Rapprochement between Instantly Decodable and Random Linear Network Coding

Mingchao Yu Neda Aboutorab Parastoo Sadeghi
Research School of Engineering, The Australian National University, Canberra, Australia
Email: {ming.yu, neda.aboutorab, parastoo.sadeghi}@anu.edu.au

***Abstract*—In this paper, a new network coding model is proposed to unify instantly decodable network coding (IDNC) and random linear network coding (RLNC), which have been considered to be incompatible in the literature. This model is based on a novel definition of generation, which is built upon optimal IDNC solutions. Under this model, IDNC and RLNC are only two extreme cases with specific generation sizes. Throughput and delay properties of this model, measured by block completion time and packet decoding delay, respectively, are studied, which fill the gap between IDNC and RLNC and thus provide a good understanding on the throughput-delay tradeoff of network coding. An efficient adaptive scheme is then designed, which allows in-block switch among IDNC and different levels of RLNC, so that the system's throughput and delay can be fine-tuned to meet the real-time requirements of the application. Extensive simulations are performed to demonstrate how the proposed generation size interacts with the number of receivers and the channel quality to affect the overall system performance.**

***Index Terms*—network coding, wireless broadcast, throughput, decoding delay**

I. Introduction

Network coding allows senders or intermediate nodes of a network to mix different data packets/flows and can enhance the throughput of many network setups [1], [2]. But this is often at the price of large decoding delay, because mixed data needs to be network decoded before delivery to the application layer [3], [4]. Understanding the tradeoff between the throughput and decoding delay in network coded systems has been the subject of research in recent years [3]–[6], where a packet-level network coding model is particularly suitable for such studies. In this paper, we consider two classic packet-level network coding techniques in wireless broadcast scenario: random linear network coding (RLNC) [7] and instantly decodable network coding (IDNC) [8]–[10]. We study their coding models and investigate the relationship between their throughput and packet decoding delay in a unified framework.

RLNC is a block-wise network coding technique, where a block, or a generation, by definition, is a set of K_T consecutive data packets. Linear combinations of all the K_T packets are sent with random coefficients from a finite field $\mathbb{F}_q$ where $q > 2$. The primary advantage of RLNC is its optimality in terms of block completion time [4], which is a measure of throughput. Another advantage is that it only requires one ACK feedback from each receiver upon successful decoding of all the K_T data packets. However, RLNC can suffer from large decoding delays since packet delivery to the application cannot generally occur until block-wise network decoding is completed.

This work was supported under Australian Research Council Discovery Projects funding scheme (project no. DP120100160).

On the contrary, IDNC is a technique aiming at minimizing packet decoding delay. By carefully choosing data packets to be coded together, it guarantees a subset (or if possible, all) of the receivers to instantly decode one of their missing date packets upon successful reception of a coded packet [8], [10]. This means that data packets can be potentially delivered to the application much faster than RLNC. Another advantage of IDNC is its simple XOR-based encoding and decoding, i.e., coding coefficients are chosen from $\mathbb{F}_2$. However, IDNC is generally not optimal in terms of block completion time, as there might exist a subset of receivers who do not receive any of their missing data packets in a transmission. Moreover, it requires feedback from all the receivers after every transmission for making coding decisions.

Obviously, the performance of IDNC and RLNC are opposite to each other in many ways, including their throughput, decoding delay, and feedback frequency [10]. Moreover, since RLNC uses a larger finite field than IDNC, the adaptive choice between IDNC and RLNC is impossible until the broadcast of a whole block of data packets is completed.

These contrasts raise the following question, which is the first out of the three questions that this work tries to answer:

1) Is there a coding scheme which provides more balanced throughput and decoding delay than IDNC and RLNC?

For example, is it possible to reduce the decoding delay of RLNC without sacrificing (or with minimum sacrifice on) the throughput? In Section II-D, we will show an example in which, by carefully partitioning a block of data packets into *sub-generations*, the decoding delay of RLNC can be reduced without sacrificing the throughput. However, as will be proven in Section III, packet partitioning following the classic definition of generation in general yields a prohibitively large block completion time. Then, the following two questions arise:

2) Is there a better definition of the sub-generations, based on which better packet partitioning can be achieved? And,
3) Is it possible to make a rapprochement between IDNC and RLNC under this definition?

These questions are answered by introducing a new definition of the sub-generations based on the optimal IDNC solutions recently found in [10]. A new coding model is then proposed, in which a block of data packets is partitioned into several such sub-generations. It successfully unifies IDNC and RLNC, because they become two extreme cases of this model with specific sub-generation sizes. Analytical and numerical studies on the throughput and decoding delay properties of this model provide a good understanding on how the throughput-delay tradeoff is affected by the proposed sub-generation size. Practically, this model allows in-block switch among IDNC and different levels of RLNC so that the system performance can be fine-tuned to meet real-time requirements of the application, including throughput, delay, and feedback frequency.

II. System Model

A. Transmission Setup

We consider wireless broadcast of K_T data packets, denoted by $\mathbf{p}_1, \cdots, \mathbf{p}_{K_T}$, from a sender to N_T receivers, denoted by $R_1, \cdots, R_{N_T}$. Time is slotted. In each time slot, a packet (either original data or coded) is sent. Wireless channels between the sender and the receivers are subjected to i.i.d. memoryless erasures with a probability of P_e. In this scenario, K_T data packets are first sent uncoded once using K_T time slots. This phase is known as the *systematic transmission phase*. [1] After this phase, a receiver might miss a number of data packets. The sender then collects feedbacks from all the receivers on their receiving states of the data packets. Feedback information can be expressed by an $N \times K$ binary state feedback matrix (SFM) [8], denoted by $\boldsymbol{A} = [a_{n,k}]$, where $a_{n,k} = 1$ means receiver R_n still wants data packet $\mathbf{p}_k$ and $a_{n,k} = 0$ otherwise. Here N is the number of receivers who have not received all the K_T data packets, and K is the number of data packets that have not been received by all the N_T receivers. The set of these K data packets is denoted by $\mathcal{P}_K$. An example of a 3×8 SFM is given in Fig. 1(a). The subset of $\mathcal{P}_K$ wanted by receiver R_n is called the *Wants* set of this receiver, denoted by $\mathcal{W}_n$. Its size is denoted by W_n and the largest W_n across all the receivers is denoted by $W_{\max}$. The subset of receivers that want $\mathbf{p}_k$ is called the *Target* set of $\mathbf{p}_k$, denoted by $\mathcal{T}_k$. Its size is denoted by T_k.

A *coded transmission phase* then starts. We are interested in the *minimum* number of coded transmissions and *minimum* average packet decoding delay needed to complete the broadcast of $\mathcal{P}_K$, denoted by U and D, respectively. The term "minimum" implies the best achievable performance, which occurs when channels are erasure-free. U is a measure of throughput, since throughput can be calculated as $K_T/(K_T+U)$. Denote by $d_{n,k}$ the *first* time slot in this phase when $\mathbf{p}_k$ could be decoded by R_n and let $d_{n,k} = 0$ if $a_{n,k} = 0$, then:

$$D = \frac{1}{\sum_{k=1}^{K} T_k} \sum_{n=1}^{N} \sum_{k=1}^{K} d_{n,k} \tag{1}$$

B. Coded Transmissions Using IDNC

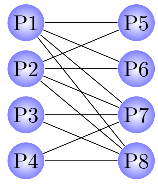
An IDNC coded packet is denoted by $\mathcal{M}$, which also represents a coding set, i.e., the subset of $\mathcal{P}_K$ to be XOR-ed together in this coded packet. It is instantly decodable if it satisfies the following *IDNC constraint* [8]:

***Definition** 1: An IDNC coding set $\mathcal{M}$ contains at most one data packet from the Wants set, $\mathcal{W}_n$, of any receiver R_n.*

[1]For IDNC, coding is unavailable at the beginning. For RLNC, it has been shown in [11], [12] that systematic RLNC, while offering better decoding delay, does not incur any throughput penalty compared with traditional RLNC.

	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{p}_4$	$\mathbf{p}_5$	$\mathbf{p}_6$	$\mathbf{p}_7$	$\mathbf{p}_8$
R_1	1	1	1	1	0	0	0	0
R_2	0	0	0	0	1	1	1	1
R_3	0	0	1	1	1	1	0	0

(a) SFM $\boldsymbol{A}$



(b) Graph $\mathcal{G}$

Fig. 1. An example of state feedback matrix and its graph representation,

According to this constraint, two data packets that are jointly wanted by at least one receiver cannot be included in the same coded packet. Such two data packets are defined as *conflicting* with each other [10]. On the other hand, two data packets do not conflict with each other if they are not jointly wanted by any receiver. Conflict states among all the K data packets can be represented by an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where vertex $v_k \in \mathcal{V}$ represents packet $\mathbf{p}_k$, and edge $e_{i,j} \in \mathcal{E}$ exists if $\mathbf{p}_i$ conflicts with $\mathbf{p}_j$ [10].

Because data packets in the same coding set $\mathcal{M}$ do not conflict with each other, the vertices representing these packets are connected to each other in $\mathcal{G}$. These vertices thus form a clique [13] in $\mathcal{G}$. Furthermore, a coding set $\mathcal{M}$ is said to be *maximal* if its corresponding clique is maximal, i.e., it is not a subset of a larger clique. Then, IDNC coding problem can be converted to the minimum clique cover problem in the graph theory [10] [13]. [10] proved that there exists a minimum number of coded transmissions using IDNC, denoted by U_{IDNC}. A set of U_{IDNC} maximal coding sets which jointly cover all the K data packets is called an *optimal IDNC solution*, denoted by $\mathcal{S} = \{\mathcal{M}_1, \cdots, \mathcal{M}_{U_{\text{IDNC}}}\}$.

The graph representation of the SFM in Figure 1(a) is shown in Fig. 1(b), for which $U_{\text{IDNC}} = 4$ and the optimal solution is $\mathcal{S} = \{\{\mathbf{p}_1, \mathbf{p}_5\}, \{\mathbf{p}_2, \mathbf{p}_6\}, \{\mathbf{p}_3, \mathbf{p}_7\}, \{\mathbf{p}_4, \mathbf{p}_8\}\}$, yielding a minimum average packet decoding delay of $D = 2.5$ according to (1). One could easily verify that every coding set in $\mathcal{S}$ is maximal.

C. Coded Transmissions Using RLNC

Coded packets in a typical RLNC scheme are linear combinations of *all* the K data packets. Coding coefficients are chosen from a finite field $\mathbb{F}_q$ and $q > 2$. In order to decode W_n data packets, R_n needs to receive W_n linearly independent coded packets. Hence, to satisfy the demand of all the N receivers, at least $W_{\max}$ linearly independent coded packets need to be broadcast. The minimum number of coded transmissions using typical RLNC is thus equal to $W_{\max}$ and is denoted by U_{RLNC}. On the receiver side, feedback is not required except an ACK from every receiver when it decodes all its missing packets.

In the next subsection, we will calculate the minimum number of coded transmissions and average packet decoding delay of the SFM in Figure 1(a) using typical RLNC, and then show that how *packet partitioning* can reduce the average packet decoding delay without sacrificing the the minimum number of coded transmissions.

D. An Example of Packet Partitioning

Consider the SFM given in Fig. 1(a). If a typical RLNC is applied, $U_{\text{RLNC}} = 4$ because $W_{\max} = 4$. The minimum average packet decoding delay is $D = 4$. Alternatively, the 8 data packets could be partitioned into two *sub-generations*: the first sub-generation is $\boldsymbol{G}_1 = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_5, \mathbf{p}_6\}$ and the second sub-generation is $\boldsymbol{G}_2 = \{\mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_7, \mathbf{p}_8\}$. RLNC is then applied to them separately. $\boldsymbol{G}_1$ requires a minimum of two coded transmissions because all the receivers want at most two data packets in it. Similarly, $\boldsymbol{G}_2$ also requires a minimum of two coded transmissions. In total, a minimum of $U = 4$ coded transmissions are required, which is the same as the typical RLNC. However, data packets in $\boldsymbol{G}_1$ can be decoded after only 2 coded transmissions and thus D is reduced to 3.

III. Partitioning Effect on Throughput and Delay

As can be seen in Section II-D, by carefully partitioning the K data packets into two sub-generations and applying RLNC to them separately, the decoding delay of RLNC is reduced without increasing the minimum number of coded transmissions. The term "sub-generations" specifically denotes the smaller generations after partitioning. A sub-generation, by classic definition of generation [3], is a set of g consecutive data packets, where g is called the sub-generation size and ranges from 1 to K here. According to this definition, sub-generations are generated by evenly partitioning K data packets into $M = \lceil K/g \rceil$ segments. Here we show that such partitioning is inefficient because it might result in a prohibitively large number of coded transmissions.

Denote by U_g the minimum number of coded transmissions under sub-generation size g, by $\boldsymbol{G}_m$ the m-th sub-generation, and let $W_{\max}^m$ be the largest number of data packets in $\boldsymbol{G}_m$ wanted by any one receiver across all receivers. Then U_g is:

$$U_g = \sum_{m=1}^{M} W_{\max}^m \tag{2}$$

It has the following property:

***Lemma** 1: If data packets are partitioned into classic sub-generations, U_g is lower bounded as:*

$$U_g \geqslant \max(W_{\max}, M) \tag{3}$$

Proof: $U_g \geqslant W_{\max}$ is obvious. Furthermore, because $W_{\max}^m \geqslant 1$ for $m \in [1, M]$, $U_g \geqslant M$ according to (2). ■

Consequently, if small values of g (e.g., $g = 1$ or 2) are applied to reduce the decoding delay, the resulted U_g could be much greater than $U_{\text{RLNC}} = W_{\max}$ and even greater than U_{IDNC}, making such partitions pointless. Hence, a new definition is needed here for the sub-generations.

A. The Proposed Definition of Generation and Coding Model

We propose a new definition of sub-generation as follows:

***Definition** 2: A sub-generation is a set of maximal coding sets in the optimal IDNC solution, and is denoted by $\boldsymbol{G}$.*

The definition of sub-generation size is changed accordingly:

***Definition** 3: Sub-generation size is the number of maximal coding sets in a sub-generation, and is denoted by g.*

According to the above definitions, the number of sub-generations becomes $M = \lceil U_{\text{IDNC}}/g \rceil$, where $g \in [1, U_{\text{IDNC}}]$. The definition of $W_{\max}^m$ of a sub-generation $\boldsymbol{G}_m$, where $m \in [1, M]$, is still the largest number of data packets in $\boldsymbol{G}_m$ wanted by any one receiver across all receivers.

We then establish our proposed coding model as follows. Given the optimal IDNC solution $\mathcal{S} = \{\mathcal{M}_1, \cdots, \mathcal{M}_{U_{\text{IDNC}}}\}$, the first sub-generation is $\boldsymbol{G}_1 = \{\mathcal{M}_1, \cdots, \mathcal{M}_g\}$, the second sub-generation is $\boldsymbol{G}_2 = \{\mathcal{M}_{g+1}, \cdots, \mathcal{M}_{2g}\}$, etc. When $g \geqslant 2$, RLNC with a field size $q > 2$ is applied to each sub-generation separately. When $g = 1$, $q = 2$ is sufficient since all the receivers want at most one data packet from a sub-generation. For all the values of g, $\boldsymbol{G}_m$ always needs a minimum of $W_{\max}^m$ coded transmissions. The relationship between U_g and $W_{\max}^m$ is thus the same as in (2).

The proposed model can be demonstrated by considering the SFM in Fig. 1(a), whose optimal IDNC solution is $\mathcal{S} = \{\{\mathbf{p}_1, \mathbf{p}_5\}, \{\mathbf{p}_2, \mathbf{p}_6\}, \{\mathbf{p}_3, \mathbf{p}_7\}, \{\mathbf{p}_4, \mathbf{p}_8\}\}$. If $g = 2$, there will be $U_{\text{IDNC}}/g = 2$ sub-generations, where $\boldsymbol{G}_1 = \{\{\mathbf{p}_1, \mathbf{p}_5\}, \{\mathbf{p}_2, \mathbf{p}_6\}\}$ and $\boldsymbol{G}_2 = \{\{\mathbf{p}_3, \mathbf{p}_7\}, \{\mathbf{p}_4, \mathbf{p}_8\}\}$. Four data packets in $\boldsymbol{G}_1$ are RLNC coded and sent using $W_{\max}^1 = 2$ transmissions, so are those in $\boldsymbol{G}_2$ using $W_{\max}^2 = 2$ transmissions. Consequently, $U_2 = 4$ and $D = 3$.

Interestingly, there are two extreme cases of this coding model, taking place when $g = 1$ and $g = U_{\text{IDNC}}$. (1) When $g = 1$, since $q = 2$, coding within a sub-generation is simplified to XOR and there are U_{IDNC} such sub-generations. Hence, the system becomes IDNC with $U_1 = U_{\text{IDNC}}$. (2) When $g = U_{\text{IDNC}}$, there will be only one sub-generation, which contains all the maximal coding sets and thus all the K data packets. Hence, the system becomes the typical RLNC with $U_{U_{\text{IDNC}}} = U_{\text{RLNC}}$. We conclude that the proposed coding model unifies IDNC and RLNC and thus is a general coding model.

We then study the throughput and decoding delay properties of this model, which further fills the gap between IDNC and RLNC, and sheds light on their throughput-delay tradeoff.

B. Throughput and Delay Properties

The throughput is measured by U_g, which, according to (2), is determined by $W_{\max}^m$. $W_{\max}^m$ has the following property:

***Lemma** 2: When $g \geqslant 2$, $W_{\max}^m \in [2, g]$ for all $m \in [1, M]$.*

Proof: 1) If all the receivers want at most one data packet in $\boldsymbol{G}_m$, then all the data packets in $\boldsymbol{G}_m$ together form a valid coding set. All the g coding sets in $\boldsymbol{G}_m$ are subsets of this coding set, which contradicts the fact that they are maximal already. Thus, $W_{\max}^m \geqslant 2$; 2) If there is a receiver who wants $W_{\max}^m > g$ data packets in $\boldsymbol{G}_m$, at least two of these $W_{\max}^m$ packets must belong to the same maximal coding set, which contradicts the IDNC constraint. Thus, $W_{\max}^m \leqslant g$. ■

Then, by considering the relationship between $W_{\max}^m$ and U_g in (2) and noting that $M = \lceil U_{\text{IDNC}}/g \rceil$, the above lemma yields an important corollary:

***Corollary** 1: For all the values of g, the minimum number of coded transmissions is bounded between U_{RLNC} and U_{IDNC}:*

$$\forall g: \ U_g \in [U_{\text{RLNC}}, U_{\text{IDNC}}] \tag{4}$$

where $U_g = U_{RLNC}$ holds when $g = U_{IDNC}$ and $U_g = U_{IDNC}$

holds when $g = 1, 2$ (because $W^m_{\max} = 2$ when $g = 2$).

Hence, in this coding model, U_g approaches U_{RLNC} when g increases gradually from 1 to U_{IDNC} and approaches U_{IDNC} the other way around. Compared with the trivial lower bound, $\max(W_{\max}, \lceil K/g \rceil)$, derived using the classis definition of generation, this result is desired because we should never sacrifice U_g to more than U_{IDNC} to gain better decoding delay.

Denote the minimum average packet decoding delay under g by D_g. Similar to the case of U_g, D_g lies between the minimum average decoding delay of using IDNC (D_1) and typical RLNC ($D_{U_{\text{IDNC}}}$). It has the following property:

***Lemma** 3:*

$$D_g \leqslant \frac{g + K}{2} \tag{5}$$

Proof: Denote the largest decoding delay of data packets in $\boldsymbol{G}_m$ by $D_g(m)$. It is equal to $\sum_{i=1}^{m} W^i_{\max}$. When $W^i_{\max}$ is maximized, i.e., when $W^i_{\max} = g$ for all $i \in [1, m]$, $D_g(m)$ is maximized with a value of mg. Since we always send $\boldsymbol{G}_m$ with more target receivers first, the largest D_g happens when all the $\boldsymbol{G}_m$ have the same number of target receivers, denoted by T. Then, as a variation of (1), this D_g is calculated as:

$$D_g = \frac{1}{MT} \sum_{m=1}^{M} Tmg = \frac{g(1+M)}{2} = \frac{g + U_{\text{IDNC}}}{2} \tag{6}$$

Since the largest U_{IDNC} is K, (5) holds. The equality takes place, e.g., when the SFM is an all-one matrix. ■

From the above lemma, one may make an intuitive deduction that the smaller the g the smaller the average packet decoding delay. However, this is not true when U_{IDNC} is much larger than U_{RLNC}. In this case, IDNC actually requires higher average packet decoding delay than typical RLNC, i.e., $D_1 > D_{U_{\text{IDNC}}}$. This is due to the fact that if IDNC is applied, data packets sent in the last few transmissions (up to U_{IDNC}) will encounter large decoding delays, which counteracts the decoding delay reduction from RLNC in earlier transmissions. Since D_g approaches D_1 when g approaches one, smaller g implies larger decoding delay. This phenomenon, as well as the upper bound in (5), will be numerically confirmed later.

C. Application: Adaptive Network Coding System

After the proposed packet partitioning, coding/decoding of data packets in different sub-generations become independent. This property enables a simple adaptive scheme where in-block switch among IDNC and different levels of RLNC becomes possible:

1) Apply systematic transmission phase to a block of packets. Collect feedbacks and find the optimal IDNC solution;
2) Calculate the throughput and decoding delay under different g using (2) and (1), respectively, and then choose the g which meets the system requirements;
3) Linear combinations (XOR when $g = 1$) of the data packets in the first sub-generation are sent until all its target receivers have sent an ACK upon decoding;
4) If the system throughput and decoding delay requirements are changed, g is adapted accordingly, and then the sub-generations are modified. Otherwise, the last step is repeated for the subsequent sub-generations.

Compared with traditional adaptive schemes which can only make adaptive decisions after the completion of a whole block of K_T data packets, the proposed system brings much more flexibility without any extra system modification, since data packets not included in a sub-generation can be treated as having zero coefficients. This system also makes feedback frequency adaptive: larger g implies less frequent feedback.

***Remark** 1: Finding optimal IDNC solutions is NP-hard. Our proposed concept of sub-generation and subsequent results apply even when U_{IDNC} and the corresponding coding sets are heuristically found (see [10] and the references therein).*

IV. Simulation Results

In this section, we investigate via simulations how the sub-generation size g interacts with the number of receivers N_T and the channel erasure probability P_e to affect the throughput and delay performance, measured by the minimum number of coded transmissions U and the minimum average packet decoding delay D, respectively. Since individually smaller U and D mean better performance, the throughput-delay tradeoff can measured by their product, denoted by $J = UD$. Lower J means better joint throughput-delay performance.

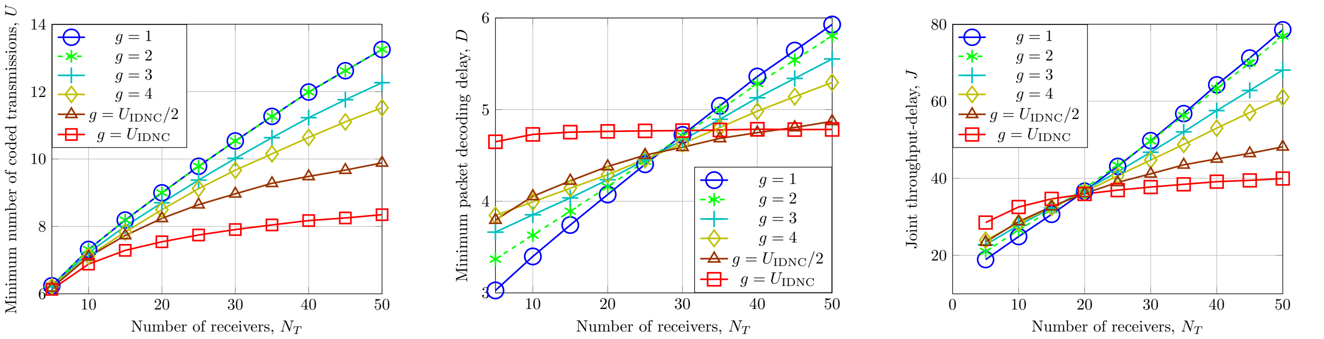
We simulate the broadcast of $K_T = 20$ data packets. After the systematic transmission phase, feedbacks are collected and then an instance of SFM is generated. Its U, D and J under $g = 1$ (i.e., IDNC), $2 \sim 4$ (if applicable), $U_{\text{IDNC}}/2$, and U_{IDNC} (i.e., RLNC) are calculated. Instances of U, D and J are then averaged, respectively. Two simulations are carried out.

In the first simulation, P_e is 0.2 and $N_T \in [5, 50]$. Results are shown in Fig. 2, from which we observe that:

1) The throughput performance under larger g is better than those under smaller g, and they are well bounded between U_{RLNC} and U_{IDNC}. This result matches Corollary 1.
2) There is no clear winner g for the delay performance across all N_T. When $N_T > 30$, U_{IDNC} is much larger than U_{RLNC}. Accordingly, the delay performance under larger g overtakes those under smaller g when $N_T > 30$. This phenomenon matches the discussion in Section III-B.
3) Due to the variation in delay performance leadership, joint throughput-delay performance, J, under smaller g is better when $N_T < K_T$ but is worse when $N_T > K_T$;
4) Overall system performance under larger g is more robust to increasing N_T than those under smaller g, since increasing number of receivers could significantly limit IDNC coding opportunities and thus increase U_{IDNC} [10].

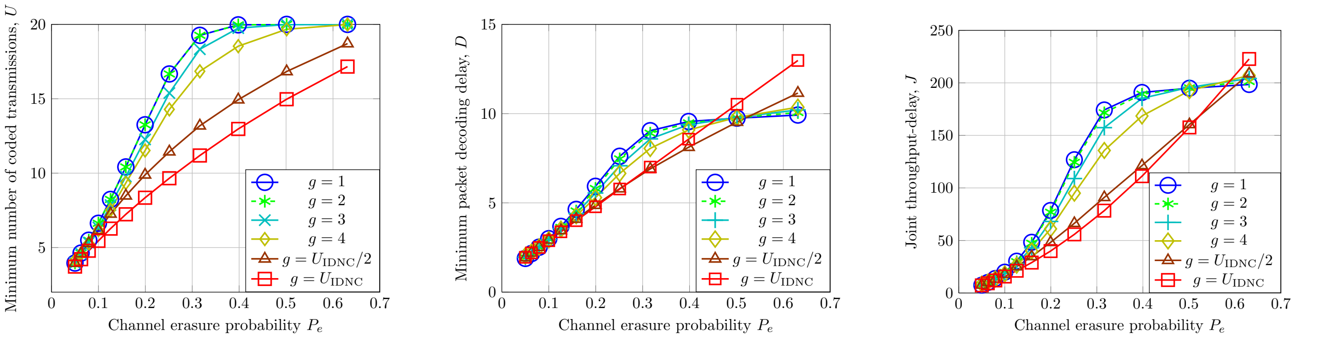
In the second simulation, a large N_T of 50 is applied to observe the decoding delay upper bound in (5). $P_e \in [0.05, 0.65]$. Results are shown in Fig. 3, from which we observe that:

1) Throughput profile is the same as in the first simulation;
2) Because U_{IDNC} is much larger than U_{RLNC} when $P_e > 0.2$, the decoding delay under larger g is expected to overtake those under smaller g when $P_e > 0.2$, which is the case in Fig. 3(b) for $P_e \in [0.2, 0.5]$;
3) Nevertheless, when P_e further increases, U_{IDNC} is almost equal to K_T, so the decoding delay starts to approach its upper bound. For example, when $g = 1$, decoding delay



(a) Throughput (b) Decoding delay (c) Joint throughput-delay performance

Fig. 2. Simulation 1: system performance with $P_e = 0.2$ and different N_T



(a) Throughput (b) Decoding delay (c) Joint throughput-delay performance

Fig. 3. Simulation 2: system performance with $N_T = 50$ and different P_e

is bounded by $(1 + 20)/2 = 10.5$. In this case, smaller g means smaller upper bound and thus better performance.

4) The profile of joint throughput-delay performance follows the profile of delay performance: performance under larger g is better when P_e is medium, while performance under smaller g is better when P_e is either small or large.
5) Increasing P_e significantly degrades the overall system performance regardless of the sub-generation size.

V. Conclusion and Future Work

In this paper, we redefined the concept of generation and proposed a new coding model which unifies IDNC and RLNC. Under this rapprochement, IDNC and RLNC are two extreme cases and there are coding schemes in between them which offer more balanced throughput and delay performance. The proposed coding model thus provided a good understanding on the throughput-delay tradeoff of packet-level network coding and enabled an adaptive network coding scheme allowing efficient in-block performance control. In the future, we will investigate that if the throughput and delay could be further improved by algorithmic partitioning of coding sets to sub-generations. Deriving closed-form relationship between sub-generation size and throughput/delay is also of interest.

References

[1] R. Ahlswede, N. Cai, S. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[3] C. Fragouli, D. Lun, M. Médard, and P. Pakzad, "On feedback for network coding," in *Proc. Annual Conf. on Inf. Sci. and Syst. (CISS)*, 2007, pp. 248–252.

[4] A. Eryilmaz, A. Ozdaglar, M. Médard, and E. Ahmed, "On the delay and throughput gains of coding in unreliable networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5511–5524, Dec. 2008.

[5] R. Costa, D. Munaretto, J. Widmer, and J. Barros, "Informed network coding for minimum decoding delay," in *Proc. IEEE Int. Conf. on Mobile Ad Hoc and Sensor Syst., (MASS)*, 2008, pp. 80–91.

[6] B. Swapna, A. Eryilmaz, and N. Shroff, "Throughput-delay analysis of random linear network coding for wireless broadcasting," in *Proc. Int. Symposium on Network Coding (NETCOD)*, Jun. 2010, pp. 1–6.

[7] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.

[8] P. Sadeghi, R. Shams, and D. Traskov, "An optimal adaptive network coding scheme for minimizing decoding delay in broadcast erasure channels," *EURASIP J. on Wireless Commun. and Netw.*, 2010.

[9] S. Sorour and S. Valaee, "On minimizing broadcast completion delay for instantly decodable network coding," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, May 2010, pp. 1–5.

[10] P. Sadeghi and M. Yu, "Instantly decodable versus random linear network coding: A comparative framework for throughput and decoding delay performance," *IEEE Trans. Inf. Theory*, (Submitted to), 2012. [Online]. Available: http://arxiv.org/abs/1208.2387

[11] L. Keller, E. Drinea, and C. Fragouli, "Online broadcasting with network coding," in *Proc. Workshop on Network Coding, Theory and Applications (NETCOD)*, 2008, pp. 1–6.

[12] J. Heide, M. V. Pederson, F. Fitzek, and T. Larsen, "Network coding for mobile devices - Systematic binary random rateless codes," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, 2009, pp. 1 – 6.

[13] J. M. Harris, J. L. Hirst, and M. J. Mossinghoff, *Combinatorics and Graph Theory, 2nd Edition*. Springer Press, 2008.