Packet Loss Analysis for Media Streaming with Network Coding in Wireless Broadcast Networks

Hancheng Lu
University of Science and Technology of China
Hefei, Anhui 230027 China

Chang Wen Chen
State University of New York at Buffalo
Buffalo, NY 14260 U.S.A.

Abstract—Network coding has been comprehensively studied and has been shown to achieve significant throughput gain in the case of wireless data broadcast. However, for media streaming, throughput are inadequate to fully determine media playback quality at receivers. Packet loss behavior is also crucial in designing a high performance media streaming system. In this paper, we carry out analysis on packet loss behavior when network coding is performed at the Base Station (BS) with limited buffer. We assume that the BS is the bottleneck in a wireless broadcast network. Two types of packet loss at the BS are considered, namely packet loss at the input introduced by buffer overflow ($L_{in}$), and packet loss at the output due to channel errors and late arrivals ($L_{out}$). Our analysis is based on a bulk-service queue model denoted as M/G(a, b)/1/B. The major contribution of this research is the successful derivation of the probability distribution of queue states at an arbitrary time by applying the supplementary variable technique. Given particular channel conditions, this probability distribution enables us to obtain the relationship between two types of packet loss ($L_{in}$-$L_{out}$) and the supportable arrival rates at the BS. Then, the maximum allowable arrival rate when $L_{in}$ and $L_{out}$ are constrained can be determined. This is critical for the design of appropriate congestion control for media streaming at the BS. Extensive simulations have been performed to demonstrate that the results from our analysis are consistent with that from simulations.

Keywords—network coding; packet loss; queueing analysis

I. INTRODUCTION

A wireless broadcast network is defined as a service-oriented network and often used as an access network where a Base Station (BS) transmits packets to all receivers within its coverage area. Therefore, it is very efficient for delivering high bandwidth-consumed media streaming services to multiple users. An example of this type of network is the Radio Access Network (RAN) in a 3G network, which can be considered as a wireless broadcast network when it delivers the Multimedia Broadcast Multicast Service (MBMS) [1] to mobile subscribers.

In a wireless broadcast network, a BS is prone to be a bottleneck while performing media streaming services. These services impose stringent Quality of Service (QoS) requirements on the BS with parameters in throughput, delay and packet loss rate. Recently, network coding [2][3][4] has been proved to be an efficient way to achieve desired throughput gain by exploiting the broadcast nature of the wireless channel. With network coding, streams can be segmented into blocks (also called generations) and processed in the unit of blocks.

In addition to throughput, packet loss also has significant impact on the playback quality at receivers and it is useful for congestion control. For example, we can drop less important packets when packet loss rate is given. There are two types of packet loss at the BS performing media streaming. One is at the input ($L_{in}$) where packets arrival and the other is at the output ($L_{out}$) where packets are transmitted. $L_{in}$ is caused by buffer overflow when the BS is equipped with limited buffer. $L_{out}$ is due to channel errors and late arrivals. In order to combat channel errors, a generation is usually transmitted in a rateless fashion with erasure coding. However, for media streaming, the playout deadline associated with each packet will limit the maximal number of transmissions for the generation resulting in $L_{out}$.

Generally, queueing analysis can be involved to study packet loss. The buffer at the BS is modeled as a bulk-service queue denoted as M/G(a, b)/1/B when network coding is performed. We assume that Random Linear Network Coding (RLNC) is adopted. The size of a generation is in the range of (a, b). Given channel conditions, $L_{out}$ will determine the service rate, and $L_{in}$ will constrain the maximal arrival rate.

Most existing schemes [4][5][6] on queueing analysis with network coding have focused on average packet delay, which cannot be adopted to obtain the packet loss rate. M. Medard et al. [7] derive the probability distribution of queue states at the time when a generation departs from the queue. However, this probability distribution at departure time cannot be applied to obtain the packet loss rate accurately with given arrival rate and service rate. Our simulations results in Section IV also verified such inaccuracy. To obtain an accurate packet loss behavior, we derive in this paper the probability distribution of queue states at an arbitrary time by applying the supplementary variable technique to the analysis. This probability distribution...
enables us to obtain the relationship between packet loss ($L_{in}$, $L_{out}$) and the supportable arrival rates at the BS. Then, the maximum allowable arrival rate when $L_{in}$ and $L_{out}$ are constrained can be determined, which is essential for congestion control at the BS. That is, the BS will calculate the allowable arrival rate for each stream based on the constraint on packet loss ($L_{in}$, $L_{out}$).

To the best of our knowledge, this paper is the first one to analyze packet loss with network coding based on the probability distribution of queue states at arbitrary time. This is critical for congestion control as well as resource allocation at the BS. Since media streaming can tolerate certain packet loss, for a given total packet loss rate, the optimal values for $L_{in}$ and $L_{out}$ are found to maximize the allowable arrival rate at the BS. It has been confirmed that the results from our analysis are consistent with that from simulation.

The rest of the paper is organized as follows. In Section II, we describe the system model for analysis. In Section III, we introduce the queue model at the BS as a bulk-service queue and derive the probability distribution of the queue length at arbitrary time. Then we can obtain the relationship between packet loss ($L_{in}$, $L_{out}$) and the supportable arrival rates. In Section IV, performance evaluation is presented. We conclude the paper in Section V with a summary.

II. SYSTEM MODEL

The media streaming system is described in Figure 1. The wireless broadcast network consists of one BS and $M$ receivers ($r_i$, $i = 1, \ldots, M$). In the wired network, the connection between the stream server and the BS is assumed to be reliable. A pre-encoded media stream is transmitted from the stream server to the BS continuously. Then, the BS broadcasts packets/generations from the stream to receivers. The BS by setting the allowable arrival rate. We assume that there exist some feedback mechanisms to ensure that the stream server can adjust its transmission rate simultaneously.

A. Packet Loss

As the wired connection is assumed to be reliable, two types of packet loss are considered in this research, namely packet loss at the input of the BS ($L_{in}$) caused by buffer overflow and packet loss at the output of the BS ($L_{out}$) due to channel errors and late arrivals. For simplicity, we also use $L_{in}$ and $L_{out}$ to denote the packet loss rates at the input and the output respectively. If heterogeneous receivers are considered, $L_{out}$ is defined as

$$L_{out} = \max_{i=1,M} L_{out}$$

where $L_{out}$ is the packet loss rate at the output from the point of view of $r_i$ in the steady state. Equation (1) means that we attempt to guarantee playback quality at the receiver with the worst channel condition.

When network coding is performed, each generation is associated with a playout deadline. The whole generation must be removed from the buffer after its deadline. Let $k_i$ denote the size of the $i$th generation and $L_{out}^{ij}$ denote the loss probability for the $j$th generation at $r_i$. At time $t_n$, we assume that $n$ generations have been transmitted by the BS. $L_{out}(t_n)$ and $L_{out}$ can be obtained as

$$L_{out}(t_n) = \frac{\sum_{j=1}^{k_i} L_{out}^{ij}}{k_i}$$

Similarly, we define $L_{in}(t)$ as

$$L_{in}(t) = \lim_{t_\rightarrow t_0} L_{in}(t)$$

where $L_{in}(t)$ denote the packet loss rate at the input of the BS at time $t$. We use the path loss rate to characterize total packet loss that can be tolerated by media streaming, which is defined as

$$L_{path} = \lim_{t_\rightarrow t_0} \left(1 - \frac{Y(t)}{Y(t_0)}\right), \ k = \arg\max_{i=1,M} L_{out}$$

where $Y(t)$ denotes the total number of packets transmitted by the stream server until time $t$ and $Y_k(t)$ denotes the number of packets received by $r_k$ until time $t$. $r_k$ is the receiver with the worst channel condition.

B. Wireless Channel Model

Time is slotted and the BS is assumed to transmit one packet in a slot. The wireless connection between the BS and the receiver $r_i$ is modeled as an erasure channel with the erasure probability $p_i$. For each receiver, perfect feedback is used to signal the event that a generation is completely received, which is also used to estimate the Channel Side Information by the BS.

In existing schemes [4][5][6][7], a reliable transmission mode with RLNC is assumed for file transfer, which means that the BS will continue to transmit random linear combinations from a generation until all receivers have received enough independent combinations to recover the original packets. However, for media streaming, each generation is associated with a playout deadline and a generation is useless after its deadline. Therefore, the maximal number of transmissions for a generation must be limited by its deadline. Let $x_i$ denote the number of transmissions for the $i$th generation. For media streaming, the $i$th generation should be transmitted within $d_i$ slots. Otherwise, it should be dropped. Therefore, we have

$$x_i = 0 \text{ if } d_i < k_i$$
We define an original packet when these packets are broadcasted in a generation after applying the supplementary variable technique. Then, given the queue at the BS as a bulk-service queue and derive the service time by \( \lambda \) packets/slot. We consider a scenario where the size of Galois Field for RLNC is large enough. Therefore, receiver 1 is able to decode the \( j \)th generation after \( k \) linear combinations has been received.

### C. Service Time Distribution

According to equation (2) and (7), if there is a constraint on \( L_{\text{out}} \) (i.e. \( L_{\text{out}} \leq L_{\text{out}}^{\text{max}} \)), \( x_j \) will be determined by the erasure probability, generation size, and deadline. Before transmitting a generation, the BS will set \( x_j \) appropriately so that \( L_{\text{out}}^{\text{max}} (t_j) \) approaches \( L_{\text{out}}^{\text{max}} \) as soon as possible with \( k = arg \max p_i \).

We consider a scenario where \( k_j = n \) and \( 1 - \sum_{k=1}^{d} \left( \frac{d}{k} \right) p_i^{d-k} (1 - p_i)^k \leq L_{\text{out}}^{\text{max}} \). Then, we can obtain the lower bound of the number of time slots for each generation.

\[
x_{\text{low}} = \min_{n=1, \ldots, M} \left[ x : \sum_{k=1}^{d} \left( \frac{d}{k} \right) p_i^{d-k} (1 - p_i)^k \geq L_{\text{out}}^{\text{max}} \right] \tag{8}
\]

Let \( X_{\text{in}} \) denote the number of time slots for \( r_j \) to recover \( n \) original packets when these packets are broadcasted in a generation with RLNC. We define

\[
X_n = \max_{i=1,2,\ldots, M} X_{\text{in}}
\]

In order to maximize the arrival rate, \( X_n \) should satisfy

\[
X_n \leq X_{\text{low}}
\]  \( \tag{9} \)

In the queue model, \( X_n \) denotes the service time (in slots) for a generation of \( n \) packets when the lower bound \( X_{\text{low}} \) is applied. From the definition of \( X_n \), we can see that the BS will remove \( n \) packets from its queue in two cases, i.e. all receivers can recover \( n \) original packets and the number of time slots for transmitting the generation reaches \( X_{\text{low}} \). We assume that a receiver can send instant feedback to the BS as soon as it decodes the generation. We define

\[
H_{\text{in}}(k) = \text{Pr}(X_n \leq k)
\]

\[
then the probability mass function of \( X_n \) \( (h_n(k)) \) can be obtained as

\[
h_n(k) = \text{Pr}(X_n \leq k) - \text{Pr}(X_n \leq k - 1)
\]

In this section, we analyze packet loss behavior at the BS when network coding is performed. First, we model the queue at the BS as a bulk-service queue and derive the probability distribution of queue states at arbitrary time by applying the supplementary variable technique. Then, given \( L_{\text{in}} \) and \( L_{\text{out}} \), the maximal arrival rate can be determined, which is critical for congestion control at the BS. Finally, as the path loss rate characterizes total packet loss that can be tolerated by media streaming, we find the optimal values for \( L_{\text{in}} \) and \( L_{\text{out}} \) to maximize the arrival rate when the path loss rate is constrained.

### A. Queue Model

Assume that packet arrival follows a Poisson process with rate \( \lambda \) (packets/slot) and RLNC is applied at the BS. The queue at the BS can be modeled as a single-server queue with a general bulk-service rule, which is denoted as the \( M/G(a,b)/1/B \) queue.

The total buffer \( B \) at the BS consists of the waiting space (queue) \( (B_q) \) and the service space \( (B_s) \). Therefore, we have \( B = B_q + B_s \). Packets enter the waiting space (queue) before they are served in the service space. After a departure, we assume there are \( n \) packets in the queue. The server performs service with network coding according to the following rules:

1. If \( n < b \), the server will wait until there are \( a \) packets. Then all packets enter service and form a generation.
2. If \( a \leq n \leq b \), the server will serve \( n \) packets in a generation.
3. If \( n > b \), the server will only serve \( b \) packets in a generation.

The service time depends on the generation size. For a generation of \( n \) packets, we denote the probability density mass of the service time by \( h_n(k) \) and denote the mean of the service time by \( h_n \). In the wireless broadcast network, \( h_n \) can be obtained based on equation (10) if the number of transmissions for the generation is given.

We analyze the queue in the steady state. Packets in the queue can be divided into two parts, i.e. \( N_q \) packets waiting in the queue and \( N_q^{-} \) packets in service. Let \( N_q^{-} \) denote the number of packets in the queue at the time when a generation departs. We define

\[
p_n^- = \text{Pr}(N_q^- = n) \quad 0 \leq n \leq B_q
\]

Thus, the probability distribution of queue states at departure time can be expressed as

\[
\bar{p}^- = (p_0^- , \ldots , p_{B_q^-})
\]

Let \( U \) denote the remaining service time of a generation in service. We define

\[
p_{n,0} = \text{Pr}(U = 0) \leq n \leq a - 1
\]

\[
p_{n,1} = \text{Pr}(U = u) \quad 0 \leq u \leq \xi_{n+1}
\]

Then, the probability of queue states at arbitrary time can be expressed as

\[
F^-_{arb} = (p_{0,0}, \ldots , p_{a-1,0}, p_{a,1}, \ldots , p_{B_q,1})
\]

Moreover, with the normalization condition, we have

\[
\sum_{n=0}^{a-1} p_{n,0} + \sum_{n=0}^{B_q} p_{n,1} = 1
\]

We can see that \( p_n^- \) and \( p_{n,0} \) should satisfy

\[
p_n^- = \frac{\sum_{n=0}^{a-1} p_{n,0}}{\sum_{n=0}^{a-1} P(0(n))} \quad 0 \leq n \leq B_q
\]  \( \tag{12} \)

### B. Probability Distribution of Queue Length

As the arrival interval follows Poisson distribution, \( \bar{p}^- \) can be obtained by applying the Markov technique based on the work carried out by Gold and Tran-Gia [8], which satisfy the following linear equations.
It can be expected that the maximum arrival rate of the queue ($\lambda_{\text{max}}$) is limited by given $L_{\text{in}}$ and $L_{\text{out}}$. $L_{\text{out}}$ will have significant impact on the service time distribution. Furthermore, to implement congestion control when network coding is performed, the generation size should also be considered. For media streaming, the BS chooses the generation size with three strategies:

**S0:** Transmit original packets without network coding. In this strategy, we have $a = b = 1$.

**S1:** Stop transmitting until there are $b$ packets in the queue. In this strategy, we have $a = b$.

**S2:** Transmit when the queue is not empty. In this strategy, we have $a = 1$.

Usually, $b$ can be determined by the format of media. For example, it can be set to the number of packets in a frame, in several frames, or even within a Group of Picture (GoP).

### D. Optimal Packet Loss Rate Allocation

The path loss rate $L_{\text{path}}$ characterizes packet loss that can be tolerated by media streaming and reflects the playback quality at the receivers. Therefore, QoS requirements from receivers will induce a constraint on $L_{\text{path}}$, i.e. $L_{\text{path}} \leq L_{\text{max}}$. In this case, $L_{\text{in}}$ and $L_{\text{out}}$ are constrained by equation (6), and there exists an optimal packet loss rate allocation for $L_{\text{in}}$ and $L_{\text{out}}$ to achieve the maximal arrival rate.

In this research, we consider a scenario where $k_i = n$ and $1 - \sum_{i=1}^{d_i} l_i f_i \leq e_{\text{path}} (i = 1, ..., M, j = 1, 2, ...)$.

This means that packet loss due to late arrivals can be eliminated. For each generation $j$, the optimal packet loss allocation can be achieved by performing the following two steps:

**Step 1** Calculate $X_{\text{low}}$ with $e_{\text{max}}$ (i.e. $L_{\text{out}} = e_{\text{path}}$) based on equation (8). Let $X$ denote the set of allowable numbers of time slots for transmitting a generation, we have $X = \{x_i | x_i = X_{\text{low}} - 1, d_i\}$

**Step 2** For each $x_i \in X$, calculate $L_{\text{out}}$. The service time distribution can be obtained based on equation (10) when $X_{\text{low}}$ is set to $x_i$. Then we can calculate the maximally allowable $L_{\text{in}}$ according to equation (6). In this case, the maximal arrival rate can be obtained based on queueing analysis in Section III.B, which is denoted as $\lambda_{\text{max}}(L_{\text{max}}, x)$. Therefore, the maximal $\lambda_{\text{max}}$ can be obtained as 

$$
\lambda_{\text{max}} = \max_{x \in X} \lambda_{\text{max}}(e_{\text{path}}, x)
$$

The optimal packet loss allocation can be expressed as

$$
S^* = \arg \max_{x \in X} \lambda_{\text{max}}(e_{\text{path}}, x)
$$

$$
e_{\text{out}}^{\text{opt}} = \max_{i=1}^{N} \left(1 - \sum_{u=1}^{X_{\text{opt}}} p_i \right)^{-1}\left(1 - p_j\right)^{u} \text{ and } L_{\text{in}}^{\text{opt}} = \frac{e_{\text{path}} - e_{\text{opt}}}{1 - e_{\text{out}}^{\text{opt}}}
$$

### IV. PERFORMANCE EVALUATION

In this section, we present the simulations carried out to evaluate our analysis. The results show that the relationship between packet loss rates and allowable arrival rates derived from our analysis are consistent with that obtained from
simulations. Therefore, our analysis can be applied to design congestion control and resource allocation at the BS.

We write an event-driven queue simulator in C language to produce simulation results and use Scilab [10] to produce numerical results based on our analysis.

Our analysis assumes that the queue is in the steady state. Surprisingly, simulation results demonstrate that the results from our analysis are still valid when the simulation time is relatively short (e.g. about $10^5$ slots) and/or the traffic intensity is greater than 1.

A. Simulation Setup

The scenario for the simulation is described in Figure 1. If not separately mentioned, the parameters are set as follows: $M = 15$, $p_i = 0.2$ ($i = 1, ..., M$) and $B = 50$ packets. For each simulation, if $L_{\text{out}}^{\text{max}}$ is set, we shall assume that $1 - \sum_{k=0}^{\infty} (1 - p_j)^k < L_{\text{out}}^{\text{max}}$ ($i = 1, ..., M$, $j = 1,2,...$), which eliminates packet loss due to late arrivals.

Packets generated by the stream server arrive at the BS follow Poisson process with rate $\lambda$ (packets/slot). For each simulation, about $10^8$ packets are generated so that the queue can reach the steady state. The BS determines (a, b) based on the selected coding strategy. When $S_0$ is applied, a and b are set to 1. When $S_1$ is applied, a and b are set to 8. When $S_2$ is applied, a is set to 1 and b is set to 8.

After the simulation, we estimate the number of packets received by the BS and the number of packets received by each receiver. Then we calculate $L_{\text{in}}$, $L_{\text{out}}$ and $L_{\text{path}}$ according to Section 2.1.

B. Results

1) Accuracy of Queueing Analysis: During the simulation, $L_{\text{out}}^{\text{max}}$ is set to 0.001. From equation (8), we can see that $X_{\text{low}}^n = 5$ when strategy $S_0$ is applied and $X_{\text{low}}^n = 17$ when strategy $S_1$ is applied. However, when strategy $S_2$ is applied, $X_{\text{low}}^n$ depends on the generation size $n$ where $n$ is varied from 1 to 8. As the BS is the bottleneck of the streaming system, it is reasonable to assume that $n = 8$ in most cases. We set $X_{\text{low}}^n$ to 8 for strategy $S_2$. Therefore, the service time distribution is determined according to equation (10). In this case, as the service time is minimized based on $L_{\text{out}}^{\text{max}}$, the maximally allowable arrival rate is referred to as the maximal arrival rate when $L_{\text{in}}$ is given.

Given the arrival rates $\lambda$, Table 1–3 compare $L_{\text{in}}$ derived from our analysis with that from the simulation by applying $S_0$, $S_1$, $S_2$ respectively, where $L_{\text{in}}^{\text{dpt}}$, $L_{\text{in}}^{\text{arb}}$ and $L_{\text{in}}^{\text{sim}}$ denote the packet loss rates at the input based on $P_{\text{dpt}}$, $P_{\text{arb}}$, and the simulation respectively.

From Table 1–3, we can see that $L_{\text{in}}^{\text{arb}}$ is consistent with $L_{\text{in}}^{\text{sim}}$, while $L_{\text{in}}^{\text{dpt}}$ deviates noticeably from $L_{\text{in}}^{\text{sim}}$ under all three coding strategies. The difference between $L_{\text{in}}^{\text{dpt}}$ and $L_{\text{in}}^{\text{sim}}$ becomes larger when RLNC is performed. This means that the probability distribution of queue states at departure time definitely cannot be used to derive packet loss rate. As expected, $L_{\text{in}}^{\text{dpt}}$, $L_{\text{in}}^{\text{arb}}$ and $L_{\text{in}}^{\text{sim}}$ increase as $\lambda$ increases.

Interestingly, results from Table 1-3 also show that $L_{\text{in}}^{\text{arb}}$ can be used to approximate the practical packet loss rate well even when the traffic intensity is greater than 1 (i.e. $\lambda > 0.39$ in Table 1 and $\lambda > 0.61$ in Table 2).

Table 1. Packet Loss Rate vs. Arrival Rate ($S_0, a=b=1$)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.36</th>
<th>0.38</th>
<th>0.40</th>
<th>0.42</th>
<th>0.44</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{\text{in}}^{\text{dpt}}$</td>
<td>0.000310</td>
<td>0.006977</td>
<td>0.050448</td>
<td>0.125398</td>
<td>0.198014</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{arb}}$</td>
<td>0.000153</td>
<td>0.003683</td>
<td>0.027738</td>
<td>0.070157</td>
<td>0.112214</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{sim}}$</td>
<td>0.000028</td>
<td>0.002614</td>
<td>0.026576</td>
<td>0.069880</td>
<td>0.112194</td>
</tr>
</tbody>
</table>

Table 2. Packet Loss Rate vs. Arrival Rate ($S_1, a=b=8$)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.60</th>
<th>0.62</th>
<th>0.64</th>
<th>0.66</th>
<th>0.68</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{\text{in}}^{\text{dpt}}$</td>
<td>0.045297</td>
<td>0.120713</td>
<td>0.220175</td>
<td>0.318422</td>
<td>0.406162</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{arb}}$</td>
<td>0.010272</td>
<td>0.028909</td>
<td>0.055094</td>
<td>0.082860</td>
<td>0.109689</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{sim}}$</td>
<td>0.010538</td>
<td>0.029212</td>
<td>0.055065</td>
<td>0.082967</td>
<td>0.109278</td>
</tr>
</tbody>
</table>

Table 3. Packet Loss Rate vs. Arrival Rate ($S_2, a=1, b=8$)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.60</th>
<th>0.62</th>
<th>0.64</th>
<th>0.66</th>
<th>0.68</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{\text{in}}^{\text{dpt}}$</td>
<td>0.035406</td>
<td>0.112888</td>
<td>0.217455</td>
<td>0.317820</td>
<td>0.406054</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{arb}}$</td>
<td>0.009297</td>
<td>0.028160</td>
<td>0.054829</td>
<td>0.082799</td>
<td>0.109677</td>
</tr>
<tr>
<td>$L_{\text{in}}^{\text{sim}}$</td>
<td>0.009839</td>
<td>0.027614</td>
<td>0.054209</td>
<td>0.082623</td>
<td>0.109597</td>
</tr>
</tbody>
</table>

2) Relationship Between maximal arrival Rate and packet loss rate at input: During the simulation, $L_{\text{out}}^{\text{dpt}}$ is set to 0.001. The service time distribution can be obtained according to equation (8) and (10).

Figure 2 shows the relationship between $L_{\text{in}}$ and the maximal arrival rate ($\lambda_{\text{max}}$) when three coding strategies are under comparison. The x-axis is on a logarithm scale. We can see that the results from our analysis coincide with that from the simulation. We can draw following conclusions.

(1) Under all coding strategies, $\lambda_{\text{max}}$ increases as $L_{\text{in}}$ increases. Given $L_{\text{in}}$, the BS can support much higher $\lambda_{\text{max}}$ with network coding than that without network coding. The reason is that network coding can increase the output service rate.

(2) When $L_{\text{in}}$ is small, $S_2$ can achieve higher $\lambda_{\text{max}}$ than $S_1$. Unlike $S_1$, the BS performing $S_2$ broadcasts packets as soon as the queue is not empty. As $L_{\text{in}}$ increases, $\lambda_{\text{max}}$
increases for S1, S2 respectively. In this case, the probability that less than b packets in the queue decreases and the probability that the BS perform network coding with b packets increases. Therefore, \( \lambda_{\text{max}} \) for S2 will approximate that for S1.

In Figure 4, RLNC is performed at the BS. It shows that \( \lambda_{\text{max}} \) is maximized when \( X_{\text{low}}^1 = 12 \). The optimal packet loss allocation is that \( X_{\text{out}}^\text{opt} = 0.0726 \) and \( L_{\text{out}}^\text{opt} = 0.0296 \).

Comparing Figure 3 with Figure 4, we can see that higher \( \lambda_{\text{max}} \) is achieved with network coding. It means that there is throughput gain at output when network coding is applied.

V. CONCLUSION

In this paper, we have investigated packet loss behavior with queueing analysis when streaming with network coding. We model the queue at the BS as a bulk-service queue denoted as M/G(a,b)/1/B and obtain the probability distribution of the queue length at arbitrary time. We are the first to obtain such distribution at an arbitrary time to accurately characterize the packet loss behavior. Then we derive the relationship between packet loss rates and supportable arrival rates, which is critical for designing congestion control. An optimal packet loss rate allocation can be found that maximizes the arrival rate. We confirm our analysis with results from various simulations.

We are working on extending current scheme to a dynamic environment. In this case, adaptation needs to be involved. This scheme can also be adopted to analyze other QoS parameters for QoS-aware routing and scheduling.

REFERENCES