Broadcast Approach Meets Network Coding for Data Streaming

Alejandro Cohen *, Muriel Médard †, and Shlomo Shamai (Shitz)*
*Faculty of Electrical and Computer Engineering, Technion, Israel, {alecohen, sshlomo}@technion.ac.il
†RLE, MIT, Cambridge, MA, USA, medard@mit.edu

Abstract—For data streaming applications, existing solutions are not yet able to close the gap between high data rates and low delay. This work considers the problem of data streaming under mixed delay constraints over a single communication channel with delayed feedback. We propose a novel layered adaptive causal random linear network coding (LAC-RLNC) approach with forward error correction. LAC-RLNC is a variable-to-variable coding scheme, i.e., variable recovered information data at the receiver over variable short block length and rate is proposed. Specifically, for data streaming with base and enhancement layers of content, we characterize a high dimensional throughput-delay trade-off managed by the adaptive causal layering coding scheme. The base layer is designed to satisfy the strict delay constraints, as it contains the data needed to allow the streaming service. Then, the sender can manage the throughput-delay trade-off of the second layer by adjusting the retransmission rate a priori and posterior as the enhancement layer, that contains the remaining data to augment the streaming service’s quality, is with the relax delay constraints. We numerically show that the layered network coding approach can dramatically increase performance. We demonstrate that LAC-RLNC compared with the non-layered approach gains a factor of three in mean and maximum delay for the base layer, close to the lower bound, and factor two for the enhancement layer.

Index Terms—Broadcast approach, ultra-reliable low-latency, network coding, in-order delivery delay, throughput.

I. INTRODUCTION

One major challenge in modern communication systems is closing the gap between high data rates and low delay for data streaming applications [1], [2]. Although high data rates can be achieved by information-theoretic solutions using coding over asymptotic block regime, low delay in ultra-reliable low-latency communications (URLLC) requires short blocks. Therefore, classical information-theoretic coding schemes do not provide the desired throughput-delay trade-offs.

Different approaches have been considered to increase data rates and close this gap. One approach, known as the Broadcast Approach [3]–[5], considers the situation of variable recovered layer information data over a fixed block length, coined variable-to-fix coding [6]. In the classical broadcast channel, a single sender transmits layered encoded data to a number of receivers [7], [8]. Each receiver, obtaining a possibly different channel condition, decodes as many encoded layers as the current channel realization allow it. The broadcast approach realizes the same fundamental principles. The sender using multi-layering virtually broadcasts all the layers to a single receiver, which can decode layers according to the actual channel condition. For erasure broadcast channels, which can be particularly considered as a degraded broadcast channel in the setting considered herein, time-sharing over long transmissions may be capacity-achieving [9], [10]. However, in a setting involving data rates and reliability, time-sharing for general channels and even for erasure channels is not optimal [4], [11]. Moreover, although the broadcast approach can increase data rate and is also considered under mixed delay constraints [12]–[16], solutions using this approach, assuming long transmission blocks with a fixed size that cannot satisfy the low delays requirements for modern URLLC applications.

Another approach of rateless and streaming codes proposed to close this gap [17]–[31]. Those solutions can be considered as fix-to-variable coding schemes, as they consider the scenario of fixed recovered information data over variable block length and rate. Recently, proposed an adaptive and causal random linear coding (AC-RLNC) approach, applied to single-path [26], multi-path and multi-hop [27], and heterogeneous networks with multi-sources and destinations managed by software-defined networking (SDN) [31]. AC-RLNC aims at mitigating the throughput-delay trade-off of non-layered data by adapting the required coded retransmissions using a sliding window with a priori and a posteriori forward error correction (FEC) mechanisms. This adaptive and causal adaptation is made by the sender, who tracks the actual channel condition as represented by the delayed feedback acknowledgments. Although AC-RLNC can reduce throughput-delay trade-off, due to the unknown channel realizations at the sender during the delayed feedback, it cannot close this gap.

In this work, we propose a novel layered adaptive causal network coding scheme for data streaming communications coined LAC-RLNC. The proposed approach combines adaptive causal network coding with the broadcast approach to obtain a new variable-to-variable coding scheme. That is, in the proposed coding scheme, we examine a new solution of variable recovered information data over variable short block length and rate under mixed delay constraints. Layering data with different priorities using an adaptive causal coded approach opens a new research area to explore that offers a new regime of a high dimensional trade-off of layers and significant gains in delay and throughput guarantees. Specifically, we propose a layering coding scheme for base and enhancement data streaming. The base layer (Layer 1) contains the amount of data needed to allow the streaming service at the receiver in real-time, and the enhancement layer (Layer 2) contains the remaining data that can augment the streaming service’s quality. In the adaptive causal solution, the high dimensional trade-off is first managed by adjusting the layering code using the tracking of the current channel condition. The base layer is designed to satisfy the strict delay constraints as it contains the data needed to allow the streaming service. Then, the sender can manage the throughput-delay trade-off of the second layer by adjusting the retransmission rate a priori and posterior as the enhancement layer is with relaxed delay constraints. The proposed solution enables layered transmissions with zero
Error probability under delay constraints [32]–[34]. We contrast the performance of the proposed approach with that of the non-layered AC-RLNC [26]. We show that the proposed LAC-RLNC achieves significant gain in delay terms as obtaining similar data rate, the mean and max order delays are reduced approximately by a factor of 2, for layer 1 and 2, respectively. Moreover, in the simulated scenarios, we show that the in-order delays of the first layer almost achieve the optimal lower-bound.

The structure of this work is as follows. In Section II, we formally describe the system model and the metrics in use. In Section III, we provide a background on adaptive causal network coding and on broadcast approach. In Section IV, we present the layered adaptive causal network coding algorithm. In Section V, we evaluate the performance of the proposed solution. Finally, we conclude the paper in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a point-to-point real-time slotted communication setting with a single forward link and feedback. Fig. 1 shows the system model. The forward link can be considered as a binary erasure channel (BEC) varying over time with average erasure probability \( \epsilon_{\text{mean}} \). The sender is interested in transmitting frames of raw data that can be divided into base layer (Layer 1) and enhancement layer (Layer 2) at each time slot \( t \), the sender transmits over the forward link a coded packet \( c_t \) that belongs to the \( f \)-th frame. The size of the frames is of \( n \) coded packets, \( k_1 \) packets for Layer 1, and the remaining \( k_2 \leq n - k_1 \) packets for Layer 2. According to the erasure realizations, the receiver sends at each time slot either an acknowledgment (ACK) or a negative-acknowledgment (NACK) message to the sender, using the feedback link. For simplicity, feedback messages are assumed to be reliable, i.e., without errors. The delay between the transmission of a coded packet and the reception of the corresponding feedback is called round trip time (RTT). Hence, for each transmitted coded packet \( c_t \), the sender receives feedback after \( t + \text{RTT} \) time slots. Let \( u_t \) denote the binary feedback received at time \( t \), where

\[
u_t = \begin{cases} 
1 & \text{received ACK for } c_t, \\
0 & \text{received NACK at time } c_t, 
\end{cases}
\]

\( t^- \triangleq t - \text{RTT} \).

Figure 1: System model and encoding process of LAC-RLNC.

Table I: LAC-RLNC algorithm: symbol definitions.

| Parameter     | Definition                                                                 |
|---------------|-----------------------------------------------------------------------------|
| \( f \)       | frame number                                                                |
| \( k_1, i \in \{1, 2\} \) | number of raw information packets in \( i \)-th layer                        |
| \( n \)       | number of layered coded packets in frame                                     |
| \( l^1, i \in \{1, 2\} \) | raw information packets of \( i \)-th layer                                   |
| \( p^j \)     | \( n \) layered linear coded information packets                             |
| \( \eta \)    | normalized throughput                                                       |
| \( D^\text{mean} \) | mean in order delivery delay of \( i \)-th layer                            |
| \( D^\text{max} \) | maximum in order delivery delay of \( i \)-th layer                          |
| \( \bar{t} \) | time slot index                                                             |
| \( c_t \)     | coded packet to transmit at time slot \( t \)                               |
| \( \mu_i \in \mathbb{F}_2 \) | random coefficients in field \( \mathbb{F}_2 \)                             |
| \( \epsilon_{\text{mean}}, \epsilon_{\text{max}} \) | mean and maximum erasure probability                                       |
| \( \frac{\text{mD}}{\text{d}} \) | number of DoFs needed to decode \( c_i \)                                    |
| \( d \)       | rate of DoF (\( \frac{md_t}{ad_t} \))                                      |
| \( r_{\text{e}} \) | estimated rate of the channel                                               |
| \( \theta \)  | throughput-delay trade-off parameter                                        |
| \( \Delta_i \geq 0 \) | retransmission criterion                                                    |
| \( \text{RTT} \) | round-trip time                                                             |
| \( w_{\text{min}} \) | index of the first information packet in \( c_t \)                          |
| \( \text{mD} \) | number of FEC to add per frame                                              |
| \( \pi \)     | \( \pi \) is the maximum number of overlap information packets             |
| \( w \in \{1, \ldots, \pi\} \) | effective window size                                                       |

Our main performance metrics are defined as follows:

1. **Normalized throughput**, \( \eta \). This is defined as the total amount of information data delivered to the receiver in units of bits per second, divided by the total amount of bits transmitted by the sender in this period.

2. **In-order delivery delay of layers**, \( D_i, i \in \{1, 2\} \). This is the difference between the time slot in which an information packet in the \( i \)-th layer is first transmitted and the time slot in which the layer is decoded in order by the receiver.

Our goal is to maximize the throughput, \( \eta \), while minimizing the in-order delivery delay of the layers, \( D^1 \) and \( D^2 \).

III. BACKGROUND

In this section, we review relevant background in single path adaptive and causal network coding [26] and in broadcast approach [5].

A. Adaptive and Causal Network Coding

In AC-RLNC, the sender decides at each time step whether to transmit a new coded linear combination or to repeat the last sent combination according to the feedback information. Here, “same” and “new” refer to the raw information packets of information contained in the linear combination. Sending the same linear combination thus means that the raw information packets are the same but with different random coefficients. Let \( \mu_i \) and \( p_i \) denote the random coefficients drawn from a sufficiently large field and the raw information packets, respectively. Using a sliding window mechanism the coded linear combination transmitted, called a degree of freedom (DoF), given by

\[
c_t = \sum_{i=w_{\text{min}}}^{w_{\text{max}}} \mu_i p_i, \quad (1)
\]

where \( w_{\text{min}} \) corresponds to the oldest raw information packet that is not yet decoded, and \( w_{\text{max}} \) is incremented each time a new raw information packet is decided to be included in the...
linear combination by the sender. We denote by $\text{DoF}(c_t)$ the number of information packets contained in $c_t$.

In this adaptive approach, the sender uses $u_t$ to track the average channel erasure probability $\epsilon_t^{\text{mean}}$, and the number of erased and repeated DoFs, denoted $md$ and $ad$, respectively. These tracked quantities are used by two suggested forward error correction (FEC) mechanisms, a priori and a posteriori, to counteract the channel erasures. The a priori mechanism transmits $[\epsilon_t^{\text{mean}} - \eta]$ repeated DoFs, with $\lfloor \cdot \rfloor$ denoting rounding to the nearest integer, periodically after $n$ transmissions of new packets of information. Note that in AC-RLNC given in [26], there are no layering. In the a posteriori mechanism, a retransmission criterion is used by the sender. As demonstrated in [26], [27], when the estimated channel rate denoted $\epsilon_t \triangleq 1 - \epsilon_t$ is higher than the rate of the DoFs $d \triangleq md/ad$, the decoder has sufficient DoFs to immediately decode the delivered packets. However, these quantities cannot be computed exactly at the sender due to the round-trip time (RTT) delay. At time step $t$, the sender can only compute these quantities for time step $t' = t - RTT$, using the delayed feedback. Hence, with a tunable parameter $th$, the DoF rate gap and the retransmission decision at each time step, are given respectively by

$$\Delta_t \triangleq 1 - \epsilon_t^{\text{mean}} - d - th,$$

and $\Delta_t > 0$. \hspace{1cm} (2)

The estimation of the erasure probability $\epsilon_t$ is given by

$$\epsilon_t^{\text{mean}} = 1 - \frac{\sum_{j=1}^{T} u_j}{T}.$$ \hspace{1cm} (3)

To manage the maximum delay, the effective sliding window size $w$ is defined, such that $w_{\text{max}} - w_{\text{min}} = w$ and is limited to $\overline{\tau}$. In AC-RLNC as $th$, $\overline{\tau}$ can be selected to obtain the desired throughput-delay trade-off. When the limit of the window is reached, the sender transmits the same packet until all the information packets in the linear combination transmitted are decoded. We refer the readers to [26], [27] for further details on the operation of AC-RLNC.

B. Broadcast Approach

Layering of raw data was considered in the literature by using Broadcast Approach [3], [4], to mitigate data rate degradation caused due to the channel temporal variations. In a virtual broadcast channel, the sender encodes multiple ordered layers. Each encoded layer is designed to match the channel observation of a virtual receiver. Then, the sender virtually broadcasts all the layers to the receiver. Thus, the receiver can decode as many layers of layered encoded raw data as the current channel realization allows it. The broadcast approach was considered and characterized for many channel settings and practical solutions. In [5] presented an extensive survey of the literature on broadcast approach. Next, we present two coding schemes that are considered in the literature for layering setting.

1) Random Codes: In the literature, random codes, and in particular random linear codes, are known to be capacity-achieving [35], [36]. In [8], layering for a broadcast channel with two layers was illustrated by using random code. Due to the space limitation, we refer the readers to [8, Chapter 15] for further details. To significantly reduce the decoding complexity, efficient guessing noise decoders, that can also operate in the short block length regime, has been recently proposed [37]–[39]. Moreover, as elaborated in [38], the results in [35] using random linear codes may also hold for network coding in heterogeneous networks [40], [41]. This motivates future studies on LAC-RLNC for heterogeneous and mesh networks.

2) Linear Nested Codes: Several nested codes are considered in the literature [42]–[46], e.g., MDS codes [42], binary parity-check codes [43], and lattice codes [44].

Let $C_1$ be an $(n, k_1)$ linear code, and $C_2$ be an $(n, k_2)$ linear code with $0 \leq k_1 \leq n$ and $0 \leq k_2 \leq n - k_1$. Let $G_1$ and $G_2$ be the generator matrices of the codes $C_1$ and $C_2$, respectively.

```
Algorithm 1 Layered adaptive and causal RLNC[1].
Init:
AC-RLNC parameters - $\theta$, $w_{\text{min}} = 1$ and $w_{\text{max}} = 1$
Layering code parameters - $n$, $k_1$ and $k_2$ as given in (5) and (6)

1: Input: Feedback $u_t$ for $c_t$
2: Update $\epsilon_t^{\text{mean}}$ and $\epsilon_t^{\text{max}}$ as given in (3) and (9)
3: Update $md$, $ad$, and $\Delta_t$ as given in (2)
4: Eliminate the layer's packets decoded from the RLNC
5: Update $w_{\text{min}}$
6: if Feedback NACK, $u_t = 0$ then
7: if $\Delta_t \leq 0$ then
8: if not EW then
9: if Entire layered frame $f$ is transmitted then
10: Encode new frame using layering code in (4)
11: end if
12: In order transmit $p_i$ packet from the current frame
13: Update $w_{\text{max}}$
14: else
15: Update $m_t$ as given in (8)
16: Transmit RLNC $m_t$ times using (1): $ad_t = ad_t + m_t$
17: end if
18: else
19: Transmit RLNC using (1) $ad_t = ad_t + 1$
20: if EW then
21: Update $m_t$ as given in
22: Transmit RLNC $m_t$ times using (1): $ad_t = ad_t + m_t$
23: end if
24: end if
25: else if Feedback ACK, $u_t = 1$ then
26: if EW then
27: Update $m_t$ as given in (8)
28: Transmit RLNC $m_t$ times using (1): $ad_t = ad_t + m_t$
29: end if
30: if $\Delta_t > 0$ then
31: Transmit RLNC using (1): $ad_t = ad_t + 1$
32: else
33: if Entire layered frame $f$ is transmitted then
34: Encode new frame using layering code in (4)
35: end if
36: In order transmit $p_i$ packet from the current frame
37: Update $w_{\text{max}}$
38: end if
39: end if
40: if $w > \overline{\tau}$ then
41: Transmit RLNC using (1) until DoF($c_t$)= 0
42: end if
```

[1] The shaded parts are for the layering coding, while the remaining parts are for the adaptive and causal coding scheme.
Figure 2: AC-RLNC simulation with and without broadcast approach for RTT = 40 [slots]. The top results are for normalized throughput (left), mean in order delay (middle), and maximum in order delay (right). The bottom results are for the frame, for layer 1 and layer 2 sizes (left), and for the histogram of the in-order delay realizations (right).

Let $l_i, i \in \{1, 2\}$ be the $k_i$ raw data packets of layers 1 and 2, respectively. At the sender, the encoding operation for $n$ packets is given by

$$p^n = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

(4)

At the receiver, decoding of the first layer can be performed with $k_1$ delivered encoded packets. The decoding of the second layer can be performed with $n$ delivered encoded packets. Efficient decoding can be done at the receiver, for example, using Reed Solomon (RS) codes as proposed in [42], [46].

IV. LAYERED ADAPTIVE AND CAUSAL CODING

In this section, we detail the layered adaptive and causal RLNC (LAC-RLNC) approach as summarised in Algorithm 1. We propose a novel solution that benefits from both, Broadcast Approach using linear layering coding and AC-RLNC solution, as described in Subsections III-B and III-A, respectively. Fig. 1 shows the system model and the layered adaptive causal encoding process. The symbol definitions are provided in Table I.

The proposed LAC-RLNC approach is designed to minimize the in-order delivery delay of the base layer (Layer 1) while closing the throughput-delay trade-off of the enhancement layer (Layer 2). The sender decides at each time step whether to transmit an FEC RLNC or new information coded packet using a sliding window mechanism as in AC-RLNC. However, the proposed layered approach differs from AC-RLNC in two main perspectives: (1) Here, the raw data is first encoded using a broadcast approach coding that allows the receiver to decode the base layer (Layer 1) with minimum in-order delay required. At the same time, the enhancement layer (Layer 2) with the a priori and posterior FEC mechanisms allow managing the desired throughput-delay trade-off. (2) RLNC coding is used to transmit a repeat layered coded packet using a linear combination with random coefficients as given in (1). That is, RLNC is used only for FEC retransmissions as controlled by the a priori and posterior mechanisms, not for all the coded packets transmitted as described in Section III-A.

The components of the proposed protocol are described next.

a) Layered Coded Packets: The raw data packets are first encoded in-order into frames of base and enhancement layers (resp. to lines 9-11 and 31-34 in Algo. 1) using the layering code as described in Section III-B2. The code and layers sizes are selected to obtain the desired high dimensional throughput-delay trade-off. The frame size is given by

$$n = \frac{\text{RTT}}{n_f},$$

(5)

where $n_f$ denotes the frame size factor, which is a tunable parameter. The code designer can control the minimum in-order delivery delay at the receiver by varying this factor. The layers’ packets number are given by

$$k_1 = n(1 - \epsilon_{\text{max}}) \quad \text{and} \quad k_2 = n - k_1$$

(6)

where $\epsilon_{\text{max}}$ denotes the maximum erasure probability estimated, as given in (7). The packets number are selected by the code designer to manage the layer 1 throughput-delay trade-off. The maximum erasure probability estimated is given by

$$\epsilon_{\text{max}} = \epsilon_{\text{mean}} + \sigma \sqrt{\frac{v}{n}}, \quad \text{where} \quad v = n(1 - \epsilon_{\text{mean}})\epsilon_{\text{mean}}$$

(7)

is the variance of the BEC during the period of $n$ transmissions, and $\sigma$ is the tunable standard deviation factor parameter, using the so-called $68 - 95 - 99.7$ rule. Note that the layer’s size can be adjusted adaptively according to the tracking of the channel, as represented by the feedback.

b) Effective Window: At each time step, the sender tracks the actual channel and DoF rates by the feedback acknowledgments $u_t$, and updates the DoF rate gap (resp. to lines 1,2,3 in Algo. 1). In LAC-RLNC, decoding at the receiver is done for coded layers, not per coded packets. The effective window, including the subset of information packets within the effective window, is updated for the decoded layer at the receiver (resp. to lines 4 and 5 in Algo. 1), as reflected at the sender by the delayed feedback. That is, $w_{\text{min}}$ and the information packets needed in the coded RLNC are updated when $k_1$ or $k_1 + k_2$ coded packets for each frame are delivered at the receiver and provide sufficient DoF’s to be decoded.
c) Layered FEC mechanisms: We use both FEC mechanisms, a priori or a posteriori, as given in Subsection III-A. The sender encodes the layered coded packets to RLNC packet for any FEC retransmission, as given in (1), using the effective window. We use the retransmission criteria as given in (2) (resp. to lines 19 and 31 in Algo. 1) for the a posteriori mechanism decisions. Finally, to obtain the desired high dimensional throughput-delay trade-off, the adaptation of the layered a priori mechanism is designed as follows: The adaptive number of the periodically a priori RLNC retransmission $m_t$, after transmission of each frame, is given by

$$m_t = \lceil \max_{\epsilon} \cdot n \rceil,$$

where

$$\max_{\epsilon} = \mean + \sigma_{\epsilon} \frac{\sqrt{D}}{n},$$

and $\sigma_{\epsilon}$ is selected to control the trade-off of layer 2.

V. PERFORMANCE EVALUATION

This section presents the results of the layered AC-RLNC approach for single path communication with feedback. In terms of mean and max in-order delivery delay and throughput, the performance is compared to the non-layered solution.

The simulation results presented are for a BEC channel. The first scenario we consider is for erasure probability varying between 0.02 to 0.2 and with RTT of 40 [Slots]. The second scenario is for erasure probability varying between 0.02 to 0.6 and with RTT of 80 [Slots]. Figures 2 and 3, show the results of the first and second scenarios, respectively, using AC-RLNC, with and without layering. The top results are for the normalized throughput $\eta$ (left), and mean $D_{\text{mean}}$ (middle) and max $D_{\text{max}}$ (right) in-order delivery of frames, as defined in Section II. The results presented on the top, have been averaged on 150 different channel realizations. The bottom results are of the frame and the layer’s sizes selected to obtain the desired high dimensional throughput-delay trade-off (left) and the histogram of the in-order delay realizations (right).

In both scenarios, the layering code parameters are selected to obtain similar performance as the non-layered AC-RLNC solution in terms of throughput with zero error probability. Thus, we compare the performance, of the layered and non-layered solutions, in terms of in-order delivery delay of frames. We can see significant gains in $D_{\text{mean}}$ and $D_{\text{max}}$ with respect to the non-layered solution. From the delay point of view, the gains are significant as both the mean and max in-order delay are reduced approximately by a factor of 3 and a factor of 3 for layers 1 and 2, respectively. Moreover, for each scenario, given the RTT delay, we show the lower bound of the delay. For relatively good forward link quality, this lower bound reducing to about half of the RTT [27]. We note that with the selected layering code parameters, the in-order delivery delay of layer 1 is very close to the optimal lower bound. In the histogram figures of the in-order delivery delay, we can see that all the realizations of layer 1 are obtained close to the optimal lower bound delay. This is critical for streaming applications, as by using the proposed approach, the base layer can be delivered in a minimum delay to enable the required service. Moreover, we can note that using the Broadcast approach, more than 90% realizations of the enhancement layer are almost achieving the lower bound.

VI. CONCLUSIONS AND FUTURE WORK

We proposed a layered adaptive causal RLNC (LAC-RLNC) algorithm for data streaming over an erasure channel. The proposed method considered variable recovered layered information over variable short block length and rate. Specifically, we show that with zero error probability LAC-RLNC can reduce the in-order delay of the layers significantly and design the high dimensional gap between data rates and delay to obtain the desired performers.

Future work includes the derivation of bounds on the mean and maximum in order delivery delay and throughput for the proposed approach. These bounds will provide theoretical guarantees for the layered code designer in the adaptive and causal network coding solution. Extensions also include the study of layered adaptive causal network codes where we have more than two layers and coded layered solutions for multi-path and multi-hop heterogeneous networks.
