GOALS AND PERSPECTIVES OF THE NETWORK CODING

Zoran Miličević[1], Zoran Bojković[2]
[1]Telecommunication and IT Department Genaralstaff of the SAF in Belgrade
[2]University of Belgrade

Abstract: *In this work the idea is to draw attention to highlights in the past decade of network coding research from the goals and perspectives point of view. In the same time, it is of importance to pave the way towards future theoretical and practical research and development in this area. In the first part, we review some network coding principles presenting in addition some deep connections between linear network coding and algebra. Also, it is pointed out that employing coding at the nodes network capacity can be increased compared with employing routing alone in a simple network topology known as butterfly network. The second part of this paper deals with network coding in peer-to-peer (P2P) networks. Here, the use of network coding is more realistic when applied in an operational on demand streaming system, operated at a large scale that accommodates millions of Internet users.*

Key words: *network coding, butterfly network, peer-to-peer (P2P) networks, computer networks.*

1. Introduction

Network coding (NC) is a new example in data transmission that combines coding with data propagation over a network. Linear network coding (LNC) adopts linear coding scheme at every node of the network and promises the optimal data transmission rate from the source to all receivers. Linearity enhances engineering simplicity which leads to wide applicability. The idea is that the bits in information flows can be mixed as long as receiving hosts have received sufficient evidence to reconstruct the original packets from the sending hosts. This idea is called network coding. Since 2000 there has been large number of high quality research papers in the open literature on the topic of network coding, attracting the interest from a diverse group of researchers and positioners in information theory communications and computer networking. There exists tremendous potentional for the theory of NC to affect the design of next generation network protocols. For example, the Chineese University of Hong Kong has recently established the Institute of Network Coding which envisions and is working towards new technological advances of network coding in the Internet. NC my find practical applications in a wide variety of communication networks and systems, starting from cloud computing systems to mobile applications. We start with presentation of NC. It will be pointed out that employing coding at the nodes of a network; the capacity can be increased compared with employing routing only. On the other hand, a key development of network coding is the fact that the maximum multicast communication capacity can be achieved using only LNC.

The second part of this work deals with coding across data which has been put to extensive use in today's communication systems at the link level. Also, it has been emphasized that applications involving peer-to-peer networks, such as video streaming, may become the most promising scenario for NC to be deployed in real – world systems.

2. Network coding presentation and important subject

Whereas the theory of the NC has been well understood for a single telecommunication session, NC for multiple concurrent communication session turned out to be an open theoretical challenge.

NC research has been focused in two general directions: (A) applications of network coding to practical networks and (B) understanding the theoretical possibilities and limitations of network coding.

For a particular network node *n*, there four types of data to deal with [1, 2]:

a) data from packets received along in edges of *n*
b) data from images that originated at *n*
c) data from packets sent along out – edges of *n*
d) data from messages demanded by *n*

Errors may occur in the communication network with the use of NC. Such errors may be random errors, erasures due to lost packets, or more seriously errors caused by intentional attacks by malicious nodes in the network. The Transmission Control Protocol (TCP) uses the lack of reordering of acknowledgement packets to indicate errors and to induce retransmission from the sending host. If NC is applied, a new mechanism needs to be designed. Rather that acknowledging the receipt of the packets in their original form, the receiving host can communicate the degree of freedom in batch of coded packets it has received so far to the sending host.

The applications that involve P2P networks, such as file sharing and video streaming, may become the most promising scenario for NC to be deployed in real-world systems. In shows protocol used to apply NC in a large scale on demand video streaming system that has already been deployed in a producing setting.

Security is important subject that deserves significant attention from both theoretical and practical point of view. There are three main approaches to security: computational, physical, and information theoretic. A system is computationally secure if it is infeasible for an adversary to break the system. Physical security depends on physical properties such as entranglement to detect or prevent eavesdropping. The information – theoretic approach, pioneered by Shannon [3] determines the maximum transmission rate such that it is impossible for the adversary to break the system regardless of the computational resources available or the physical embodiment of the data [4]. A code is perfectly secure if the wiretapper obtains no information about the transmitted message from the eavesdropped messages. A secure network code is called ***r***-

secure if a wiretapper, how can eavesdrop any ***r*** channels in the network, gains no knowledge about the source message.

3. Butterfly network

The concept of network coding (NC) will be presented using the example of butterfly network. Remember that the conventional mode of information delivery through a communication network is routing, where an intermediate node performs store-and-forward, just like a post office in the process of forwarding a mail. Thus, every outgoing message is a copy of a received message.

The NC solution for the butterfly network suggests that for a general single source NC problem, the max-flow bound can be achieved by linear network coding (LNC) [5, 6]. In the classical butterfly network shown in Figure 1, source A intends to multicast bits x and y to receivers B and C, simultaneously. Every arrow represents a channel for transmitting one data bit, while two bits x and y are to be transmitted from node A to both nodes A and C.

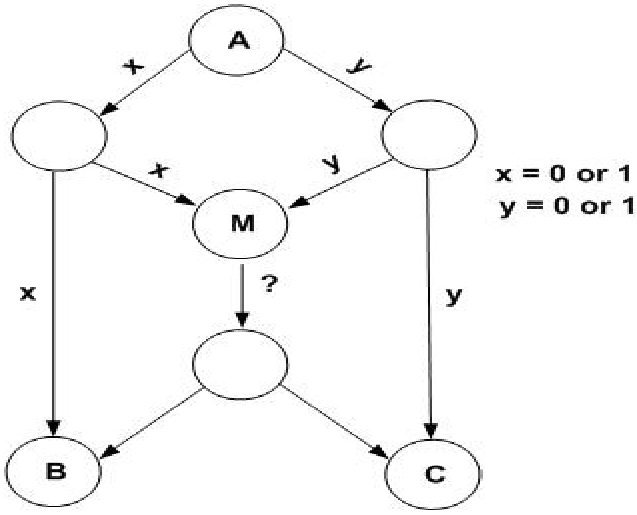

Figure 1. Classical butterfly network. Every channel transmits one bit.

When node M receives x and y, it can only forward one bit. If x is forwarded, then B cannot receive y. If y is forwarded, then C cannot get x. This dilemma can be avoided if the data transport node is not store-and-forwarded. Then exists the traffic jam of the middle channel. A simple solution is for M to send down the binary symbol x+y, which represents weather x and y are equal as shown in Figure 2. When node B receives both x and x+y, it can decode y. Similarity, node C can decode x. The binary symbol x+y is a mathematical function of x and y. Calculation of a function from received data is called coding. Also, this shows the merit of mixed coding among multiple messages at an intermediate node and is called NC. In algebra x+y is called the binary sum of x and y. In general terms of linear algebra, this is the linear sum 1x+1y over the binary field. Thus, the calculation of x+y is not only a form of coding but also belongs to the more restricted form of linear coding.

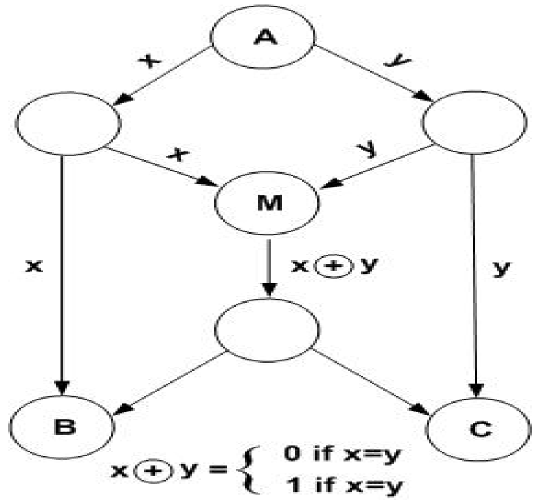

Figure 2. The network coding solution.

An application of the same technique in wireless communication is shown in Figure 3. In fact, the some technique as in the butterfly network example to message exchange between two wireless devices via a middle relay is used. Two devices intend to exchange one-bit information with each other through an intermediate relay. This relay can only either receive or broadcast one bit at a time. With store-and-forward, it takes the relay four steps to complete the exchange. With network coding, it takes only three steps.

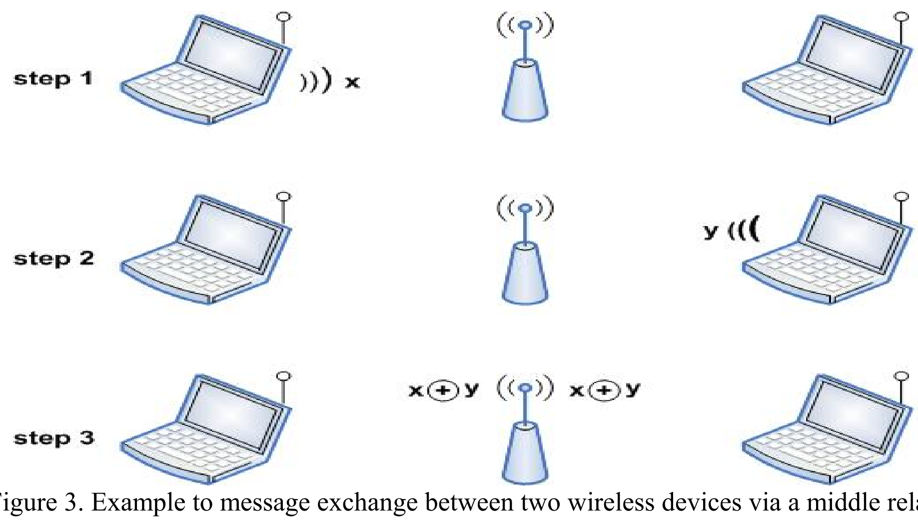

Figure 3. Example to message exchange between two wireless devices via a middle relay.

4. Network coding in practice

The theory of NC performance promises significant benefits in network, especially in lossy networks and in multicast and multipath scenarios. To realize these benefits in practice, we need to see how coding across packets interact with the acknowledgment (ACK) – based flow control mechanism that forms a central part of today's Internet protocols such as transmission control protocol (TCP). Current approaches such as rate-less codes and batch-based coding are not compatible with TCP's retransmission and sliding-window mechanisms.

The concept of coding across data has been put to expensive use in today's communication systems at the link level, due to practical coding schemes that are known to achieve data rates very close to the fundamental limit (capacity) of the additive white Gaussian noise channel. Although the fundamental limits for many multi user information theory problems have yet to be established, it is well known that there are significant benefits to coding beyond the link level.

As an example, consider multicasting over a network of broadcast – mode link in wireless systems. Due to the broadcast nature of the medium, a transmitted packet is likely to be received by several nodes. If one of the nodes experienced a bed channel state and thereby lost the packet, then a simple retransmission strategy may not be the best option, since the retransmission is useless from the view point of the other receivers that have already received the packet. In Figure 4, node A broadcasts 2 packets to nodes B and C. In time – slot 1, only node B receives packet p_1, and in the second slot, only node C receives packet p_2. At this point, if instead of retransmission p_1 and p_2, node A is allowed to mix the information and send a single packet containing the bitwise of packet p_1 and p_2. Both B and C receive their missing packet in just one additional time slot. It can be seen that if we allow coding across packets, it is possible to convey new information simultaneously to connected receivers.

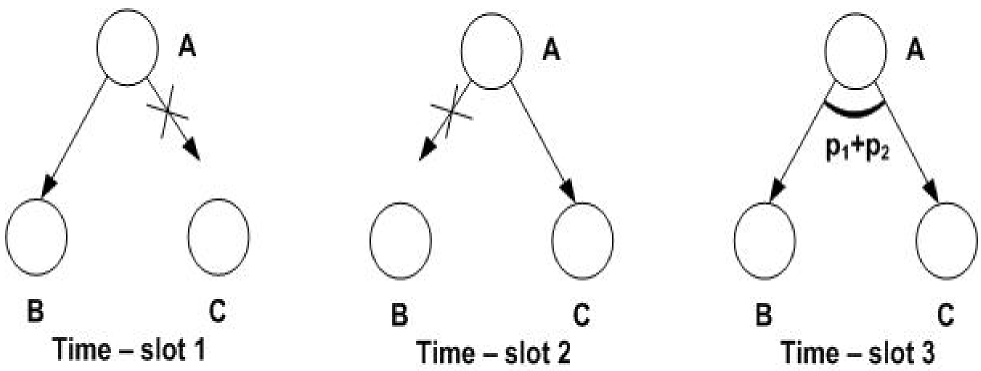

Figure 4. Coding through a broadcast – mode link.

Despite the potential of NC, we still seem far from seeing wide spread implementation of NC across networks. A major reason for this is the incremental deployment problem. The question often arises is how to naturally add NC to existing systems, and to understand ahead of time the actual efforts of NC. There have been several advances in bridging the gap between theory and practice in this space. For example, the distributed linear coding idea is a significant step toward a robust implementation. The next idea was of embedding the coefficients used in the linear combination in the packet header, together with the notation of grouping packets into batches for coding together [7]. The use of network deployment remains to increase adoption of NC in practical settings. A protocol that brings out the benefits of NC while requiring very little change in the existing protocol/stack is welcome.

5. Network coding in peer-to-peer networks

In NC network nodes between the source and the receivers are able not only to relay and replicate data packets, but also code them using randomly generated coding coefficients. Random network coding minimizes the network flow rates in multicast sessions. Random network coding has got practical and real applications in peer-to-peer (P2P) networks, in which overlay network topologies are formed among participating and hosts, called peers.

Now, the question is whether theoretical benefits of NC may become useful in practical real networking systems. To practically implement NC, one needs to address the challenges of computing coding coefficients to be used by each of the intermediate nodes in a session, so that coded packets at the receivers are guaranteed to be decoded. This process is called code assignment. Although deterministic code assignment algorithm have been proposed and shown to be polynomial time algorithms [10], they require expensive exchanges of control packets.

An excellent scenario where NC may be applied in practice is peer-to-peer (P2P) network. In P2P networks, end hosts, from servers to smart phones called "peers" organize themselves in overlay topologies, in which packet transmission on each of the overlay links is free of errors, thanks to transport protocols used in the Internet, such as transmission control protocol (TCP). Also, peers are computing devices that are potentially capable of coding in software without the need of revising existing switches and routers in the Internet.

The P2P block content distribution systems allow peers to collaborate with one another so that large files can be distributed from one peer to a large number of subscribing receivers, without the aid of dedicated servers. A file is divided in blocks, and peers connect with one another in a random mesh topology, exchanging these blocks with random "*gossiping*". In gossiping, each peer transmits a subset of the blocks it has obtained to a subset of its neighbors that are selected using randomized algorithms. Such random gossiping on random mesh topologies is simple to implement, resilient to the level of volatility caused by peer arrivals and departures, and has been shown to achieve good performance from a theoretical perspective. For these reasons, real P2P media streaming systems, currently used by millions of users, are also designed by following such a design principle.

P2P network topologies are formed by hosts who connect with one another using overlay links. These technologies are formed for a particular objective. One of the objectives is to distribute some bulk content, such as a large file, from one peer to other subscribing peers in a topology.

Unlike bulk content distribution live and on-demand streaming systems require the timely delivery and playback of time – sensitive data streams, typically video streams. The motivation of using P2P paradigm is to conserve the bandwidth consumed at dedicated streaming servers since peers are able to contribute their uplink bandwidth to the system by uploading media streams to one another. In P2P live streaming systems, where media channels are broadcast live to participating peers, the only fundamental difference is that a dynamic sliding window of blocks over time needs to be distributed in a streaming fashion, rather than a fixed number of blocks in a static file.

There are limitations to the benefits that random NC may bring the design of P2P content distribution protocols. Since random NC can only be performed on blocks within the same generation, reconciliation between a pair of neighboring peers may still be necessary across the boundary of generations. It means that, rather than collecting fine-grained blocks, now a receiver only needs to collect all the distinct coarse-grained generations that constitute the file to be distributed. It mitigates the problem of locating rare portions of a file that may be less available in the entire P2P network.

6. Conclusion

In this work important milestones toward bringing theoretical benefits of network coding (NC) to practical implementation are introduced. In NC, intermediate nodes forward linear contributions of information packets, while destination nodes collect a sufficient number of linearly independent functions so as to be able to recover their desired information packets. Physical layer network coding is an intuitively pleasing enhancement to network coding. It starts from the insight that it can be much more efficient for a node to directly learn the linear function of several packets, rather than having to first learn each packet separately and then evaluating the function. Also, it is presented that employing coding at the network nodes, the capacity is increased comparing the case where routing alone is employed. On the other hand, applications such as file sharing and video streaming that involve peer-to-peer networks may be the most promising area for using network coding. Thanks to optimized implementations of NC in software, it is feasible to perform NC on a wide range of devices including servers, desktops, notebook computers and mobile phones. Of course, the additional energy consumed to perform NC can be well justified.

References

[1] R. Ahlswede, H. Cai, S.-Y.R. Li, R.W. Yeung, ”Network information flow”, *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[2] R. Dougherty, C. Freiling, K. Zeger, ”Network coding and matroid theory”, *Proc. of the IEEE*, vol. 99, no. 3, pp. 388-405, March 2011.

[3] C. E. Shannon, ”Communication theory of secrecy systems”, Bell Syst. Tech. Journ., vol. 28, pp. 656-715, 1949.

[4] N. Cai, T. Chan, ”Theory of secure network coding”, *Proc. of the IEEE*, vol. 99, no. 3, pp. 421-437, March 2011.

[5] S.-Y. R. Li, R. W. Yeung, N. Cai, ”Linear network coding”, *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371-381, February 2003.

[6] R. Koelter, M. Medrad, "An algebraic approach to network coding", *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782-795, October 2003.
[7] P. A. Chou, Y. Wu, K. Jain, "Practical network coding", *Proc. Allerton. Conf. Commun. Control Comput.*, pp. 44-49, 2003.
[8] S. Katty, et al., "x OR s in the air: Practical wireless network coding", *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 497-510, June 2008.
[9] S. Chachulski, et al., "Trading structure for randomness in wireless opportunistic routing", *Proc. ACM SIE COMM.*, pp. 169-180, August 2007.
[10] P. Sanders, S. Egner, L. Tolhnizen, "Polynomial time algorithm for network information flow", in *Proc. 15th ACM Symp. Parallelism Algorithms Architectures*, pp. 286-294, June 2003.

Sadržaj: *Ideja u radu je da se prikažu istraživanja u oblasti mrežnog kodovanja sa aspekta ciljeva i dalje perspektive. Istraživanja koja se sprovode u poslednjih desetak godina u eminentnim institucijama u svetu ukazuju na buduće pravce kojima treba ići kako u teorijskom, tako i u praktičnom smislu. U prvom delu rada prikazani su neki principi mrežnog kodovanja. Ukazuje se na čvrstu povezanost između linearnog mrežnog kodovanja i algebre. Takođe je naglašeno da primena kodovanja u čvorovima žičnih i bežičnih mreža utiče na povećanje njihovog kapaciteta u poređenju sa slučajevima kada se koristi samo rutiranje. Kao primer uzeta je jednostavna mrežna topologija (butterfly mreža). Drugi deo rada bavi se mrežnim kodovanjem u slučaju mreže tipa P2P (peer-to-peer). U ovom slučaju, primena mrežnog kodovanja ima vrlo praktičan značaj kod tzv. operacionih "streaming" sistema kojima se služe milioni Internet korisnika.*

Ključne reči: *mrežno kodovanje, butterfly mreža, P2P mreža, računarske mreže.*

CILJEVI I PERSPEKTIVE MREŽNOG KODOVANJA

Zoran Miličević, Zoran Bojković