sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]$ and $\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L] \geq 1$ as follows

$$\begin{aligned}
&\sum_{n(1)} \cdots \sum_{n(K)} \sum_{n_A(1)} \cdots \sum_{n_A(K)} \prod_{\mu=1}^{K} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A) \\
&\times \mathbf{1}\left[\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]\, \mathbf{1}[n_A(\mu) \geq 1] = \sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]\right] \\
&\times \mathbf{1}\left[\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L] \geq 1\right]
\end{aligned} \tag{10}$$

Now the MGF of random variables $\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]\, \mathbf{1}[n_A(\mu) \geq 1]$ and $\sum_{\mu=1}^{K} \mathbf{1}[n(\mu) = L]$ is given by

$$\begin{aligned}
\Gamma[\lambda_1, \lambda_2] &= \sum_{n(1)} \cdots \sum_{n(K)} \sum_{n_A(1)} \cdots \sum_{n_A(K)} \prod_{\mu=1}^{K} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A) \\
&\quad \times \mathrm{e}^{\lambda_1 \sum_{\mu=1}^{K} \mathbf{1}[n(\mu)=L]\, \mathbf{1}[n_A(\mu)\geq 1] + \lambda_2 \sum_{\mu=1}^{K} \mathbf{1}[n(\mu)=L]} \\
&= \prod_{\mu=1}^{K} \sum_{n(\mu)} \sum_{n_A(\mu)} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A) \\
&\quad \times \mathrm{e}^{\mathbf{1}[n(\mu)=L]\, \{\lambda_1 \mathbf{1}[n_A(\mu)\geq 1] + \lambda_2\}} \\
&= \prod_{\mu=1}^{K} \left[\sum_{n(\mu)=0}^{L-1} \sum_{n_A(\mu)} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A) \right. \\
&\quad \left. + \mathrm{P}(L|q_F) \sum_{n_A(\mu)} \mathrm{P}(n_A(\mu)|q_A) \mathrm{e}^{\{\lambda_1 \mathbf{1}[n_A(\mu)\geq 1] + \lambda_2\}} \right] \\
&= \prod_{\mu=1}^{K} \left[1 - \mathrm{P}(L|q_F) \right. \\
&\quad \left. + \mathrm{P}(L|q_F) \sum_{n_A(\mu)} \mathrm{P}(n_A(\mu)|q_A) \mathrm{e}^{\{\lambda_1 \mathbf{1}[n_A(\mu)\geq 1] + \lambda_2\}} \right] \\
&= \left[1 - \mathrm{P}(L|q_F) \right.
\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{P}(L|q_F)\Big\{\mathrm{P}(0|q_A)\,\mathrm{e}^{\lambda_2} + [1-\mathrm{P}(0|q_A)]\,\mathrm{e}^{\lambda_1+\lambda_2}\Big\}\Big]^K \\
&= \Bigg[1 - \mathrm{P}(L|q_F) + \mathrm{P}(L|q_F)\,\mathrm{P}(0|q_A)\,\mathrm{e}^{\lambda_2} \\
&\qquad + \mathrm{P}(L|q_F)\,[1-\mathrm{P}(0|q_A)]\;\mathrm{e}^{\lambda_1+\lambda_2}\Bigg]^K \\
&= \sum_{n=0}^{K}\binom{K}{n}\left[1 - \mathrm{P}(L|q_F) + \mathrm{P}(L|q_F)\,\mathrm{P}(0|q_A)\,\mathrm{e}^{\lambda_2}\right]^n \\
&\qquad \times \left[\mathrm{P}(L|q_F)\,[1-\mathrm{P}(0|q_A)]\;\mathrm{e}^{\lambda_1+\lambda_2}\right]^{K-n} \\
&= \sum_{n=0}^{K}\binom{K}{n}\sum_{\tilde{n}=0}^{n}\binom{n}{\tilde{n}}\,[1-\mathrm{P}(L|q_F)]^{\tilde{n}} \\
&\quad \times \mathrm{P}^{n-\tilde{n}}(L|q_F)\,\mathrm{P}^{n-\tilde{n}}(0|q_A) \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F)\,[1-\mathrm{P}(0|q_A)]^{K-n} \\
&\quad \times \mathrm{e}^{\lambda_1(K-n)}\mathrm{e}^{\lambda_2(K-\tilde{n})}
\end{aligned}
\tag{11}
$$

From above follows that the prob. distribution of random variables $n_1 = \sum_{\mu=1}^{K}\mathbf{1}[n(\mu)=L]\,\mathbf{1}[n_A(\mu)\geq 1]$ and $n_2 = \sum_{\mu=1}^{K}\mathbf{1}[n(\mu)=L]$ is given by

$$
\begin{aligned}
\mathrm{P}(n_1,n_2) = &\sum_{n=0}^{K}\binom{K}{n}\sum_{\tilde{n}=0}^{n}\binom{n}{\tilde{n}}\,[1-\mathrm{P}(L|q_F)]^{\tilde{n}} \\
&\times \mathrm{P}^{n-\tilde{n}}(L|q_F)\,\mathrm{P}^{n-\tilde{n}}(0|q_A) \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F)\,[1-\mathrm{P}(0|q_A)]^{K-n} \\
&\times \delta_{n_1;K-n}\delta_{n_2;K-\tilde{n}}.
\end{aligned}
\tag{12}
$$

From above follows that the probability of *adversarial* broadcast failure is given by

$$
\begin{aligned}
&\sum_{n_1}\sum_{n_2}\mathrm{P}(n_1,n_2)\,\delta_{n_1;n_2}\mathbf{1}[n_2\geq 1] \\
&= \sum_{n=0}^{K}\binom{K}{n}\sum_{\tilde{n}=0}^{n}\binom{n}{\tilde{n}}\,[1-\mathrm{P}(L|q_F)]^{\tilde{n}} \\
&\times \mathrm{P}^{n-\tilde{n}}(L|q_F)\,\mathrm{P}^{n-\tilde{n}}(0|q_A) \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F)\,[1-\mathrm{P}(0|q_A)]^{K-n} \\
&\times \delta_{K-n;K-\tilde{n}}\,\mathbf{1}[K-\tilde{n}\geq 1] \\
= &\sum_{n=0}^{K}\binom{K}{n}\sum_{\tilde{n}=0}^{n}\binom{n}{\tilde{n}}\,[1-\mathrm{P}(L|q_F)]^{\tilde{n}} \\
&\times \mathrm{P}^{n-\tilde{n}}(L|q_F)\,\mathrm{P}^{n-\tilde{n}}(0|q_A) \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F)\,[1-\mathrm{P}(0|q_A)]^{K-n} \\
&\times \delta_{n;\tilde{n}}\,\mathbf{1}[K-1\geq\tilde{n}]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{n=0}^{K} \binom{K}{n} \sum_{\tilde{n}=0}^{n} \binom{n}{n} \left[1 - \mathrm{P}(L|q_F)\right]^n \\
&\quad \times \mathrm{P}^{n-n}(L|q_F)\, \mathrm{P}^{n-n}(0|q_A) \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F) \left[1 - \mathrm{P}(0|q_A)\right]^{K-n} \\
&\quad \times \delta_{n;\tilde{n}}\, \mathbf{1}[K-1 \geq n] \\
&= \sum_{n=0}^{K} \binom{K}{n} \left[1 - \mathrm{P}(L|q_F)\right]^n \\
&\qquad \times \mathrm{P}^{K-n}(L|q_F) \left[1 - \mathrm{P}(0|q_A)\right]^{K-n} \\
&\quad \times \mathbf{1}[K-1 \geq n] \qquad (13)
\end{aligned}
$$

Now consider

$$
\begin{aligned}
&\sum_{n=0}^{K} \binom{K}{n} \left[1 - \mathrm{P}(L|q_F)\right]^n \mathrm{P}^{K-n}(L|q_F) \\
&\qquad \times \left[1 - \mathrm{P}(0|q_A)\right]^{K-n} \mathbf{1}[K-1 \geq n] \\
&= \left[1 - \mathrm{P}(0|q_A)\right]^K \sum_{n=0}^{K} \binom{K}{n} \left[1 - \mathrm{P}(L|q_F)\right]^n \mathrm{P}^{K-n}(L|q_F) \\
&\qquad \times \left[1 - \mathrm{P}(0|q_A)\right]^{-n} \mathbf{1}[K-1 \geq n] \\
&= \left[1 - \mathrm{P}(0|q_A)\right]^K \left[\mathrm{P}(L|q_F) + \frac{1 - \mathrm{P}(L|q_F)}{1 - \mathrm{P}(0|q_A)}\right]^K \\
&\quad - \left[1 - \mathrm{P}(0|q_A)\right]^K \left[1 - \mathrm{P}(L|q_F)\right]^K \left[1 - \mathrm{P}(0|q_A)\right]^{-K} \\
&= \left[1 - \mathrm{P}(0|q_A)\right]^K \left[\mathrm{P}(L|q_F) + \frac{1 - \mathrm{P}(L|q_F)}{1 - \mathrm{P}(0|q_A)}\right]^K \\
&\quad - \left[1 - \mathrm{P}(L|q_F)\right]^K \\
&= \left[\left(1 - \mathrm{P}(0|q_A)\right) \mathrm{P}(L|q_F) + 1 - \mathrm{P}(L|q_F)\right]^K - \left[1 - \mathrm{P}(L|q_F)\right]^K \\
&= \left[1 - \mathrm{P}(0|q_A)\mathrm{P}(L|q_F)\right]^K - \left[1 - \mathrm{P}(L|q_F)\right]^K \\
&= \left[1 - (1-q_A)^L (1-q_F)^L\right]^K - \left[1 - (1-q_F)^L\right]^K \qquad (14)
\end{aligned}
$$

and hence the probability of adversarial broadcast failure is given by

$$
\begin{aligned}
\mathrm{P}_{ab} &= \sum_{n_1} \sum_{n_2} \mathrm{P}(n_1, n_2)\, \delta_{n_1;n_2} \mathbf{1}[n_2 \geq 1] \\
&= \left[1 - (1-q_A)^L (1-q_F)^L\right]^K - \left[1 - (1-q_F)^L\right]^K . \qquad (15)
\end{aligned}
$$

We note that in above $(1-q_A)^L(1-q_F)^L$ is the probability that a communication path doesn't have failed and adversarial nodes.

- To compute the probability of event $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell_1=1}^{L} \sigma_{\ell_1}(\mu) = L]\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu) = L] \geq 1$ we consider MGF of the random variable $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell_1=1}^{L} \sigma_{\ell_1}(\mu) = L]\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu) = L]$ as follows

$$
\Gamma[\lambda] = \sum_{n(1)} \cdots \sum_{n(K)} \sum_{n_A(1)} \cdots \sum_{n_A(K)} \prod_{\mu=1}^{K} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A)
$$

$$\times \mathrm{e}^{\lambda \sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell_1=1}^{L} \sigma_{\ell_1}(\mu)=L]\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu)=L]}$$

$$= \prod_{\mu=1}^{K} \sum_{n(\mu)} \sum_{n_A(\mu)} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A)$$

$$\times \mathrm{e}^{\lambda \mathbf{1}[\sum_{\ell_1=1}^{L} \sigma_{\ell_1}(\mu)=L]\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu)=L]}$$

$$= \prod_{\mu=1}^{K} \Bigg[\sum_{n(\mu)=0}^{L-1} \sum_{n_A(\mu)} \mathrm{P}(n(\mu)|q_F)\, \mathrm{P}(n_A(\mu)|q_A)$$

$$+ \mathrm{P}(L|q_F) \sum_{n_A(\mu)} \mathrm{P}(n_A(\mu)|q_A) \mathrm{e}^{\lambda\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu)=L]} \Bigg]$$

$$= \Bigg[1 - \mathrm{P}(L|q_F)$$

$$+ \mathrm{P}(L|q_F) \left\{ 1 - \mathrm{P}(L|q_A) + \mathrm{P}(L|q_A)\, \mathrm{e}^{\lambda} \right\} \Bigg]^{K}$$

$$= \left[1 - \mathrm{P}(L|q_F)\mathrm{P}(L|q_A) + \mathrm{P}(L|q_F)\mathrm{P}(L|q_A)\, \mathrm{e}^{\lambda} \right]^{K} \tag{16}$$

hence

$$\Gamma[\lambda] = \sum_{n=0}^{K} \binom{K}{n} \left[\mathrm{P}(L|q_F)\mathrm{P}(L|q_A)\right]^{n} \times \left[1 - \mathrm{P}(L|q_F)\mathrm{P}(L|q_A)\right]^{K-n} \mathrm{e}^{\lambda n}. \tag{17}$$

Thus the prob. distribution of random variable $n = \sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell_1=1}^{L} \sigma_{\ell_1}(\mu) = L]\, \mathbf{1}[\sum_{\ell_2=1}^{L} s_{\ell_2}(\mu) = L]$ is given by the binomial

$$\mathrm{P}(n|q_F, q_A) = \binom{K}{n} \left[\mathrm{P}(L|q_F)\mathrm{P}(L|q_A)\right]^{n} \times \left[1 - \mathrm{P}(L|q_F)\mathrm{P}(L|q_A)\right]^{K-n}$$

$$= \binom{K}{n} \left[(1-q_F)q_A\right]^{Ln} \times \left[1 - (1-q_F)^{L} q_A^{L}\right]^{K-n}. \tag{18}$$

From above follows that the probability of the event $n \geq 1$ is given by

$$\sum_{n=1}^{K} \mathrm{P}(n|q_F, q_A) = 1 - \mathrm{P}(0|q_F, q_A) = 1 - \left[1 - (1-q_F)^{L}\, q_A^{L}\right]^{K}. \tag{19}$$

- From above follows that the probability of *anonymity* failure is given by $\mathrm{P}_a = 1 - \left[1 - (1-q_F)^{L}\, q_A^{L}\right]^{K}$.
- We are also interested in probability of the event $\sum_{\mu=1}^{K} \mathbf{1}[\sum_{\ell=1}^{L} \sigma_{\ell}(\mu) < L] = K$, i.e. the *broadcast failure* event due to presence of faulty nodes. To compute prob. of the latter we consider MGF of the random variable $n =$

$\sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} \sigma_\ell(\mu) < L\right]$ as follows

$$\begin{aligned}
\Gamma[\lambda] &= \sum_{\boldsymbol{\sigma}(1)} \cdots \sum_{\boldsymbol{\sigma}(K)} \prod_{\mu=1}^{K} \mathrm{P}(\boldsymbol{\sigma}(\mu)) \\
&\quad \times \mathrm{e}^{\lambda \sum_{\mu=1}^{K} \mathbf{1}\left[\sum_{\ell=1}^{L} \sigma_\ell(\mu) < L\right]} \\
&= \prod_{\mu=1}^{K} \sum_{\boldsymbol{\sigma}(\mu)} (1-q_F)^{\sum_{\ell=1}^{L} \sigma_\ell(\mu)} q_F^{L-\sum_{\ell=1}^{L} \sigma_\ell(\mu)} \mathrm{e}^{\lambda \mathbf{1}\left[\sum_{\ell=1}^{L} \sigma_\ell(\mu) < L\right]} \\
&= \left[\left[1-(1-q_F)^L\right] \mathrm{e}^{\lambda} + (1-q_F)^L\right]^K \\
&= \sum_{n=0}^{K} \binom{K}{n} \left[1-(1-q_F)^L\right]^n (1-q_F)^{L(K-n)} \mathrm{e}^{\lambda n} \qquad (20)
\end{aligned}$$

From above follows that distribution of the random variable n is the binomial $\binom{K}{n} \left[1-(1-q_F)^L\right]^n (1-q_F)^{L(K-n)}$ and hence probability of the broadcast failure event $n = K$ is given by

$$\mathrm{P}_b = \left[1-(1-q_F)^L\right]^K . \qquad (21)$$

- Thus we obtain the following set of equations

$$\begin{aligned}
\mathrm{P}_b &= \left[1-(1-q_F)^L\right]^K \\
\mathrm{P}_a &= 1-\left[1-(1-q_F)^L q_A^L\right]^K \\
\mathrm{P}_{ab} &= \left[1-(1-q_A)^L(1-q_F)^L\right]^K - \left[1-(1-q_F)^L\right]^K \qquad (22)
\end{aligned}$$

  for the probability of broadcast failure P_b, prob. of anonymity failure P_a and prob. of adversarial broadcast failure P_{ab}.

- Let us first consider the prob. of broadcast failure P_b.
- We note that P_b is (monotonic) decreasing function of K and (monotonic) increasing function of L. This result is intuitive as increasing number of communications paths (of fixed length) increases chances that at least one of these paths is functional. Also increasing length of paths (for a fixed number of paths) increases chances that in each path at least one node is faulty.
- For $K \to \infty$, with L fixed, the prob. $\mathrm{P}_b \to 0$ and for $L \to \infty$, with K fixed, the prob. $\mathrm{P}_b \to 1$.
- Let us assume that $K = f(L)$, where $f(L) \in \mathbb{N}$, and consider the prob. P_b. For the latter we have the following inequality

$$\begin{aligned}
\mathrm{P}_b &= \left[1-(1-q_F)^L\right]^K \\
&= \mathrm{e}^{f(L)\log\left(1-(1-q_F)^L\right)} \leq \mathrm{e}^{-f(L)(1-q_F)^L}, \qquad (23)
\end{aligned}$$

  where in above we used $\log(x) \leq x-1$ to obtain inequality.

- Hence for $K = f(L)$ we can have $\mathrm{P}_b \to 0$ when $f(L)(1-q_F)^L \to \infty$ as $L \to \infty$.
- Second, we consider the prob. of anonymity failure P_a.
- We note that P_a is (monotonic) increasing function of K and (monotonic) decreasing function of L.
- For $K \to \infty$, with L fixed, the prob. $\mathrm{P}_a \to 1$ and for $L \to \infty$, with K fixed, the prob. $\mathrm{P}_a \to 0$.

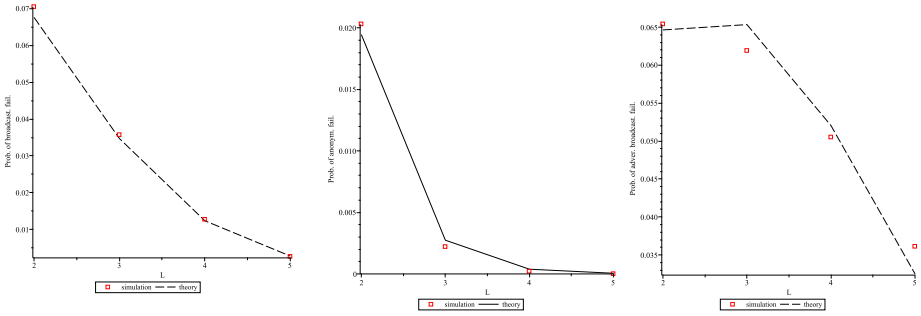

Figure 5. Analysis of failures in 2^L linear trees with L layers (see Figure 4). Left: The probability of broadcast failure plotted as a function of number of layers L for the (average) fraction $q_F = 0.3$ of faulty nodes. Center: The probability of anonymity failure plotted as a function of the number of layers L for the (average) fraction $q_A = 0.1$ of adversarial nodes and $q_F = 0.3$. Right: The probability of adversarial broadcast failure plotted as a function of the number of layers L for $q_F = 0.3$ and $q_A = 0.1$. In simulation probabilities were computed from $M = 10^4$ samples.

- Let us assume that $K = f(L)$, where $f(L) \in \mathbb{N}$, and consider the prob. P_a. For the latter we have the following inequality

$$\mathrm{P}_a = 1 - \left[1 - (1-q_F)^L\, q_A^L\right]^K = 1 - \mathrm{e}^{f(L)\log\left(1-(1-q_F)^L\, q_A^L\right)} \qquad \leq 1 - \mathrm{e}^{-f(L)\frac{(1-q_F)^L\, q_A^L}{1-(1-q_F)^L\, q_A^L}}$$

- Hence for $K = f(L)$ we can have $\mathrm{P}_a \to 0$ when $f(L)\frac{(1-q_F)^L\, q_A^L}{1-(1-q_F)^L\, q_A^L} \to 0$ as $L \to \infty$. We note that $\frac{(1-q_F)^L\, q_A^L}{1-(1-q_F)^L\, q_A^L} = (1-q_F)^L\, q_A^L + O((1-q_F)^{2L}\, q_A^{2L})$ when $L \to \infty$ and hence $f(L)(1-q_F)^L\, q_A^L$ is the dominant term in $f(L)\frac{(1-q_F)^L\, q_A^L}{1-(1-q_F)^L\, q_A^L} \to 0$. Thus $\mathrm{P}_a \to 0$ when $f(L)(1-q_F)^L\, q_A^L \to 0$ as $L \to \infty$.

2. Broadcasting on Branching Trees

- We consider broadcasting on a tree $\mathcal{T}_L$ with layers‖ labeled, from leaf nodes to the root node, by the set $\{0, 1, \ldots, L\}$ (see Figure 7).
- We assume that the root node of $\mathcal{T}_L$ is labeled by 0 and the rest of nodes are labeled by the set of natural numbers $\mathbb{N}$.
- We assume that the root node of $\mathcal{T}_L$ is sending a message to leaf nodes.
- A node inside $\mathcal{T}_L$ is relaying a message to b nodes.
- The set of all leaf node is the "boundary" $\partial\mathcal{T}_L$ of the tree $\mathcal{T}_L$.

‖ All nodes in a tree at the same distance from its root constitute a *layer*.

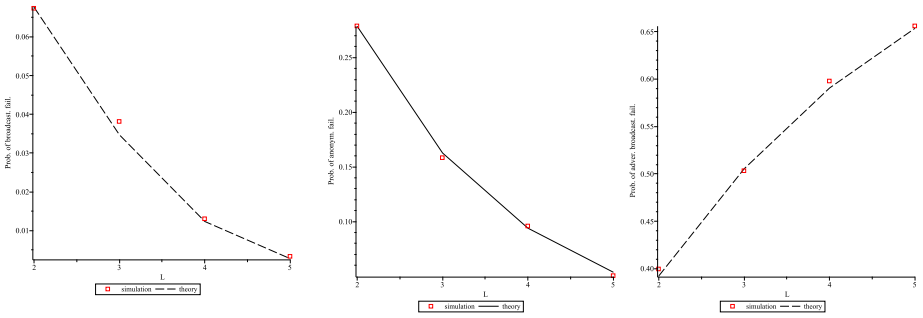

Figure 6. Analysis of failures in 2^L linear trees with L layers (see Figure 4). Left: The probability of broadcast failure plotted as a function of number of layers L for the (average) fraction $q_F = 0.3$ of faulty nodes. Center: The probability of anonymity failure plotted as a function of the number of layers L for the (average) fraction $q_A = 0.4$ of adversarial nodes and $q_F = 0.3$. Right: The probability of adversarial broadcast failure plotted as a function of the number of layers L for $q_F = 0.3$ and $q_A = 0.4$. In simulation probabilities were computed from $M = 10^4$ samples.

- The number of leaf nodes $|\partial\mathcal{T}_L| = b^L$ is also the number of paths from the root node to leaf nodes.
- We assume that each node in the tree $\mathcal{T}_L$, but the root, has associated with it binary random variable which models some internal state, such as "failed"/"not failed", etc., of this node.

2.1. Single-variable model of failures

2.1.1. Analysis of communication failure

- We consider broadcast on the tree $\mathcal{T}_L$ (see Figure 7).
- We assume that the root node of $\mathcal{T}_L$ sends a message to all leaf nodes.
- We assume that each node can fail to relay the message with probability q independently from other nodes.
- The internal state of node $i > 0$, s_i, is 1 when it failed to relay the message and 0 otherwise.
- If the sum of all s_i variables of nodes on the path from the root to some leaf node j, $\sum_{k\in 0\to j\setminus 0} s_k$ is greater than 0, then node j didn't receive the message, i.e. *communication* failure occurred.
- If $\min_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} s_k > 0$ then all leaf nodes didn't receive the message, i.e. *broadcast* failure occurred.
- The $h_0 = \min_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} s_k$ can be computed recursively as follows

$$h_0 = \min_{i\in\partial 0} h_i$$

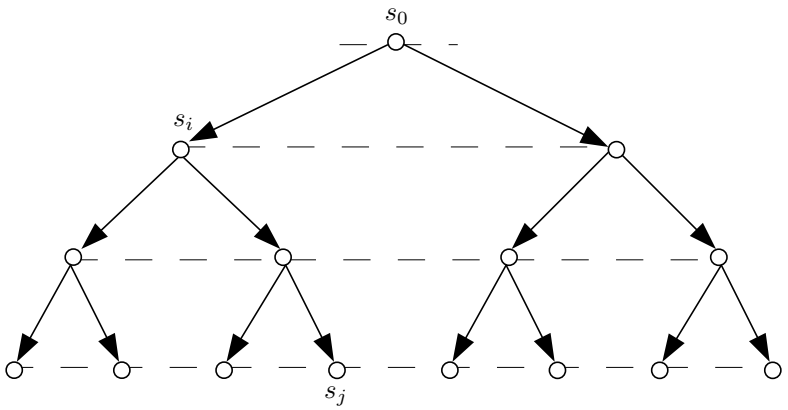

Figure 7. Broadcast on a (balanced and complete) tree $\mathcal{T}_L$. The layers in the tree are labeled by the set $\{0, 1, \ldots, L\}$ (bottom to top). A message is sent from the root node (layer L) to the leaf nodes (layer 0). All leaf nodes of the tree $\mathcal{T}_L$ constitute its boundary $\partial\mathcal{T}_L$. Each node in the tree, but the root, has associated with it binary random variable.

$$h_i = s_i + \min_{k \in \partial i \setminus 0} h_k, \tag{24}$$

where $h_i = \min_{j \in \partial \mathcal{T}_L} D_{i \to j}[\mathcal{T}_L]$ and for the path $i \to j$ we defined the "distance" $D_{i \to j}[\mathcal{T}_L] = \sum_{k \in i \to j} s_k$.

- We note that for $j \in \partial\mathcal{T}_L$ we have $h_j = s_j$.
- Assuming that M message were broadcast on $\mathcal{T}_L$, i.e. M copies of the tree $\mathcal{T}_L$ were created with 0/1 distributed randomly in each copy, we expect that the prob. distributions of random variables h_0 and h_i are governed by the equations

$$\begin{aligned} \mathrm{Q}_L(h) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\, \delta_{h;\, \min_{k \in [b]} h_k} \\ \mathrm{P}_{\ell+1}(h) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{\ell}(h_k) \sum_{s} \mathrm{P}(s)\, \delta_{h;\, s+\min_{k \in [b]} h_k} \\ \mathrm{P}_0(h) &= \sum_{s} \mathrm{P}(s)\, \delta_{h;\, s}, \end{aligned} \tag{25}$$

where in above $\mathrm{P}(s) = q\, \delta_{s;1} + (1-q)\, \delta_{s;0}$, when $M \to \infty$¶.

¶ We need to show that in the limit $M \to \infty$ the (empirical) distribution $\hat{\mathrm{Q}}_L(h) = \frac{1}{M} \sum_{\mu=1}^{M} \delta_{h;h^0(\mu)}$ converges (almost surely) to the average $\mathrm{Q}_L(h) = \frac{1}{M} \sum_{\mu=1}^{M} \langle \delta_{h;h^0(\mu)} \rangle_{\mathbf{s}(\mu)}$ generated by the probability distribution $\mathrm{P}(\mathbf{s}(\mu)) = \prod_{i \in \mathcal{T}_L \setminus 0} \mathrm{P}(s_i(\mu))$, where $\mathrm{P}(s) = q\, \delta_{s;1} + (1-q)\, \delta_{s;0}$, i.e. the random variable $\hat{\mathrm{Q}}_L(h) = \frac{1}{M} \sum_{\mu=1}^{M} \delta_{h;h^0(\mu)}$ is *self-averaging* [1].

- We note that above are density evolution equations [1].
- The domain of $\mathrm{P}_\ell(h)$ is the set $\{0, 1, \ldots, \ell+1\}$.
- We are interested in the probability $\mathrm{Q}_L(h>0) = 1 - \mathrm{Q}_L(h=0)$, i.e. the prob. of broadcast failure, which can be computed as follows

$$\begin{aligned}\mathrm{Q}_L(0) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\, \delta_{0;\, \min_{k\in[b]} h_k} \\ &= 1 - [1 - \mathrm{P}_{L-1}(0)]^b \\ &= 1 - \mathrm{P}_{L-1}^b(h>0)\end{aligned} \tag{26}$$

  Furthermore, we consider

$$\begin{aligned}\mathrm{P}_{\ell+1}(0) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{\ell}(h_k) \sum_{s} \mathrm{P}(s)\, \delta_{0;\, s+\min_{k\in[b]} h_k} \\ &= \mathrm{P}(0) \left[1 - [1 - \mathrm{P}_\ell(0)]^b\right] \\ &= \mathrm{P}(0) \left[1 - \mathrm{P}_\ell^b(h>0)\right], \end{aligned} \tag{27}$$

  where $\mathrm{P}_0(h>0) = q$ and $\mathrm{P}(0) = 1-q$.
- Hence, we have

$$\begin{aligned}\mathrm{Q}_L(h>0) &= \mathrm{P}_{L-1}^b(h>0) \\ \mathrm{P}_{\ell+1}(h>0) &= 1 - (1-q)\left[1 - \mathrm{P}_\ell^b(h>0)\right] \\ \mathrm{P}_0(h>0) &= q\end{aligned} \tag{28}$$

- We note that $\mathrm{P}_\ell(h>0) = 1$ is always solution of the equation

$$\mathrm{P}_{\ell+1}(h>0) = 1 - (1-q)\left[1 - \mathrm{P}_\ell^b(h>0)\right]. \tag{29}$$

  Let us assume that $\mathrm{P}_\ell(h>0) = 1 - \epsilon_\ell$, where $0 < \epsilon_\ell \ll 1$, then above equation gives us

$$\begin{aligned}\epsilon_{\ell+1} &= (1-q)\left[1 - (1-\epsilon_\ell)^b\right] \\ &= (1-q)\, b\, \epsilon_\ell + O(\epsilon_\ell^2)\end{aligned} \tag{30}$$

  Hence $\mathrm{P}_\ell(h>0) = 1$ is a *stable* solution when $(1-q)\, b < 1$ which is equivalent to $q > (b-1)/b$.
- The function $1 - (1-q)\left[1 - \mathrm{P}^b\right]$, i.e. the RHS of the equation (29), is monotonic increasing function of $\mathrm{P} > 0$. Furthermore, it is also *convex* function of P. The latter implies $\mathrm{P}_\ell(h>0) = 1$ is *unique* solution when $q > (b-1)/b$ but when the latter inequality is violated then a second (stable) solution, which is strictly less than 1, appears.
- Let us define $\mathrm{P}_\infty = \lim_{\ell\to\infty} \mathrm{P}_\ell(h>0)$ then from above follows that $\mathrm{P}_L(h>0) = \mathrm{P}_{L-1}^b(h>0) \le \mathrm{P}_\infty^b$, where P_∞ is the solution of the equation

$$\mathrm{P} = 1 - (1-q)\left[1 - \mathrm{P}^b\right], \tag{31}$$

  i.e. the fixed point equation of the equation (29).

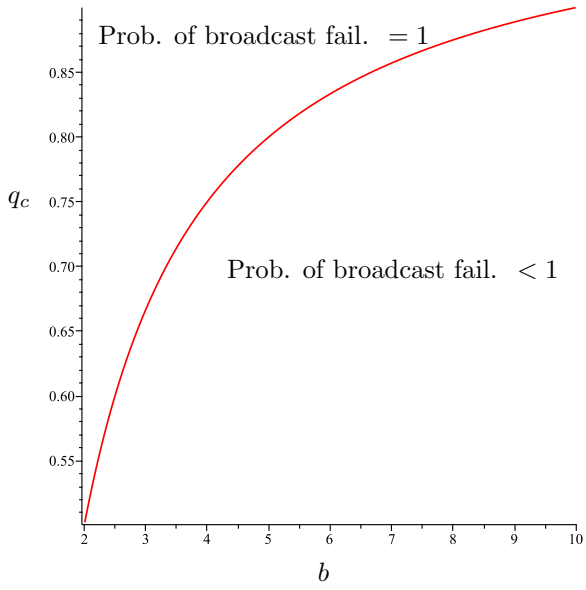

Figure 8. The (critical) probability that a node is faulty, $q_c = (b-1)/b$, as a function of tree branching factor b. For $q > q_c$ broadcast on a tree is only possible for a small number of layers. For $q < q_c$ broadcast is possible for *infinite* number of layers.

- The non-trivial $\mathrm{P}_\infty < 1$ solution of the equation (31) can be computed analytically for small values of the branching parameter b. In particular for $b = 2$ we have $\mathrm{P}_\infty = q/(1-q)$. The latter implies that the prob. of comm. failure $\mathrm{P}_L(h > 0) \leq \mathrm{P}_\infty^2 = q^2/(1-q)^2$ when $q < 1/2$ and $\mathrm{P}_L(h > 0) \leq \mathrm{P}_\infty^2 = 1$ when $q > 1/2$ (here $\mathrm{P}_\infty = 1$). We note that $\mathrm{P}_\infty = \lim_{\ell\to\infty} \mathrm{P}_\ell(h > 0)$ and hence the prob. of comm. failure remains bounded from 1 when $q < 1/2$ even for trees with *infinite* number of layers.

2.1.2. Analysis of anonymity failure

- We consider broadcast on the tree $\mathcal{T}_L$ (see Figure 7).
- We assume that the root node of $\mathcal{T}_L$ sends a message to all leaf nodes.
- We assume that any node in $\mathcal{T}_L$, but the root, can be "curious" with probability q independently from other nodes.
- The internal state of node, s_i, is 1 when node is "curious" and 0 otherwise.
- If the sum of all s_i variables of nodes on the path from the root to some leaf node j, $\sum_{k\in 0\to j\setminus 0} s_k$ is equal to L, then a message sent by node 0 can be associated with this node, i.e. *anonymity* failure occurred[+].
- If $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} s_k = L$ then anonymity failure occurred.

[+] We note that for anonymity failure to happen a some degree of "cooperation" between curious nodes is also required.

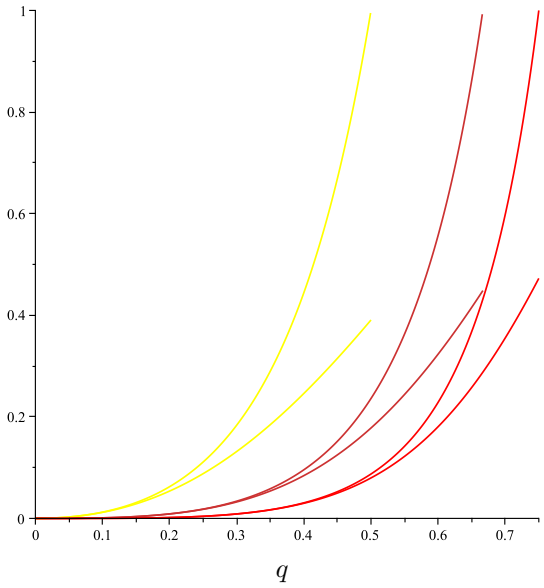

Figure 9. The probability of broadcast failure (lower and upper bound) plotted as a function of probability that a node is faulty, q, for the values of tree branching factor $b = \{2, 3, 4\}$ (yellow, orange, red). Here $q < q_c$ and the lower bound corresponds to a branching tree with 3 layers. The upper bound corresponds to a branching tree with an *infinite* number of layers.

- The $h_0 = \max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} s_k$ can be computed recursively as follows

$$\begin{aligned} h_0 &= \max_{i\in\partial 0} h_i \\ h_i &= s_i + \max_{k\in\partial i\setminus 0} h_k, \end{aligned} \tag{32}$$

  where $h_i = \max_{j\in\partial\mathcal{T}} D_{i\to i}[\mathcal{T}_L]$ and for the path $i \to j$ we defined the "distance" $D_{i\to j}[\mathcal{T}_L] = \sum_{k\in i\to j} s_k$. We note that for $j \in \partial\mathcal{T}$ we have $h_j = s_j$.
- From above follows the density evolution [1] equation

$$\begin{aligned} \mathrm{Q}_L(h) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b}\mathrm{P}_{L-1}(h_k)\,\delta_{h;\,\max_{k\in[b]}h_k} \\ \mathrm{P}_{\ell+1}(h) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b}\mathrm{P}_{\ell}(h_k)\sum_{s}\mathrm{P}(s)\,\delta_{h;\,s+\max_{k\in[b]}h_k} \\ \mathrm{P}_0(h) &= \sum_{s}\mathrm{P}(s)\,\delta_{h;\,s}, \end{aligned} \tag{33}$$

  where in above $\mathrm{P}(s) = q\,\delta_{s;1} + (1-q)\,\delta_{s;0}$.
- The domain of $\mathrm{P}_\ell(h)$ is the set $\{0, 1, \ldots, \ell+1\}$.
- We are interested in the probability $\mathrm{Q}_L(h = L)$ which can be computed as follows

$$\mathrm{Q}_L(h=L) = \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b}\mathrm{P}_{L-1}(h_k)\,\delta_{L;\,\max_{k\in[b]}h_k}$$

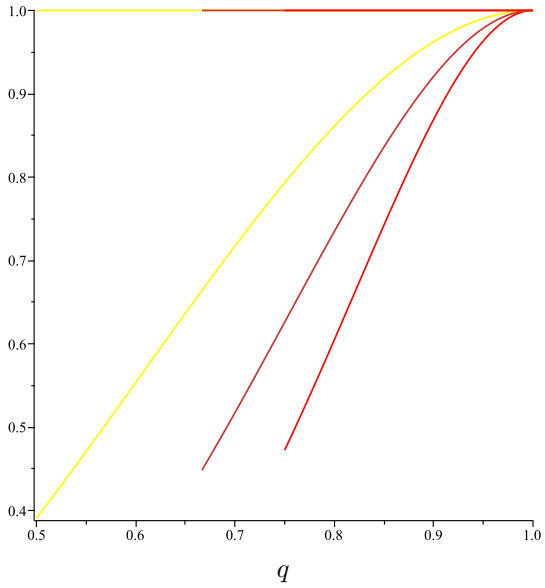

Figure 10. The probability of broadcast failure (lower and upper bound) plotted as a function of probability that a node is faulty, q, for the values of tree branching factor $b = \{2, 3, 4\}$ (yellow, orange, red). Here $q > q_c$ and the lower bound corresponds to a branching tree with 3 layers. The upper bound corresponds to a branching tree with an *infinite* number of layers.

$$= 1 - [1 - \mathrm{P}_{L-1}(h = L)]^b \tag{34}$$

Furthermore, we consider

$$\begin{aligned}
\mathrm{P}_{\ell+1}(h = \ell + 2) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_\ell(h_k) \sum_s \mathrm{P}(s)\, \delta_{\ell+2;\, s+\min_{k\in[b]} h_k} \\
&= \mathrm{P}(1) \left[1 - [1 - \mathrm{P}_\ell(h = \ell + 1)]^b\right] \\
&= q \left[1 - [1 - \mathrm{P}_\ell(h = \ell + 1)]^b\right], \qquad (35)
\end{aligned}$$

where $\mathrm{P}_0(h = 1) = q$. Hence we have the following set of equations

$$\begin{aligned}
\mathrm{Q}_L(h = L) &= 1 - [1 - \mathrm{P}_{L-1}(h = L)]^b \\
\mathrm{P}_{\ell+1}(h = \ell + 2) &= q \left[1 - [1 - \mathrm{P}_\ell(h = \ell + 1)]^b\right] \\
\mathrm{P}_0(h = 1) &= q \qquad (36)
\end{aligned}$$

- Let us define $\mathrm{P}_{\ell+1} = \mathrm{P}_{\ell+1}(h = \ell + 2)$ and consider the equation

$$\mathrm{P}_{\ell+1} = q \left[1 - [1 - \mathrm{P}_\ell]^b\right] \tag{37}$$

We note that $\mathrm{P}_\ell = 0$ is a *fixed point* of above equation. Let us assume $\mathrm{P}_\ell = \epsilon_\ell$, where $0 < \epsilon_\ell \ll 1$, and consider

$$\begin{aligned}
\epsilon_{\ell+1} &= q \left[1 - [1 - \epsilon_\ell]^b\right] \\
&= q\, b\, \epsilon_\ell + O(\epsilon_\ell^2). \qquad (38)
\end{aligned}$$

Hence the fixed point $\mathrm{P}_\ell = 0$ is *stable* when $q < 1/b$. Let us consider the function

$$f(\mathrm{P}) = q\left[1 - [1 - \mathrm{P}]^b\right] \tag{39}$$

on the RHS of the equation (37). This function is monotonic non-decreasing and *concave*. The latter implies that for $q < 1/b$ the equation (37) has only *one* solution $\mathrm{P}_\ell = 0$. However, the fixed point $\mathrm{P}_\ell = 0$ becomes *unstable* when $q > 1/b$ and a second (stable) solution $\mathrm{P}_\ell > 0$ emerges.

2.2. Two-variable model of failures

- We consider broadcast on a tree $\mathcal{T}_L$ (see Figure 11) constructed from nodes sampled (with replacement) from the N nodes of the network.
- We assume that M_F nodes in the network are "faulty"* and the probability that a sampled node is faulty is $q_F = M_F/N$.
- We assume that M_A nodes in the network are "adversarial"♯ and the probability that a sampled node is adversarial is $q_A = M_A/N$.
- We assume that the root node of $\mathcal{T}_L$ sends a message to *all* leaf nodes.
- The root node is labeled by 0 and all leaf nodes constitute the set $\partial\mathcal{T}_L$.
- We assume that each node can fail to relay the message with probability q_F independently from other nodes.
- We assume that a node can be adversarial with probability q_A independently from other nodes.
- Let us define the binary variable $\sigma_i \in \{0, 1\}$ for a node i in some communication path. A node is faulty/not-faulty when $\sigma_i = 0/1$ with probability $q_F/1 - q_F$.
- If the sum of all σ_i variables of nodes on the path from the root 0 to some leaf node j, $\sum_{k\in 0\to j\setminus 0}\sigma_k$ is less than L, i.e. there is at least one faulty node on this path, then node j didn't receive the message, i.e. *communication* failure occurred.
- If *all* nodes in a communication path are *non-faulty* then this is a *functioning* communication path.
- If $\max_{j\in\partial\mathcal{T}_L}\sum_{k\in 0\to j\setminus 0}\sigma_k < L$, i.e. each comm. path contains at least one faulty node, then all leaf nodes didn't receive the message, i.e. *broadcast* failure occurred.
- Equivalently the *broadcast* failure occurred when

$$\sum_{j\in\partial\mathcal{T}_L}\mathbf{1}\left[\sum_{k\in 0\to j\setminus 0}\sigma_k < L\right] = |\partial\mathcal{T}_L|$$

.

- Also we define the binary variable $s_i \in \{0, 1\}$. A node is "honest"/"adversarial" when $s_i = 0/1$ with probability $1 - q_A/q_A$.
- If

$$\sum_{j\in\partial\mathcal{T}_L}\mathbf{1}\left[\sum_{k\in 0\to j\setminus 0}\sigma_k = L\right]\mathbf{1}\left[\sum_{k\in 0\to j\setminus 0}s_k \geq 1\right] = \sum_{j\in\partial\mathcal{T}_L}\mathbf{1}\left[\sum_{k\in 0\to j\setminus 0}\sigma_k = L\right]$$

* Faulty node is unable to relay a message.

♯ Adversarial nodes are controlled by an adversary which can make nodes faulty, use them for traffic analysis, etc.

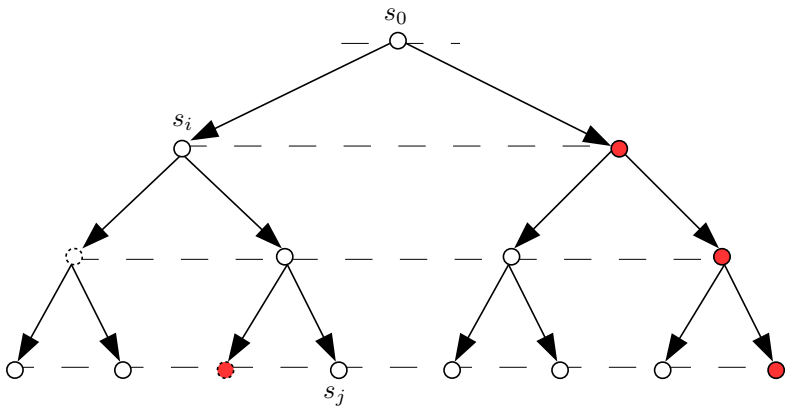

Figure 11. Broadcast on a (balanced and complete) tree $\mathcal{T}_L$. The layers in the tree are labeled by the set $\{0, 1, \ldots, L\}$ (bottom to top). A message is sent from the root node (layer L) to the leaf nodes (layer 0). All leaf nodes of the tree $\mathcal{T}_L$ constitute its boundary $\partial\mathcal{T}_L$. Each node in the tree, but the root, has associated with it binary random variable. A node could be faulty (circle with dashed boundary), or adversarial (red circle). Presence of faulty node leads to communication failures. Presence of adversarial nodes could lead to communication and anonymity failures.

and $\sum_{j\in\partial\mathcal{T}_L} \mathbf{1}[\sum_{k\in 0\to j\setminus 0} \sigma_k = L] \geq 1$, i.e. *all functioning* communication paths have at least *one* adversarial node, then adversary has *opportunity* to cause *broadcast* failure.

- We note that above is equivalent to the event

$$\min_{j\in\partial\mathcal{T}_L \wedge \sum_{k\in 0\to j\setminus 0}\sigma_k = L} \sum_{k\in 0\to j\setminus 0} s_k \geq 1$$

.

- If $\sum_{j\in\partial\mathcal{T}_L} \mathbf{1}[\sum_{k\in 0\to j\setminus 0} \sigma_k = L]\, \mathbf{1}[\sum_{k\in 0\to j\setminus 0} s_k = L] \geq 1$, i.e. there is at least *one* functioning communication paths where *all* nodes are adversarial, then adversary has *opportunity* to cause *anonymity* failure.
- We note that above definition of anonymity failure is equivalent to the event $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k s_k = L$. Here the $\sum_{k\in 0\to j\setminus 0} \sigma_k s_k$ counts number of adversarial nodes on the path $0 \to j$.

2.2.1. Analysis of broadcast failure

- We are interested in the probability of the event $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k < L$.

- The $h_0 = \max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k$ can be computed recursively as follows

$$\begin{aligned} h_0 &= \max_{i\in\partial 0} h_i \\ h_i &= s_i + \max_{k\in\partial i\setminus 0} h_k, \end{aligned} \tag{40}$$

  where $h_i = \max_{j\in\partial\mathcal{T}} \sum_{k\in i\to j} \sigma_k$. We note that for $j \in \partial\mathcal{T}$ we have $h_j = \sigma_j$.
- From above follows the density evolution [1] equation

$$\begin{aligned} \mathrm{Q}_L(h) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\,\delta_{h;\,\max_{k\in[b]}h_k} \\ \mathrm{P}_{\ell+1}(h) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b} \mathrm{P}_{\ell}(h_k)\sum_{\sigma}\mathrm{P}(\sigma)\,\delta_{h;\,\sigma+\max_{k\in[b]}h_k} \\ \mathrm{P}_0(h) &= \sum_{\sigma}\mathrm{P}(\sigma)\,\delta_{h;\,\sigma}, \end{aligned} \tag{41}$$

  where in above $\mathrm{P}(\sigma) = q_F\,\delta_{\sigma;0} + (1-q_F)\,\delta_{\sigma;1}$.
- The domain of $\mathrm{P}_\ell(h)$ is the set $\{0, 1, \ldots, \ell+1\}$.
- The probability of the event $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k < L$ is given by $\mathrm{Q}_L(h < L) = 1 - \mathrm{Q}_L(h = L)$. Let us consider

$$\begin{aligned} \mathrm{Q}_L(h=L) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\,\delta_{L;\,\max_{k\in[b]}h_k} \\ &= 1 - \left[1 - \mathrm{P}_{L-1}(h=L)\right]^b \end{aligned} \tag{42}$$

  Furthermore, we consider

$$\begin{aligned} \mathrm{P}_{\ell+1}(h=\ell+2) &= \sum_{h_1}\cdots\sum_{h_b}\prod_{k=1}^{b} \mathrm{P}_{\ell}(h_k)\sum_{s}\mathrm{P}(s)\,\delta_{\ell+2;\,s+\max_{k\in[b]}h_k} \\ &= \mathrm{P}(1)\left[1 - \left[1-\mathrm{P}_\ell(h=\ell+1)\right]^b\right] \\ &= (1-q_F)\left[1 - \left[1-\mathrm{P}_\ell(h=\ell+1)\right]^b\right], \end{aligned} \tag{43}$$

  where $\mathrm{P}_0(h=1) = 1 - q_F$. Hence we have the following set of equations

$$\begin{aligned} \mathrm{Q}_L(h=L) &= 1 - \left[1-\mathrm{P}_{L-1}(h=L)\right]^b \\ \mathrm{P}_{\ell+1}(h=\ell+2) &= (1-q_F)\left[1 - \left[1-\mathrm{P}_\ell(h=\ell+1)\right]^b\right] \\ \mathrm{P}_0(h=1) &= 1 - q_F \end{aligned} \tag{44}$$

- Let us define $\mathrm{P}_{\ell+1} = \mathrm{P}_{\ell+1}(h=\ell+2)$ and consider the equation

$$\mathrm{P}_{\ell+1} = (1-q_F)\left[1 - \left[1-\mathrm{P}_\ell\right]^b\right] \tag{45}$$

  We note that $\mathrm{P}_\ell = 0$ is a *fixed point* of above equation. Let us assume $\mathrm{P}_\ell = \epsilon_\ell$, where $0 < \epsilon_\ell \ll 1$, and consider

$$\begin{aligned} \epsilon_{\ell+1} &= (1-q_F)\left[1 - \left[1-\epsilon_\ell\right]^b\right] \\ &= (1-q_F)\,b\,\epsilon_\ell + O(\epsilon_\ell^2). \end{aligned} \tag{46}$$

Hence the fixed point $\mathrm{P}_\ell = 0$ is *stable* when $1 - q_F < 1/b$. Let us consider the function

$$f(\mathrm{P}) = (1 - q_F)\left[1 - [1 - \mathrm{P}]^b\right] \tag{47}$$

on the RHS of the equation (45). This function is monotonic non-decreasing and *concave*. The latter implies that for $1 - q_F < 1/b$ the equation (37) has only *one* solution $\mathrm{P}_\ell = 0$. However, the fixed point $\mathrm{P}_\ell = 0$ becomes *unstable* when $1 - q_F > 1/b$ and a second (stable) solution $\mathrm{P}_\ell > 0$ emerges.

2.2.2. Analysis of anonymity failure

- We are interested in the probability of the event $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k s_k = L$.
- The $h_0 = \max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k s_k$ can be computed recursively as follows

$$\begin{aligned} h_0 &= \max_{i\in\partial 0} h_i \\ h_i &= \sigma_i s_i + \max_{k\in\partial i\setminus 0} h_k, \end{aligned} \tag{48}$$

  where $h_i = \max_{j\in\partial\mathcal{T}} \sum_{k\in i\to j} \sigma_k s_k$. We note that for $j \in \partial\mathcal{T}$ we have $h_j = \sigma_j s_j$.
- From above follows the density evolution [1] equation

$$\begin{aligned} \mathrm{Q}_L(h) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\, \delta_{h;\,\max_{k\in[b]} h_k} \\ \mathrm{P}_{\ell+1}(h) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_\ell(h_k) \\ &\quad \times \sum_\sigma \mathrm{P}(\sigma) \sum_s \mathrm{P}(s)\, \delta_{h;\,\sigma s + \max_{k\in[b]} h_k} \\ \mathrm{P}_0(h) &= \sum_\sigma \mathrm{P}(\sigma) \sum_s \mathrm{P}(s)\, \delta_{h;\,\sigma s}, \end{aligned} \tag{49}$$

  where in above $\mathrm{P}(\sigma) = q_F\, \delta_{\sigma;0} + (1 - q_F)\, \delta_{\sigma;1}$ and $\mathrm{P}(s) = q_A\, \delta_{s;1} + (1 - q_A)\, \delta_{s;0}$.
- The domain of $\mathrm{P}_\ell(h)$ is the set $\{0, 1, \ldots, \ell + 1\}$.
- The probability of the event $\max_{j\in\partial\mathcal{T}_L} \sum_{k\in 0\to j\setminus 0} \sigma_k s_k = L$ is given by

$$\begin{aligned} \mathrm{Q}_L(h = L) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_{L-1}(h_k)\, \delta_{L;\,\max_{k\in[b]} h_k} \\ &= 1 - [1 - \mathrm{P}_{L-1}(h = L)]^b \end{aligned} \tag{50}$$

  Furthermore, we consider

$$\begin{aligned} \mathrm{P}_{\ell+1}(h = \ell + 2) &= \sum_{h_1} \cdots \sum_{h_b} \prod_{k=1}^{b} \mathrm{P}_\ell(h_k) \sum_\sigma \\ &\quad \times \mathrm{P}(\sigma) \sum_s \mathrm{P}(s)\, \delta_{\ell+2;\,\sigma s + \max_{k\in[b]} h_k} \\ &= (1 - q_F)\, q_A \left[1 - [1 - \mathrm{P}_\ell(h = \ell + 1)]^b\right], \end{aligned} \tag{51}$$

where $\mathrm{P}_0(h=1) = (1-q_F)q_A$. Hence we have the following set of equations

$$\begin{aligned}
\mathrm{Q}_L(h=L) &= 1-[1-\mathrm{P}_{L-1}(h=L)]^b \\
\mathrm{P}_{\ell+1}(h=\ell+2) &= (1-q_F)\,q_A\left[1-[1-\mathrm{P}_\ell(h=\ell+1)]^b\right] \\
\mathrm{P}_0(h=1) &= (1-q_F)\,q_A
\end{aligned} \tag{52}$$

- Let us define $\mathrm{P}_{\ell+1} = \mathrm{P}_{\ell+1}(h=\ell+2)$ and consider the equation

$$\mathrm{P}_{\ell+1} = (1-q_F)\,q_A\left[1-[1-\mathrm{P}_\ell]^b\right] \tag{53}$$

We note that $\mathrm{P}_\ell = 0$ is a *fixed point* of above equation. Let us assume $\mathrm{P}_\ell = \epsilon_\ell$, where $0 < \epsilon_\ell \ll 1$, and consider

$$\begin{aligned}
\epsilon_{\ell+1} &= (1-q_F)\,q_A\left[1-[1-\epsilon_\ell]^b\right] \\
&= (1-q_F)\,q_A\,b\,\epsilon_\ell + O(\epsilon_\ell^2).
\end{aligned} \tag{54}$$

Hence the fixed point $\mathrm{P}_\ell = 0$ is *stable* when $(1-q_F)\,q_A < 1/b$. Let us consider the function

$$f(\mathrm{P}) = (1-q_F)\,q_A\left[1-[1-\mathrm{P}]^b\right] \tag{55}$$

on the RHS of the equation (45). This function is monotonic non-decreasing and *concave*. The latter implies that for $(1-q_F)\,q_A < 1/b$ the equation (37) has only *one* solution $\mathrm{P}_\ell = 0$. However, the fixed point $\mathrm{P}_\ell = 0$ becomes *unstable* when $(1-q_F)\,q_A > 1/b$ and a second (stable) solution $\mathrm{P}_\ell > 0$ emerges.

2.2.3. Analysis of adversarial broadcast-failure

- We note that (adversarial) broadcast failure event can be written as $\min_{j\in\partial\mathcal{T}_L \wedge \sum_{k_1\in 0\to j\setminus 0}\sigma_{k_1}=L}\sum_{k_2\in 0\to j\setminus 0} s_{k_2} \geq 1$ and $\max_{j\in\partial\mathcal{T}_L}\sum_{k\in 0\to j\setminus 0}\sigma_k = L$. Furthermore, the equality $\sum_{k\in 0\to j\setminus 0}\sigma_k = L$ implies that $\prod_{k\in 0\to j\setminus 0}\sigma_k = 1$ and hence

$$\mathbf{1}\left[\sum_{k_1\in 0\to j\setminus 0}\sigma_{k_1} = L\right]\sum_{k_2\in 0\to j\setminus 0} s_{k_2} = \prod_{k_1\in 0\to j\setminus 0}\sigma_{k_1}\sum_{k_2\in 0\to j\setminus 0} s_{k_2} \tag{56}$$

- The $h_0 = \min_{j\in\partial\mathcal{T}_L\wedge\sum_{k_1\in 0\to j\setminus 0}\sigma_{k_1}=L}\sum_{k_2\in 0\to j\setminus 0} s_{k_2}$ can be computed recursively as follows

$$\begin{aligned}
h_0 &= \min_{i\in\partial 0} h_i \\
h_i &= \min_{j\in\partial\mathcal{T}_L\wedge\sum_{k_1\in i\to j}\sigma_{k_1}=L}\sum_{k_2\in i\to j} s_{k_2} \\
h_i &= \min_{j\in\partial\mathcal{T}_L\wedge\sum_{k_1\in i\to j}\sigma_{k_1}=L}\left[s_i + \sum_{k_2\in i\to j\setminus i} s_{k_2}\right] \\
h_i &= \sigma_i s_i + \max_{k\in\partial i\setminus 0} h_k
\end{aligned} \tag{57}$$

We note that for $j\in\partial\mathcal{T}$ we have $h_j = \sigma_j s_j$.

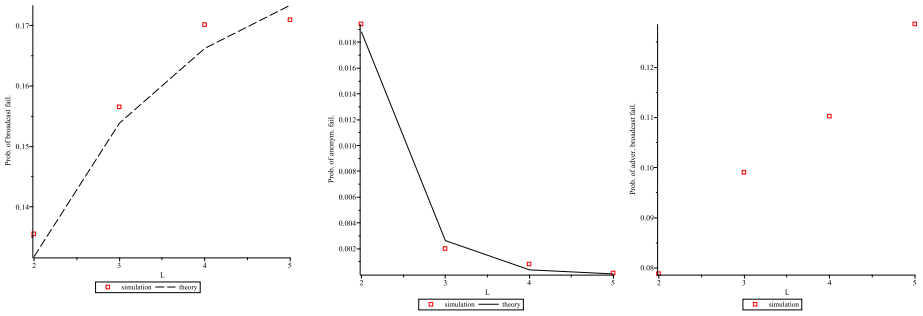

Figure 12. Analysis of failures in branching trees with branching factor $b = 2$ (see Figure 11). Left: The probability of broadcast failure plotted as a function of number of layers L for the (average) fraction $q_F = 0.3$ of faulty nodes. Center: The probability of anonymity failure plotted as a function of the number of layers L for the (average) fraction $q_A = 0.1$ of adversarial nodes and $q_F = 0.3$. Right: The probability of adversarial broadcast failure plotted as a function of the number of layers L for $q_F = 0.3$ and $q_A = 0.1$. In simulation probabilities were computed from $M = 10^4$ samples.

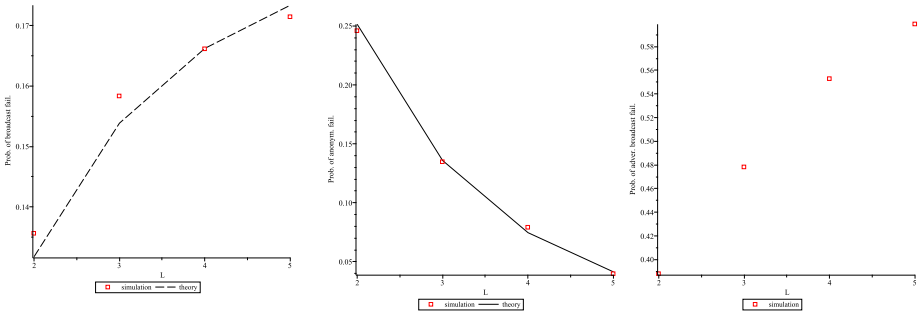

Figure 13. Analysis of failures in branching trees with branching factor $b = 2$ (see Figure 11). Left: The probability of broadcast failure plotted as a function of number of layers L for the (average) fraction $q_F = 0.3$ of faulty nodes. Center: The probability of anonymity failure plotted as a function of the number of layers L for the (average) fraction $q_A = 0.4$ of adversarial nodes and $q_F = 0.3$. Right: The probability of adversarial broadcast failure plotted as a function of the number of layers L for $q_F = 0.3$ and $q_A = 0.4$. In simulation probabilities were computed from $M = 10^4$ samples.

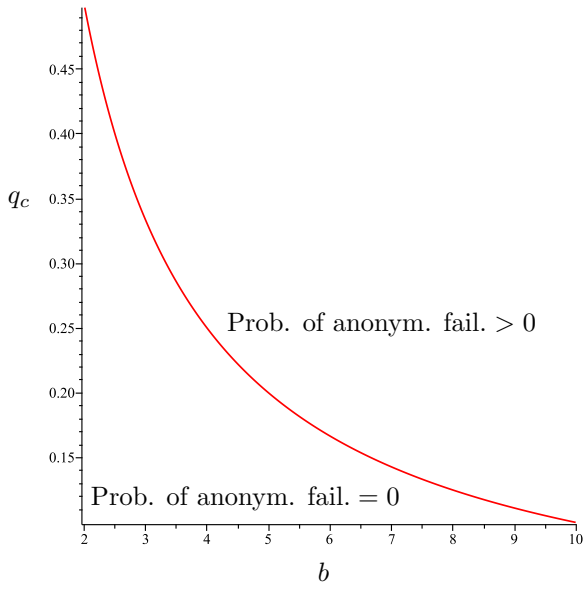

Figure 14. The (critical) probability that a node is "curious", $q_c = 1/b$, as a function of tree branching factor b. For $q_c < q < 1$ the probability of anonymity failure is bounded away from 0 and 1 for *infinite* number of layers, i.e. the probability of anonymity failure is approaching a non-zero value with increasing number of layers in a tree. For $q < q_c$ the probability of anonymity failure is exactly 0 for infinite number of layers.

Acknowledgements

Authors would like to thank...

References

[1] M Mézard and A Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

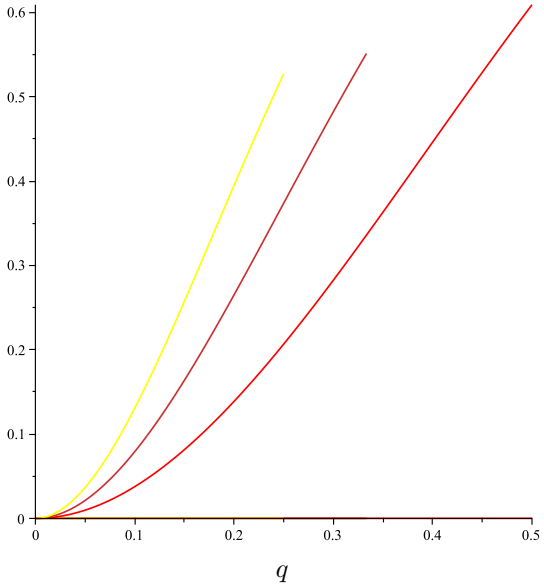

Figure 15. The probability of anonymity failure (lower and upper bound) plotted as a function of probability that a node is "curious", q, for the values of tree branching factor $b = \{2, 3, 4\}$ (red, orange, yellow). Here $q < q_c = 1/b$ and the lower bound, given by 0, corresponds to a branching tree with an *infinite* number of layers. The upper bound corresponds to a branching tree with 3 layers.

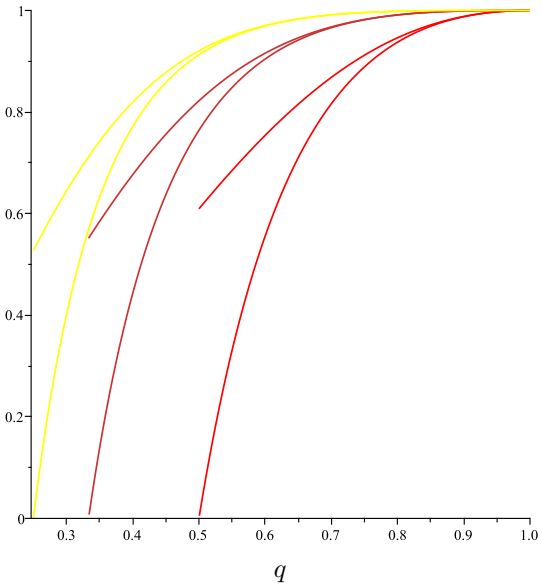

Figure 16. The probability of anonymity failure (lower and upper bound) plotted as a function of probability that a node is "curious", q, for the values of tree branching factor $b = \{2, 3, 4\}$ (red, orange, yellow). Here $q > q_c = 1/b$ and the lower bound corresponds to a branching tree with an *infinite* number of layers. The upper bound corresponds to a branching tree with 3 layers.

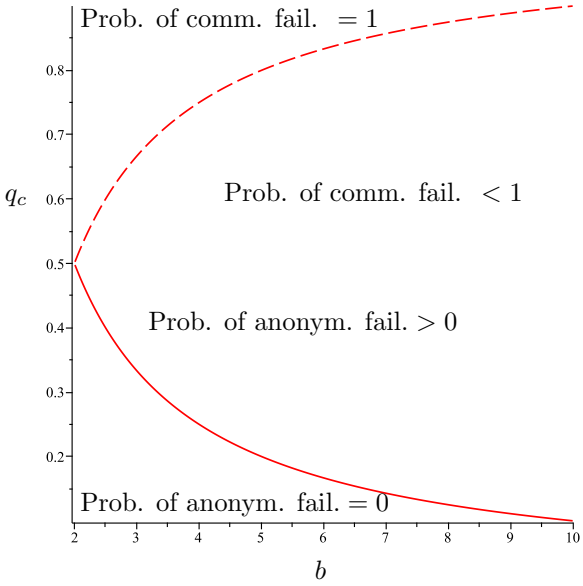

Figure 17. The (critical) probability that a node is faulty (top dashed line), $q_c = (b-1)/b$, and that a node is "curious" (bottom solid line), $q_c = 1/b$, as a function of tree branching factor b.

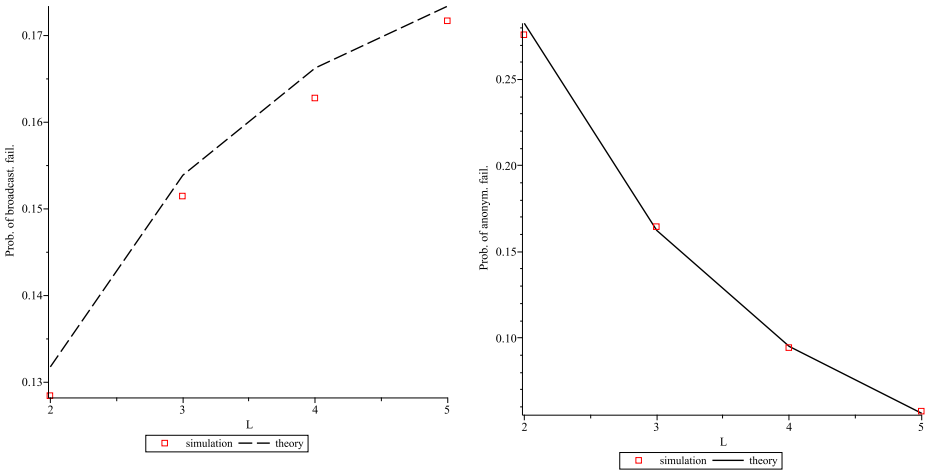

Figure 18. Analysis of failures in branching trees with branching factor $b = 2$ (see Figure 7). Left: The probability of broadcast failure plotted as a function of number of layers L for the fraction $q = 0.3$ of faulty nodes. Right: The probability of anonymity failure plotted as a function of the number of layers L for the fraction $q = 0.3$ of "curious" nodes. In simulation probabilities were computed from $M = 10^4$ samples.

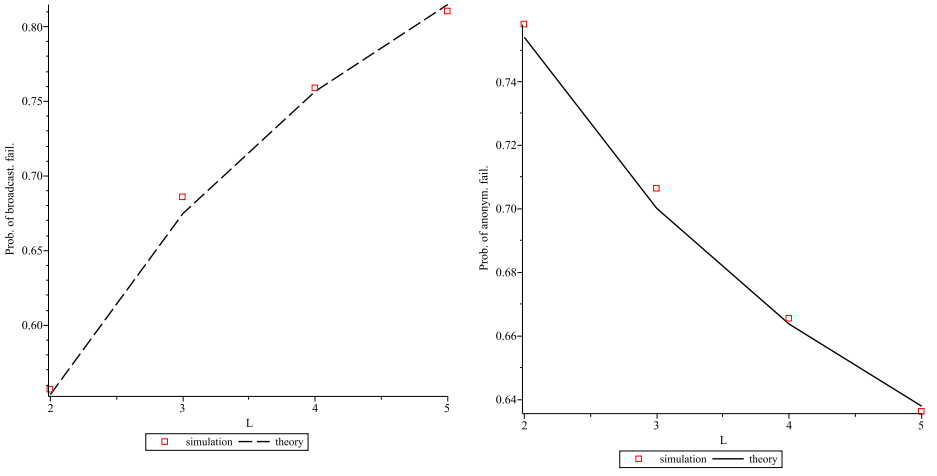

Figure 19. Analysis of failures in branching trees with branching factor $b = 2$ (see Figure 7). Left: The probability of broadcast failure plotted as a function of number of layers L for the fraction $q = 0.6$ of faulty nodes. Right: The probability of anonymity failure plotted as a function of the number of layers L for the fraction $q = 0.6$ of "curious" nodes. In simulation probabilities were computed from $M = 10^4$ samples.