Decoding delay in network coded multipath transmissions

Berksan Serbetci, Jasper Goseling,
Jan-Kees van Ommeren and Richard J. Boucherie

Stochastic Operations Research, University of Twente, The Netherlands

{b.serbetci, j.goseling,
j.c.w.vanommeren, r.j.boucherie}@utwente.nl

Abstract

We investigate the decoding delay performance of a communication network in which a single source is transmitting data packets to a single receiver via multiple routers. Network coding is applied to all data packets at the source at each transmission opportunity. Receiver receives network coded packets from routers and decodes them. We define the delay as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. We show that the delay can be expressed in closed-form.

1 Introduction

In modern communication networks, data packets are transmitted from the gateway to user equipments via base stations. In principle, each base station is responsible of transmitting data packets to users that are present in its coverage. In practice, there are many areas that are covered by multiple base stations. Depending on channel conditions, it may be more viable for the user equipment to receive data packets from different base stations at different transmission opportunities. At some occasions, it is possible that same data packets are requested by multiple users. Then, we can come up with an alternative way of transmitting data packets to these users as in [1], [2], [3]. In these works, it is shown that sending random linear combinations of all data packets is another way of transmitting all data packets and this alternative data transmission scheme is called network coding.

The system consists of a single source transmitting data packets to a single receiver via multiple routers. The source refers to gateway, routers refer to base stations and receiver refers to user equipment. New data packets arrive at the source according to a Poisson process. The intermediate network consists of two routers that receive packets from the source and forward these to the receiver. The source and the routers have exponential service rates. The source transmits network coded packets through the network. In particular, at each transmission opportunity, the source transmits a random linear combination over all data packets that are present at the source at that time. Each network coded packet is then transmitted to one of the routers with probabilistic routing. Once a network coded packet is transmitted to one of the routers, the source drops the data packet that is located at the head of the queue as proposed in [4], [5], [6].

As the receiver obtains network coded packets from multiple routers, it is necessary to decode these network coded packets in order to retrieve the data packets. Decoding is only possible when the number of network coded packets is at least equal to the number of data packets involved in the received linear combinations. We show that once the source queue becomes empty and all network coded packets that have been generated so far have been received by the receiver, it decodes all these network coded packets and retrieve data packets.

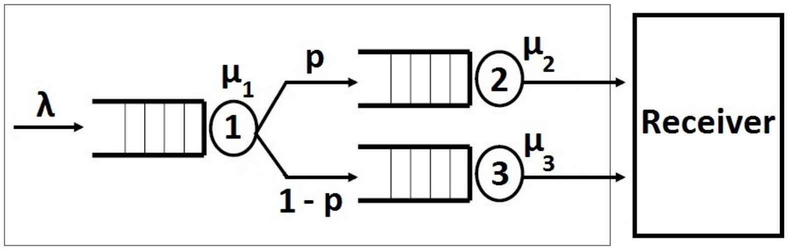

Figure 1: Queueing network for the system.

This work mainly focuses on analyzing the delay where the delay is defined as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. Note that even though sending network coded packets do not save any resource over sending data packets for unicast transmissions, it is still useful to analyze the delay for the unicast system in order to prepare a baseline for future research.

2 Model and Problem Statement

We define the source and the routers as independent $M/M/1$ queues. Data packets arrive at the source according to a Poisson process with rate λ. The source queue is called *Queue* 1 ($Q1$) and has an exponential service rate μ_1. At each transmission opportunity, network coding is applied to all data packets at the source, namely each data packet that is present at the source at that transmission opportunity is multiplied with a random coefficient and the sum of them forms a network coded packet. Then the network coded packet is routed to one of the two routers called *Queue* 2 ($Q2$) and *Queue* 3 ($Q3$) with probabilistic routing with parameter p. Namely, the network coded packet that is ready to be transmitted from the source is routed to $Q2$ with probability p, and to $Q3$ with probability $1-p$. The system is shown in Figure 1.

As the routers transmit network coded packets to the receiver, the receiver must decode these network coded packets in order to retrieve the data packets. $Q2$ and $Q3$ have exponential service rates μ_2 and μ_3 respectively. The receiver can decode the data packets when it receives as many network coded packets as at least equal to the number of data packets involved in the received linear combinations. Once a network coded packet is transmitted to one of the routers, the source drops the data packet that is located at the head of the queue. Then, when the source queue becomes empty and all network coded packets that have been generated so far have been received by the receiver, this condition is satisfied. We assume that all linear combinations that have been generated are independent. Probability of receiving identical network coded packets is neglected. This probability can be made arbitrarily small by making the field size over which network coding is performed sufficiently large. In this work, the delay which is defined as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet is analyzed.

3 Preliminaries

At this section, we will list the specifications and necessary tools that we will use to analyze the system that is the shown in Figure 1. The source and the routers form a Jackson network. From the traffic equations of the Jackson network, it follows from [7] that arrival rates to the routers are $\lambda_2 = p\lambda$ and $\lambda_3 = (1-p)\lambda$. We denote $\rho_1 = \lambda/\mu_1$, $\rho_2 = p\lambda/\mu_2$ and $\rho_3 = (1-p)\lambda/\mu_3$. Throughout the paper, we assume $\rho_1 < 1$, $\rho_2 < 1$ and $\rho_3 < 1$ for stability.

Lemma 1. *It follows directly from [7] that equilibrium distribution of the system is defined as*

$$\pi(n_1, n_2, n_3) = \prod_{i=1}^{3}(1-\rho_i)\rho_i^{n_i}$$

where n_1, n_2 and n_3 are the number of packets located at $Q1$, $Q2$ and $Q3$ respectively.

Based on the proposed performance parameter, we need to define the probability of a departure from $Q1$ when $n_1 = 1$ so that $Q1$ becomes empty after this departure and then the receiver can decode the packets that it has received from the source in $Q1$'s last busy period. In order to do so, we need to define *Palm probabilities*. *Palm probability* is used on defining a specific transition by characterizing the past and future of a Continuous Time Markov Chain (CTMC) at such a transition. The issue deals with how to evaluate any probability for a CTMC conditioned that a specific transition occurs. Since the occurrence of any specific transition at any time has probability 0, conventional conditional probabilities cannot be used. Instead these conditional probabilities must be formulated as *Palm probabilities*. Then, the *Palm probability* of a stationary CTMC conditioned that a specific transition occurs at any time is the ratio of the expected number of that specific transition at which that specific transition occurs in a fixed time interval divided by the expected number of all possible transitions in the interval.

Theorem 1. *Palm probability $P_H(C)$* of event C given that H occurs for an $M/M/1$ queue follows directly from [8] as:

$$P_H(C) = \frac{\sum_{(n,n')\in C} \pi(n)q(n,n')}{\sum_{(n,n')\in H} \pi(n)q(n,n')}, \quad C \subseteq H$$

where n is the current state, n' is the next state, $\pi(n)$ is the equilibrium distribution and $q(n, n')$ is the transition rate.

4 Analysis

The system can be defined as a three dimensional Markov chain with state space $S = (n_1, n_2, n_3)$ where each non-negative value corresponds to number of customers in $Q1$, $Q2$ and $Q3$ respectively. $Q1$, $Q2$ and $Q3$ become busy and idle sequentially as shown in Figure 2. Transitions between these states are the crucial moments as specified earlier. At time A_1, busy period of $Q1$ started. At B_1, last packet is served from $Q1$ and it becomes empty again. Hence, $B_1 - A_1$ is a busy period duration for $Q1$. At B_1, all coded packets that have been served in the busy period $[A_1, B_1]$ are routed to $Q2$ and $Q3$. Then, $Q2$ becomes empty for the first time after $Q1$ finishes its busy period at B_2' and $Q3$ becomes empty at B_1''. Hence, we define two time parameters defined as $T_2 = B_2' - B_1$ and $T_3 = B_1'' - B_1$. When $Q1$ finishes its busy period, $Q2$ becomes empty after T_2 and $Q3$ becomes empty after T_3. To conclude, the receiver can decode all packets served in $[A_1, B_1]$ in $T_{dec} = B_1 - A_1 + \max\{T_2, T_3\}$. Once the distribution of the number of packets at $Q2$ and $Q3$ is known at the end of the busy period of $Q1$, we can find the maximum of T_2 and T_3.

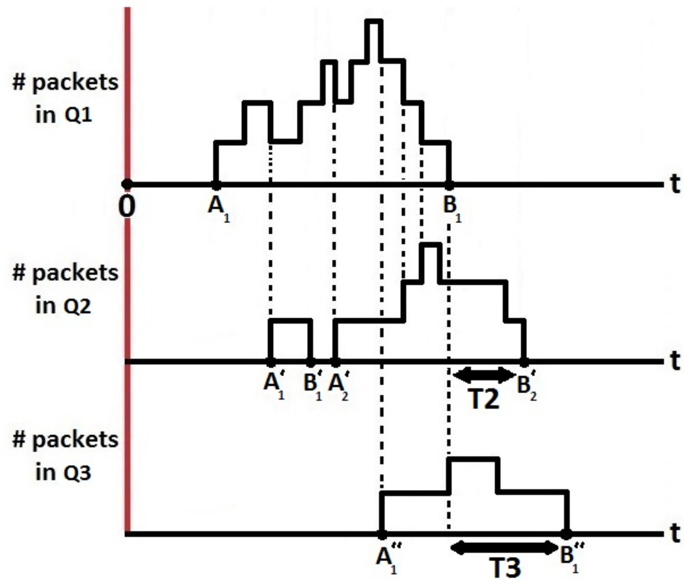

Figure 2: Timeline for the system.

Lemma 2. *The distribution of the number of packets at Q2 and Q3 at the end of the busy period of Q1 is equal to*

$$P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) = \begin{cases} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) & \text{if } n_2 \geq 1, n_3 \geq 1 \\ Rp\rho_2^{n_2-1} & \text{if } n_2 \geq 1, n_3 = 0 \\ R(1-p)\rho_3^{n_3-1} & \text{if } n_2 = 0, n_3 \geq 1 \end{cases} \tag{1}$$

where $R = (1-\rho_2)(1-\rho_3)$.

Proof. We will use Theorem 1 here. We define the event C as one packet from the source is transmitted to one of the two routers when $n_1 = 1$ and the event H as all possible transitions from $Q1$ when $n_1 = 1$.
For $n_2 \geq 1, n_3 \geq 1$,

$$\begin{aligned} P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) &= \frac{\pi(1, n_2 - 1, n_3)p\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} + \frac{\pi(1, n_2, n_3 - 1)(1-p)\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} \\ &= \frac{\rho_1\rho_2^{n_2-1}\rho_3^{n_3}p}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} + \frac{\rho_1\rho_2^{n_2}\rho_3^{n_3-1}(1-p)}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} \\ &= \frac{\rho_2^{n_2-1}\rho_3^{n_3}p}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}} + \frac{\rho_2^{n_2}\rho_3^{n_3-1}(1-p)}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}} \\ &= R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right). \end{aligned}$$

For $n_2 \geq 1, n_3 = 0$,

$$\begin{aligned} P(N_2 = n_2, N_3 = 0 | N_1 \to 0) &= \frac{\pi(1, n_2 - 1, 0)p\mu_1}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\pi(1,j,k)\mu_1} \\ &= \frac{p\rho_1\rho_2^{n_2-1}}{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\rho_1\rho_2^j\rho_3^k} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p\rho_2^{n_2-1}}{\frac{1}{1-\rho_2}\frac{1}{1-\rho_3}} \\
&= Rp\rho_2^{n_2-1}.
\end{aligned}
$$

And similarly, for $n_2 = 0, n_3 \geq 1$,

$$
P(N_2 = 0, N_3 = n_3 | N_1 \to 0) = R(1-p)\rho_3^{n_3-1}.
$$

□

Theorem 2. *We know that for each queue every single service time is exponentially distributed and service times between n occurrences are Erlang-distributed with the number of packets n, rate parameter λ_{Er} with mean $\mu_{Er} = n/\lambda_{Er}$ and cdf*

$$
F(t) = 1 - \sum_{j=0}^{n-1} \frac{(\lambda_{Er} t)^j}{j!} e^{-\lambda_{Er} t}, \quad t \geq 0.
$$

We have $T_2 \sim Er\{n_2\}$ and $T_3 \sim Er\{n_3\}$ where n_2 and n_3 are number of customers at any moment in $Q2$ and $Q3$ respectively. We need to determine the expected value of $T_{max} = \max\{T_2, T_3\}$.

Lemma 3. *We have*

$$
E[T_{max}] = \frac{p\left(\mu_2 + (1-p)\lambda\right)}{\mu_2\left(\mu_2 - p\lambda\right)} + \frac{(1-p)\left(\mu_3 + p\lambda\right)}{\mu_3\left(\mu_3 - (1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.
$$

Proof.

$$
\begin{aligned}
E[T_{max}] &= E\left[E[T_{max}|N_2, N_3]\right] \\
&= E\left[E[\max\{T_2, T_3\}|N_2, N_3\right] \\
&= \sum_{n_2=0}^{\infty}\sum_{n_3=0}^{\infty} E[\max\{T_2, T_3\}|N_2 = n_2, N_3 = n_3] P(N_2 = n_2, N_3 = n_3 | N_1 \to 0) \\
&= \underbrace{\sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1}\int_0^{\infty} P(T_2 > m)dm}_{S_1} + \underbrace{\sum_{n_3=1}^{\infty}(1-p)R\rho_3^{n_3-1}\int_0^{\infty} P(T_3 > m)dm}_{S_2} \\
&+ \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right)\int_0^{\infty} P(T_2 > m)dm}_{S_3} \\
&+ \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right)\int_0^{\infty} P(T_3 > m)dm}_{S_4} \\
&- \underbrace{\sum_{n_2=1}^{\infty}\sum_{n_3=1}^{\infty} R\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right)\int_0^{\infty} P(T_2 > m)P(T_3 > m)dm}_{S_5}
\end{aligned}
\tag{2}
$$

We split (2) into 5 pieces and compute these terms separately. We start with computing S_1 as follows

$$\begin{aligned}
S_1 &= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \int_0^{\infty} P(T_2 > m)dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \int_0^{\infty} \left[\sum_{i=0}^{n_2-1} \frac{1}{i_1!} e^{-\mu_2 m}(\mu_2 m)^i\right] dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \sum_{i=0}^{n_2-1} \frac{1}{i_1!} \left[\int_0^{\infty} e^{-\mu_2 m}(\mu_2 m)^i\right] dm \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \sum_{i=0}^{n_2-1} \frac{1}{\mu_2} \\
&= \sum_{n_2=1}^{\infty} pR\rho_2^{n_2-1} \frac{n_2}{\mu_2} \\
&= \frac{p(1-\rho_3)}{\mu_2(1-\rho_2)}.
\end{aligned}$$

Similarly,

$$S_2 = \frac{(1-p)(1-\rho_2)}{\mu_3(1-\rho_3)}.$$

Due to space constraints, computation steps of S_3 and S_4 are skipped and the results of the two parameters are given as

$$S_3 = \frac{(1-p)\rho_2 + p\rho_3}{\mu_2(1-\rho_2)},$$

and

$$S_4 = \frac{(1-p)\rho_2 + p\rho_3}{\mu_3(1-\rho_3)}.$$

Finally S_5 can be computed as follows

$$\begin{aligned}
S_5 &= \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} R \underbrace{\left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right)}_{h(n_2,n_3)} \int_0^{\infty} P(T_2 > m)P(T_3 > m)dm \\
&= R \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} h(n_2,n_3) \sum_{i=0}^{n_2-1} \sum_{j=0}^{n_3-1} \frac{1}{i!j!} \int_0^{\infty} e^{-(\mu_2+\mu_3)m}(\mu_2 m)^i (\mu_3 m)^j dm \\
&= \underbrace{\frac{R}{(\mu_2+\mu_3)}}_{K} \sum_{n_2=1}^{\infty} \sum_{n_3=1}^{\infty} h(n_2,n_3) \sum_{i=0}^{n_2-1} \sum_{j=0}^{n_3-1} \frac{(i+j)!}{i!j!} \frac{\mu_2^i \mu_3^j}{(\mu_2+\mu_3)^{i+j}} \\
&= K \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \frac{\mu_2^i \mu_3^j}{(\mu_2+\mu_3)^{i+j}} \sum_{n_2=i+1}^{\infty} \sum_{n_3=j+1}^{\infty} \left(p\rho_2^{n_2-1}\rho_3^{n_3} + (1-p)\rho_2^{n_2}\rho_3^{n_3-1}\right) \\
&= K \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \left(\frac{\mu_2}{\mu_2+\mu_3}\right)^i \left(\frac{\mu_3}{\mu_2+\mu_3}\right)^j \left(\frac{\rho_2^i \rho_3^j \left[(1-p)\rho_2 + p\rho_3\right]}{(1-\rho_2)(1-\rho_3)}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{(i+j)!}{i!j!} \left(\frac{p\lambda}{\mu_2+\mu_3}\right)^i \left(\frac{(1-p)\lambda}{\mu_2+\mu_3}\right)^j \\
&= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \sum_{i=0}^{\infty} \frac{1}{i!} \left(\frac{p\lambda}{\mu_2+\mu_3}\right)^i \left(1-\frac{(1-p)\lambda}{\mu_2+\mu_3}\right)^{-1-i} i! \\
&= \frac{(1-p)\rho_2 + p\rho_3}{\mu_2+\mu_3} \left(\frac{\mu_2+\mu_3}{\mu_2+\mu_3-\lambda}\right) \\
&= \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.
\end{aligned}
$$

Then we use (2) to compute $E[T_{max}]$ as follows

$$
\begin{aligned}
E[T_{max}] &= S_1 + S_2 + S_3 + S_4 - S_5 \\
&= \frac{p\left(\mu_2 + (1-p)\lambda\right)}{\mu_2\left(\mu_2 - p\lambda\right)} + \frac{(1-p)\left(\mu_3 + p\lambda\right)}{\mu_3\left(\mu_3 - (1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.
\end{aligned}
$$

□

Now we are ready to state the main result.

Theorem 3. *The total expected delay for the time that is needed to decode all the packets served in a single busy period of $Q1$ is equal to*

$$
E[T_{dec}] = \frac{1}{\mu_1-\lambda} + \frac{p\left(\mu_2 + (1-p)\lambda\right)}{\mu_2\left(\mu_2 - p\lambda\right)} + \frac{(1-p)\left(\mu_3 + p\lambda\right)}{\mu_3\left(\mu_3 - (1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.
$$

Proof. All packets served in a busy period of $Q1$ will certainly be decoded in $E[T_{dec}]$ which is computed as follows:

$$
\begin{aligned}
E[T_{dec}] &= E[BP_{Q1}] + E[T_{max}] \\
&= \frac{1}{\mu_1-\lambda} + \frac{p\left(\mu_2 + (1-p)\lambda\right)}{\mu_2\left(\mu_2 - p\lambda\right)} + \frac{(1-p)\left(\mu_3 + p\lambda\right)}{\mu_3\left(\mu_3 - (1-p)\lambda\right)} - \frac{p(1-p)\lambda(\mu_2+\mu_3)}{\mu_2\mu_3(\mu_2+\mu_3-\lambda)}.
\end{aligned}
$$

□

For $\lambda = 1$, $\mu_1 = 4$, $\mu_2 = 2$, $\mu_3 = 0.5$, expected delay vs. probabilistic routing parameter p graph is shown in Figure 3. Delay can be computed only for the case when all queues are stable and it is infinity otherwise. For this specific example, $Q3$ has a lower service rate. $Q3$ receives more packets as p decreases. When $(1-p)\lambda > \mu_3$, $Q3$ is not stable anymore and the queue is exploded. This means that receiver will not be able to decode data packets. As p increases, $Q2$ starts receiving more packets and delay decreases since $Q3$ has a lower service rate compared to $Q2$.

5 Discussion & Conclusion

In this work, we have presented a network scenario containing a source transmitting network coded packets via multiple routers to a receiver. The receiver must receive enough number of packets to decode network coded packets and retrieve data packets. We define the delay as the time between arrival of a data packet at the source and decoding of all the packets served in the busy period of the source queue starting from the arrival of that data packet. We show that for the proposed network scenario, the

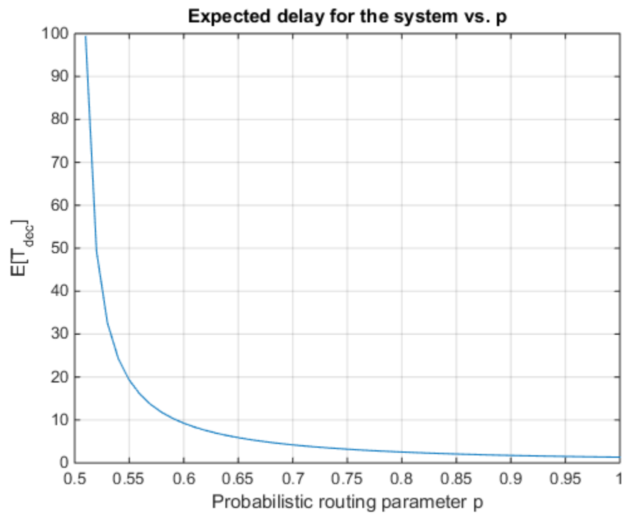

Figure 3: Expected delay vs. p for $\lambda = 1$, $\mu_1 = 4$, $\mu_2 = 2$, $\mu_3 = 0.5$.

delay can be expressed in closed-form. In practice, as the service rates of the routers change due to channel conditions, it is possible to minimize the delay by changing the probabilistic routing parameter p.

Even though sending network coded packets do not save any resources over sending data packets for unicast transmission scenarios, it is still useful to analyze the delay for the unicast system in order to prepare a baseline for future research. Various systems with different network coding techniques and comparisons between coded and non-coded systems will be analyzed in the future.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow", *IEEE Trans, Inform. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[2] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding", *Proc. 41st Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2003.

[3] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast", *IEEE Trans, on Information Theory*, vol. 52, iss. 10, pp. 4413-4430, October 2006.

[4] C. Fragouli, D. Lun, M. Medard, and P. Pakzad, "On feedback for network coding", *CISS 2007*, March 2007.

[5] J. K. Sundararajan, D. Shah, and M. Medard, "ARQ for Network Coding", *ISIT 2008*, Toronto, Canada, July 2008.

[6] J. K. Sundararajan, D. Shah, M. Medard, M. Mitzenmacher, and J. Barros, "Network coding meets TCP", *INFOCOM 2009*, April 2009.

[7] L. Lipsky, "Queueing Theory: A Linear Algebraic Approach", Second Edition, Springer, 2009.

[8] R. Serfozo, "Basics of Applied Stochastic Processes", Springer, Berlin, 2009.