End-to-end Delay Analysis for Reliable Communications over Lossy Channels: Integrating Network Coding and ARQ Schemes

Francesco Chiti*, Romano Fantacci*, Russell Allan Johnson†, Vladimir Crnojević‡, Dejan Vukobratović‡

*Department of Electronics and Telecommunication - University of Florence
†Department of Systems and Informatics - University of Florence
‡Department of Power, Electronics and Communication Engineering - University of Novi Sad

Abstract—This paper deals with the evaluation of the end-to-end packet delay for a lossy network where network coding is adopted to achieve better performance. Differently from previous published papers on this subject, connection oriented services are assumed, this means that packets have to be in-order received and wait at the receiver buffer till all the previous ones have been correctly delivered. The focus is on a network model where two source nodes broadcast packets to a group of two sink nodes over lossy wireless channels. Three different alternatives have been considered in order to assure a reliable data multicasting, namely: a classical random linear network coding scheme, a linear network coding combined with a basic ARQ or, alternatively, with a soft combined ARQ scheme. Performance comparisons provided by means of analytical and numerical results clearly highlight that the better solution is to adopt the latter alternative.

Index Terms—Network Coding, (Soft)ARQ, Erasure Channels, Delivery Delay.

I. INTRODUCTION

Classical data networks rely on store-and-forward principle for packets transmission. According to it, intermediate nodes receive, process and eventually repeat incoming packets, while maintaining their individuality. However, this approach has been shown [1] to be suboptimal, as for as the capacity allocation over the entire network graph.

To tackle this drawback, the network coding (NC) has been proposed in that paper: according to it routers (or relays) might combine several incoming (or even generated) packets into one outgoing packet. In particular, this provides capacity gain in the case of multicast transmissions.

Many other papers have demonstrated the advantages of NC in different scenarios [2], [3]. However, original investigations mainly focused on wireline networks; only recently NC has received attention for applications in wireless networks as a way of improving network capacity and coping with unreliable wireless links [4], [5]. This is a crucial topic for delivering real-time media applications with QoS constraints in terms of available bandwidth and delay sensitivity. In fact, wireless channels are inherently error prone with a time varying packet error rate (PER) resulting in packet erasures. As a result, many solutions have been proposed in the past to cope with reliability, ranging from basic retransmission schemes as Automatic Repeat reQuest (ARQ) [6] or even Hybrid ARQ (HARQ) [7] types. Nevertheless, the exact evaluation of the reliability gain of NC, or the comparison with (S)ARQ schemes for different load and error conditions are still open issues.

This paper focus on the study of the overall end-to-end delay performance of different schemes for reliable communications over a lossy wireless network, on the basis of the same model proposed in [4]. Differently from previous papers on this subject [8], in-order communication is assumed, this meaning that correctly received packets are stored in the sink buffer and sent to the application layer as soon the original sequence has been completed. Thus, this buffer is properly referred to as resequencing buffer, and the additional latency experienced by the packets is consequently called the resequencing delay. The proposed performance evaluation takes into account the following reliability schemes:

1) basic linear NC scheme: intermediate nodes whenever possible send out packets that are random linear combination of previously received packets belonging to independent data streams.

2) combined ARQ-NC scheme: intermediate nodes linearly combine packets belonging to independent streams with the same sequential number. In particular an ideal ARQ scheme, where the outcome of any packet transmission attempt is immediately known has been considered, thus implying that, differently from the previous scheme, intermediate nodes continue to send out exactly the same combined packet till an error-free reception occurs.

3) combined Soft ARQ-NC scheme (SARQ-NC): it is analogous to ARQ-NC except for the soft combining of the last received copy of a packet with all possible previous incorrect copies, according to the Chase’s principle [9], [10], instead of merely discarding them.

The rest of this paper is organized as follows. In Section II, the network model adopted as basis for our evaluations and an analytical approach to derive asymptotic bounds on performance are presented. The performance comparisons for the three different mechanisms under considerations, are discussed in Section III, while conclusions are drawn in Section IV.
together with addressing further developments.

II. PERFORMANCE ANALYSIS

A. System Model

The network model adopted in deriving our analytical and numerical results is shown in Fig. 1 and it is consistent with the classical scheme introduced in [1] (a one-source two-sink network) to point out the advantages of the network coding approach. In particular, it consists of two source nodes that have to broadcast independent data to two sink nodes. In our model all the links are bidirectional and nodes are ideally capable of simultaneously receiving packets\(^1\). Time reference for each link is slotted and all links are assumed to be synchronous. In addition, links are affected by independent errors, which makes a packet erroneously received to be retransmitted until an error-free reception occurs. The acknowledgment messages (ACK) are assumed to be instantaneously received after the packet transmission completion and error-free.

Moreover, the packet loss probability \(P_e\) has been assumed equal for all the wireless links and independent from links or transmission attempts. As a consequence, the random variables \(p\) representing the probability of success in delivering a packet (such that \(P_e = 1 - p\)) are identically distributed whatever the link is involved and the same is valid for the associated delivery delay.

Fig. 1. Reference topology.

\(^1\)It could be possible by adopting multiradio interfaces or exploiting antenna diversity.

B. ARQ-NC Delay Analysis

The probability distribution of the overall end-to-end delay is derived hereinafter, under the assumption of a negligible queuing and processing delay at the transit nodes. This means that network is low loaded (i.e., \(\rho \to 0\), where \(\rho\) is the so called loading factor [11]), as in more general scenario each node can be properly modeled as a G/G/1 system.

Note that, according to our model, intermediate nodes have to perform combinations of packets belonging to individual input data flows having the same sequence number. As a consequence, packets are sent out to the application layer at the sink nodes according to their original order, hence there is no need of resequencing.

Finally, an ideal medium access control scheme (MAC) has been considered, in which packets could be received simultaneously without collisions. In fact, an eventual delay in accessing the shared medium should affect in the same way all the investigated reliable schemes.

1) Path \(1 \to 3\) : The distribution of delay \(\tau^{(3)}\) \(\sim\) \(\tau\) in delivering packet \(\text{a}\) from node 1 (source) to node 3 follows a shifted geometric distribution\(^2\) such that:

\[
\Pr \{\tau = m\} = p(1 - p)^{m-1}.
\]

with \(m \geq 1\). The same occurs for delay distribution over path \(2 \to 3\).

2) Path \((1 \to 3) \cup (2 \to 3)\) : The latency \(\tau^{(3)}\) for collecting both packets \(\text{a}\) and \(\text{b}\) at node 3 can be expressed as it follows:

\[
\Pr \{\tau^{(3)} = m + 1\} = \Pr \{\max \{\tau, \tau^{\text{a+b}}\} = m + 1\} =
\]

\[
= \sum_{l=1}^{m} (\tau_{m+l} + \tau_{l}\tau_{m+1}) + \tau_{m+1}\tau_{m+1}
\]

(2)

where \(m \geq 1\). It is worth noticing that (2) does not take into account collisions occurring within the same time slot.

3) Path \(1 \to 3 \to 4 \to 5\) : The latency \(\theta\) for collecting packet \(\text{a+b}\) at node 5 (sink) is a random variable such that:

\[
\theta = \tau^{(3)} + \tau + \tau = \tau^{(3)} + \tau^{\Sigma}
\]

where the latency \(\tau^{\Sigma}\) can be expressed as the addition of delays over two independent links as it follows:

\[
\tau^{\Sigma} = \tau + \tau.
\]

In addition, it is possible to show that:

\[
\Pr \{\tau^{\Sigma} = m\} = \sum_{l=2}^{m} \tau_{m-l} \tau_{l}
\]

(5)

with \(m \geq 2\). Thus, it follows that:

\[
\Pr \{\theta = m\} = \sum_{l=1}^{m} \tau_{m-l} \tau_{l}^{\Sigma}
\]

(6)

where \(m \geq 2\) and \(\tau_0^{\Sigma} = \tau_1^{\Sigma} = 0\).

4) Path \((1 \to 3 \to 4 \to 5) \cup (1 \to 5)\) : Finally, the overall delivering latency \(\tau^{(5)}\) for collecting packets \(\text{a}\) and \(\text{a+b}\) at node 5 (sink) is a random variable such that:

\[
\Pr \{\tau^{(5)} = m + 1\} = \Pr \{\max \{\theta, \tau^{\text{a+b}}\} = m + 1\} =
\]

\[
= \sum_{l=1}^{m} (\theta_{m+l} + \theta_{l}\theta_{m+1}) + \theta_{m+1}\theta_{m+1}
\]

(7)

\(^2\)Delay is normalized to the slot duration \(T\) that is equal to round trip time (RTT) if ACK are instantaneously received.
The above delay delivery analysis under the low load assumption can be easily generalized to any directed acyclic graph (DAG) resorting to the concept of subtree decomposition introduced in [12]. According to this approach, a multi-source multi-sink network coding problem over any DAG can be partitioned into subtrees through which the same information flows. NC for achieving the multicast capacity is sufficient to be applied only on a small subset of nodes in the network called coding points, which are intersections of the subgraphs. Each subgraph is a tree, rooted either at the coding point, or at the source node. Clearly, the delay delivery distribution can be “built” along the DAG by applying the rule of maximization of random variables inside the coding points (as applied in node 3 of the reference topology, see Eq. (2)), and the rule of summation of random variables inside the “interior” subtree nodes (as applied in node 4 and 5 of the reference topology, see Eq. (5)). However, we leave the details of this generalization for our future work.

Extension of the above analysis to derive the end-to-end closed form delay distribution by taking into account processing and queuing time at each intermediate node (in case 0 < ρ < 1) is a extremely complex task, even admitting it feasible [11], as the model to be considered is a network of G/G/1 queues. As a consequence, a simulation campaign is employed in delivery delay analysis.

The simulation model applied is described as follows. For simplicity, Poisson arrival process of arrival rate λ is assumed, where ρ=λ/RTT such that 0 < ρ < 1. The same arrival process fills the incoming queues of both source nodes 1 and 2, i.e., the k-th arrival simultaneously delivers two packets, a_k and b_k, to the respective incoming queues of nodes 1 and 2. Incoming queues are assumed in front of each link, as well as two additional queues, one for the network coding processing in node 3 and one for the decoding in node 5. Each link serves its respective incoming queue on a FCFS (first-come first-served) basis, where the link service is determined by an ideal ARQ scheme. NC processing queue in node 3 performs linear combining (NC) on the pair of packets {a_k, b_k} as soon as both packets a_k and b_k are available. The same holds for the decoding process applied on the pair of packets {a_k, a_k + b_k} in the decoding queue of node 5. Ideal MAC protocol is assumed while accessing NC processing and decoding queues in nodes 3 and 5, respectively. The size of all queues is assumed to be infinite.

For each pair of packets {a_k, b_k} transmitted over the network, the simulation measures the overall delivery delay τ_k = t_d(k) - t_g(k) (that includes link transmission, MAC scheduling and queuing delay) between the time instant t_g(k) of packet pair generation at node 1, and the time instant of packet pair decoding t_d(k) at node 5. The average delivery delay is obtained by averaging the packet delivery delays τ_k over sufficiently long time period. Clearly, for a fixed network topology, τ = τ(ρ, P_e).

C. SARQ-NC Delay Analysis

The analytical approach outlined before in Section II-B for the ARQ-NC case is extended here to the SARQ-NC. The main modification is that in the SARQ-NC case all the received copies of a same packet are used at any receiving node in order to improve data reliability. Previously received copies with errors of a same packet are combined bit by bit with the last received one according to the Chase’s principle [9]. Interested readers are referred to [13], [14] for an analysis concerning the performance advantages of soft combining ARQ schemes with respect to classical alternatives. From above and [10], it follows that P_e depends upon the number of transmission attempts, in particular for the i-th transmission, it is equal to P_e = 1 − p_i. As a result, (1) becomes:

\[
\Pr \{\tau = m\} = p_m \prod_{i=1}^{m-1} (1 - p_i). \tag{8}
\]

The derivation of the delivery delay (Eq. (7)) is straightforward, following the same rationales. Finally, for a more general case (in which 0 ≤ ρ ≤ 1) the performance evaluation can be conducted only via numerical simulations for the reasons explained in Section II-B.

D. NC Latency Bound

In previous Sections, delivery delay for simple linear network coding scheme in conjunction with basic ARQ or sARQ mechanisms is analyzed. When (S)ARQ techniques are applied, sequential delivery of data packets across network links is guaranteed, which is why coding in network node 3 of the reference topology applies on the pair of data packet {a_i, b_j} having the same indices i = j. Whenever (S)ARQ is no longer applied, i.e., if link layer service is unreliable, network coding is still applicable in node 3 over the pair of arriving data packets {a_i, b_j} for which, in general, i ≠ j. In this case, the decoding process at node 5 is not instantaneous upon reception of the pair {a_i + b_j, a_j} as, in general, i ≠ j ≠ k. Therefore, successful decoding may require significant delays until the received system becomes at least partially solvable, which is why arbitrary long waiting time and infinite queue size at the decoding node have to be assumed.

To compare the delivery delay performance of linear NC with and without (S)ARQ, the simulation model described in Section II-B is extended: (S)ARQ procedures are disabled at each network link, and the decoding procedure at the receiving node is changed accordingly. To admit fair comparison between the two schemes, the overall delivery delay τ_k for the pair of packets {a_k, b_k} has to be defined in a different way than is Section II-B. This is because in non-(S)ARQ setup, a_k and b_k are independently decoded and the decoding of a_k and/or b_k may not be successful, and due to the fact that the decoding of a_k is possible immediately after a_k succeeds in traversing a single link (1 → 5). For this reason, the delivery delay τ_k in non-(S)ARQ setup is defined only with respect to the decoding of data packet b_k at the destination node 5, for those indices k for which the packet b_k is recovered. In
other words, \( \tau_k = t_{d,b}(k) - t_{g,b}(k) \) where \( t_{g,b}(k) \) is the time instant of the packet \( b_k \) generation at node 2, and \( t_{d,b}(k) \) is the time instant of the packet \( b_k \) decoding at node 5 (whenever \( t_{d,b}(k) \) exists). The average delivery delay \( \bar{\tau} \) is obtained by averaging the packet delivery delays \( \tau_k \) over sufficiently long time period.

### III. Numerical results

In accomplishing the performance evaluation, it has been supposed the presence of independent errors affecting packet transmissions, with a Signal to Noise Ratio (SNR) in the range [4-9] dB.

First of all, the analytical results derived for low loaded network (i.e., \( \rho \to 0 \)) are presented. In particular, the normalized delay cumulative distribution function (CDF) is depicted in Fig. 2 for basic ARQ scheme for different SNR values.

In addition to this, it has been taken into account also a HARQ scheme based on the adoption of Chase combining [9]. The normalized delivery delay CDF is shown in Fig. 3 again within the same SNR range. It is worth pointing out the gain provided by the improved retransmission mechanism especially for low-to-medium SNRs. For instance, no more than 10 transmissions are needed adopting SARQ (with SNR = 4dB), instead of 40.

For addressing the more general case, in which the queuing delay is also considered, simulations have been performed, relying on operative hypotheses addressed in Section II-B. First of all, the normalized delivery delay CDF is shown in Fig. 4 for low loaded network \( \rho = 0.01 \), pointing out a good agreement with the theoretical predictions.

In addition to that, parameter \( \bar{\tau} \) has been evaluated for low-to-medium loading factor with SNR = 9 dB to highlight the impact of \( \rho \) on the delivery delay for basic ARQ scheme in moderate propagation conditions (Fig. 5). Not surprisingly the performance decreases at the increase of \( \rho \) as the queuing delay is not negligible.

Finally, the average delivery delay \( \bar{\tau} \) of NC with ARQ or SARQ schemes has been compared in Fig. 6 as a function of both SNR and \( \rho \). It is interesting pointing out the better performance provided by SARQ approach, while it saturates at the increasing of SNR values as the combining gain reduces (being in the fall region of FER curve). The converse is true with regard to the effect of traffic load \( \rho \) as previously explained.

To complete the investigation, the average delay \( \bar{\tau} \) comparison for NC with (S)ARQ or without (S)ARQ is reported in for SNR = 9dB, as depicted in Fig. 7. It could be noticed that NC+(S)ARQ always outperforms NC alone due to minor impact of packet retransmissions to the decoding process. However, NC profits of increased traffic load as the larger buffer occupancy provides more packet to be combined.

### IV. Discussion and Conclusions

In this paper, it has been investigated the delivery delay performance of NC with or without ARQ protocol for connection oriented services. The focus has been on a network model where two source nodes broadcast packets to a group of two sink nodes over lossy wireless channels. Throughout the proposed analytical framework and numerical results it has clearly pointed out the advantage of adopting packet retransmission. In particular, a remarkable gain is provided for low-to-medium SNRs which can be further improved by jointly adopting a simple soft combined ARQ scheme (SARQ-NC). As a result, NC alone performs worse than other schemes for all SNR and loading factor values, so that it is always preferable to retransmit uncorrect packet rather than merely relying on linear combination.

Possible developments of the present work could be focused on generalizing the analysis to generic topologies. In addition a MAC layer model might be taken into account, possibly including the queuing delay.
Fig. 4. Simulated normalized delay CDF for basic ARQ scheme ($\rho = 0.01$).

Fig. 5. Simulated normalized delay DF for basic ARQ scheme with $0.01 \leq \rho \leq 0.5$ and SNR = 9 dB.

Fig. 6. Average normalized delay for both basic ARQ and SARQ schemes with $\rho \in [0.01, 0.5]$ and SNR $\in [5, 9]$ dB.

Fig. 7. Average normalized delay for both basic ARQ and SARQ schemes with $\rho \in [0.01, 0.5]$ and SNR = 9 dB.


