Physical-Layer Network Coding

Shengli Zhang
Dept. of Information Engineering,
The Chinese University of HK,
Hong Kong SAR, P.R.China
slzhang5@ie.cuhk.edu.hk

Soung Chang Liew
Dept. of Information Engineering,
The Chinese University of HK,
Hong Kong SAR, P.R.China
soung@ie.cuhk.edu.hk

Patrick P. Lam
Dept. of Information Engineering,
The Chinese University of HK,
Hong Kong SAR, P.R.China
lampk3@ie.cuhk.edu.hk

Abstract

A main distinguishing feature of a wireless network compared with a wired network is its broadcast nature, in which the signal transmitted by a node may reach several other nodes, and a node may receive signals from several other nodes simultaneously. Rather than a blessing, this feature is treated more as an interference-inducing nuisance in most wireless networks today (e.g., IEEE 802.11). The goal of this paper is to show how the concept of network coding can be applied at the physical layer to turn the broadcast property into a capacity-boosting advantage in wireless ad hoc networks. Specifically, we propose a physical-layer network coding (PNC) scheme to coordinate transmissions among nodes. In contrast to "straightforward" network coding which performs coding arithmetic on digital bit streams after they have been received, PNC makes use of the additive nature of simultaneously arriving electromagnetic (EM) waves for equivalent coding operation. PNC can yield higher capacity than straight-forward network coding when applied to wireless networks. We believe this is a first paper that ventures into EM-wave-based network coding at the physical layer and demonstrates its potential for boosting network capacity. PNC opens up a whole new research area because of its implications and new design requirements for the physical, MAC, and network layers of ad hoc wireless stations. The resolution of the many outstanding but interesting issues in PNC may lead to a revolutionary new paradigm for wireless ad hoc networking.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication.

General Terms

Algorithms, Performance, Design, Theory.

Keywords

network coding; wireless networks; ad hoc netrworks; cooperative transmission; relay networks; multiple-access networks.

1. INTRODUCTION

At the physical layer of wireless networks, all data are transmitted through electromagnetic (EM) waves. Wireless communications have many characteristics not found in its wired counterpart. One of them is the broadcast nature of wireless links: transmission of the EM signals from a sender is often received by more than one node. At the same time, a receiver may be receiving EM signals transmitted by multiple nodes simultaneously.

These characteristics may cause interference among signals. While interference has a negative effect on wireless networks in general, the effect on the throughput of *multi-hop* ad hoc networks is particularly noticeable. For example, in conventional 802.11 networks, the theoretical throughput of a multi-hop flow in a linear network is less than 1/4 of the single-hop case due to the "self interference" effect, in which the packet of a hop collides with another packet of a nearby hop [1, 2] for the same traffic flow.

Most communication system designs try to either reduce or avoid interference (e.g., through receiver design or transmission scheduling [3]). Instead of treating interference as a nuisance to be avoided, we can actually embrace interference to improve throughput performance. To do so in a multi-hop network, the following goals must be met:

1. A relay node must be able to convert simultaneously received signals into interpretable output signals to be relayed to their final destinations.
2. A destination must be able to extract the information addressed to it from the relayed signals.

Network coding's capability of combining and extracting information through simple Galois field GF(2^n) additions [4, 5] provides a good foundation to meet such goals. Network coding arithmetic is generally only applied on bits that have already been detected. Specifically, it cannot be used to resolve the interference of simultaneously arriving EM signals at the receiver. So, criterion 1 above cannot be met.

This paper proposes the use of Physical-layer Network Coding (PNC). The main idea of PNC is to create an apparatus similar to that of network coding, but at the lower physical layer that deals with EM signal reception and modulation. Through a proper modulation-and-demodulation technique at relay nodes, additions of EM signals can be mapped to GF(2^n) additions of digital bit streams, so that the interference becomes part of the arithmetic operation in network coding.

2. ILLUSTRATING EXAMPLE: A THREE-NODE WIRELESS LINEAR NETWORK

Consider the three-node linear network in Fig. 1. N_1 (Node 1) and N_3 (Node 3) are nodes that exchange information, but they are out of each other's transmission range. N_2 (Node 2) is the relay node between them.

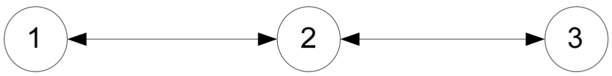

Figure 1. A three-node linear network

This three-node wireless network is a basic unit for cooperative transmission and it has previously been investigated extensively [6-8]. In cooperative transmission, the relay node N_2 can choose different transmission strategies, such as Amplify-and-Forward or Decode-and-Forward [6], according to different Signal-to-Noise (SNR) situations. This paper focuses on the Decode-and-Forward strategy. We consider frame-based communication in which a time slot is defined as the time required for the transmission of one fixed-size frame. Each node is equipped with an omni-directional antenna, and the channel is half duplex so that transmission and reception at a particular node must occur in different time slots.

Before introducing the PNC transmission scheme, we first describe the traditional transmission scheduling scheme and the "straightforward" network-coding scheme for mutual exchange of a frame in the three-node network [8, 9].

2.1 Traditional Transmission Scheduling Scheme

In traditional networks, interference is usually avoided by prohibiting the overlapping of signals from N_1 and N_3 to N_2 in the same time slot. A possible transmission schedule is given in Fig. 2. Let S_i denote the frame initiated by N_i. N_1 first sends S_1 to N_2, and then N_2 relays S_1 to N_3. After that, N_3 sends S_3 in the reverse direction. A total of four time slots are needed for the exchange of two frames in opposite directions.

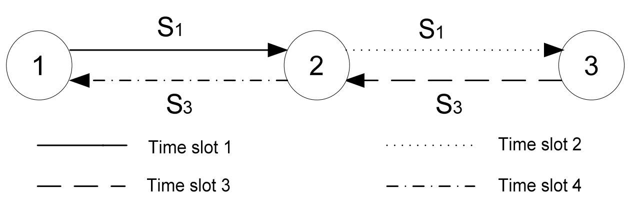

Figure 2. Traditional scheduling scheme

2.2 Straightforward Network Coding Scheme

Ref. [8] and [9] outline the straightforward way of applying network coding in the three-node wireless network. Fig. 3 illustrates the idea. First, N_1 sends S_1 to N_2 and then N_3 sends frame S_3 to N_2. After receiving S_1 and S_3, N_2 encodes frame S_2 as follows:

$$S_2 = S_1 \oplus S_3 \tag{1}$$

where $\oplus$ denote bitwise exclusive OR operation being applied over the entire frames of S_1 and S_3. N_2 then broadcasts S_2 to both N_1 and N_3. When N_1 receives S_2, it extracts S_3 from S_2 using the local information S_1, as follows:

$$S_1 \oplus S_2 = S_1 \oplus (S_1 \oplus S_3) = S_3 \tag{2}$$

Similarly, N_2 can extract S_1. A total of three time slots are needed, for a throughput improvement of 33% over the traditional transmission scheduling scheme.

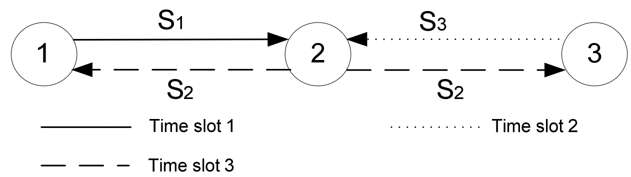

Figure 3. Straightforward network coding scheme

2.3 Physical-Layer Network Coding (PNC)

We now introduce PNC. Let us assume the use of QPSK modulation in all the nodes. We further assume symbol-level and carrier-phase synchronization, and the use of power control, so that the frames from N_1 and N_3 arrive at N_2 with the same phase and amplitude (Additional discussions on synchronization issues can be found in section 6). The combined bandpass signal received by N_2 during one symbol period is

$$\begin{aligned} r_2(t) &= s_1(t) + s_3(t) \\ &= [a_1\cos(\omega t) + b_1\sin(\omega t)] + [a_3\cos(\omega t) + b_3\sin(\omega t)] \\ &= (a_1 + a_3)\cos(\omega t) + (b_1 + b_3)\sin(\omega t) \end{aligned} \tag{3}$$

where $s_i(t)$, i = 1 or 3, is the bandpass signal transmitted by N_i and $r_2(t)$ is the bandpass signal received by N_2 during one symbol period; a_i and b_i are the QPSK modulated information bits of N_i; and ω is the carrier frequency. Then, N_2 will receive two baseband signals, in-phase (I) and quadrature phase (Q), as follows:

$$\begin{aligned} I &= a_1 + a_3 \\ Q &= b_1 + b_3 \end{aligned} \tag{4}$$

Note that N_2 cannot extract the individual information transmitted by N_1 and N_3, i.e., a_1, b_1, a_3 and b_3 , from the combined signal I and Q. However, N_2 is just a relay node. As long as N_2 can transmit the necessary information to N_1 and N_3 for extraction of a_1, b_1, a_3, b_3 over there, the end-to-end delivery of information will be successful. For this, all we need is a special modulation/demodulation mapping scheme, referred to as *PNC mapping* in this paper, to obtain the equivalence of GF(2) summation of bits from N_1 and N_3 at the physical layer.

Table 1 illustrates the idea of PNC mapping. Recall that a QPSK data stream can be considered as two BPSK data streams: an in-phase stream and a quadrature-phase stream. In Table 1, $s_j^{(I)} \in \{0, 1\}$ is a variable representing the in-phase data bit of N_j and $a_j \in \{-1, 1\}$ is a variable representing the BPSK modulated bit of $s_j^{(I)}$ such that $a_j = 2s_j^{(I)} - 1$. A similar table (not shown here) can also be constructed for the quadrature-phase data by letting $s_j^{(Q)} \in \{0,\ 1\}$ be the quadrature data bit of N_j, and $b_j \in \{-1,\ 1\}$ be the BPSK modulated bit of $s_j^{(Q)}$ such that $b_j = 2s_j^{(Q)} - 1$.

Table 1. PNC Mapping: modulation mapping at N_1, N_2; demodulation and modulation mappings at N_3

Modulation mapping at N_1 and N_3,				Demodulation mapping at N_2		
				Input	Output	
Input		Output			Modulation mapping at N_2	
					Input	Output
$s_1^{(I)}$	$s_3^{(I)}$	a_1	a_3	$a_1 + a_3$	$s_2^{(I)}$	a_2
1	1	1	1	2	0	-1
0	1	-1	1	0	1	1
1	0	1	-1	0	1	1
0	0	-1	-1	-2	0	-1

With reference to Table 1, N_2 obtains the information bits:

$$s_2^{(I)} = s_1^{(I)} \oplus s_3^{(I)}; \qquad s_2^{(Q)} = s_1^{(Q)} \oplus s_3^{(Q)} \tag{5}$$

It then transmits

$$s_2(t) = a_2 \cos(\omega t) + b_2 \sin(\omega t) \tag{6}$$

Upon receiving $s_2(t)$, N_1 and N_3 can derive $s_2^{(I)}$ and $s_2^{(Q)}$ by ordinary QPSK demodulation. The successively derived $s_2^{(I)}$ and $s_2^{(Q)}$ bits within a time slot will then be used to form the frame S_2. In other words, the operation $S_2 = S_1 \oplus S_3$ in straightforward network coding can now be realized through PNC mapping.

As illustrated in Fig. 4, PNC requires only two time slots for the exchange of one frame (as opposed to three time slots in straightforward network coding and four time slots in traditional scheduling).

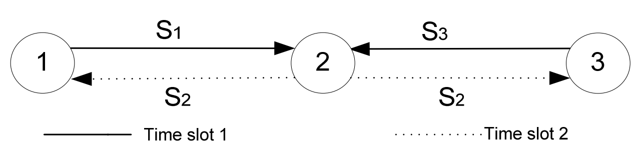

Figure 4. Physical layer network coding

2.4 Performance Comparison

We now analyze the bit error rate (BER) performance of PNC as described above. Suppose the received signal energy for one bit is unity, and the noise is Gaussian white with density $N_0/2$. For frames transmitted by N_2, the BER generated by PNC is simply the standard BPSK modulation $Q(\sqrt{2/N_0})$ [10], where $Q(.)$ is the complementary cumulative distribution function of the zero-mean, unit-variance Gaussian random variable, which is identical to the BER in traditional transmission or straightforward network coding.

For frames transmitted by N_1 and N_3, the BER performance of PNC can be derived as follows. According to Table 1, the in-phase signal space is {-2, 0, 2} with corresponding probabilities of 25%, 50%, 25% respectively. Applying the maximum posterior probability criterion [10], we can obtain the optimal decision rule: when the received signal is less than $\gamma_1 = -1 - \frac{N_0}{4}\ln(1+\sqrt{1-e^{-8/N_0}})$, we declare $a_1 + a_3$ to be –2; when the received signal is more than $\gamma_2 = 1 + \frac{N_0}{4}\ln(1+\sqrt{1-e^{-8/N_0}})$, we declare $a_1 + a_3$ to be 2; otherwise, it is assumed to be 0. According to Table 1, a_2 is –1 for $a_1 + a_3 = 2$ or $a_1 + a_3 = -2$. Thus, the BER can be derived as follows:

$$\begin{aligned} BER = {} & \frac{1}{2}\int_{-\infty}^{\gamma_1} \frac{1}{\sqrt{\pi N_0}} \exp(\frac{-r^2}{N_0})\, dr + \frac{1}{2}\int_{\gamma_2}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp(\frac{-r^2}{N_0})\, dr \\ & + \frac{1}{4}\int_{\gamma_1}^{\gamma_2} \frac{1}{\sqrt{\pi N_0}} \exp(\frac{-(r+2)^2}{N_0})\, dr + \frac{1}{4}\int_{\gamma_1}^{\gamma_2} \frac{1}{\sqrt{\pi N_0}} \exp(\frac{-(r-2)^2}{N_0})\, dr \end{aligned}$$

where r is the received in-phase signal at N_2. We plot the BER performance of PNC modulation and regular BPSK modulation in Fig. 5. We can see that the PNC modulation scheme

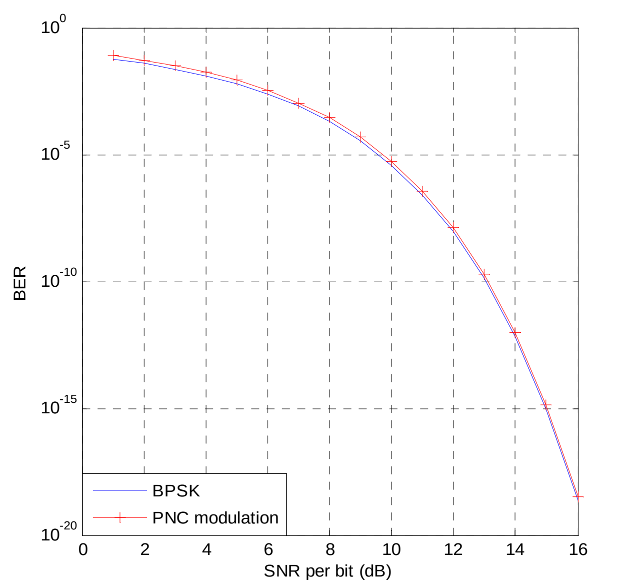

Figure 5. BER for standard BPSK and PNC modulation

has a slightly worse BER. However, when the SNR is larger than 10dB, the SNR penalty of PNC modulation is less than 0.1 dB. For the sake of simplicity, henceforth, we will ignore this small SNR penalty and assume PNC to have the same BER performance as the traditional 802.11 and straightforward network coding schemes.

The last paragraph relates to the BER for the reception at N_3. Let us assume the per-hop BER P_e is small. The end-to-end PNC BER for the transmission in one direction is approximately $2P_e$. The traditional transmission scheme has the same end-to-end BER. The straightforward network coding scheme, however, has a larger BER of approximately $3P_e$, since the integrity of the three transmissions by N_1, N_3, and N_2 must be intact for the extraction of information in one direction.

For simplicity, let us assume similar BER performance for the three schemes. For a frame exchange, PNC requires two time slots, 802.11 requires four, while straightforward network coding requires three. Therefore, PNC can improve the system throughput of the three-node wireless network by a factor of 100% and 50% relative to traditional transmission scheduling and straightforward network coding, respectively.

3. GENERAL PNC MODULATION-DEMODULATION MAPPING PRINCIPLE

A specific example of PNC mapping scheme has been constructed in Table 1 for the relay node in a 3-node linear network. We now generalize the PNC mapping principle.

Let us consider the three-node linear network scenario depicted in Fig. 4 again, but now look deeper into its internal operation as shown in Fig. 6. Let M denote the set of digital symbols, and let $\oplus$ be the general binary operation for network-coding arithmetic (note that $\oplus$ is not necessary the bitwise XOR hereinafter). That is, applying $\oplus$ on $m_i, m_j \in M$ gives $m_i \oplus m_j = m_k \in M$. Next, let E denote the set of modulated symbols in the EM-wave domain. Each $m_i \in M$ is mapped to a modulated symbol $e_i \in E$. Let $f : M \to E$ denote the modulation mapping function such that $f(m_i) = e_i$, $\forall m_i$. Note that $f : M \to E$ is a one-to-one mapping.

In the EM-wave domain, two signals may combine to yield a composite signal at the receiver. Let $\odot$ represent the binary combination operation. That is, combination of $e_i, e_j \in E$ yields $e_i \odot e_j = e'_k \in E'$, where E' is the domain after the

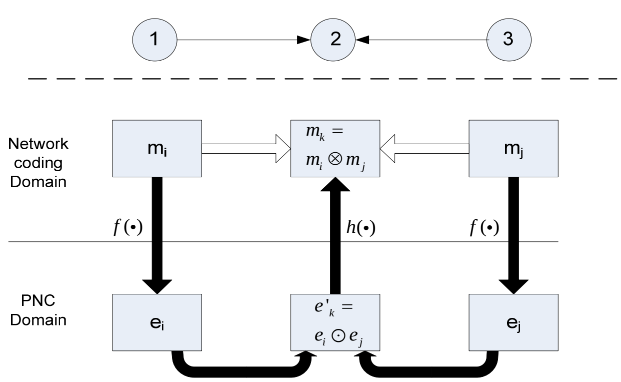

Figure 6. Illustration of PNC mapping

binary operation $\odot$. Note that E' is not the same as E and has a higher cardinality than A. For example, for 4-PAM, $E = \{-3, -1, 1, 3\}$, and $E' = \{-6, -4, -2, 0, 2, 4, 6\}$. For BFSK, $E = \{f_1, f_2\}$, and $E' = \{f_1, f_1 \text{ and } f_2, f_2\}$, where f_1 and f_2 are the constituent frequencies.

Each $e'_k \in E'$ received by the relay node must be mapped to a demodulated symbol $m_k \in M$. Let $h : E' \to M$ denote the demodulation mapping function such that $h(e'_k) = m_k$. Note that $h : E' \to M$ is a many-to-one mapping since the cardinality of E' is larger than that of M.

To summarize, a PNC transmission scheme consists of the following:

a) Network code specified by M and $\oplus$.

b) One-to-one modulation mapping, $f : M \to E$.

c) Many-to-one demodulation mapping, $h : E' \to M$.

Note that while the choices of M, $\oplus$, $f : M \to E$, and $h : E' \to M$ are up to the network designer, $\odot$ and E' are not because they relate to the fundamental characteristics of EM-wave. Now, there are many possibilities for 1 and 2 above. An interesting question is that, given (M, $\oplus$, $f : M \to E$), whether we can find an appropriate $h : E' \to M$ to realize PNC. More precisely, for a network code and a modulation scheme, we have the following PNC mapping requirement:

PNC Mapping Requirement: Given (M, $\oplus$, $f : M \to E$), there exists $h : E' \to M$ such that for all $m_i, m_j \in M$, if $m_i \oplus m_j = m_k$, then $h(e_i \odot e_j) = m_k$. That is, $h(f(m_i) \odot f(m_j)) = m_k$.

Fig. 6 illustrates the above requirement, in which the network-coding operation (white arrows) is realized by the PNC operation (dark arrows).

The following proposition specifies the characteristics that the modulation scheme $f : M \to E$ must possess in order that an appropriate $h : E' \to M$ can be found.

Proposition 1: Consider a modulation mapping $f : M \to E$. Suppose that f has the characteristic that $e_i \odot e_j = e_p \odot e_q$ implies $m_i \oplus m_j = m_p \oplus m_q$. Then a demodulation mapping $h : E' \to M$ can be found such that the PNC Mapping Requirement is satisfied. Conversely, if $e_i \odot e_j = e_p \odot e_q$ but $m_i \oplus m_j \neq m_p \oplus m_q$, then $h : E' \to M$ that satisfies the PNC Mapping Requirement does not exist.

Proof: For a given $e'_k \in E'$, one or more pairs of (e_i, e_j) can be found such that $e_i \odot e_j = e'_k$. If the condition "$e_i \odot e_j = e_p \odot e_q$ implies $m_i \oplus m_j = m_p \oplus m_q$" is satisfied, for any pair of such (e_i, e_j), $f^{-1}(e_i) \oplus f^{-1}(e_j)$ has the same value as $m_i \oplus m_j$, where $f^{-1}(\cdot)$ is the reverse mapping of the one-to-one mapping $f(\cdot)$. Therefore, $h(e'_k)$ can simply be $f^{-1}(e_i) \oplus f^{-1}(e_j)$ to satisfy the PNC Mapping Requirement. Conversely, suppose that "$e_i \odot e_j = e_p \odot e_q$ but $m_i \oplus m_j \neq m_p \oplus m_q$". According to the *PNC Modulation-Demodulation Requirement*, the appropriate mapping $h : E' \to M$ must produce $m_i \oplus m_j = h(e_i \odot e_j) = h(e'_k) = h(e_p \odot e_q) = m_p \oplus m_q$, which contradicts the condition.

4. PNC IN GENERAL REGULAR LINEAR NETWORK

In the preceding sections, we have illustrated the basic idea of PNC with a three-node linear wireless network. In this section, we consider the general regular linear network with more than three nodes. For simplicity, we assume the distance between any two adjacent nodes is fixed at *d*.

As will be detailed later, when applying PNC on the general linear network, each node transmits and receives alternately in successive time slots; and when a node transmits, its adjacent nodes receive, and vice versa (see Fig. 7). Let us briefly investigate the signal-to-inference ratio (SIR) given this transmission pattern to make sure that it is not excessive. Consider the worst-case scenario of an infinite chain. We note the following characteristics of PNC from a receiving node's point of view:

a) The interfering nodes are symmetric on both sides.

b) The simultaneous signals received from the two adjacent nodes do not interfere due to the nature of PNC.

c) The nodes that are two hops away are also receiving at the same time, and therefore will not interfere with the node.

Therefore, the two nearest interfering nodes are three hops away. We have the following SIR:

$$SIR = \frac{P_0 / d^{\alpha}}{2 * \sum_{l=1}^{\infty} P_0 / [(2l+1)d]^{\alpha}}$$

where P_0 is the common transmitting power of nodes and α is the path-loss exponent. Assume the two-ray transmission model where $\alpha = 4$. The resulting SIR is about 16dB and based on Fig. 5, the impact of the interference on BER is negligible for BPSK. More generally, a thorough treatment should take into account the actual modulation scheme used, the difference between the effects of interference and noise, and whether or not channel coding is used. However, we can conclude that as far as the SIR is concerned, PNC is not worse than *traditional scheduling* (see Section II) when generalized to the *N*-node network. This is because for the generalized traditional scheduling, the interferers are 2, 2, 3, 5, 6, 6, 7, 9, 10, 10,…, hops away and the total interference power is larger than that in the PNC case above. To limit our scope, we leave the thorough SIR investigation to future work.

We now describe the PNC scheme under the general regular linear network more precisely. In the following we first consider the operation of PNC in the simple uni-directional case, followed by the bi-directional case.

4.1 Uni-Directional Transmission

Consider a regular linear network with *n* nodes. Label the nodes as node 1, node 2, …, node *n*, successively with nodes 1 and *n* being the two source and destination nodes, respectively. Fig. 7 shows a network with $n = 5$.

Divide the time slots into two types: odd slots and even slots. In the odd time slots, the odd-numbered nodes transmit and the even-numbered nodes receive. In the even time slots, the even-numbered nodes transmit and the odd-numbered nodes receive. Suppose that node 1 is to transmit frames X_1, X_2, …. to the destination node *n*.

Fig. 7 shows the sequence of frames being transmitted by the nodes in a 5-node network. In slot 1, node 1 transmits X_1 to node 2. In slot 2, node 2 transmits X_1 to node 3; node 2 also stores a copy of X_1 in its buffer. In slot 3, node 1 transmits X_2 to node 2, and node 3 transmits X_1 to node 4, but the transmission also reaches node 2; node 3 stores a copy of X_1 in its buffer. Thus, node 2 receives $X_1 \oplus X_2$. Node 2 then "adds" the inverse of its stored copy of X_1, X_1^{-1}, to $X_1 \oplus X_2$ to obtain $X_1^{-1} \oplus X_1 \oplus X_2 = X_2$. In slot 4, node 2 transmits X_2 and node 4 transmits X_1. In this way, node 5 receives a copy of X_1 in slot 4. Also, in slot 4, node 3 receives $X_1 \oplus X_2$ and then use X_1^{-1} to obtain $X_1^{-1} \oplus X_2 \oplus X_1 = X_2$.

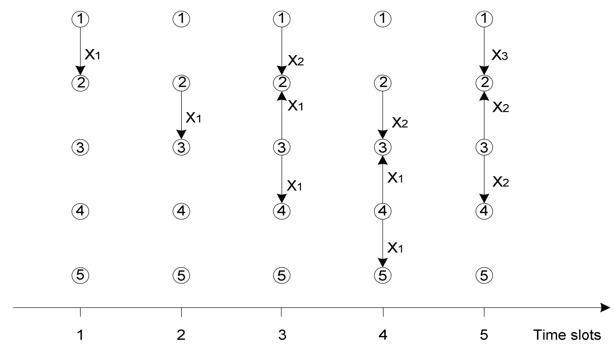

Figure 7. Uni-direct PNC transmission in linear network

Theorem 1: For the regular linear network, PNC can achieve the upper-bound capacity, 0.5 frame/time slot, for uni-directional transmission from one end of the network to the other end.

Proof: In a multi-hop transmission, each half-duplex relay node must use one time slot to receive a frame and another to send it out. So, it can at most relay one frame in two time slots (i.e., the upper bound is 0.5 frame/time slot). On the other hand, in PNC, each relay node transmits and receives frames in alternative time slot with no idle time, and it relay information contained in a frame in every two time slots. So, it achieves this upper-bound capacity.

4.2 PNC for Bi-directional Transmission

Let us now consider the situation when the two end nodes (i.e., nodes 1 and *n*) transmit frames to each other with the same rate via multiple relay nodes. Suppose that node 1 is to transmit frames X_1, X_2, …. to node *n*, and node *n* is to transmit frames Y_1, Y_2, …. to node 1.

Fig. 8 shows the sequence of frames being transmitted by the nodes in a 5-node network. As in the uni-directional case, a relay node stores a copy of the frame it sends in its buffer. It "adds" the inverse of this stored frame to the frames that it receives from the adjacent nodes in the next time slot to retrieve the "new information" being forwarded by either side. With reference to Fig. 8, we see that a relay node forwards two frames, one in each direction, every two time slots. So, the throughput is 0.5 frame/time slot in each direction.

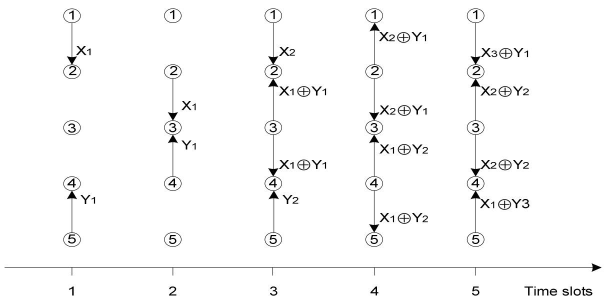

Figure 8. Bi-direct PNC transmission in linear network

Theorem 2: For the regular linear network, PNC can achieve the upper-bound capacity, 0.5 frame/time slot in each direction, for bi-directional transmissions between two end nodes.

Proof: If the rates from both sources are identical, the proof is similar to the one given for Theorem 1. In general, let us denote the data rate in one direction by V_X and the data rate in another direction by V_Y. First, we note that it is simply not feasible for either V_X or V_Y to exceed 0.5 frame/slot because it would exceed the capability of the half-duplex channel. Define the slacks as $S_X = 0.5 - V_X$, and $S_Y = 0.5 - V_Y$. We insert dummy null frames $\varnothing$ into the buffers at nodes 1 and *n*, so that nothing is transmitted during a slot when only a null frame comes up in the buffers (more detailed discussion of null frame can be found in the section, “formal description of PNC frame-forwarding mechanism”, below). The rate at which null frames appear correspond to the slacks S_X and S_Y. So, essentially the transmission rates are V_X and V_Y

In the regular linear network, if all the frames to be delivered are already available at the sources at the inception of the transmission, there is no incentive to use rates lower than 0.5. Rates smaller than 0.5 is relevant in two situations: 1) the source node generates frames in real-time at a rate smaller than 0.5; 2) a link between two nodes is used by many bi-directional PNC flows. The latter is particularly relevant in a general network topology, in which the per-directional link capacity have to be shared among all the flows that traverse the link.

4.3 Formal Description of PNC Frame-Forwarding Mechanism

This section may be skipped without sacrificing continuity. The time slots are divided into odd and even slots, and during odd slots, odd nodes transmit and during even slots, even nodes transmit. For generality, we allow for the possibility of a null frame, denoted by $\varnothing$ (this is relevant to the proof of Theorem 2 above and also for capacity allocation in a general network with many PNC flows). When we say an odd (even) node transmits a null frame in an odd (even) slot, we mean the node keeps silence and transmits nothing; similarly, when we say an odd (even) node receives a null frame in an even (odd) slot, we mean the node receives nothing. The null frame has the following property:

$$X_l \oplus \varnothing = X_l \quad \text{for all } X_l$$

$$X_l \oplus X_l^{-1} = \varnothing \quad \text{for all } X_l$$

$$\varnothing^{-1} = \varnothing$$

In terms of protocol implementation, if a transmitter intends to keep silence during one of its assigned transmission time slots, it should inform its two adjacent receivers at the beginning of the time slot, so that the receivers can revert back to ordinary non-PNC demodulation scheme to effect the above operational outcome. There is no need to inform the adjacent nodes during a reception (unassigned) slot of a node because it is understood that nothing will be transmitted by the node.

We now give the formal description of the PNC frame-forwarding mechanism for a general situation. The data rates in the two directions are not necessarily the same in this general scheme. We assume that each node *i* has a buffer B_i containing alternately the frame “to be transmitted” and the frame “just transmitted” by node *i* in successfully time slots. Initially, B_i is empty for all *i*. Let $S_i[j]$ and $R_i[j]$ denote the frames transmitted and received by node *i* in the time slot *j*, respectively. Let $B_i[j]$ be the buffer content of B_i in time slot *j*. Assuming the transmissions start in time slot 1, we have the following initial condition for node *i*:

$$\begin{aligned} &S_i[j] = R_i[j] = B_i[j] = \varnothing, \quad j \le 0, \forall i \\ &X_l = Y_l = \varnothing \quad l \le 0 \end{aligned} \tag{7}$$

Without loss of generality, let us assume that *n* is odd. The case of even *n* can be easily extrapolated from the same procedure presented here. The following equations describe the operation at node 1:

$$\begin{aligned} S_1[j] &= \begin{cases} B_1[j] & \text{for } j = 1,3,5,\cdots \\ \varnothing & \text{for } j = 2,4,6,... \end{cases} \\ R_1[j] &= \begin{cases} \varnothing & \text{for } j = 1,3,5,\cdots \\ S_2[j] & \text{for } j = 2,4,6,... \end{cases} \\ B_1[j] &= \begin{cases} X_{(j+2)/2} \oplus B[j-1] & \text{for } j = 1,3,5,\cdots \\ X_{j/2}^{-1} \oplus R_1[j] & \text{for } j = 2,4,6,... \end{cases} \end{aligned} \tag{8}$$

The following equations describe the similar operation at node *n*:

$$\begin{aligned} S_n[j] &= \begin{cases} B_n[j] & \text{for } j = 1,3,5,\cdots \\ \varnothing & \text{for } j = 2,4,6,... \end{cases} \\ R_n[j] &= \begin{cases} \varnothing & \text{for } j = 1,3,5,\cdots \\ S_{n-1}[j] & \text{for } j = 2,4,6,... \end{cases} \\ B_n[j] &= \begin{cases} Y_{(j+1)/2} \oplus B_n[j-1] & \text{for } j = 1,3,5,\cdots \\ Y_{j/2}^{-1} \oplus R_n[j] & \text{for } j = 2,4,6,... \end{cases} \end{aligned} \tag{9}$$

For odd nodes $i \in \{3,5,...,n-2\}$, we have

$$\begin{aligned} S_i[j] &= \begin{cases} B_i[j] & \text{for } j = 1,3,5,\cdots \\ \varnothing & \text{for } j = 2,4,6,... \end{cases} \\ R_i[j] &= \begin{cases} \varnothing & \text{for } j = 1,3,5,\cdots \\ S_{i-1}[j] \oplus S_{i+1}[j] & \text{for } j = 2,4,6,... \end{cases} \\ B_i[j] &= \begin{cases} B_i[j-1] & \text{for } j = 1,3,5,\cdots \\ B_i^{-1}[j-1] \oplus R_i[j] & \text{for } j = 2,4,6,... \end{cases} \end{aligned} \tag{10}$$

For even nodes $i \in \{2,4,...,n-1\}$, we have

$$S_i[j] = \begin{cases} \varnothing & \text{for } j = 1,3,5,\cdots \\ B_i[j] & \text{for } j = 2,4,6,... \end{cases}$$

$$R_i[j] = \begin{cases} S_{i-1}[j] \oplus S_{i+1}[j] & \text{for } j = 1,3,5,\cdots \\ \varnothing & \text{for } j = 2,4,6,... \end{cases} \quad (11)$$

$$B_i[j] = \begin{cases} B_i^{-1}[j-1] \oplus R_i[j] & \text{for } j = 1,3,5,\cdots \\ B_i[j-1] & \text{for } j = 2,4,6,... \end{cases}$$

It can be shown from the above that $B_1[j] = Y_{(j-n+3)/2}$, and $B_n[j] = X_{(j-n+3)/2}$, for $j = 2,4,6,...$ That is, after some delay, the information from one end reaches the other end and can be decoded there based on the above procedure.

5. RESOURCE ALLOCATION WITH PNC: AN ARCHITECTURAL OUTLINE

Our discussions so far has only focused on a single flow. How to support multiple PNC flows in a general network is an issue that deserves attention. We briefly outline a possible architecture for using PNC in a general network in this section. With this architecture, time slots are pre-assigned to PNC flows in an exclusive manner, so that a particular time slot will only be used by one bidirectional PNC flow. In this way, a receiver will always know which bidirectional flow it is dealing with in the demodulation process.

5.1 Partitioning of Time Resources

By nature, PNC is suitable for flows with bidirectional isochronous traffic with implied rate requirements (i.e., traffic with predictable bandwidth requirements that do not fluctuate much); it is not as suitable for uni-directional best-effort bursty flows. Based on this observation, we can divide time into periodically repeating intervals. Within each interval, there are two subintervals. The first subinterval is dedicated to PNC traffic and the second subinterval is dedicated to non-PNC traffic. The second subinterval may contain best-effort traffic as well as isochronous traffic that does not make use of PNC. The first subinterval, however, contains only PNC isochronous traffic.

Each PNC flow passing through a node is dedicated specific time slots within the first subinterval. In the parlance of the previous discussion, an "odd" node will only transmit data frames in the "odd" time slots within the first subinterval. With multiple PNC flows, the odd time slots are further partitioned so that different PNC flows will use different odd time slots.

The relative lengths of the first and second subintervals can be adjusted dynamically based on the traffic demands and the relative portions of the isochronous traffic that can exploit PNC. Some isochronous flows passing through a node cannot make use of PNC. This will be the case, for example, when the end-to-end path of a flow consists of several PNC chains, in between of which the conventional multi-hop scheme is used (see Part B of this section for further details). It may be necessary to break a long end-to-end path into multiple PNC chains to simplify resource management as well as to limit the synchronization overhead (see Section 6 for discussion of synchronization issues). The conventional multi-hop scheme is also needed in portions of the network in which PNC is not possible due to physical constraints.

Conceptually, the rates of the isochronous traffic can be described by a traffic matrix $[T_{i,j}]$. The (i, j) entry, $T_{i,j} = \sum_n f_{i,j}^{(n)}$, contains the total traffic originating from node *i* that is destined for node *j*, where $f_{i,j}^{(n)}$ is traffic flow *n* from node *i* to node *j*. The problem of joint routing and scheduling of the traffic flow in conventional multi-hop networks has been formulated in [11] as an integer linear programming problem. In assigning time slots, two nearby links cannot transmit together if they can mutually interfere with each other. This falls within the framework of a coloring problem.

With PNC, the coloring problem takes on a new angle: the traffic of a PNC flow at alternate links must adopt the same color (same time slots). In addition, as far as PNC is concerned, the individual make-ups of the flows between node *i* and *j*, $f_{i,j}^{(n)}$, is not important. It is the aggregate traffic $T_{i,j}$ that matters. Also, it is conceivable that PNC can also be used for uni-directional individual flows as long as there is bidirectionality in the aggregate flow. That is, the amount of bidirectional traffic at the "aggregate" level is $\min(T_{i,j}, T_{j,i})$ and they can leverage PNC. The rest, $\max(T_{i,j}, T_{j,i}) - \min(T_{i,j}, T_{j,i})$, may use the conventional scheme. We believe routing and resource allocation in PNC is a topic of much interest for more in-depth future research.

5.2 Flow Decomposition

Due to various reasons, including interference and synchronization, some of the nodes on the flow path can leverage PNC while others cannot. In general, an end-to-end path may need to be decomposed into several paths, some using PNC while other using the conventional scheme. With such decomposition, a flow essentially becomes a sequence of sub-flows. Fig. 9 depicts an example of decomposition of a flow into three sub-flows, where PNC is used by sub-flow1 and sub-flow3, and the conventional scheme is used by sub-flow2. With respect to the resource allocation problem mentioned in Part A, the decomposition will also alter the constraints in the optimization problem.

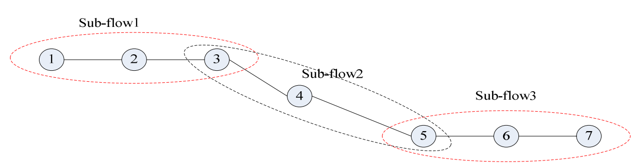

Figure 9. Illustration of flow decomposition

6 DISCUSSION OF SYNCHRONIZATION ISSUES

In the previous sections, synchronization between all the nodes has been taken for granted. However, synchronization for ad hoc networks more than three nodes is generally considered as difficult. In this section, we briefly discuss how synchronization beyond three-node PNC networks can be practically performed, and what penalty will be faced by PNC when synchronization is imperfect or even unavailable. Interested readers are referred to [12] for a more detailed quantitative treatment.

Three-node synchronization has been extensively studied by others. Time and carrier-frequency synchronizations have been actively investigated by researchers in the fields of OFDMA [13], wireless-sensor network [14], and/or cooperative transmission. Carrier-phase synchronization has also been studied by researchers in the fields of coherent cooperation [15] and/or distributed beam forming [16].

N-node synchronization can be built on top of the basic mechanism of 3-node synchronization, as illustrated in Fig. 10. We could partition the *N*-node chain into multiple three-node basic groups (BGs), and perform the synchronization group after group successively, starting from BG1. Note that Fig. 10 only depicts the BGs in the odd time slots; in the even time slots, the grouping will begin at the second node. With this scheme, given small synchronization errors in BG1, it is conceivable that by the end of the synchronization process, the synchronization errors of node *N* with respect to node 1 become very large. At first glance, this may seem horrendous. It is important to note, however, that with PNC, a particular receiving node only receives signals simultaneously from the two immediate adjacent nodes. As far as PNC demodulation is concerned, only the relative synchronization errors of the two adjacent nodes are important. That is, any synchronization error within a particular BG is always enclosed within the BG, and is transparent to any nodes outside the BG. The demodulation performance penalty due to synchronization errors in an *N*-node network is the same as that of a 3-node network.

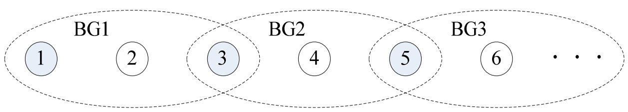

Figure 10. Synchronization for multiple nodes

In [12], we provide an analysis of the impact of 3-node sychronization errors assuming the use of BPSK. There are three types of sychronization errors that may affect the performance of PNC demodulation: carrier-phase, carrier-frequency, and symbol-time-alignment errors. When carrier-phase sychronization is not performed at all, it can be shown that the SNR penalty is less than 1dB on average, and 3dB at worst. When the carrier-frequency error is as high as 10% relative to the symbol frequency, the maximum SNR penalty is still less than 0.6dB. When symbol-time alignment is not performed at all, the maximum SINR penalty is less than 2.5dB when the SNR is 10dB. The above penalties are within the acceptable range given that more than 100% throughput improvement can be obtained by PNC.

7 CONCLUSIONS

This paper has introduced a novel scheme called *Physical-layer Network Coding* (PNC) that significantly enhances the throughput performance of multi-hop wireless networks. Instead of avoiding interference caused by simultaneous electromagnetic waves transmitted from multiple sources, which has been the major research direction in the past, PNC embraces interference to effect network-coding operation directly at the physical-layer signal modulation and demodulation. With PNC, signal scrambling due to interference, which causes packet collisions in the MAC layer protocol of traditional wireless networks (e.g., IEEE 802.11), can be eliminated.

For PNC to be feasible, network-coding arithmetic must be realized with direct electromagnetic-wave mixing, coupled with appropriate modulation and demodulation schemes. This paper has presented the fundamental condition for the equivalence of the conventional network-coding operation and PNC operation.

We have shown that PNC can achieve 100% improvement in physical-layer throughput over the traditional multi-hop transmission scheduling scheme, and 50% over the straightforward network coding scheme. In addition, the throughput achieved by PNC in a regular linear multi-hop network is that of the theoretical upper-bound throughput.

REFERENCES

[1] J. Li, C. Blake, D . D. Couto, H . Lee, and R. Morris, "Capacity of Ad Hoc Wireless Networks," In ACM MobiCom'01, pp. 61-69, July 2001.

[2] P. C. Ng, and S. C. Liew, "Offered Load Control in IEEE 802.11 Multi-hop Ad-hoc Networks," The 1st IEEE International Conference on Mobile Ad-hoc and Sensor System, Nov. 2004, Florida, USA.

[3] T. Ojanpera, and R. Prasad, "An Over View of Air Interface Multiple Access For IMT 2000/UMTS," IEEE Communication Magazine, pp. 82-91, Sep. 1998.

[4] S.-Y. R. Li, R.W. Yeung and N.Cai, "Linear Network Coding," IEEE Trans. Inform. Theory, vol. 49, no.2, pp. 1204-1216, Feb. 2003.

[5] R.Ahlswede, N.Cai, S.-Y. R. Li, and R.W. Yeung, "Network information flow," IEEE Trans. on Information theory, vol. 46, no. 4, pp. 1204-1216, Jul 2000.

[6] J. N. Laneman, D. N. Tse, G. W. Wornell, "Cooperative diversity in Wireless Networks: Efficient Protocols and Outage Behavior," IEEE Trans on Information theory, vol. 50, pp.3062-3080, Dec.2004.

[7] T.M. Cover and A.A. El Gamal, "Capacity theorems for the relay channel," IEEE Trans on Information theory, pp.572-584, Sept.1979.

[8] Y. Wu, P. A. Chou, and S. Y. Kung, "Information Exchange in Wireless Networks with Network Coding and Physical Layer Broadcast," Technical Report MSR-TR-2004-78, Microsoft Research, Redmond WA, Aug. 2004.

[9] C. Fragouli, J. Y. Boudec, and J. Widmer, "Network Coding: An Instant Primer," ACM SIGCOMM Computer Communication Review, pp. 63-68, Vol. 36, No. 1, Jan. 2006

[10] J. R. Proakis, Digital Communication. New York: McGraw Hill, 1989.

[11] K. Jain, J. Padhye , V. N. Padmanabhan and L. Qiu, "Impact of Interference on Multi-hop Wireless Network Performance," Proceedings of the 9th annual international conference on Mobile computing and networking (MobiCom), September 14-19, 2003, San Diego, CA, USA

[12] S. Zhang, S. Liew and P. Lam, "Physical Layer Network Coding (PNC)", http://personal.ie.cuhk.edu.hk/~slzhang5/, Technical Report, July. 2006.

[13] M. Morelli, "Timing and frequency sunchronization for the uplink of an OFDMA system," IEEE Trans. Commun., vol. 52, no. 2, pp. 296-306, Feb. 2004.

[14] J. Elson, L. Girod, and D. Estrin. Fine-grained network time synchronization using reference broadcasts. In OSDI, 2002.

[15] Y. Tu and G. J.Pottie, "Coherent Cooperative Transmission from Multiple Adjacent Antennas to a Distant Stationary Antenna through AWGN Channels," VTC'02, pp. 130-134, 2002.

[16] R Mudumbai, G Barriac, U Madhow , "On the feasibility of distributed beamforming in wireless networks", IEEE Trans. on Wireless Commun (under review).