



**HAL**  
open science

## **MICN: a network coding protocol for ICN with multiple distinct interests per generation**

Hirah Malik, Cedric Adjih, Claudio Weidmann, Michel Kieffer

### ► **To cite this version:**

Hirah Malik, Cedric Adjih, Claudio Weidmann, Michel Kieffer. MICN: a network coding protocol for ICN with multiple distinct interests per generation. 2020. hal-02887550v1

**HAL Id: hal-02887550**

**<https://hal.science/hal-02887550v1>**

Preprint submitted on 2 Jul 2020 (v1), last revised 14 Jan 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MICN: a network coding protocol for ICN with multiple distinct interests per generation

H. Malik<sup>a,c,\*</sup>, C. Adjih<sup>a</sup>, C. Weidmann<sup>b</sup>, M. Kieffer<sup>c</sup>

<sup>a</sup>*Inria Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France*

<sup>b</sup>*ETIS UMR8051, CY University, ENSEA, CNRS, 95000 Cergy, France*

<sup>c</sup>*Université Paris-Saclay - CNRS - CentraleSupélec, Laboratoire des Signaux et Systèmes,  
3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France*

---

## Abstract

In Information-Centric Networking (ICN), consumers send interest packets to the network and receive data packets as a response to their request without taking care of the producers, which have provided the content, contrary to conventional IP networks. ICN supports the use of multiple paths; however, with multiple consumers and producers, coordination among the nodes is required to efficiently use the network resources. Network coding (NC) is a promising tool to address this issue. The challenge in the case of NC is to be able to get independent coded content in response to multiple parallel interests by one or several consumers. In this work, we propose a novel construction called MILIC (Multiple Interests for Linearly Independent Contents) that impose constraints on how the replies to interests are coded, intending to get linearly independent contents in response to multiple interests. Several protocol variants, called MICN (MILIC-ICN), built on top of NDN (Named Data Networking), are proposed to integrate these interest constraints and NC of data packets. Numerical analysis and simulations illustrate that the MILIC construction performs well and that the MICN protocols are close to optimal throughput on some scenarios. MICN protocols compare favorably to existing protocols, and show significant benefits

---

\*Corresponding author

*Email addresses:* [hirah.malik@inria.fr](mailto:hirah.malik@inria.fr) (H. Malik), [cedric.adjih@inria.fr](mailto:cedric.adjih@inria.fr) (C. Adjih), [claudio.weidmann@ensea.fr](mailto:claudio.weidmann@ensea.fr) (C. Weidmann), [michel.kieffer@l2s.centralesupelec.fr](mailto:michel.kieffer@l2s.centralesupelec.fr) (M. Kieffer)

when considering the total number of transmitted packets in the network, and in the case of high link loss rate.

*Keywords:* Information centric networking, ICN, Network coding, Named Data Networking

---

## 1. Introduction

Content distribution has become the primary task for today's Internet. According to CISCO's forecast, video traffic will be accounting for 79 percent of total mobile data traffic by 2022 [1]. The communication network's traditional paradigm has some drawbacks, especially when dealing with large-scale content distribution because of the point-to-point nature of communication and location dependence. The consumers, however, care about the content itself and not about its origin.

Information-Centric Networking (ICN) has recently been proposed as an alternative to the traditional point-to-point communication to make content the center of the communication network [2]. The ICN principle is based on receiving data through names by performing a named-based routing. It removes the need to establish a connection between endpoints and allows caching throughout the network. Named data networking (NDN) [3, 4] is one of the ICN architectures.

The basic NDN framework is a pull-based mechanism where the clients send interest packets that contain the name of the requested content. These interest packets are routed based on their names. A node holding a copy of the requested content replies to the interest with the content in a data packet.

Devices nowadays come with multiple network interfaces that can be used to retrieve content, e.g., WiFi, 3G/LTE. Traditional networking requires to establish a session among endpoints and hence does not allow simultaneous use of all available interfaces. NDN, however, enables the use of multiple interfaces. Nevertheless, in a dynamic network with multiple clients, some coordination is required to take advantage of multiple paths. Montpetit *et al.* introduced an

alternative to the coordination approach by utilizing network coding (NC) over NDN [5].

NC is a communication paradigm which, unlike traditional networking, allows the nodes to perform operations on the packets (computing algebraic combination of packets) [6, 7]. Decoding is performed by solving a linear system of equations once enough linearly independent combinations/packets are received. NC helps to exploit the network’s capacity, minimize delays and may help recover from link failures. Traditional routers that could only forward or replicate the packets are replaced by coding routers that can mix packets of the same content.

In this work, we aim to integrate more efficiently NC within NDN. For that purpose, special construction of the interests called MILIC (Multiple Interests for Linearly Independent Contents) is introduced that imposes some constraints on the content these interests bring. Several protocol variants, called MICN (MILIC-ICN), are then built on top of MILIC and NDN. MICN protocols allow parallel processing of multiple interests send by nodes and ensure linearly independent content with each of the multiple interests. Numerical analysis and simulations illustrate that the MILIC construction performs well and that the MICN protocols manage to get close to the optimal throughput on the considered scenarios. The performances obtained with MICN compare favorably to existing protocols and show significant benefits when considering the total number of transmitted packets in the network, and in the case of high link loss rate.

Section 2 summarizes some related work. Section 3 details the special construction of the interests MILIC. In Section 4, we detail the MICN protocol that uses MILIC construction to integrate NC over NDN. Section 4.7 introduces some optimizations to improve the performance of the protocol further. Section 5 presents the simulation setup and results. Section 6 concludes this paper and introduces future work.

## 2. Related Work

### 2.1. NDN

In this section, we briefly explain the basic concepts of the NDN architecture [3, 4]. Communication in NDN is *consumer-driven*, with two basic types of communication packets: *interest* and *data* packets. Consider an NDN network consisting of a set  $\mathcal{N}$  of nodes. Nodes can be *sources* that generate content, intermediate nodes or *caching routers*, or *clients* that request content. A node can have any of these roles at a given time. Each node  $r \in \mathcal{N}$  is connected in the network through a set of faces  $\mathcal{F}_r$ . The term *face* is a generalization of the interface that corresponds to various communication links.

The clients request the network to find the content by sending an NDN interest packet that carries the name of the requested content. The interest is forwarded in the network until it reaches a node holding a copy of the content with the requested name. The content is then sent back in a data packet. Both interest and data packets carry the name of the content but there is no information regarding the client or source.

Each NDN node has a *Pending Interest Table* (PIT), a *Forwarding Information Base* (FIB), and a *Content Store* (CS) for the transport of the named content in the network [2]. The PIT keeps a record of pending interests forwarded by the node that are not satisfied yet. Along with interest, the PIT stores the face where each interest arrives (in-faces) and the faces to which it was forwarded (out-faces) to record the reverse link for the data packet. The FIB stores routing information used to forward interest packets toward potential sources of matching content ; it can be populated by self-learning or by a routing protocol. Routing and forwarding strategies to efficiently perform the named based routing are presented in [8] and [9]. Finally, the CS is a cache memory. An intermediate node can decide to cache the content that it forwards downstream for replying to future interests. Consequently, content is stored in source nodes and in caching routers [3].

Each interest packet brings back one data packet if a copy of the requested

content is found. If the content object is large, it may be partitioned into smaller segments to fit into data packets. In a classical NDN, the client requests a content segment by sending the name prefix with the segment identifier [10]. For example, the interest `<content-name>/<i>` is requesting the  $i^{th}$  segment of a content. Each interest also carries a random identifier, named *nonce*, which helps to prevent interest forwarding loops [3]. Interests are forwarded using the information in the FIB.

A node that receives interest for a segment verifies that there is no similar interest pending in its PIT. Having a pending interest means that the requested content is not in the CS, so the node updates the pending entry in the PIT by adding the receiving face of the new interest. If there is no pending entry with the same name, the node then checks its CS for a copy of the requested content. If there is a *cache hit*, *i.e.*, the requested content is available in the cache, the node replies to the interest with a data packet. In the case of the unavailability of requested content or *cache miss*, a new PIT entry is created, and the interest is forwarded to the available faces in the FIB. Once the content is received from upstream, it is routed back to the requesting node using the information in the PIT. The node also decides whether the content should be cached locally [3].

## 2.2. Network Coding and NDN

NC and NDN both inherently tend to address the content delivery and focus on the improvement of content distribution over the network. NC and NDN can work jointly to exploit network capacity better (by exploiting caching, multi-path delivery, *etc.*). The idea of applying NC over ICN/NDN was first introduced in NC3N by Montpetit *et al.* to take advantage of NDN and NC's inherent features to improve content delivery [5]. Currently, there is ongoing work in standardization on the precise topic of mixing NC and ICN [11].

In an NC scenario, the original content is partitioned into smaller groups of segments, called *generations*. NC is only allowed among the segments of a generation to reduce the decoding complexity. In a given generation, segments may be linearly combined within a source or at any intermediate node in the

network. The linear combinations are performed in some Galois field  $\mathbb{F}_q$  to get *coded* segments. The coefficients in  $\mathbb{F}_q$  of the linear combination form the *encoding vector* [12] of each coded segment. In the NDN context, the source nodes and caching routers may store original and coded segments.

The client nodes send interests requesting coded segments instead of a specific segment. The name carried by interest packets for coded content is adapted to indicate that a coded content is expected (*e.g.*, by setting a flag that indicates the retrieval of coded segments [5]). Requesting content like this allows the intermediate or source nodes to send different coded segments generated by combining the original segments of the content in their cache, instead of one particular segment.

The pull-based request and response mechanism of NDN allows one interest to bring back one content segment. A client node sends at least  $n$  interests to be able to decode a generation of  $n$  segments. Based on the interest processing in NDN, it is challenging to ensure retrieval of *innovative* (linearly independent) content with each interest as is required to keep minimal decoding delay and network load.

In the coded NDN schemes proposed in [13, 14, 15], encoding vectors of all the received coded segments are sent in the interests. The encoding vectors help the nodes to either generate coded content that will be innovative for the requesting node or forward the interest to their next-hop neighbors. However, this approach introduces an overhead in the interest packets that increases with each coded segment the requesting node receives. The size of the overhead is limited by keeping the generation size small. This approach also introduces a delay as the client node waits to receive the replies for previously sent interests to arrive before it can issue interests for more coded segments to ensure retrieval of linearly independent content with each interest.

Zhang *et al.* compare the approach of sending all received coefficients (precise matching) to rank-based matching, *i.e.*, sending only the client node's rank. They observe that rank-based matching achieves slightly lower performance but has much lower computation and communication overhead [16].

Liu *et al.* [17] introduced an interest coding and forwarding strategy that allow splitting and joining of interests for the same content and generation at intermediate nodes. The interests request a subset of segments by indicating the number of required coded segments to get a decodable generation. This scheme implements a one interest-multiple replies strategy, which is contrary to the NDN principle of one interest-one reply.

NetCodCCN [18], tries to address the shortcomings of previous approaches by sending undifferentiated interests for coded segments of a generation. The client node implicitly states that it requires another coded combination by sending additional interests for coded segments. The intermediate routers that have previously sent coded segments keep track of the number of coded segments forwarded on each face and the rank of the set of linear combinations in their CS. The node only replies to interest if the rank of its CS is bigger than the number of coded segments it has sent on a particular face. NetCodCCN also supports the transmission and handling of multiple interests at one time (pipelining), to allow nodes to request content more efficiently. With pipelining, a burst of interests is sent first by a client. Each time content is received, a new interest is sent.

This approach increases the amount of information stored per router. The other disadvantage of this approach is unnecessary data traffic in the network that flows in the network after the clients have received a decodable generation.

Liu *et al.* [19] add a parameter  $r$  indicating the number of desired coded segments in the interest requesting more than one coded segment, which is again contrary to the NDN principle of one interest-one reply. Matsuzono *et al.* [20] proposed L4C2 a low loss, low latency, network coding enabled video streaming over CCN. In L4C2 nodes, request network coded packets in case of data packet losses only. Bilal *et al.* [21] proposed an algebraic framework of Network coding over NDN.



### 3. Our approach for NC-NDN

In the classical NDN framework, there is a one-to-one mapping between the requested content and interests. For a content split in  $n$  segments, the client sends  $n$  interests, each requesting a specific segment by stating its unique name and segment number. The replies to these interests are the requested segments.

With NC, coded segments are stored at various places of the network. As observed, *e.g.*, in [5], many different coded segments could serve as a reply to a given interest. The main challenge when sending interests for coded segments is to ensure that linearly independent segments are sent as replies. This problem is challenging since any intermediate node that has cached coded segments from a generation can generate one or more coded segments and hence can reply to several interests from the same client, but linear independence among the replies is not ensured.

As mentioned, in some prior work [13, 14, 15], only one outstanding interest is allowed at the expense of delay. Alternately, [19] reduces the delay and overhead by requesting multiple coded segments in one interest. Saltarin *et al.* in [18, 22] overcome the delay problem by pipelining multiple identical interests, at the cost of additional information that needs to be stored at intermediate nodes in the network for proper interest processing.

In this work, we propose a NC-based NDN protocol with minimum overhead. The client nodes indicate to the network what they require to decode a generation by pipelining multiple distinct interests. The distinct interests allow parallel processing of the multiple interests and ensure that replies to each of the pipelined interests are not redundant.

#### 3.1. MILIC

The main idea of MILIC starts from this previously described pipelining idea, where a subset of interests for a content generation  $g$  are sent in a burst, with the goal that each of them brings innovative coded content.

For a generation of size  $n$ , we propose to use  $n$  distinct interests. Interest  $i \in \{1, \dots, n\}$  can be satisfied by any coded segment whose encoding vector

belongs to a predefined subset  $\mathcal{A}_i$  of the set of all possible encoding vectors. In the following, these constraints will be such that the set of all non-zero encoding vectors will be partitioned into  $n$  non-overlapping subsets  $\mathcal{A}_1, \dots, \mathcal{A}_n$  satisfying some additional constraints.

First, to ensure that the answer to any of the  $n$  interests is linearly independent of the other answers, the subsets must satisfy the following property.

**Property 1.** *For any  $a_1 \in \mathcal{A}_1, \dots, a_n \in \mathcal{A}_n$ , the vectors  $a_1, \dots, a_n$  should be linearly independent, i.e.,  $\sum_{i=1}^n \alpha_i a_i = 0$  iff  $\alpha_1 = \dots = \alpha_n = 0$ .*

An additional condition can be imposed on subsets  $\mathcal{A}_1, \dots, \mathcal{A}_n$  to benefit from the observation that when a node sends the same interests over  $\ell$  faces,  $\ell$  answers to each of these interests will likely be received. Ideally, these replies should be linearly independent. This leads to a property of subsets that is not mandatory but desirable to improve the efficiency of the proposed solution.

**Property 2.** *Consider  $k$  distinct subsets  $\mathcal{A}_{\pi(1)}, \dots, \mathcal{A}_{\pi(k)}$ . Consider  $\ell \geq 1$  vectors  $a_{\kappa}^1, \dots, a_{\kappa}^{\ell}$  chosen uniformly at random from each subset  $\mathcal{A}_{\pi(\kappa)}$ ,  $\kappa = 1, \dots, k$  such that  $\ell k \leq n$ , then  $\text{rank}(a_1^1, \dots, a_k^{\ell}) = \ell k$  with high probability.*

Finally, one may try to exploit the fact that segments are coded with possible re-encoding at intermediate nodes. Intermediate nodes may have received several segments belonging to the same subset. It may be of interest to combine these to generate a coded segment belonging to another subset to satisfy interest for that subset. This translates into the following additional desirable property for the subsets.

**Property 3.** *Consider the subset  $\mathcal{A}_i$ ,  $i \in \{1, \dots, n-1\}$ . For any pair  $(a_i^1, a_i^2)$  of linearly independent vectors belonging to  $\mathcal{A}_i$ , then with high probability, there exist  $\alpha_1 \in \mathbb{F}_q^*$  and  $\alpha_2 \in \mathbb{F}_q^*$  such that  $\alpha_1 a_i^1 + \alpha_2 a_i^2 \in \mathcal{A}_k$  with  $k \neq i$ .*

### 3.2. Proposed construction

Here we propose a construction of the sets  $\mathcal{A}_1, \dots, \mathcal{A}_n$ , called MILIC, that partly satisfies the above properties. Consider

$$\mathcal{A}_i = \{(v_1, \dots, v_n) \in \mathbb{F}_q^n \mid v_i \neq 0 \text{ and } \forall j < i, v_j = 0\}, \quad (1)$$

with  $i = 1, \dots, n$ . With this construction the sets  $\mathcal{A}_1, \dots, \mathcal{A}_n$  form a partition of  $\mathbb{F}_q^n \setminus \{(0, \dots, 0)\}$ . The construction (1) implicitly imposes an ordering among sets, when considering their cardinal number.

We will now prove that the proposed MILIC construction satisfies Property 1, Property 2 for the  $k$  first subsets  $\mathcal{A}_1, \dots, \mathcal{A}_k$ , and Property 3 for all  $k > i$ .

Property 1 is satisfied by construction: consider any  $a_1 \in \mathcal{A}_1, \dots, a_n \in \mathcal{A}_n$ . The matrix whose rows are  $a_1, \dots, a_n$  is in row echelon form, and thus of full rank. The vectors  $a_1, \dots, a_n$  are thus linearly independent. Then, we have the following property of the size of each subset  $\mathcal{A}_k$ .

**Proposition 1.** *The cardinal number of  $\mathcal{A}_k$  verifies  $|\mathcal{A}_k| = (q - 1)q^{n-k}$ .*

*Proof.* Consider first the case of  $\mathcal{A}_1$ :  $\forall a_i \in \mathcal{A}_1$ , one has  $a_{i,1} \neq 0$ . There are  $q^{n-1}$  vectors with leading zeros in  $\mathbb{F}_q^n$  hence  $|\mathcal{A}_1| = (q - 1)q^{n-1}$ . Then consider the more general case of  $\mathcal{A}_k$  for  $k > 1$ :  $\forall a_i \in \mathcal{A}_k$ , one has  $a_{i,j} = 0$  for  $j = 1, \dots, k - 1$  and  $a_{i,k} \neq 0$ . For  $a_{i,k}$ , we have  $q - 1$  possible choices. Then each  $a_{i,j}$ ,  $j = k + 1, \dots, n$  may take  $q$  possible values. Consequently  $(a_{i,k+1}, \dots, a_{i,n})$  may take  $q^{n-k}$  possible values and  $|\mathcal{A}_k| = (q - 1)q^{n-k}$ .  $\square$

We start proving Property 2 for a single subset  $\mathcal{A}_k$  provided that  $\ell \leq n - k + 1$ , evaluating the probability of having  $\text{rank}(a_k^1, \dots, a_k^\ell) = \ell$ .

**Lemma 1.** *Consider  $\ell$  vectors  $a_k^1, \dots, a_k^\ell$  chosen uniformly at random from the set  $\mathcal{A}_k$ ,  $k = 1, \dots, n$ , and with  $1 \leq \ell \leq n$ . The probability that  $a_k^1, \dots, a_k^\ell$  are linearly independent is*

$$\Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell) = \prod_{\ell=1}^{\ell} \left(1 - \frac{q^{\ell-1} - 1}{(q - 1)q^{n-k}}\right).$$

*Proof.* Consider first  $\ell = 2$ , and  $a_k^1 \in \mathcal{A}_k$ . The set of non-zero vectors collinear to  $a_k^1$  and included in  $\mathcal{A}_k$  is  $\text{span}(a_k^1) \cap \mathcal{A}_k = \text{span}(a_k^1) \setminus \{(0, \dots, 0)\}$ , whose size is  $q - 1$ . When choosing a second vector  $a_k^2 \in \mathcal{A}_k$  uniformly at random, the probability that  $a_k^1$  and  $a_k^2$  are linearly dependent is equal to the probability

that  $a_k^2 \in \text{span}(a_k^1) \setminus \{(0, \dots, 0)\}$ . Consequently, the probability that  $a_k^1$  and  $a_k^2$  are linearly independent is

$$\begin{aligned} \Pr(\text{rank}(a_k^1, a_k^2) = 2) &= 1 - \frac{|\text{span}(a_k^1) \setminus \{(0, \dots, 0)\}|}{|\mathcal{A}_k|} \\ &= 1 - \frac{1}{q^{n-k}}. \end{aligned}$$

Assume now that the  $j-1$  first vectors  $a_k^1 \in \mathcal{A}_k, \dots, a_k^{j-1} \in \mathcal{A}_k$  are linearly independent. The set of vectors that are linearly dependent with  $a_k^1, \dots, a_k^{j-1}$  and included in  $\mathcal{A}_k$  is  $\text{Span}(a_k^1, \dots, a_k^{j-1}) \cap \mathcal{A}_k = \text{span}(a_k^1, \dots, a_k^{j-1}) \setminus \{(0, \dots, 0)\}$ . Its size is  $q^{j-1} - 1$ . Then, when choosing  $a_k^j \in \mathcal{A}_k$  uniformly at random, the probability that  $a_k^1, \dots, a_k^j$  are linearly dependent is equal to the probability that  $a_k^j \in \text{span}(a_k^1, \dots, a_k^{j-1}) \setminus \{(0, \dots, 0)\}$ . Consequently

$$\Pr(\text{rank}(a_k^1, \dots, a_k^j) = j \mid \text{rank}(a_k^1, \dots, a_k^{j-1}) = j-1) = 1 - \frac{q^{j-1} - 1}{(q-1)q^{n-k}}. \quad (2)$$

Then one has

$$\begin{aligned} \Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell) &= \Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell, \text{rank}(a_k^1, \dots, a_k^{\ell-1}) = \ell-1) \\ &= \Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell \mid \text{rank}(a_k^1, \dots, a_k^{\ell-1}) = \ell-1) \\ &\quad \Pr(\text{rank}(a_k^1, \dots, a_k^{\ell-1}) = \ell-1). \end{aligned} \quad (3)$$

Applying this recursively and using (2), one gets

$$\begin{aligned} \Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell) &= \prod_{j=2}^{\ell} \Pr(\text{rank}(a_k^1, \dots, a_k^j) = j \mid \text{rank}(a_k^1, \dots, a_k^{j-1}) = j-1) \\ &\quad \Pr(\text{rank}(a_k^1) = 1) \\ &= \prod_{j=1}^{\ell} \left(1 - \frac{q^{j-1} - 1}{(q-1)q^{n-k}}\right). \end{aligned}$$

□

**Example 1.** Table 1 provides  $P_F(\ell, 1) = 1 - \Pr(\text{rank}(a_k^1, \dots, a_k^\ell) = \ell)$  for vectors of  $n = 10$  elements in  $\mathbb{F}_{256}$  for different subsets  $\mathcal{A}_k$  and different values of  $\ell$ . One observes that choosing 5 vectors at random from any of the subsets  $\mathcal{A}_k$ ,  $k = 1, \dots, 5$ , results in a very high probability of getting linearly independent vectors. Consequently, if a client sends 5 interest packets for elements in  $\mathcal{A}_k$  over different faces, it is likely, provided that these interests follow different paths in the network, to get 5 linearly independent data packets.

	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
$\mathcal{A}_1$	0	$2.11 \times 10^{-22}$	$5.46 \times 10^{-20}$	$1.39 \times 10^{-17}$	$3.58 \times 10^{-15}$
$\mathcal{A}_2$	0	$5.46 \times 10^{-20}$	$1.39 \times 10^{-17}$	$3.58 \times 10^{-15}$	$9.16 \times 10^{-13}$
$\mathcal{A}_3$	0	$1.39 \times 10^{-17}$	$3.58 \times 10^{-15}$	$9.16 \times 10^{-13}$	$2.34 \times 10^{-10}$
$\mathcal{A}_4$	0	$3.58 \times 10^{-15}$	$9.16 \times 10^{-13}$	$2.34 \times 10^{-10}$	$6.01 \times 10^{-8}$
$\mathcal{A}_5$	0	$9.16 \times 10^{-13}$	$2.34 \times 10^{-10}$	$6.01 \times 10^{-8}$	$1.53 \times 10^{-5}$

Table 1: Probability of getting linearly dependent coded vectors chosen at random from  $\mathcal{A}_k \subset \mathbb{F}_{256}^{10}$

We now prove Property 2 for the  $k$  first subsets  $\mathcal{A}_1, \dots, \mathcal{A}_k$ .

**Lemma 2.** Consider  $\ell \geq 1$  vectors  $a_\kappa^1, \dots, a_\kappa^\ell$  chosen uniformly at random from each subset  $\mathcal{A}_\kappa$ ,  $\kappa = 1, \dots, k$  such that  $\ell k \leq n$ . The probability that  $a_1^1, \dots, a_k^\ell$  are linearly independent is

$$\Pr(\text{rank}(a_1^1, \dots, a_k^\ell) = \ell k) = \prod_{j=1}^{(\ell-1)k} \left(1 - \frac{q^{j-1}}{q^{n-k}}\right).$$

*Proof.* According to Property 1, the vectors  $a_1^1, \dots, a_k^1$  are linearly independent. Consider the matrix  $A$ , whose first  $k$  rows are the vectors  $a_1^1, \dots, a_k^1$  and the  $(\ell - 1)k$  remaining rows are  $a_\kappa^2, \dots, a_\kappa^\ell$ ,  $\kappa = 1, \dots, k$ . The first  $k$  rows are used to perform Gaussian elimination on the  $(\ell - 1)k$  remaining rows to get a matrix  $A_1$  of the form

$$A_1 = \begin{bmatrix} 1 & * & \cdots & & * \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \\ & & & 1 & * & \cdots & * \\ & & & 0 & & & \\ \vdots & & & \vdots & & B & \\ 0 & \cdots & & 0 & & & \end{bmatrix}.$$

In  $A_1$ ,  $B$  is a matrix of  $(\ell - 1)k$  rows and  $n - k$  columns. Since all vectors chosen in the subset  $\mathcal{A}_\kappa$ ,  $\kappa = 1, \dots, k$ , have been selected uniformly at random, the  $n - k$  last entries of each vector are independently and uniformly distributed. The  $i$ -th row of  $B$  results in a linear combination of  $a_1^1, \dots, a_k^1$  with one of the remaining vectors  $a_\kappa^2, \dots, a_\kappa^\ell$ ,  $\kappa = 1, \dots, k$ . Consequently, the  $n - k$  components of the  $i$ -th row of  $B$  are still independently and uniformly distributed. Since all  $n - k$  last components of  $a_\kappa^2, \dots, a_\kappa^\ell$ ,  $\kappa = 1, \dots, k$  are independently and uniformly distributed; all components of the matrix  $B$  are independently and uniformly distributed.

The matrix  $A$  is of full row rank  $\ell k$  iff the matrix  $B$  is full row rank  $(\ell - 1)k$ . The first row  $b_1 \in B$  is non-zero with probability  $1 - \frac{1}{q^{n-k}}$ . The second row  $b_2 \in B$  has components that are uniformly and independently distributed from the other entries of  $B$  and thus of  $b_1$ . The vectors  $(b_1, b_2)$  are linearly independent if  $b_2$  does not belong to the space spanned by  $b_1$ . Since  $\text{span}(b_1)$  is of size  $q$ , one has

$$\Pr(\text{rank}(b_1, b_2) = 2) = 1 - \frac{q}{q^{n-k}}.$$

Assume now that the  $j - 1$  first row vectors  $b_1, \dots, b_{j-1}$  of  $B$  are linearly independent. Under this assumption, the probability that  $b_j$  is such that the  $j$  first row vectors  $b_1, \dots, b_j$  of  $B$  are linearly independent is equal to the probability that  $b_j$  does not belong to the subspace of dimension  $q^{j-1}$  spanned by

$b_1, \dots, b_{j-1}$ . Consequently,

$$\Pr(\text{rank}(b_1, \dots, b_j) = j \mid \text{rank}(b_1, \dots, b_{j-1}) = j - 1) = 1 - \frac{q^{j-1}}{q^{n-k}}.$$

Then similarly as 3, the probability that  $B$  is of full rank is given by

$$\begin{aligned} \Pr(\text{rank}(B) = (\ell - 1)k) &= \prod_{j=1}^{(\ell-1)k} \left(1 - \frac{q^{j-1}}{q^{n-k}}\right) \\ &= \prod_{j=1}^{(\ell-1)k} \left(1 - \frac{1}{q^{n-k-j+1}}\right) \\ &\approx 1 - \frac{1}{q^{n-lk+1}} \text{ when } q \text{ large} \end{aligned}$$

□

**Example 2.** Table 2 provides  $P_F(\ell, k) = 1 - \Pr(\text{rank}(a_1^1, \dots, a_k^\ell) = \ell k)$  for vectors of  $n$  elements in  $\mathbb{F}_q$  when choosing at random  $\ell$  vectors from each subset  $\mathcal{A}_\kappa$ ,  $\kappa = 1, \dots, k$ . One observes that when a node receives 2 random packets from each of the  $\mathcal{A}_\kappa$ ,  $\kappa = 1, \dots, 50$  subsets, provided that NC is performed in  $\mathbb{F}_{256}$ , the probability of getting a linearly independent packet is above 99.6%. The same result is obtained when 4 packets are obtained from each of the  $k = 25$  first subsets. The constraints introduced on the subsets do not degrade significantly the generation recovery performance compared to plain NC. This result is mainly obtained due to the fact that one considers packets received from the first (largest) subsets.

$k$	$\ell$	$\mathbb{F}_2$	$\mathbb{F}_{256}$
50	2	0.71	0.0039
25	4	0.71	0.0039
33	3	0.42	$1.53 \times 10^{-5}$
49	2	0.23	$5.98 \times 10^{-8}$
48	2	0.06	$9.13 \times 10^{-13}$
32	3	0.06	$9.13 \times 10^{-13}$
24	4	0.06	$9.13 \times 10^{-13}$
47	2	0.015	$1.39 \times 10^{-17}$
45	2	0.00097	$3.24 \times 10^{-27}$

Table 2: Probability  $P_{\mathbb{F}}(\ell, k)$  of getting  $\ell$  linearly dependent vectors chosen at random from consecutive subsets  $\mathcal{A}_1$  to  $\mathcal{A}_k$ .

To prove Property 3 for  $k > i$ , consider an intermediate node that received two linearly independent segments  $a_i^1 \in \mathcal{A}_i$  and  $a_i^2 \in \mathcal{A}_i$ . The  $i-1$  first entries of  $a_i^1$  and  $a_i^2$  are zero, and their  $i$ -th entries  $a_{i,i}^1$  and  $a_{i,i}^2$  are non-zero. Then, as  $\mathbb{F}_q$  is a group for multiplication, considering any  $\alpha_1 \in \mathbb{F}_q^*$ , there exists  $\alpha_2 \in \mathbb{F}_q^*$  such that  $\alpha_1 a_{i,i}^1 + \alpha_2 a_{i,i}^2 = 0$ . Moreover, since  $a_i^1$  and  $a_i^2$  are linearly independent, one has  $b = \alpha_1 a_i^1 + \alpha_2 a_i^2 \neq 0$ . Let  $k$  be the smallest index such that  $b_k \neq 0$ . Necessarily  $k > i$  and  $b \in \mathcal{A}_k$ .

**Remark 1.** *Imposing an ordering in the subsets  $\mathcal{A}_k$  might look inefficient. Nevertheless, due to pipelining behavior, this is not a problem. When there is a single path between a client and a source, Property 1 ensures that all contents are innovative. If  $\ell$  distinct paths connect the client to one or several sources, the first interest packet in the pipeline should bring back  $\ell$  linearly independent data packets thanks to Property 2. Then again, thanks to Property 2, the  $k$  first pipelined interest packets are likely to bring back  $k\ell$  linearly independent data packets, see Table 2. Consequently, when  $\ell$  distinct paths connect the client to one or several sources, it is unlikely that this client will need to send interests for contents in the subsets  $\mathcal{A}_k$  with  $k$  close to  $n$ . This opens the potential for an*



optimization of the size of the pipeline.

#### 4. MICN protocol

This section describes the MICN protocol, focusing on the interest and content processing using the MILIC construction presented in Section 3.1 to recover linearly independent content with each interest in the context of ICN.

##### 4.1. Content Segmentation and Naming

The original content  $C$  is partitioned into  $G$  smaller groups of segments, called *generations*  $C = [c_1, c_2, \dots, c_G]$ . Each generation  $c_g$ ,  $g = 1, \dots, G$  contains  $n$  equally-sized segments

$$c_g = [c_{g,1}, c_{g,2}, \dots, c_{g,n}].$$

The NC operations are restricted to segments that belong to the same generation and assumed to be performed in  $\mathbb{F}_q$ .

A MILIC-compliant coded segment, whose encoding vector in the subset  $\mathcal{A}_i$ ,  $i = 1, \dots, n$ , is defined as

$$\tilde{c}_{g,i} = \sum_{j=i}^n a_j c_{g,j}$$

with  $a_i \neq 0$ . The entries of  $c_{g,j}$ ,  $j = 1, \dots, n$  and  $\tilde{c}_{g,i}$  are represented as elements of  $\mathbb{F}_q$ . Any coded segment  $\tilde{c}_{g,i}$  is identified by a prefix, a generation ID  $g$ , a MILIC index  $i$ , and the encoding vector  $a = (0, \dots, 0, a_i, \dots, a_n) \in \mathbb{F}_q^n$  to indicate the weight of each original segment in  $\tilde{c}_{g,i}$ . Consequently, we propose to identify  $\tilde{c}_{g,i}$  by the NDN name  $\langle \text{prefix} \rangle / \text{micn} / \langle g \rangle / \langle i \rangle / \langle a_i, \dots, a_n \rangle$  (`micn` indicates that the content is network coded). Other naming conventions are possible with MICN.

##### 4.2. Requesting MILIC-compliant contents

According to the naming convention of data packets, see Section 4.1, the name carried by the interest  $I_{g,i}$  for a coded content from  $C$  belonging to the generation  $g$  and with an encoding vector in  $\mathcal{A}_i$  is  $\langle \text{prefix} \rangle / \text{micn} / \langle g \rangle / \langle i \rangle$ .

Contrary to other proposals integrating NC to ICN, this interest format allows the client nodes to pipeline multiple interests for the same generation, provided that different  $\mathcal{A}_i$  are specified in the names.

In practice, a client starts sending successive interests for contents in a given generation  $g$ , starting from packets with encoding vectors in  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_\rho$ , where  $\rho$  is the pipeline size. Additional interests are sent once the content starts flowing back. The pipeline size  $\rho$  limits the number of outstanding interests from a client node at any time.

Each interest  $I_{g,i}$  has an associated time-out. If no innovative content in response to  $I_{g,i}$  is received before time out, the interest  $I_{g,i}$  is sent again. Time out may occur, *e.g.*, in case of losses of the interest or data packets.

#### 4.3. MICN-compliant PIT

Compared to the classical NDN PIT, a MICN-compliant PIT identifies interests requesting coded segments with the same prefix and generation ID as *related interests*. PIT entries for related interests are grouped in a sub-table (identified by prefix and generation ID  $g$ ). Each entry itself includes the associated index  $i$ , *nonce*  $\nu$ , as well as the *in* and *out* faces. The PIT entries are sorted by order of arrival.

Figure 1 illustrates a part of a MICN-compliant PIT at a given node with three faces  $f_1, f_2$ , and  $f_3$ . Three interests have been received and have been forwarded. The two interests associated with  $\mathcal{A}_1$  are considered as different since they have different nonce, which implies that different clients sent them.

#### 4.4. Just-in-Time Content Re-encoding / Replying

In plain NDN, whenever a node can satisfy an interest, a copy of the requested content is sent immediately. In MICN, as in some other NC-NDN protocols, nodes do not just forward a copy of the cached coded content as an answer to the matching interests. They linearly combine cached contents from the same generation to generate a new coded segment.

PIT			
<prefix>/micn/<g>			
Index	Nonce	in-faces	out-faces
1	$\nu_1$	$f_1$	$f_2, f_3$
1	$\nu_2$	$f_1$	$f_2, f_3$
2	$\nu_3$	$f_2$	$f_1, f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
<prefix>/micn/<g'>			

Figure 1: MICN compliant PIT

PIT			
<prefix>/micn/<g>			
Index	Nonce	in-faces	out-faces
1	$\nu_1$	$f_1$	$f_2, f_3$
1	$\nu_2$	$f_1$	$f_2, f_3$
2	$\nu_3$	$f_2$	$f_1, f_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k$	$\nu_k$	$f_1$	
<prefix>/micn/<g'>			

Figure 2: MICN compliant PIT: the interest with index  $k$  lead to cache hit and is temporarily stored in the PIT until the queue of face  $f_1$  is empty to send back the associated data packet

In MICN, the reply strategy is further modified, compared to plain NDN. A node waits until the queue of a face is empty before generating a coded segment that satisfies a pending interest on this face. This allows the node to use its latest cached contents when replying, hence sending more diverse content through the network. To achieve this one packet queue is considered at the faces that is filled only when the packet in transit is completely delivered. The process to achieve this just-in-time re-encoding is detailed in Sections 4.5.1 and 4.6.

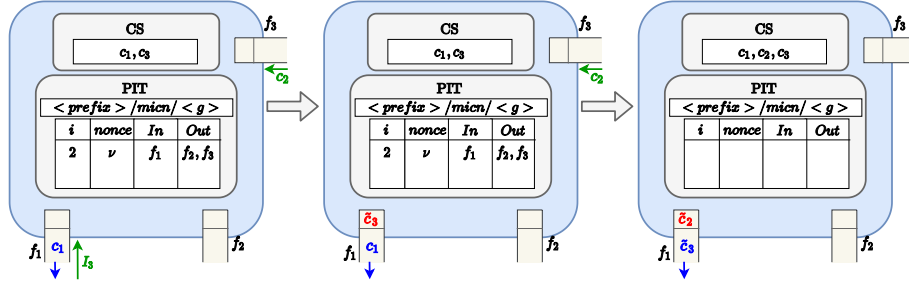
#### 4.5. Interest processing

When a node receives an interest  $I_{g,i}$ , it initially performs loop detection. If an interest with the same nonce has already been received,  $I_{g,i}$  is considered as a looping interest. Otherwise, the node processes the interest. It can either reply using a content generated from its CS or further forwards the interest to the network.

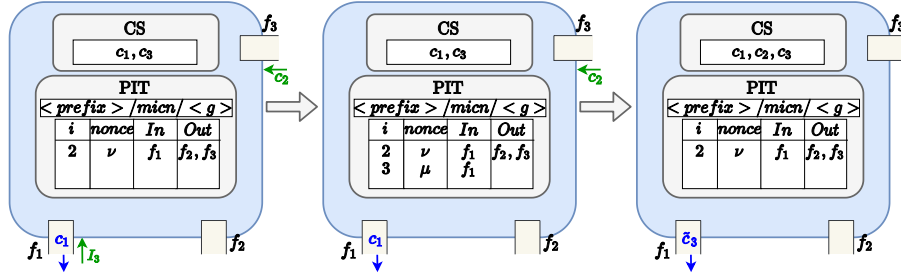
##### 4.5.1. CS lookup

Like the PIT, the *related contents* (*i.e.*, contents with the same prefix and generation ID) are grouped in the CS. The *CS lookup* starts by identifying the related content matching the received interest. A *cache hit* occurs when this cached content can be used to generate a coded segment belonging to the subset requested in the interest.

In case of a cache hit, the node schedules a reply for the interest. The node first checks the outgoing queue of the face where the interest arrived. If the queue is empty, the content is immediately sent in a data packet. Otherwise, a reply is generated only when the queue becomes empty. In our implementation, this scheduling is achieved by creating a *volatile* PIT entry to store the incoming face, nonce, *etc.*, but without specifying an outgoing face, since the interest does not require to be forwarded. See, for example, the interest with index  $k$  in Figure 2.



(a) Immediate re-encoding:  $\tilde{c}_3 = \alpha_1 c_1 + \alpha_3 c_3$  is put in the outgoing queue of face  $f_1$  before the reception and processing of content packet  $c_2$ .



(b) Just-in-time re-encoding with MICN:  $\tilde{c}_3 = \alpha_1 c_1 + \alpha_2 c_2 + \alpha_3 c_3$  is put in the outgoing queue of face  $f_1$  only once this queue is empty; this gives the opportunity to the later received  $c_2$  on face  $f_2$  to be included in  $\tilde{c}_3$ .

Figure 3: Re-encoding cached content

Fig. 3a illustrates the state of a node that has enough content in its CS to respond to the incoming interest  $I_3$ , it immediately uses the cached related content to generate a response  $\tilde{c}_3$ , but the content remains in the queue until the content  $c_1$  is transmitted. While the node in Fig. 3b waits until  $c_1$  is transmitted since it may receive more content and have a more diverse CS (since more content from the same generation is requested). So a volatile PIT entry is generated that is replied as soon as the queue becomes empty.

#### 4.5.2. Interest Forwarding

In case of a cache miss, the node forwards the interest to its next hop neighbors on the faces in the FIB (except the incoming face) and creates a PIT entry, which records the incoming and outgoing faces. Unlike classical NDN, different nonces result in different entries, see Figure 1.

According to the management of the FIB, multiple interest forwarding strategies can be implemented depending on the subset of chosen faces to forward the content. In this paper, to take advantage of the multiple paths to the source(s) and to have the opportunity to receive multiple linearly independent segments, the FIB is filled with all faces that can lead to a source *without looping back* to the node. Then, the *multicast forwarding strategy* where the interests are forwarded on all faces in the FIB is used, as suggested in [3, Section 5.2.2].

#### 4.6. Content Processing

When a coded segment arrives at a node from one of its faces, it adds it in its cache if it is linearly independent with the already cached related contents. The updated cache might then satisfy some additional/new interests.

The node uses its updated cache to reply to the pending interests. Whenever the queue of a face  $f$  is empty, the node checks if any pending interest on that face can be satisfied utilizing the current state of the cache. It answers the oldest PIT entry, that may be satisfied and removes the entry.

#### 4.7. Optimizations

In this section, we introduce some optimization compared to the classical NDN to improve the performance of MICN in an NC-NDN scenario.

##### 4.7.1. Content Redirection

A node can receive an interest on an alternate face while the same interest (same nonce) is still pending at the first face due to the Just-in-Time content re-encoding of MICN, see Section 4.4. Such interest brings information that there exists an alternate path to the client. If the queue associated with this

alternate/second face is empty and the node has matching content, the content is immediately redirected to the client via this alternate face. This redirection is likely to improve the network utilization, by benefiting from all paths leading to the client.

Fig. 4 depicts the state of a node that receives interest  $I_3$  with the same nonce  $\nu$  from an alternate face  $f_2$  with an empty queue. Since the node has enough content to generate a reply for the interest but the face  $f_1$  is busy, the node redirects the content via the alternate face to send the reply immediately and possibly benefit from a second path to the client.

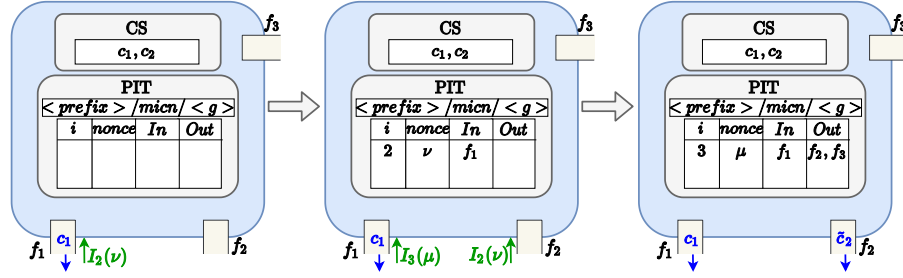


Figure 4: Content redirection on face  $f_2$ : during the transmission of  $c_1$ , an interest for content associated to  $\mathcal{A}_2$  has been received from face  $f_1$  (left) and then from face  $f_2$  (middle); since the outgoing queue of face  $f_1$  is still occupied,  $\tilde{c}_2$  is transmitted on face  $f_2$  (right).

#### 4.7.2. Interest Cancellation (MICN-IC)

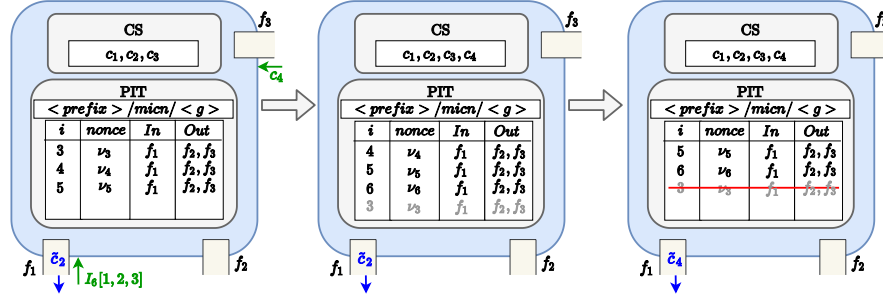


Figure 5: Interest Cancellation: An interest for packets associated to  $\mathcal{A}_6$  is coming from face  $f_1$  indicating that the source has already access to content associated to  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ , and  $\mathcal{A}_3$  (left); The pending interest for content associated to  $\mathcal{A}_3$  is first tagged with low priority (middle); This pending interest is canceled as soon as an interest associated to a subset of higher index (here  $\mathcal{A}_4$ ) is replied to (right).

We observe that the content continues to flow in the network due to delay differences in different parts of the network even after the client nodes have received a decodable generation. In order to reduce the traffic represented by redundant contents, we introduce the concept of *interest cancellation*.

For that purpose, client nodes add information about the content they already received. To achieve this, the optional *client identifier* and *state* fields are added in the interest packets. The client identifier field is a hash of the client node identifier, while the state field bears the information of subsets as defined by MILIC for which that client has already available content. Such content may have been directly obtained or deduced after Gaussian elimination involving several data packets. The state field may, *e.g.*, be represented by a bitmap indicating the available indices.

A node, when receiving interest with the state of a client, may ignore the pending interests referencing subsets  $\mathcal{A}_i$  for the indices  $i$  for which content is already available. Nevertheless, this node does not immediately delete them. Instead, they get low priority for replies, contrary to other interests in the PIT, which have a normal priority. Keeping and answering these low-priority interests



may still be useful, according to Properties 2 and 3: NC contents sent as replies even for subsets from which content is already available may bring information with a *high probability*.

A reply to interest with a low priority index is sent only if the outgoing face is empty, and the node cannot generate content as a reply to interest with a normal priority index. The deletion of low-priority interests occurs when the node has sent content for an interest with a higher index to the client. This version of MICN with Interest Cancellation is referred to as MICN-IC.

Fig. 5 illustrates the state of a node that receives an interest for some content associated to  $\mathcal{A}_6$ . The interest also carries the state of requesting node indicating that it already access to contents associated to  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ , and  $\mathcal{A}_3$ . Using this information, the node sets the pending interest for content associated to  $\mathcal{A}_3$  to low priority (interest in gray). This low priority interest is deleted once a content associated to  $\mathcal{A}_4$  has been sent to the considered client.

## 5. Evaluation

### 5.1. Simulation setup

We implemented our simulator in Python. It includes a generic packet network simulator (scheduler, link, packet transmission), on top of which we developed an implementation of the proposed MICN protocol, and lightweight reimplementations of NDN and NetCodCCN (capturing the main semantics of these protocols as described in [3] and [18]). We experimented protocols over two topologies: one simple illustrative butterfly topology and a second, more elaborate topology close to the PlanetLab topology from NetCodCCN [18]. At the link level, the parameters of our simulations are a propagation delay of 0.1 time unit for each packet, a transmission time of 1 time unit for data packets, and a very small transmission delay for interest packets ( $1/2^{14} \simeq 6 \times 10^{-5}$ ). A small amount of uniformly distributed transmission jitter (between 0 and  $1/2^{18} \simeq 3.8 \times 10^{-6}$ ) was also introduced.

In each topology, we consider the following scenario. Several clients request coded content, divided into generations of 100 segments each. We study the transmission of one generation. Each source stores a complete copy of the coded content. We assume that the intermediate nodes have enough cache space to store all segments of a generation. All the coding operations are performed in the finite field  $\mathbb{F}_{2^8}$ .

In MICN, the interest pipeline size  $\rho = 10$  is considered, the FIB and the interest forwarding are as described in Section 4.5.2. At the client, each interest packet has a time out of 10 time units (*i.e.*, equivalent the transmission delay of 10 data packets, that is a bit longer than the longer round-trip delay). If a client does not receive innovative content for an interest after this time interval, it will resend the interest.

The performance is evaluated in terms of download time, *i.e.*, the time it takes for a client to download and decode a generation. The time needed to perform Gaussian elimination is neglected. An upper bound of the throughput (content/time unit) received by a client is given by the maximum flow on the graph from the sources to the node. From this max-flow, one can derive a lower bound of the download time. In similar settings, it had been proven that network coding could approach the max-flow bound [23], hence representing a meaningful benchmark. Another metric of interest is the total number of data packets exchanged in the network until all clients have retrieved the generation with no interest or data packet is present in the network anymore.

### 5.2. Results with the butterfly topology

We first analyze the behavior of MICN on a simple butterfly topology with two sources  $S_1$  and  $S_2$  and two clients  $U_1$  and  $U_2$  connected through a set of intermediate caching routers as represented in Fig. 6.

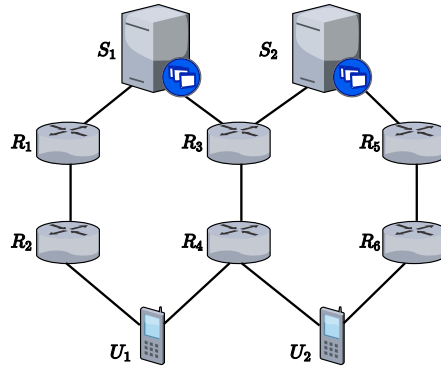


Figure 6: Butterfly topology

The performance of MICN mainly depends on how the bottleneck link ( $R_3 \leftrightarrow R_4$ ) is used. With classical NDN, the two clients  $U_1$  and  $U_2$  should request precisely the same segments on the middle link to improve performance. Nevertheless, the clients would require topology knowledge and coordination to do so. However, with NC, this is not required, and the clients can simultaneously send their interests to all their available faces.

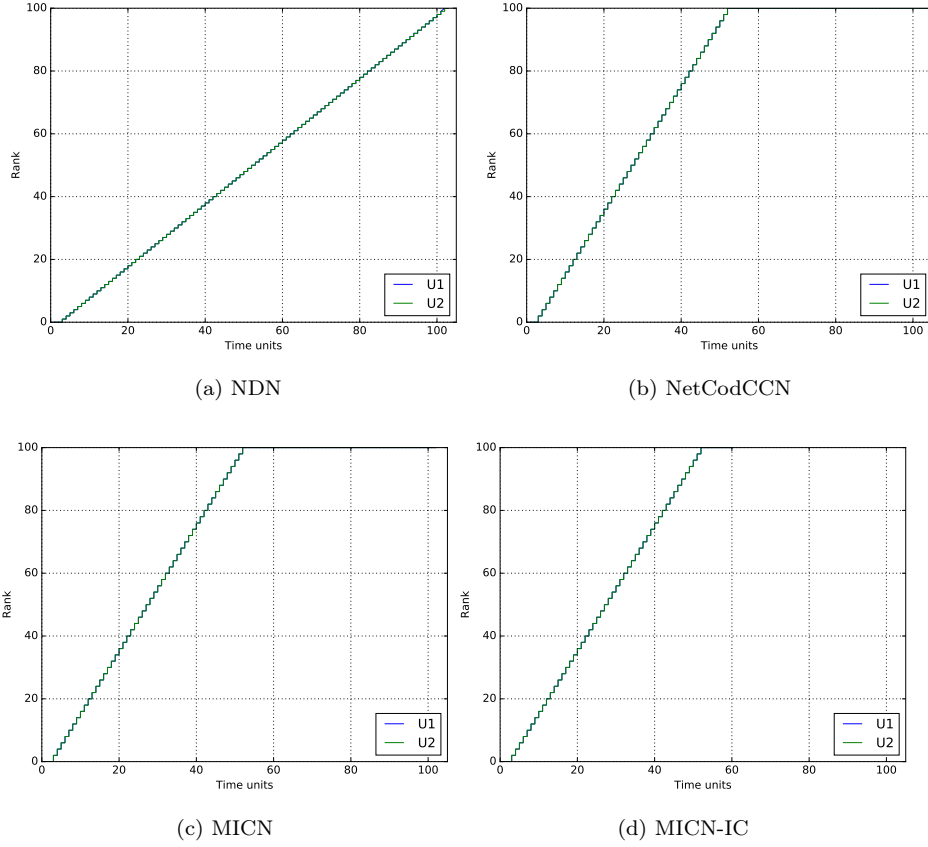
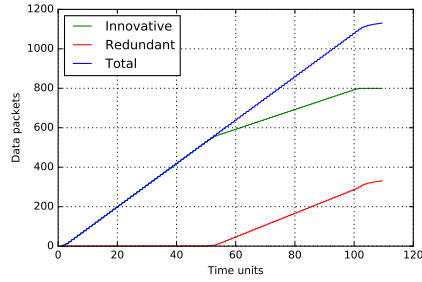
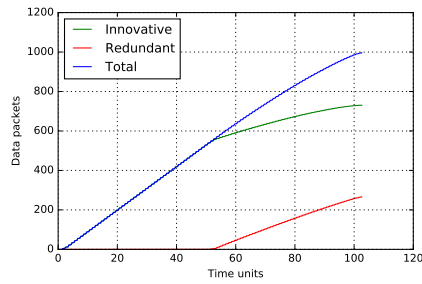


Figure 7: Butterfly topology: Rank evolution as a function of time

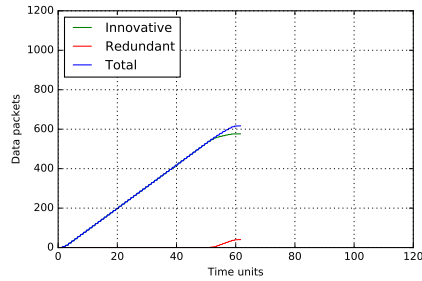
Fig. 7 shows the rank evolution of the client nodes over time for MICN, MICN-IC, NetCodCCN, and NDN. The three protocols retrieve content at both the clients at the max-flow rate, *i.e.*, each data packet received at the client is innovative. After some initial delay, due to propagation, with all protocols, the clients receive 2 linearly independent data packets every 10 time units. Nevertheless, there are significant differences in the volume of data traffic that each protocol generates, as shown in Fig. 8.



(a) NetCodCCN



(b) MICN



(c) MICN-IC

Figure 8: Data traffic in the butterfly topology

Fig. 8 depicts the evolution with time of the cumulative number of data packets transmitted on all the links of the network, counted from time  $t = 0$ . The curves end when transmission of data packets stops. In the beginning, there is only innovative traffic in the network, *i.e.*, data packets which are innovative for the routers or the client node receiving them. While towards the end, even after

the clients have received the entire generation, the outlying interests continue to generate data traffic that is just redundant. MICN-IC deletes the interests tagged with low priority, which are pending even if a client has access to content for those interests. Canceling such interests reduces the redundant data traffic, at the price of some signaling overhead. Precisely, in the butterfly topology (Fig. 6), 10 transmissions of data packets over various links are necessary for delivering 2 data packets to the clients  $U_1$  and  $U_2$ , *i.e.*, 5 transmissions per packet. For a generation of size 100, a minimum of 500 transmissions are required for both clients to receive the entire generation. Fig. 8 shows that with MICN-IC, a slightly larger amount of transmissions are required. NetCodCCN achieves similar download performance, but interests are not canceled and several data packets are redundant, leading to increased traffic.

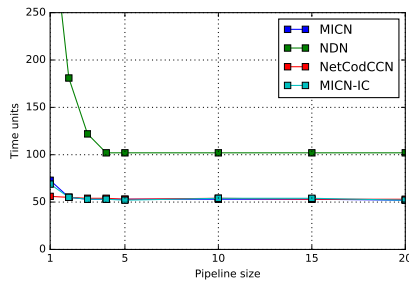


Figure 9: Butterfly topology: Download time vs pipeline size, no losses

The effect of sending consecutive interests by clients is analyzed in Fig. 9. To have a continuous flow of content in the butterfly network (in the absence of losses), the clients need to have at least two outstanding interests at any time (because there are two paths) and usually even more because of the propagation delays. In the case of plain NDN with multicast strategy, the link  $R_3 \leftrightarrow R_4$  becomes a bottleneck due to no coordination among the clients. Even when the pipeline size increases, the performance cannot reach the one obtained with NC. MICN, MICN-IC, and NetCodCCN, however, with a sufficient pipeline size (here as small as  $\rho = 5$ ), can reach the maximum capacity.

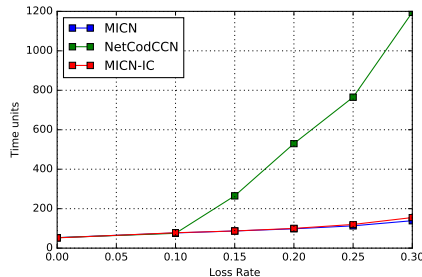


Figure 10: Butterfly Topology:Download time vs transmission error rate

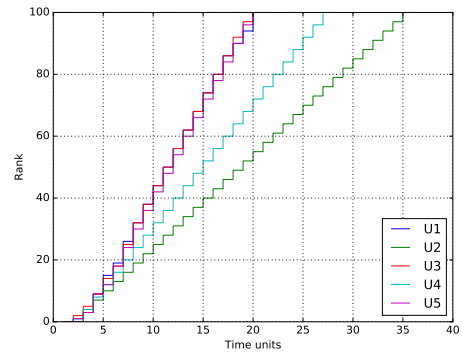
Next, we evaluate the performance of MICN in case of losses. Fig. 10, depicts the effect of losses on the performance of the protocols. We consider transmission losses modeled with a fixed loss probability for both interest and data packets.<sup>1</sup> MICN and MICN-IC appear to have much better performance compared to NetCodCCN. MICN has the advantage of precisely identifying which interest (pointing to a subset  $\mathcal{A}_i$ ) has timed-out (no matching content received). In NetCodCCN, even if a downstream data packet is lost, the router will consider the interest satisfied. An interest repeated due to time-out is considered a new interest, and the router will typically forward it. In MICN, the repeated interest will be immediately satisfied by the router’s cache.

### 5.3. Results with the PlanetLab topology

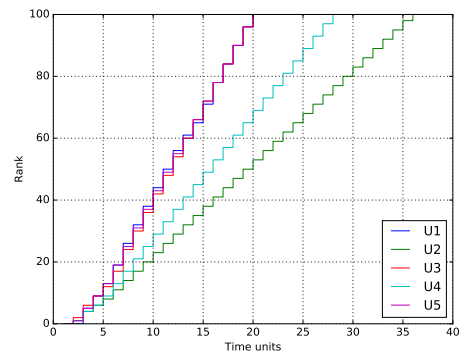
The behavior of MICN is then analyzed considering the PlanetLab topology from [18], with one source and five client nodes connected through a set of 20 intermediate caching routers.

---

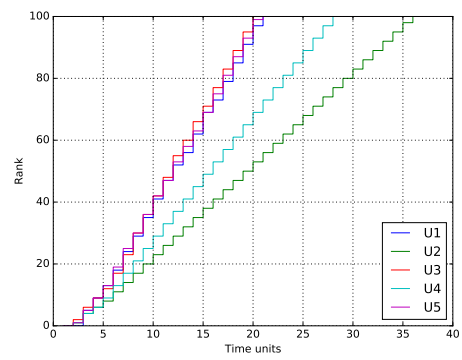
<sup>1</sup>Notice that NetCodCCN simulations in [18] consider only segment (data packets) losses, but here both interest and data packets are prone to losses.



(a) NetcodCCN



(b) MICN



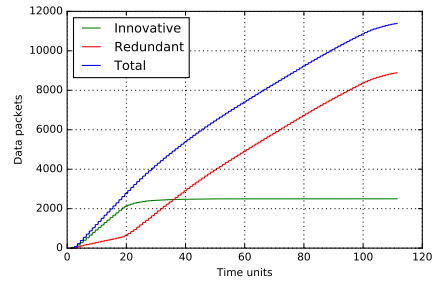
(c) MICN-IC

Figure 11: PlanetLab topology: Rank evolution as a function of time

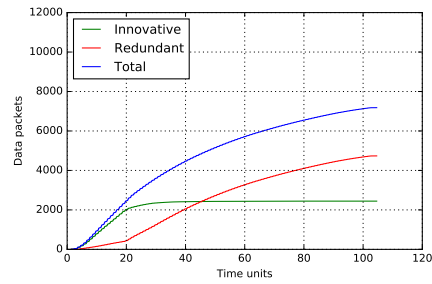


As seen in Fig. 11, with MICN, MICN-IC, and NetCodCCN, clients receive enough content to decode a generation at a rate above 95% of the maximum rate (provided by the min-cut between the source and the clients), as observed for the butterfly topology.

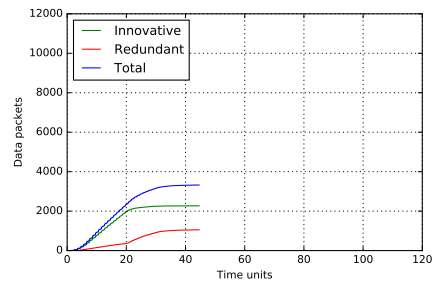
Fig. 12 illustrates that the cumulative number of data packets transmitted on all the links of the network as a function of time. NetCodCCN generates the most data traffic (also for the longer duration), followed by MICN. MICN-IC performs the best in terms of traffic, with respectively 2.16 and 3.42 times fewer transmitted data packets compared to MICN and NetCodCCN. In the PlanetLab topology, with the considered scenario, the amount of non-innovative packets dominates: about 80% of the content traffic with NetCodCCN is redundant (non-innovative). Some innovative packets might not be useful for the client because intermediate nodes of the network are unable to detect when the client has received all packets required to decode a generation.



(a) NetCodCCN



(b) MICN



(c) MICN-cancel

Figure 12: Data traffic in PlanetLab topology

In the PlanetLab topology, the pipeline size impacts the performance only when it is too small, as shown in Fig. 13. Increasing the pipeline size above 2 (MICN), 3 (MICN-IC), and 5 (NetCodCCN) do not bring additional benefit.

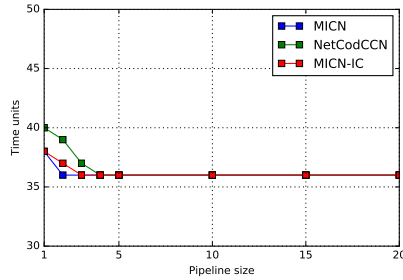


Figure 13: PlanetLab topology: Download time vs pipeline size, no losses

Fig. 14 shows the effect of transmission losses. The download time with MICN and MICN-IC increases almost linearly with the loss rate, compared to NetCodCCN, which increases faster when the loss rate is above 10%. In the PlanetLab topology, compared to the butterfly topology, MICN, MICN-IC, and NetCodCCN are all more robust to packet losses due to the more significant amount of redundant content traffic in the network which helps to compensate for the losses.

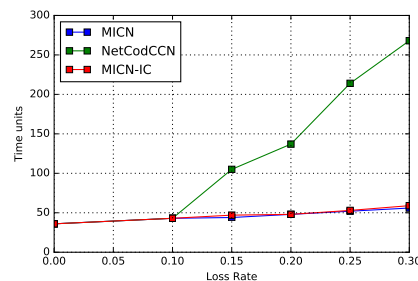


Figure 14: PlanetLab Topology: Download time vs transmission error rate

## 6. Conclusion

In this work, we propose a novel way of integrating NC and information-centric networking. The proposed MICN protocol is built around the MILIC construction that allows the clients to request content that belongs to predefined subsets by adding an index in the interest, that indicates the subset. This

interest naming allows the nodes to send multiple interests in parallel and ensures that linearly independent content satisfies each interest. In the considered scenarios, the clients download content close to their maximum capacity (like NetCodCCN). Nevertheless, thanks to interest cancellation, MILIC-IC limits the redundant data traffic considerably. This reduces the network load and leaves earlier free network resources to fetch contents from consecutive generations.

Our future research includes investigating improved interest forwarding algorithms to use the multiple active links better and reduce the data traffic by adjusting the number of outgoing interests.

#### *Acknowledgements*

This research was partly supported by Labex DigiCosme (project ANR11 LABEX0045DIGICOSME) operated by ANR as part of the program « Investissement d’Avenir » Idex Paris-Saclay (ANR11IDEX000302).

#### **References**

- [1] Cisco visual networking index: Global mobile data traffic forecast update, 2017-2022, Cisco White Paper (2019).
- [2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, R. L. Braynard, Networking named content, in: Proc. 5th ACM International Conference on Emerging Networking Experiments and Technologies, ACM, 2009, pp. 1–12.
- [3] A. Afanasyev, J. Shi, B. Zhang, L. Zhang, I. Moiseenko, Y. Yu, W. Shang, Y. Huang, J. P. Abraham, S. Dibenedetto, C. Fan, D. Pesavento, G. Grassi, G. Pau, H. Zhang, T. Song, H. B. Abraham, P. Crowley, S. O. Amin, V. Lehman, L. Wang, NFD Developer’s Guide, NDN, Technical Report NDN-0021 Rev. 10 (July) (2018).

- [4] L. Zhang, D. Estrin, J. Burke, V. Jacobson, J. D. Thorton, D. K. Smetters, B. Zhang, G. Tsudik, K. Claffy, D. Krioukov, D. Massey, C. Papadopoulos, T. Abdelzaher, L. Wang, P. Crowley, E. Yeh, Named Data Networking (NDN) Project, NDN, Technical Report NDN-0001 (October) (2010).
- [5] M.-J. Montpetit, C. Westphal, D. Trossen, Network coding meets information-centric networking: An architectural case for information dispersion through native network coding, in: Proc. 1st ACM Workshop on Emerging Name-Oriented Mobile Networking Design-Architecture, Algorithms, and Applications, 2012, pp. 31–36.
- [6] R. Ahlswede, N. Cai, S.-y. R. Li, R. W. Yeung, Network information flow, *IEEE Transactions on Information Theory* 46 (4) (2000) 1204–1216.
- [7] R. Koetter, M. Médard, An algebraic approach to network coding, *IEEE/ACM transactions on networking* 11 (5) (2003) 782–795.
- [8] R. Ahmed, M. F. Bari, S. R. Chowdhury, M. G. Rabbani, R. Boutaba, B. Mathieu,  $\alpha$ Route: Routing on Names, *IEEE/ACM Transactions on Networking* 24 (5) (2016) 3070–3083.
- [9] D. Posch, H. Hellwagner, B. Rainer, SAF: Stochastic adaptive forwarding in named data networking, *IEEE/ACM Transactions on Networking* 25 (2) (2017) 1089–1102.
- [10] P. Gusev, J. Burke, NDN-RTC: Real-time videoconferencing over named data networking, in: Proc. of the 2nd ACM Conference on Information-Centric Networking, 2015, pp. 117–126.
- [11] K. Matsuzono, H. Asaeda, C. Westphal, Network Coding for Content-Centric Networking / Named Data Networking: Requirements and Challenges, Internet-Draft draft-irtf-nwcr-g-nwc-ccn-reqs-03, Internet Engineering Task Force, work in Progress (Mar. 2020).  
URL <https://datatracker.ietf.org/doc/html/draft-irtf-nwcr-g-nwc-ccn-reqs-03>

- [12] P. A. Chou, Y. Wu, K. Jain, Practical network coding, *Proc. Annual Allerton Conference on Communication Control and Computing* 41 (1) (2003) 40–49.
- [13] K. Lei, S. Zhong, F. Zhu, K. Xu, H. Zhang, A NDN IoT Content Distribution Model with Network Coding Enhanced Forwarding Strategy for 5G, *IEEE Transactions on Industrial Informatics* 14 (6) (2017) 2725–2735.
- [14] Q. Wu, Z. Li, G. Xie, CodingCache: multipath-aware CCN cache with network coding, in: *Proc. 3rd ACM SIGCOMM Workshop on Information-centric Networking*, 2013, pp. 41–42.
- [15] Q. Wu, Z. Li, G. Tyson, S. Uhlig, M. A. Kaafar, G. Xie, Privacy-aware multipath video caching for content-centric networks, *IEEE Journal on Selected Areas in Communications* 34 (8) (2016) 2219–2230.
- [16] G. Zhang, Z. Xu, Combing CCN with network coding: An architectural perspective, *Computer Networks* 94 (2016) 219–230.
- [17] Y. Liu, S. Z. Yu, Network coding-based multisource content delivery in Content Centric Networking, *Journal of Network and Computer Applications* 64 (2016) 167–175.
- [18] J. Saltarin, E. Bourtsoulatze, N. Thomos, T. Braun, NetCodCCN: A network coding approach for content-centric networks, in: *Proc. of IEEE INFOCOM*, 2016, pp. 1–9.
- [19] W. X. Liu, S. Z. Yu, G. Tan, J. Cai, Information-centric networking with built-in network coding to achieve multisource transmission at network-layer, *Computer Networks* 115 (2017) 110–128.
- [20] K. Matsuzono, H. Asaeda, T. Turetli, Low latency low loss streaming using in-network coding and caching, in: *Proc. of IEEE INFOCOM*, IEEE, 2017, pp. 1–9.

- [21] M. Bilal, S.-G. Kang, Network-Coding Approach for Information-Centric Networking, *IEEE Systems Journal* 13 (2) (2018) 1376–1385.
- [22] J. Saltarin, E. Bourtsoulatze, N. Thomos, T. Braun, Adaptive Video Streaming With Network Coding Enabled Named Data Networking, *IEEE Transactions on Multimedia* 19 (10) (2017) 2182–2196.
- [23] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, B. Leong, A random linear network coding approach to multicast, *IEEE Transactions on Information Theory* 52 (10) (2006) 4413–4430.