Generalized Instantly Decodable Network Coding for Relay-Assisted Networks

Adel M. Elmahdy¹, Sameh Sorour², and Karim G. Seddik³

¹Electrical Engineering Department, Alexandria University, Alexandria 21544, Egypt.
²Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia.
³Electronics Engineering Department, American University in Cairo, AUC Avenue, New Cairo 11835, Egypt.

Email: adel.elmahdy@ieee.org, sameh.sorour@kaust.edu.sa, kseddik@aucegypt.edu

Abstract—In this paper, we investigate the problem of minimizing the frame completion delay for Instantly Decodable Network Coding (IDNC) in relay-assisted wireless multicast networks. We first propose a packet recovery algorithm in the single relay topology which employs generalized IDNC instead of strict IDNC previously proposed in the literature for the same relay-assisted topology. This use of generalized IDNC is supported by showing that it is a super-set of the strict IDNC scheme, and thus can generate coding combinations that are at least as efficient as strict IDNC in reducing the average completion delay. We then extend our study to the multiple relay topology and propose a joint generalized IDNC and relay selection algorithm. This proposed algorithm benefits from the reception diversity of the multiple relays to further reduce the average completion delay in the network. Simulation results show that our proposed solutions achieve much better performance compared to previous solutions in the literature.

Index Terms—Network Coding; Relay Assisted Networks; Wireless Multicast.

I. INTRODUCTION

Since the introduction of Network Coding (NC) in the seminal paper by Ahlswede et al. [1], it has been widely employed to improve the transmission efficiency in many network settings (2, 3 and ref. therein). Moreover, several works focused on using opportunistic network coding (ONC) to enhance the packet recovery process in point-to-multipoint (PMP) wireless broadcast erasure channels (4, 5 and ref. therein). One of these ONC schemes that has been widely considered in this context is Instantly Decodable Network Coding (IDNC) due to its many desirable properties such as the instant decoding of the received packets at the selected set of receivers, simple XOR encoding and decoding, and the needlessness for decoding buffers at the receiver side to store the coded packets. These properties allow simple and power efficient designs of IDNC-capable receivers, which are more suitable for hand-held mobile terminal.

Recently, NC was extended to enhance packet recovery in relay assisted networks (RANs) [6, 7]. This extension responded to the consideration of relay nodes (RNs) in the standardization process of next generation mobile broadband communication systems such as 3GPP LTE-Advanced [8, 9] and [10]. Deploying relay nodes is a smart solution to increase the capacity or to extend the cell coverage area by enhancing the coverage at the cell edge where the receivers suffer from low Signal to Interference plus Noise Ratio (SINR). The problem of minimizing the number of recovery transmissions (a.k.a. completion delay) using IDNC was considered in the framework of one decode-and-forward relay-assisted broadcast networks [7]. The proposed IDNC-based solution was shown to achieve a better completion delay and thus transmission efficiency, as compared to IDNC with no relay case as well as automatic repeat-request (ARQ).

Even though the authors of [7] considered a strict IDNC (S-IDNC) approach when solving the completion delay problem in RANs, the senders are constrained to encode packets that include at most one lost packet for each receiver in this IDNC approach. In [11], it has been shown that this S-IDNC approach is too restrictive in enhancing the decoding delay of the different receivers. The authors thus proposed a more relaxed IDNC approach which allowed any packet encoding at the sender and forced each of the receivers to discard any coded packets combining two or more non-received source packets at that receivers. This generalized IDNC (G-IDNC) approach was shown to provide much more coding flexibility at the sender and thus achieved much better completion delay performance.

This paper extends the works in [6, 7] in two directions. First, we propose the use of G-IDNC as a more efficient NC scheme to further reduce the completion delay in the one-relay setting. Although the G-IDNC approach was proven to outperform S-IDNC in decoding delay only [11], the coding flexibility of G-IDNC and its previous use for completion delay reduction in simple PMP networks [12, 13] can be used to outperform the completion delay achieved by S-IDNC in the one-relay scenario. We then extend our study to the case of multiple-relay scenario, in which joint decisions on the best encoding packet and best transmitting relay should be made. We thus propose a solution for this joint decision problem, which is based on the same methodology to minimize the completion delay in PMP networks with one relay node. We finally compare this proposed solution to other famous approaches in the PMP NC literature.
II. SYSTEM MODEL

The proposed system consists of a base station (BS) that is required to deliver a frame \( \mathcal{N} \) of \( N \) source packets to a set \( \mathcal{M} \) of \( M \) terminal nodes (TNs) over erasure channels. Each of the receivers is interested in a subset or all of the packets of \( \mathcal{N} \), which defines a generic multicast scenario. The packets requested by receiver \( i \) are referred to as the primary packets, while the undesired packets are referred to as the secondary packets. In addition, there is a set \( \mathcal{R} \) of \( R \) decode-and-forward RNs to improve the coverage of the BS. An example of such network with 3 RNs is shown in Fig. 1.

Similar to the assumptions and constraints of [7], we assume that the channels between the BS and RNs are better than those between the BS and the TNs. We also assume that the channel between the RNs and the TNs are better than those between the BS and the TNs. RNs operate in half-duplex mode and cannot transmit simultaneously with the BS, even if they have received all the packets in \( \mathcal{N} \).

In the initial transmission phase, the BS sequentially sends all the source packets of the frame. These transmissions are subject to possible erasures at the RNs and TNs. For each received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent received packet, each RN/TN feeds back a positive acknowledgement (ACK) to the BS. RNs also store the feedback sent.

For the non-targeted TNs and most importantly the RNs (i.e. terminals that do not want any packets. The sender stores this information in a state feedback matrix (SFM) \( F = [f_{ij}] \), \( \forall i \in \mathcal{M}, j \in \mathcal{N} \), such that \( f_{ij} = 0 \) if \( j \in \mathcal{H}_i \), \( f_{ij} = 1 \) if \( j \in \mathcal{W}_i \), and \( f_{ij} = -1 \) if \( j \in \mathcal{L} \setminus \mathcal{W}_i \).

Afterwards, a recovery phase begins in which both the BS and RNs should collaborate in order to deliver the missing packets of all TNs in the smallest number of transmissions. Both BS and RNs employ IDNC packet that is able to serve many TNs simultaneously in each recovery transmission. Note that each recovery transmission from the BS or a RN is still subject to erasures. Thus, the TNs (and RNs if the BS is transmitting) should respond to each received recovery transmission with an ACK, which is used to update the SFM for subsequent transmissions. The method with which these IDNC coded packets are generated to target specific TNs (and possibly RNs) and the algorithms for BS/RN selection for each recovery transmission will be described in details in Sections III and V, respectively.

III. G-IDNC GRAPH

The IDNC graph [11, 12] provides a framework to determine the packets that can be combined together and simultaneously recovered by any given set of TNs using the same transmission. This graph \( \mathcal{G} \) is constructed by first generating a vertex \( v_{ij} \) in \( \mathcal{G} \) for each packet \( j \in \mathcal{L}_i \), and for all users. Two vertices \( v_{ij} \) and \( v_{kl} \) in \( \mathcal{G} \) are adjacent if one of the following conditions is true:

- C1: \( j = l \Rightarrow \) The two vertices represent the loss of the same packet \( j \) by two different TNs \( i \) and \( k \).
- C2: \( j \in \mathcal{H}_k \) and \( i \in \mathcal{H}_i \Rightarrow \) The requested packet of each vertex is in the Has set of the TN of the other vertex.

Consequently, each edge between two vertices \( v_{ij} \) and \( v_{kl} \) in the graph means that the missing packets \( j \) and \( l \) at TNs \( i \) and \( k \) can be served simultaneously and instantly by sending either packet \( j \) or \( j \oplus l \) if \( j = l \) or \( j \neq l \), respectively. This property extends from two adjacent vertices to all cliques in the graph. A clique in a graph is a subset of that graph whose vertices are all adjacent to one another. Thus, each clique in \( \mathcal{G} \) defines a packet combination that can instantly serve all the users inducing this clique’s vertices.

We can classify the vertices of a graph into two layers:

- Primary graph \( \mathcal{G}_p \): It includes all the vertices from the Wants sets of all users.
- Secondary graph \( \mathcal{G}_s \): It includes all the vertices that are not in the Wants set of any user.

For any transmission, we first select a clique \( \kappa_\rho \) from the primary graph, which will target a set of TNs \( \mathcal{T}_{\kappa_\rho} \subseteq \mathcal{M} \). For the non-targeted TNs and most importantly the RNs (i.e. \( \mathcal{M} \setminus \mathcal{T}_{\kappa_\rho} \)), the adjacent secondary subgraph to \( \kappa_\rho \) can be used to deliver unwanted packets to them without violating the instant decidability of the packets \( \kappa_\rho \) at the TNs in \( \mathcal{T}_{\kappa_\rho} \)

We start our analysis by proving the following lemma.

IV. WHY G-IDNC INSTEAD OF S-IDNC

In this section, we compare the performance of G-IDNC against S-IDNC and show the ability of the former to achieve a better step towards a lower completion delay, given any feedback matrix and for any scheduling approach.
Defining the S-IDNC constraint on the vertex connectivity conditions. Since packet combination possibilities of any S-IDNC scheme can be formulated as a graph $G_s$ that is a subgraph of the G-IDNC graph $G$.

**Proof**: For any feedback matrix, we can generate a graph $G_s$ for S-IDNC packet transmission, which will have the same vertices of the corresponding G-IDNC graph $G$ for the same feedback matrix. Since packet combination possibilities are represented by the edges in the graph, we can enforce the S-IDNC constraint on the vertex connectivity conditions. Defining $M_j$ and $M_l$ as the sets of receivers that lack packets $j$ and $l$, respectively, we can express the S-IDNC conditions to serve two vertices simultaneously as follows:

- $C1$: $j = l$
- $C2$: $j \in H_k$ and $l \in H_l$ and $M_j \cap M_l = \emptyset$.

Indeed, for condition $C1$ ($j = l$), the edges do not represent coding combinations since both end vertices of the edge request the same packet. Consequently, this edge on its own represents the transmission of the source packet $j$, which can never be a non-instantly decodable for any TN, and thus does not violate the coding constraints in S-IDNC. As for condition $C2$, it does indeed involve a packet combination $j \oplus l$ but again the last constraint $M_j \cap M_l = \emptyset$ guarantees that no coding of two packets missed by the same receiver can be done.

Comparing these conditions to those in Section III (C1 and C2), it is obvious that both C1 and C1 are equivalent. Thus, all the adjacent vertices using condition C1 in $G_s$ will also be adjacent in $G$. On the other hand, C2 is a relaxed condition of C2, and thus any two vertices satisfying C2 also satisfy C2 but not vice versa. Consequently, all edges generated in $G_s$ using C2 will also be generated in $G$ using C2. However, $G$ will have extra edges between vertices that satisfy C2 but not C2.

We can thus conclude that $G_s$ is a subgraph of $G$ for any feedback matrix.

From [13], it has been shown that a transmission aiming to reduce the completion delay should target the largest number of receivers with largest Wants sets and erasure probabilities, as such transmission brings the system the closest to completion by:

- Reducing the largest individual completion delays of worst case receivers.
- Maximizing the resulting number of coding opportunities for subsequent transmissions.

Such strategy, termed the WoRLT strategy, can be applied by attributing the weights $\left(\frac{|W_i|}{1-p_i}\right)^n$ to the vertices of each TN $i$ in the IDNC graph, and choosing the maximum weight clique (MWC) ($p_i$ is the erasure probability of TN $i$). For large $n$, the chosen MWCs are always the ones satisfying the WoRLT strategy.

After the initial transmission phase (i.e. at time $t = 0$), both $G_s$ and $G$ can be generated for the corresponding feedback matrix. Lemma 1 shows that $G_s$ will be a subset of $G$. Consequently, the MWC $\kappa_s^*$ chosen from $G_s$ will definitely be a subset of or equal to a maximal clique $\kappa_g(\kappa_g^*)$ in $G$, which need not be the MWC in $G$. Thus, the weight $\omega(\kappa_g(\kappa_g^*))$ of $\kappa_g(\kappa_g^*)$ will be definitely smaller than or equal to that of the MWC $\kappa_g^*$ in $G$. Consequently, $\omega(\kappa_s^*) \geq \omega(\kappa_g(\kappa_g^*)) \geq \omega(\kappa_g^*)$, and thus $\kappa_s^*$ can bring the system closer to completion compared to $\kappa_g^*$.

Assume that we send the packet corresponding to $\kappa_g^*$ at $t = 0$. It is easy to see by repeating the same analysis that sending the MWC of the resulting G-IDNC graph at $t = 1$ will bring the system even closer to completion. Consequently, we can infer from the previous discussion that the use of the G-IDNC graph all along the recovery phase will bring the system to completion at least as fast as the use of S-IDNC graph.

V. PROPOSED ALGORITHM

A. ONE-RN SCENARIO

The algorithm of the recovery transmission phase for the one-RN case is performed on two steps as follows:

**Step 1**: The BS carries out the re-transmission process to the TNs and the RN as follows:

1) The weight of the vertices of each TN $i$ is set to be $\left(\frac{|W_i|}{1-p_{(BS,i)}}\right)^n$, where $p_{(BS,i)}$ is the erasure probability from the BS to TN $i$. Note that the RN has an empty Wants set and all its vertices are in the secondary graph. We thus assign a very low weight to its vertices in the secondary graph, which makes their selection the last priority. This is similar to the assumption used in [7], except that we do care about RN decodability when possible.

2) Select the primary MWC $\kappa_p$ in the IDNC primary graph. By doing so, the TNs that have the largest Wants sets and erasure probability are targeted.
3) Extract the secondary subgraph \( G(\kappa_{\rho}) \) consisting of secondary vertices each of which being adjacent to all the vertices in \( \kappa_{\rho} \).
4) Select the MWC \( \kappa_\sigma \) in \( G(\kappa_{\rho}) \). By doing so, the TNs that have the largest Wants sets but lower erasure probability of receiving a packet are targeted with unwanted information. If possible, the RN is also targeted with extra packets.
5) Send the packet corresponding to the combined clique \( \kappa^* = \kappa_{\rho} \cup \kappa_\sigma \).
6) Update the feedback matrix according to the received feedback and repeat this step again until either of these conditions occur:

- The Wants sets of all TNs are depleted, which means that they received all their wanted packets.
- The RN receives all the packets that are still in the Wants set of any of the TNs.

In the former condition, the BS declares frame delivery and start the transmission of subsequent frames, if any. In the latter, the system moves to Step 2.

**Step 2:** The RN continues the recovery process by following the same procedures of Step 1, with the following two changes:

- There are no vertices belonging to the RN in the graph.
- The weights of the vertices of each TN \( i \) become \( \left( \frac{|W_i|}{1-p(h_{R(i,i)})} \right)^n \), where \( p(h_{R(i,i)}) \) is the erasure probability from the RN to TN \( i \).

This step is repeated until all TNs receive their wanted packets.

**B. Multiple-RN Scenario**

In the multiple-RN scenario, the algorithm should perform a selection of both the transmitting RN and the clique determining the coded transmission, in order to achieve the least possible completion delay in the whole network. Thus, the algorithm operates as follows:

**Step 1:** The BS sends NC recovery packets to the TNs and the RNs, similar to the one-RN case described above. Nonetheless, the algorithm moves to Step 2 when:

\[
\bigcup_{i \in M} W_i \subseteq \bigcup_{h \in \mathcal{R}} \mathcal{H}_h. \tag{1}
\]

In other words, the algorithm moves to Step 2 when each packet that is still needed by any receiver is received by at least one RN. This condition clearly shortens Step 1 compared to the one-RN scenario because of the diversity of the Has sets at the different RNs.

**Step 2:** All \( R \) RNs continue the recovery process together, under the management of the BS. For each recovery transmission, the chosen RN and clique for transmission are jointly selected as follows:

1) Each RN \( h \in \mathcal{R} \) separately selects its MWC \( \kappa^*(h) \) similar to the one-RN scenario, except that it employs the weight \( \left( \frac{|W_i|}{1-p(h_{R(i,i)})} \right)^n \) for the vertices of TN \( i \).
2) The BS selects the RN \( h^* = \arg \max_{h \in \mathcal{R}} \omega(h^*) \) (i.e. the RN having the MWC with the highest weight)

and gives permission to this RN to transmit the packet combination defined by \( \kappa^*(h^*) \) in this recovery transmission.

This step is repeated until all TNs receive their wanted packets. Note that Step 2 in this case may take longer than its one-RN version. Indeed, the condition in (1) makes it very unlikely that any of the \( R \) RNs will have all the packets still wanted by all the TNs by the end of Step 1. Thus, each of the \( R \) RNs will need to deliver its received subset of packets in separate transmissions, which may take longer than Step 2 in the one-RN case.

**C. Maximum Vertex Search Algorithms**

Since the MWC algorithm has high computational complexity, we propose a modification to the above algorithm by replacing the MWC algorithm with the maximum vertex search (MVS) algorithm, proposed in [12], [13], and having complexity of \( O(M + R)^2 N \). This algorithm constructs the transmission clique by greedily choosing in each step the vertex that has both the largest weight and is connected to the largest number of vertices having high weights. The details of this algorithm can be found in [12], [13].

**VI. Simulation Results**

**A. G-IDNC vs S-IDNC**

For the one-RN and three-RN scenarios, Fig. 3 depicts the average completion delay against the number of TNs \( M \) for G-IDNC and S-IDNC schemes. For each of the two schemes, the WoRLT strategy is employed for vertex weighting and both MWC and MVS are tested. The number of packets used in the simulation is \( N = 30 \) and each TN requires on average 80% of these packets. The packet erasure probabilities on the BS-TN, BS-RN and RN-TN channels are randomly selected from the ranges \([0.3,0.5]\), \([0.1,0.2]\) and \([0.05,0.15]\), respectively.
For both the one-RN and three-RN scenarios, it is obvious that the performance of G-IDNC is superior to S-IDNC for both the MWC and MVS algorithms. Similar results have been obtained for all ranges of $N$ and TN average demand but weren’t included due to space limitation. Furthermore, we can see that the three-RN case achieves a significant improvement compared to the one-RN case. This means that, compared to the one-RN algorithm, the shortening that occurs in Step 1 of the three-RN case, due to the RN reception diversity, has a much bigger impact than the lengthening of Step 2 due to the separate RN transmissions (resulting from the incomplete Has sets of individual RNs at the end of Step 1). Moreover, the results show that the MVS algorithm performance gets closer to the one of MWC algorithm at larger numbers of RNs.

**B. Joint Clique-RN Selection for G-IDNC**

For the G-IDNC scheme and $R = 3$, Fig. 4 compares the average completion delay achieved by the WoRLT joint clique-RN selection strategy with the following joint approaches:

**Approach 1**: Choosing the clique-RN combination that maximizes the service of the maximum clique. To find the maximum clique in each RN, the weight of each vertex is set to $1$.

**Approach 2**: Choosing the clique-RN combination that serves the maximum expected number of TNs. To find the maximum expected number of TNs that can be served by each RN, the weight of each vertex $v_{ij}$ is set to $(1 - p_{h,i})$.

**Approach 3**: Choosing the clique-RN combination that maximizes the service of the most wanted packets. The weight of each vertex $v_{ij}$ is set to be $P_j$, defined as the number of TNs wanting packet $j$.

After finding $\kappa^*(h)$ for any recovery transmission using any of the above approaches, the RN $h^*$ having $h^* = \arg\max_{h \in R} \left\{ \sum_{i,j \in \kappa^*_h} (1 - p_{h,i}) \right\}$ is selected by the BS to send the packet combination defined by $\kappa^*(h^*)$ in this recovery transmission. The BS thus selects the RN that has higher chance of delivering its selected clique.

It is obvious that WoRLT based joint clique-RN scheduling approach achieves a significant improvement in the average completion delay over all these different approaches that are common in the PMP NC literature.

**VII. CONCLUSION**

In this paper, we investigated the problem of minimizing the frame completion delay for instantly decodable network coding in RN-assisted wireless multicast networks. To achieve a lower completion delay, we first proposed the use of the the generalized version of the IDNC scheme to the one-RN scenario instead of the previously proposed strict IDNC version in the literature. We supported this exchange by showing that S-IDNC provides coding combinations that are always a subset of the G-IDNC combinations for the same feedback matrix. We then designed G-IDNC based WoRLT algorithm for the one-RNs scenario, and extended it to the multiple-RN scenario. Simulation results showed that our proposed solutions achieve much better performance compared to the S-IDNC solutions in the literature. They also showed that the use of the WoRLT strategy achieves a significantly better completion delay compared to other commonly used strategies in the PMP NC literature.

**REFERENCES**

[1] R. Ahlswede, Cai N., S.-Y.R. Li, and R.W. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] S. Katti, D. Katabi, W. Hu, H. S. Rahul, and M. Médard, “The importance of being opportunistic: practical network coding for wireless environment,” in *Proc. Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2005.

[3] J-S. Park, M. Gerla, D.S. Lun, Yunjung Yi, and M. Médard, “Codecast: a network-coding-based ad hoc multicast protocol,” *IEEE Wireless Communications*, vol. 13, no. 5, pp. 76–81, Oct. 2006.

[4] L. Keller, E. Drinea, and C. Fragouli, “Online broadcasting with network coding,” *Fourth Workshop on Network Coding, Theory and Applications (NetCod’08)*, Jan. 2008.

[5] P. Sadeghi, R. Shams, and D. Traskov, “An optimal adaptive network coding scheme for minimizing decoding delay in broadcast erasure channels,” *EURASIP Journal of Wireless Communications and Networking*, vol. 2010, pp. 1–14, Apr. 2010.

[6] P. Fan, C. Zhi, C. Wei, and K. Ben Letaief, “Reliable relay assisted wireless multicast using network coding,” *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 5, pp. 749–762, Jun. 2009.

[7] L. Lu, M. Xiao, and L.K. Rasmussen, “Design and analysis of relay-aided broadcast using binary network codes,” *Journal of Communications*, vol. 6, no. 8, pp. 610–617, Nov. 2011.

[8] S. Parkvall, E. Dahlman, A. Furuskar, Y. Jading, M. Olsson, S. Wanstedt, and K. Zangi, “LTE-advanced - evolving LTE towards IMT-advanced,” in *IEEE Vehicular Technology Conference (VTC’08)*, Sep. 2008.

[9] Y.Yang, H. Hu, J. Xu, and G. Mao, “Relay technologies for wimax and LTE-advanced mobile systems,” *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100–105, Oct. 2009.

[10] E. Dahlman, S. Parkvall, and Sköld J., “4G: LTE-Advanced for Mobile Broadband,” Academic Press, Waltham, MA, May 2011.

[11] S. Sorour and S. Valaee, “Minimum broadcast decoding delay for generalized instantly decodable network coding,” *IEEE Global Communications Conference (GLOBECOM 2010)*, pp. 1–5, 2010.

[12] S. Sorour and S. Valaee, “On minimizing broadcast completion delay for instantly decodable network coding,” in *Proc. IEEE International Conference on Communications (ICC’10)*, Cape Town, May 2010.

[13] S. Sorour and S. Valaee, “Completion delay minimization for instantly decodable network codes,” submitted to IEEE/ACM Transactions on Networking, http://arxiv.org/pdf/1201.4768v1.pdf.