On RANC ARQ for Wireless Relay Networks: From the Transmission Perspective

Jung-Chun Kao, *Member, IEEE,* and Fu-Wen Chen, *Student Member, IEEE*

***Abstract*—The relay-assisted, network-coding (RANC) automatic repeat request (ARQ) protocols are ARQ protocols that leverage both opportunistic retransmission and network coding for wireless relay networks. This paper proposes two single-relay RANC ARQ protocols, the listen-and-supersede (LS) protocol and the hold-and-proceed (HP) protocol. LS offers a fundamental limit to any single-relay RANC ARQ protocol. HP is a simple yet efficient RANC ARQ protocol with near-zero overhead. Moreover, we analyze saturation throughput and segment delay for both LS and HP. Their performances are compared with a representative cooperative ARQ protocol, the opportunistically forwarding (OF) protocol. Through extensive analysis and simulation results, we show that HP has a performance close to LS and outperforms OF significantly.**

***Index Terms*—Cooperative ARQ, opportunistic retransmission, network coding, queueing model, wireless network.**

I. Introduction

WIRELESS communication is inherently error-prone due to path loss, fading, noise, interference, etc. In such an error-inclined environment, it is of paramount importance to achieve efficient and reliable data delivery. Advanced wireless communication and networking systems such as cellular networks (e.g., 3G cellular networks, WiMAX, LTE, and LTE-Advanced) often utilize automatic repeat request (ARQ) methods along with relaying mechanisms, at the link layer. By relaying, low-rate (or unreliable) direct communication between source and destination can be replaced by high-rate (or reliable) source-relay and relay-destination transmissions, thus enhancing coverage, reliability, and throughput. Recently, link-layer network coding [1], [2] and opportunistic retransmission [3], [4] have been hot research topics to further improve network performance.

Opportunistic retransmission presented in [3] is a link-layer technique that leverages the broadcasting nature of wireless communication and the benefit of multi-path transmission: Data packets not reaching the destination node will be retransmitted by close-by relay nodes that overhear the data packets. Such an opportunistic use of multi-path transmission can extend coverage and bring performance gain because appropriately chosen relay nodes have stronger connectivity to destination nodes than source nodes have. To enable source and relay nodes aware of missed data packets, opportunistic retransmission relies on the per-packet acknowledgement (ACK)

Manuscript received September 12, 2012; revised January 13, 2013; accepted March 13, 2013. The associate editor coordinating the review of this paper and approving it for publication was J. Wu.

The authors are with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan (e-mail: jungchuk@cs.nthu.edu.tw, richarticle.chen@gmail.com).

Digital Object Identifier 10.1109/TWC.2013.050613.121394

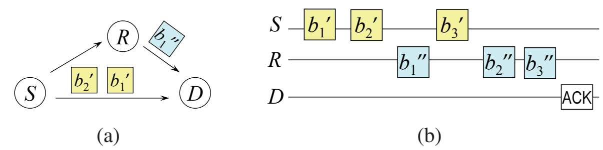

Fig. 1. (a) A three-node network containing source S, destination D, and relay R. The source node and relay node send out coded blocks, denoted by b_i', and recoded blocks, denoted by b_i'', respectively. (b) A high-level overview of RANC ARQ protocols.

function. However, the experiments in [5] suggests that 802.11 ACKs contribute to over 20% overhead in 802.11 networks. In addition, according to [1], ACK loss due to poor or varying channel condition may cause the ARQ timeout, which triggers a redundant retransmission of the same data packet and causes a significant communication overhead.

Network coding offers an elegant solution to this challenge, because of its ability to removing the need for per-packet ACKs. Indeed, a link-layer network coding technique is on a *per-segment* basis, where a segment consists of a number of blocks. Instead of simply sending out individual blocks, a sender mixes all or a subset of the blocks it has by a network coding technique and then sends out these mixed blocks. This way, regardless which packets (i.e., mixed blocks) are lost in transit, as long as a sufficient number of packets reach the receiver, the receiver can retrieve the entire segment. No specific packet will be retransmitted. Per-segment ACKs, rather than per-packet ACKs, are used to notify senders of the ends of segment transmissions. The number of needed ACKs is reduced to one per segment from one per packet.

A. Introduction to RANC ARQ

This paper aims to analyze performance and fundamental limit of relay-assisted, network-coding (RANC) ARQ protocol, which is a technique combining opportunistic retransmission and link-layer network coding. Particularly, the focus is on one-way wireless relaying systems in order to be readily implementable in a wide variety of today's technologies without requiring any specific traffic pattern. Fig. 1 gives a high-level overview of a RANC ARQ protocol. Messages are transmitted on a per-segment basis, where a segment consists of K original blocks, denoted by $b_1, b_2, \ldots, b_K$. The original blocks are not transmitted over wireless channels. Instead, a source node keeps sending out coded blocks, denoted by b_1', b_2', and so forth. The coded blocks are produced using a coding technique, for example, by taking linear combinations of the original blocks over a finite field GF(2^m).

A relay node does not forward the coded blocks it overhears. Instead, when needed, the relay node produces a number of recoded blocks (denoted by b_1'', b_2'', and so forth) by, for example, taking linear combinations of all or a subset of the overheard coded blocks over a finite field GF(2^m). After that, the recoded blocks are sent out to the destination node.

Any coded block and recoded block may or may not reach the destination node, depending on the varying channel condition. Regardless which blocks are lost, the destination node can retrieve the entire segment back from the received (coded and recoded) blocks, for example, by Gaussian elimination, as long as a sufficient number of blocks reach the destination node. Once the destination node retrieves the entire segment, a segment transmission completes. Then a per-segment ACK is sent out by destination, notifying the source node of the termination of the segment transmission.

B. Our Contributions

So far, little effort has been devoted to studying RANC ARQ protocols designed for one-way wireless relaying systems at the link layer, especially from an analysis perspective. In this paper, we propose and analyze two single-relay RANC ARQ protocols—LS and HP. LS offers a fundamental limit to any single-relay RANC protocol. HP is a simple yet efficient RANC protocol that achieves a performance close to the fundamental limit, while having near-zero overhead.

We derive saturation throughput and average segment delay for both LS and HP. The segment delay is defined as the time elapsed from the moment when the source node generates a segment to the moment when the destination node can retrieve the entire segment. A segment in reality often represents a message that could be, for example, a video frame at the application layer or a datagram at the network layer. Original blocks typically are fragmented transmission units at the link layer, each of size not exceeding the maximum transmission unit (MTU). From this perspective, decoding of a complete segment is often more meaningful than decoding of a single block.

Saturation throughput is defined as the expected number of blocks successfully decoded at the destination per second, assuming that source's transmission buffer is always full. In a time division multiple access (TDMA) system, which is inherently collision-free, saturation throughput can be regarded as the maximum service rate in the sense that an increase in the segment arrival rate increases the actual throughput until saturation throughput is achieved.

Our analysis on saturation throughput and average segment delay applies to collision-free wireless networks, in which all the nodes share the radio medium using their own orthogonal resources. The orthogonal resources can be exclusive time slots (in TDMA), non-overlapping frequency bands (in FDMA), or a combination thereof. For a contained presentation, saturation throughput and average segment delay are specifically derived for the TDMA case only, although they can be adjusted to fit other schemes.

In our analysis framework, any single-relay ARQ protocol for the TDMA case essentially behaves as an M/G/1 queue with vacation—various protocols differ in their distributions of service time. Multiple flows are allowed to deliver their data concurrently. Each node can play multiple roles, acting as source, destination, and/or relay. For example, a source node can generate and send out its own packets in some time slot of a time frame, while helping other nodes for their information deliveries at other time slots.

A constraint in our analysis framework is that at most one relay is involved in each segment transmission. The reason for focusing on the single-relay case is to minimize control overhead and protocol complexity. The case of having multiple relays, although might be approximated by the single-relay case with some modifications, is out of the scope of this paper.

For comparison purpose, the proposed RANC ARQ protocols are evaluated and compared against the OF protocol, which represents cooperative ARQ protocols with a single relay. OF is essentially the Type-I/II cooperative ARQ protocol [4] with an explicit ACK loss handling mechanism added. OF can be also regarded as a modification of the decode-and-forward scheme [6] by adding an ACK function and by not redundantly forwarding the packets that have been successfully delivered to destination. Numerical results obtained from our analysis and simulation help to learn how much RANC ARQ protocols yield superior performance.

The remainder of this paper is organized as follows. In Section II, we review related work. Section III introduces the network model we consider. Section IV presents two RANC ARQ protocols we develop for wireless relaying systems and one cooperative ARQ protocol used for comparison purpose. Saturation throughput and segment delay for these protocols are derived in Section V. Section VII shows the performance of the protocols by theoretical and simulation results. Section VIII presents some concluding remarks.

II. Related Work

Network coding has been an active research topic in wireless networks. Many studies have shown that network coding brings performance gains [7], [8], particularly in specific scenarios with multicast traffic [9]–[11], with two-way flows [12], [13], or with multiple unicast flows [14]. Only few work in the field of network coding are designed for the one-way relaying scenario. For one-way relaying systems, Li et al. in [1] showed that the MAC-layer random network coding (MRNC) protocol excels hybrid ARQ in WiMAX systems. According to [2], the MRNC protocol is a special case of the N-in-1 retransmission with network coding scheme (abbreviated as N-in-1 ReTX). By utilizing random linear network coding [15], [16], N-in-1 ReTX [2] distributes overhead of retransmission over multiple frames, which makes N-in-1 ReTX outperform MRNC.

Link-layer protocols exploiting opportunistic retransmission are also known as cooperative ARQ protocols or C-ARQ protocols. Fundamental concepts, theoretical bounds, and practical protocols of C-ARQ mechanisms have been extensively studied in the literature. The concept of cooperative diversity—source and relay nodes cooperate to form a virtual antenna array that exploits transmissions over a statistically independent relay path—is proposed by Laneman et al. [6]. In [6], several C-ARQ protocols with single relay are developed and analyzed in terms of the mutual information and outage probability. Note that all the protocols in [6] do not really take link-layer retransmissions into account.

Extending the work in [6], Zimmermann et al. [17] introduced a number of modified versions of the C-ARQ protocols proposed in [6] and derived the outage probabilities and SNR gains for these protocols. Rong and Ephremides [18] studied stability region and throughput region for a simplified case where two source nodes send packets to a common destination and the source closer to the destination helps to relay packets for the other source. Still these above works do not consider link-layer retransmissions.

A few works have considered the aspect of link-layer retransmissions. For example, Zimmermann et al. [19] simulated the average number of retransmissions of a single-relay C-ARQ protocol. Cerutti et al. [4] derived the first delay model for C-ARQ protocols with single relay and proved that C-ARQ protocols have performance gain over hybrid ARQ in terms of throughout and delay. Sadek et al. [20] analyzes the maximum stable throughput region and the delay performance of two cognitive multiaccess protocols for a TDMA system, in which the cognitive relay common to all source nodes senses whether time slots are being used and retransmits lost packets in the unused time slots. Si et al. [21] proposed relay selection protocols with interference taken into account. All the above works consider a three-node topology in which there is one relay node for each source-destination pair.

Having more than one relay node for a source-destination pair often improves overall performance. Indeed, Maham et al. [22], [23] derived the exact outage probability of a multi-relay delay-limited system and showed full diversity in the proposed multi-relay networks. However, multi-relay systems also bring additional challenges. To have multiple relay nodes cooperate in a distributed fashion, there have been a number of C-ARQ protocols developed in the literature [3], [24]–[27]. For example, Steenkiste et al. [3] proposed a distributed relay selection algorithm to choose a set of eligible relays and a relay prioritization mechanism to make sure high-priority relays transmit with high probability. These works are based on the IEEE 802.11 Distributed Coordination Function (DCF) and thus inherit the inefficiency of DCF when the contention window is not well adjusted to the number of relays. To solve this problem, Alonso-Zarate et al. proposed in [28] a multiple-relay C-ARQ scheme whose operation is almost independent of the number of relays and analyzed this scheme in [29].

Leveraging either network coding or opportunistic retransmission is beneficial; nevertheless, it would be more beneficial for an ARQ protocol to utilize both. The ARQ protocols that utilize both techniques are referred to as RANC ARQ protocols. As stated by Skianis et al. [30] in 2012, there are few works that study the link-layer aspect of network coding-based cooperative communication.

In the following, related works for the two-way relaying, multicast, and one-way relaying scenarios are reviewed respectively. For the two-way relaying scenario, several IEEE 802.11-compatible RANC ARQ protocols with single or multiple relays are proposed in [30]–[32], which are shown to outperform C-ARQ protocols in terms of throughput and total delay through simulation and analytical results. These analyses in [30]–[32] focus on the saturation situation where a certain number of active and saturated sources attempt to send packets.

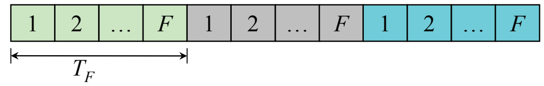

Fig. 2. TDMA scheme with F flows.

For the multicast scenario, Fan et al. [33] in 2009 studied the network coding gains with the use of XOR retransmissions, in terms of throughput, delay and queue length. Fanous and Ephremides [34] in 2010 investigated the impact of random linear network coding on the stable throughput. Song et al. [35] in 2012 studied the coding gains for applying XOR retransmissions at both source and relay nodes in the case with two sources and many destinations. Shrader et al. [36] in 2011 focused on the proxy case, in which destinations are formed into multiple clusters and destinations can act as relays for other destinations in the same cluster. In the proxy case, the benefit of cooperative strategies for multicasting over wireless lossy links is explored, particularly in terms of the number of packets sent by source and relays.

The above works focus on either the two-way relaying scenario or the multicast scenario; none of them involves one-way relaying systems. Indeed, there has been little effort devoted to the one-way relaying scenario. Among the very few related works, Kao et al. [37] in 2011 proposed an IEEE 802.11-compatible RANC ARQ protocol and evaluated its performance through simulation; while Abuzainab and Ephremides [38] in 2012 analyzed energy efficiency and minimum stable throughput, considering the use of RANC ARQ at the link layer on top of Alamounti coding at the physical layer. Nevertheless, the delay aspect of RANC protocols, particularly when nodes are not always saturated, has not been analyzed yet. This motivates the main objective of this paper—developing and analyzing RANC ARQ protocols at the link layer for the one-way relaying scenario. To our best knowledge, we present the first analysis of RANC protocols in terms of segment delay and buffer occupancy.

III. System Model

We consider a wireless network in which all the nodes are in close proximity. There are F flows (or source-destination pairs) sharing the same channel in a TDMA manner. As shown in Fig. 2, exclusive and cyclic time slots are assigned to each flow. A single data packet (i.e., a block) is transmitted in a time slot. ACKs are short enough to fit in the same time slots allotted to the associated blocks; or they are sent in the reverse direction in separate channels (FDD) or time slots (TDD). These are often seen settings in various wireless communication systems. For example, LTE provides information about successful/failed reception by a single-bit feedback acknowledgment with a fixed-timing relation to the corresponding transmission attempt [39].

A flow in the network involves a source and a destination, possibly with help of the third device, a relay. A node is allowed to play more than one role simultaneously, acting as source, destination, and/or relay. For example, a node can be a source node generating its own traffic; meanwhile it can be a

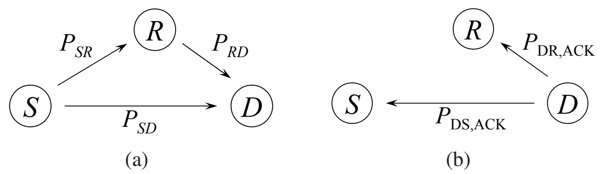

Fig. 3. (a) Link reception probabilities of blocks (denoted by P_{SD}, P_{SR}, and P_{RD}), for a given source-destination pair. (b) Reception probabilities of per-packet ACKs originated from the destination node (denoted by $P_{\mathrm{DS,ACK}}$ and $P_{\mathrm{DR,ACK}}$).

relay node helping other flow(s). We consider the scenario that a relay helps a certain flow using the time slots allocated to that flow: After overhearing packets sent by the source, the relay may opportunistically send some coded version of overheard information in the flow's own time slots. In this way, multiple concurrent flows are allowed to choose the same relay node.

Because an appropriately chosen relay node has better connectivity to the destination node than the source node has, the relay is assumed to have higher priority than the source: If both relay and source have packets waiting to be sent, the relay acquires access to the next time slot allocated to the flow. Such source-relay cooperation can be implemented, for example, by a single-bit feedback from relay to source or by the capture effect. The capture effect is a phenomenon that stronger signal (relay's transmission) suppresses weaker signal (source's transmission) and thus the stronger signal (relay's transmission) will be demodulated.

Original blocks are assumed to be generated at each source node in accordance with a Poisson arrival process with random-sized segments. That is, arrival times of segments follow a Poisson process and each segment consists of a random number of blocks. The segment size in blocks, denoted by K, can be arbitrarily distributed with finite mean and finite variance. The mean of K is denoted by $\mathbf{E}[K]$ or $\overline{K}$.

Considering the error-prone nature of wireless links over which data packets (i.e., blocks) may or may not be delivered successfully, link receptions for blocks are modeled by erasure channels. The channels are assumed to undergo independent block fading across different links. Fig. 3(a) illustrates the link reception probabilities from source to destination (denoted by P_{SD}), from source to relay (denoted by P_{SR}), and from relay to destination (denoted by P_{RD}), respectively. It is assumed that the relay node is appropriately chosen[1] and hence the relay has stronger connectivity to the destination than the source has. Equivalently, $P_{RD} > P_{SD}$.

The destination node could indicate success or failure of a reception by broadcasting an ACK to the source and relay. There are two types of ACKs—per-segment ACKs (which are used by all RANC ARQ protocols to notify the ends of segment transmissions) and per-packet ACKs (which are used by conventional ARQ protocols at the link layer but may or may not be used by a RANC ARQ protocol). Reproduced from [6] in which a justification is given, it is assumed that all (per-packet and per-segment) ACKs are perfectly received by the relay. For the same reasons, per-segment ACKs are assumed to be perfectly received by the source. Per-packet ACKs instead could be lost (i.e., either undetected or corrupt) in transit to the source. In a word, per-segment ACKs are perfectly received by both the relay and the source; while per-packet ACKs are perfectly received by the relay but they could fail to reach the source. Fig. 3(b) illustrates the reception probabilities for per-packet ACKs, where $P_{\mathrm{DR,ACK}} = 1$ and $P_{\mathrm{DS,ACK}}$ is a fractional number.

[1]Although relay selection is beyond the scope of this paper, we note that for the LS and HP protocols, a brute force algorithm that selects the optimal relay (in terms of either saturation throughput or segment delay) can be of a linear complexity O(n), where n is the total number of possible candidates for relay. This is because all the formulas in theorems 1, 2, 3, and 4 are independent of n and consequently the brute force algorithm simply uses the given formulas to calculate the throughput or delay for all possible candidates and then selects the optimal one. Such an optimal and low-complexity algorithm for relay selection can be used by the LS and HP protocols.

IV. Protocols Description

In this section, we describe one C-ARQ protocol, OF, and two RANC ARQ protocols, LS and HP. The OF protocol is presented for comparison purpose. The LS protocol offers a fundamental limit to any single-relay RANC ARQ protocol. The HP protocol is a simple yet efficient RANC ARQ protocol with near-zero overhead. These protocols are described respectively in the following subsections. Their performance will be analyzed in Section V and compared in Section VII.

A. Comparative C-ARQ Protocol: The OF Protocol

The opportunistically forwarding protocol, abbreviated as *OF*, represents C-ARQ protocols belonging to the stop-and-wait family and can be regarded as the decode-and-forward protocol [6] or the Type-I/II C-ARQ protocol [4], with a per-packet acknowledgement function added at the destination: A source sends a block and, every time when the relay successfully overhears a block but does not hear a corresponding (positive) per-packet ACK, the relay will forward the block to the destination in the next time slot allocated to the flow. The destination indicates each success reception by broadcasting a per-packet ACK to the source and relay. As described in Section III, it is assumed that per-packet ACKs are reliably received by the relay but might be lost in transit to the source. It is also assumed that the relay has higher priority than the source and hence collisions caused by simultaneous transmissions from any source-relay pair never happen. In case the source does not receive a (positive) ACK, regardless because the ACK was lost or no ACK was sent, the source will retransmit the same block later in the time slots allocated to the flow. Retransmission may take one or several times[2] until the source receives a (positive) ACK. After that, the source transmits the next block if it needs.

B. RANC ARQ Protocol 1: The LS Protocol

Among all possible single-relay RANC ARQ protocols, the listen-and-supersede protocol, abbreviated as *LS*, is optimal in terms of efficiency of channel usage. The basic idea behind this protocol is that since the relay has better connectivity to the destination than the source has, it would be optimal

[2]The maximum retry limit in the mathematical analysis later in Section V is set to infinity. In practice, the maximum retry limit can be any appropriate integer.

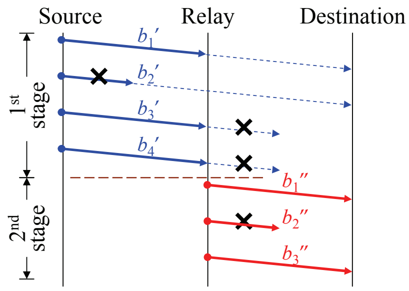

Fig. 4. An illustrative example of the LS protocol with the segment size of four. Dotted lines depict direct communications from source to destination.

for the relay to take over from the source, once the relay and destination together become able to retrieve the entire segment cooperatively.

Based on the aforementioned idea, the segment transmission process of LS consists of two stages. In the first stage, the source keeps sending out coded blocks for the segment without waiting for per-packet ACKs back. Meanwhile, the relay listens to the coded blocks sent by the source and learns immediately which blocks have successfully reached the destination node through feedback (i.e., per-packet ACKs) from the destination. Once the union of the blocks overheard by relay and destination is sufficient to retrieve the entire segment, the first stage finishes.

In the second stage, the relay takes over from the source. The source no longer sends any blocks. Instead the relay keeps sending out recoded blocks, which are encoded over all the blocks it has overheard, until the destination can retrieve the entire segment on its own. Then the destination sends out a per-segment ACK, which notifies the source and relay of the end of the segment transmission.

Fig. 4 illustrates a segment transmission process with a segment size of four. In the first stage, the source node sends out coded blocks, $b'_1, b'_2, \ldots$, one by one. The source might send out more than four blocks due to packet loss. At the moment the dashed horizontal line indicates, the destination has received $\{b'_1, b'_2\}$ and the relay has overheard $\{b'_1, b'_3, b'_4\}$. With their union $\{b'_1, b'_2, b'_3, b'_4\}$, relay and destination nodes together become able to decode the entire segment, because the union has a cardinality equal to the segment size. At this point, the second stage starts, in which the relay sends out recoded blocks, b''_1, b''_2, and so forth. Upon receiving b''_3, the destination has accumulated four (coded and recoded) blocks, $\{b'_1, b'_2, b''_1, b''_3\}$, and thus becomes able to decode the entire segment on its own. This completes the segment transmission.

The source and relay send coded/recoded blocks instead of original blocks; the below presents an optimal coding scheme that is essentially a Reed-Solomon based algorithm. In the first stage, the source generates coded blocks as $b'_n = \sum_{i=1}^{K} n^i b_i$. Each coded block corresponds to a linear equation in the K variables. Having any distinct K coded blocks can decode the entire segment. This is because the corresponding system of linear equations has coefficients forming a Vandermonde matrix and thus its determinant is non-zero.

Define M_{SD} as the number of coded blocks successfully received by the destination in the first stage. Because the first stage finishes once the relay and destination have K distinct coded blocks in all, the number of the coded blocks overheard by the relay that are missed by the destination in the first stage is $K - M_{SD}$. The relay learns these blocks by constantly hearing per-packet ACKs sent by the destination through the feedback channel.[3] Denote these missing blocks by $c_1, c_2, \ldots, c_{K-M_{SD}}$. The relay generates recoded blocks in the second stage as $b''_n = \sum_{i=1}^{K-M_{SD}} n^i c_i$. For the same reason aforementioned, reception of any distinct $K - M_{SD}$ recoded blocks enables the destination to decode $c_1, c_2, \ldots, c_{K-M_{SD}}$. With these $K - M_{SD}$ coded blocks obtained in the second stage and the M_{SD} coded blocks received in the first stage, the destination itself can decode the entire segment.

We note that under the LS protocol, relay nodes have to continuously learn what blocks have been delivered to the destination through a feedback channel. This requires a tight cooperation among relay and destination nodes, which may incur control overhead to some extent and increases protocol complexity. To minimize control overhead and dependency on the destination-to-relay feedback, we develop the HP protocol.

C. RANC Protocol 2: The HP Protocol

The hold-and-proceed protocol, abbreviated as *HP*, is a RANC ARQ protocol. It aims to cause near-zero overhead and to reduce protocol complexity by removing the dependency of destination-to-rely feedback, while retaining a performance close to the optimum. HP does not need per-packet ACKs (sent by the destination) at all. The basic idea behind this protocol is for relay nodes to hold quiescent for a while before proceeding to send recoded blocks in a way independent of destination-to-relay feedback. Whether or not the relay sends a recoded block after overhearing a coded block depends only on the relay's own state.

The segment transmission process of HP consists of two stages. In the first stage, a source node keeps sending coded blocks in the time slots allotted to the flow; meanwhile, the associated relay node holds silently and saves the overheard blocks in its buffer, until the relay has (successfully) received h coded blocks. After that, the second stage starts, in which each successful reception of coded blocks at the relay triggers a transmission of recoded block in the next time slot allocated to the flow. The source still attempts to send coded blocks in the time slots allocated to the flow. However, because the relay node has higher priority over the source node, the source has to wait while the relay is sending. The second stage continues until the destination itself becomes able to retrieve the entire segment with its received (coded and recoded) blocks. Then the destination sends out a per-segment ACK, which notifies the source of the end of the segment transmission.

As illustrated in Fig. 5, the relay keeps silent in the first stage. At the moment the dashed horizontal line indicates, the first stage of HP completes because the relay has overheard $h = 1$ coded blocks. After that, every time when overhearing a coded block, the relay sends out a recoded block. Upon receiving b''_3, the destination has accumulated four blocks, $\{b'_1, b'_2, b''_1, b''_3\}$, and thus becomes able to retrieve the entire segment. This completes the segment transmission.

[3]The LS protocol requires the destination to send per-packet ACKs in the first stage. On the contrary, HP does not need per-packet ACKs at all.

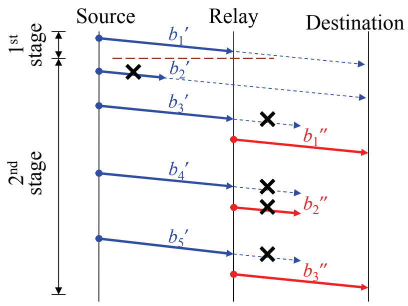

Fig. 5. An illustrative example of the HP protocol with the segment size of four and $h = 1$. Dotted lines depict direct communications from source to destination.

In the below, we present a coding scheme at source and at relay, which is a modification of the Reed-Solomon based algorithm used in the LS protocol (cf. Section IV-B). The source generates coded blocks as $b'_n = \sum_{i=1}^{K} n^i b_i$, in both the first and second stages. Denote the coded blocks the relay receives in chronological order by $c'_1, c'_2, \ldots$. The relay keeps silent in the first stage. Upon each successful reception of coded blocks in the second stage, the relay generates recoded blocks using all the coded blocks it has overheard by that moment. The n-th recoded block is $b''_n = \sum_{i=1}^{n+h} n^i c'_i$.[4]

Ideally, it would be great for a coding scheme to hold the innovation property—any K (distinct) blocks received by the destination are innovative to each other and hence K received blocks allows the destination to decode the entire segment. However, in reality any coding scheme for the HP protocol, including the one aforementioned and even an optimal coding scheme, may fail to hold the innovative property in some circumstances. The reason is irrelevant to the coding scheme itself; instead, it is coupled with what blocks are received at relay and at destination.

Take the extreme case in which $P_{SD} = 1$ and $h = 0$ as an example. In this extreme case, because all blocks sent by the source reach the destination successfully, the relay cannot really provide any help. All recoded blocks generated upon receptions of coded blocks at the relay are non-innovative, regardless of what coding scheme is used.

To make as many blocks received at destination innovative as possible and to make channel usage efficient, the value of h must be appropriately chosen with channel quality taken into account. The way to choosing an appropriate value for h is discussed later in Section VI. Through simulation, it is observed that using the coding scheme aforementioned with such an appropriate value h, almost all of the blocks received by the destination are innovative.

V. Performance Analysis

In this section, the following performance metrics are derived for all the ARQ protocols described in Section IV.

- η: The saturation throughput (in blocks per second)
- T: The average segment delay (i.e., the expected time elapsed between a segment's arrival at node S and the segment's complete retrieval at node D)
- L: The buffer occupancy (i.e., the expected number of segments buffered at node S)

The above performance metrics of all the ARQ protocols are analyzed using a unified framework: Each ARQ protocol in the TDMA system described in Section III essentially behaves as an M/G/1 queue with vacation. Different ARQ protocols has distinct distributions of service times, thereby resulting in different performances. Indeed, the service time of an ARQ protocol depends on the first and second moments of N, where N is the random variable representing the total number of blocks sent by either source or relay for a segment (until the segment is fully decoded at destination).

$\overline{N}$ and $\overline{N^2}$ for each ARQ protocol described in Section III will be derived in later subsections. In the following, Theorem 1 lists the fundamental formulas used commonly in the unified analysis framework. Its proof can be found in Appendix A.

Theorem 1: Given the arrival rate of segments (λ_s) and the first and second moments of N ($\overline{N}$ and $\overline{N^2}$), the segment delay T, the buffer occupancy L, and the saturation throughput η for the TDMA system are

$$T = \frac{\lambda_s \overline{N^2} T_F{}^2}{2(1-\lambda_s \overline{N} T_F)} + (\overline{N} - \frac{F-1}{F} + \frac{1}{2})T_F$$

$$L = \lambda_s T = \frac{\lambda_s{}^2 \overline{N^2} T_F{}^2}{2(1-\lambda_s \overline{N} T_F)} + \lambda_s(\overline{N} - \frac{F-1}{F} + \frac{1}{2})T_F$$

$$\eta = \frac{\overline{K}}{\overline{N}\, T_F}$$

where F is the number of time slots per time frame; T_F is the length of a time frame; $\overline{K}$ is the average segment size.

With the above formulas commonly used when calculating segment delay, saturation throughput and buffer occupancy for any specific protocol, what is left is to derive $\overline{N}$ and $\overline{N^2}$ for that protocol. Before deriving $\overline{N}$ and $\overline{N^2}$ for the OF, LS, and HP protocols in the following subsections, we list below a few symbols, each representing a random variable frequently used in this section.

- N: The number of blocks per segment sent by either source or relay (until a segment is fully decoded at destination)
- N_S: The number of blocks per segment sent by source
- N_R: The number of blocks per segment sent by relay
- M_{SD}: The number of blocks per segment sent by source that successfully reach the destination
- M_{RD}: The number of blocks per segment sent by relay that successfully reach the destination
- N_i: The number of blocks per segment sent in the i-th stage of a RANC ARQ protocol

A. The OF Protocol

This subsection aims to analyze the saturation throughput, average segment delay, and buffer occupancy for the OF protocol. With the commonly used formulas listed in Theorem 1, what we need to do is to derive $\overline{N}$ and $\overline{N^2}$ for OF.

Under the OF protocol, a block can reach the destination either through the direct link $\overline{SD}$ or along the relaying path $S \rightarrow R \rightarrow D$. After transmitting a block, the source learns the transmission successful if it receives an ACK. The probability

[4]The information about which blocks are combined is sent to the destination as a field of the packet.

that the source learns a block transmission successful is $p = (1-(1-P_{SD})(1-P_{SR}P_{RD}))P_{\text{DS,ACK}}$. After a failed block transmission, the source will retransmit the same block until it learns the block transmission successful.

Because a segment consists of K original blocks and the probability that the source receives a (positive) ACK after sending a block is p, the number of blocks sent by source, N_S, is a negatively binomially distributed random variable. More precisely, given the value of K, $N_S \sim \text{NB}(p, K)$. Using the well-known formulas for the mean and the second moment of a negative binomial distribution, $\overline{N_S}$ and $\overline{N_S{}^2}$ can be derived as follows:

$$\overline{N_S} = \mathbf{E}\big[\mathbf{E}[N_S|K]\big] = \mathbf{E}\big[K/p\big] = \frac{\overline{K}}{p} \tag{1}$$

$$\begin{aligned}\overline{N_S{}^2} &= \mathbf{E}\big[\mathbf{E}[N_S{}^2|K]\big]\\ &= \mathbf{E}\Big[\frac{K^2+K(1-p)}{p^2}\Big] = \frac{\overline{K^2}+\overline{K}(1-p)}{p^2}\end{aligned} \tag{2}$$

The relay forwards an overheard block only if the corresponding direct transmission from source to destination is failed. The probability that a direct transmission is failed is $1-P_{SD}$ and the probability that the relay overhears a block sent by source is P_{SR}. Therefore, every time after the source sends a block, the probability that the relay will forward the block is $p' = P_{SR}(1-P_{SD})$. Since the source sends N_S blocks in all, the total number of blocks the relay forwards is $N_R = \sum_{i=1}^{N_S} X_i$, where $X_i \sim \text{Bernoulli}(p')$ is a Bernoulli random variable with the success probability of p'. X_is are independent and identically distributed (iid).

Because $N_R = \sum_{i=1}^{N_S} X_i$, N_R is a Binomial random variable given the value of N_S. That is, given the value of N_S, $N_R \sim \text{Binomial}(p', N_S)$. By the well-known formulas for the mean and the second moment of a binomial distribution, $\overline{N_R}$ and $\overline{N_R{}^2}$ can be derived as follows:

$$\overline{N_R} = \mathbf{E}[N_R|N_S] = \mathbf{E}[p' \cdot N_S] = p' \cdot \overline{N_S} \tag{3}$$

$$\begin{aligned}\overline{N_R{}^2} &= \mathbf{E}[N_R{}^2|N_S] = \mathbf{E}[N_S \cdot p'(N_S \cdot p' + 1 - p')]\\ &= (p')^2\overline{N_S{}^2} + p'(1-p')\overline{N_S}\end{aligned} \tag{4}$$

Now we are ready to derive $\overline{N}$ and $\overline{N^2}$ for the OF protocol.

Theorem 2: Given the first and second moments of segment size ($\overline{K}$ and $\overline{K^2}$), $\overline{N}$ and $\overline{N^2}$ for the OF protocol are

$$\overline{N} = \frac{\overline{K}}{p}(1+p')$$

$$\overline{N^2} = (1+p')^2\frac{\overline{K^2}+\overline{K}(1-p)}{p^2} + p'(1-p')\frac{\overline{K}}{p}$$

where $p = (1-(1-P_{SD})(1-P_{SR}P_{RD}))P_{\text{DS,ACK}}$ is the probability of the source learning a block transmission successful and $p' = P_{SR}(1-P_{SD})$ is the probability of the relay forwarding blocks.

Proof:

By definition, $N = N_S + N_R$. Summing up (1) and (3) gets

$$\overline{N} = \overline{N_S} + \overline{N_R} = \frac{\overline{K}}{p}(1+p')$$

Next we derive $\overline{N^2} = \overline{N_S{}^2} + \overline{N_R{}^2} + 2\overline{N_S N_R}$. Since $\overline{N_S{}^2}$ and $\overline{N_R{}^2}$ have been derived in (2) and (4), what is left is to derive $\overline{N_S N_R}$. To this end, we define the centralized random variable $Y_i \triangleq X_i - \overline{X} = X_i - p'$. Y_is are iid. $\overline{N_S N_R}$ can be derived with the help of Wald's second equation. Applying Wald's second equation to $(\sum_{i=1}^{N_S} Y_i)^2$, we get

$$\text{var}(Y_i) \cdot \overline{N_S} = \mathbf{E}\left[\left(\sum_{i=1}^{N_S} Y_i\right)^2\right]$$

Because $Y_i = X_i - p'$ and $X_i \sim \text{Bernoulli}(p')$, we know $\text{var}(Y_i) = \text{var}(X_i) = p'(1-p')$. So

$$\begin{aligned}p'(1-p') \cdot \overline{N_S} &= \mathbf{E}\left[\left(\sum_{i=1}^{N_S}(X_i - p')\right)^2\right]\\ &= \mathbf{E}\left[\left(\sum_{i=1}^{N_S} X_i - p'N_S\right)^2\right]\\ &= \mathbf{E}\left[(N_R - p'N_S)^2\right]\\ &= \overline{N_R{}^2} - 2p'\overline{N_S N_R} + (p')^2\overline{N_S{}^2}\end{aligned}$$

Substituting (4) into the above equation, we get $\overline{N_S N_R}$ after some arithmetic manipulation:

$$\begin{aligned}\overline{N_S N_R} &= \frac{(p')^2\overline{N_S{}^2} + \overline{N_R{}^2} - p'(1-p')\overline{N_S}}{2p'}\\ &= p'\overline{N_S{}^2}\end{aligned} \tag{5}$$

By (4) and (5), we get

$$\begin{aligned}\overline{N^2} &= \overline{N_S{}^2} + \overline{N_R{}^2} + 2\overline{N_S N_R}\\ &= (1+p')^2\overline{N_S{}^2} + p'(1-p')\overline{N_S}\end{aligned}$$

After substituting (1) and (2) into the above equation, we have derived $\overline{N^2} = (1+p')^2\frac{\overline{K^2}+\overline{K}(1-p)}{p^2} + p'(1-p')\frac{\overline{K}}{p}$. ■

Here we give an interpretation of how $\overline{N}$ for the OF protocol comes from. The source keeps sending blocks until it learns K transmissions successful; so the source takes $\frac{\overline{K}}{p}$ transmissions on average. The probability of the relay forwarding blocks is p'; therefore, the relay takes $\frac{\overline{K}}{p}p'$ transmissions on average. Summing up the two numbers of transmissions, we know that the source and relay nodes in all take $\overline{N} = \frac{\overline{K}}{p}(1+p')$ transmissions. This interpretation can be regarded as a verification of the above proof.

With Theorem 2, it is easy to obtain the saturation throughput, average segment delay, and buffer occupancy for the OF protocol, simply by substituting $\overline{N}$ and $\overline{N^2}$ given in Theorem 2 into the formulas in Theorem 1.

B. The LS Protocol

This subsection aims to analyze the saturation throughput, average segment delay, and buffer occupancy for the LS protocol. With the commonly used formulas listed in Theorem 1, what we need to do is to derive $\overline{N}$ and $\overline{N^2}$ for LS.

To derive $\overline{N}$ and $\overline{N^2}$ for the LS protocol, we consider the behavior of source and relay nodes. The source sends out coded blocks in the first stage and then keeps silent in the second stage. The first stage finishes when relay and destination nodes collectively receive K distinct blocks. As a result, given the value of K, the number of blocks sent by source, N_S, is a random variable following the negative binomial distribution:

$$N_S \sim \mathrm{NB}(q, K)$$

where $q = P_{SD} + P_{SR} - P_{SD}P_{SR}$ is the probability that a single transmission sent by source reaches either relay or destination (or both). We call such a transmission *helpful*.

Among helpful transmissions, a part of them reach the relay node but fail to reach the destination node. Denote the number of such transmissions by W. Given the value of K, W is a binomially-distributed random variable:

$$W \sim \mathrm{Binomial}(q', K)$$

where $q' = P_{SR}(1-P_{SD})/q$ is the conditional probability that source's transmission fails to reach the destination given the transmission is helpful. q' can be also regarded as the fraction of helpful transmissions that fail to reach the destination.

At the end of the first stage, the destination missed the W blocks. Because the source does not send any block in the second stage, it is relay's responsibility to deliver the information carried by these W blocks to the destination. (The relay supplements the missing information by sending out recoded blocks.) The value of W affects the actual number of recoded blocks sent by relay, N_R. Obviously, given the value of W, N_R is negatively binomially distributed:

$$N_R \sim \mathrm{NB}(P_{RD}, W)$$

Using the well-known formulas of the means and the second moments of a binomial distribution and a negatively binomial distribution, we can derive the first and second moments of N_S and W as:

$$\overline{N_S} = \mathbf{E}[\mathbf{E}[N_S|K]] = \mathbf{E}[K/q] = \overline{K}/q \tag{6}$$

$$\begin{aligned}\overline{{N_S}^2} &= \mathbf{E}\left[\mathbf{E}[{N_S}^2|K]\right] = \mathbf{E}\left[\frac{K^2 + K(1-q)}{q^2}\right]\\ &= \frac{1}{q^2}\overline{K^2} + \frac{1-q}{q^2}\overline{K}\end{aligned} \tag{7}$$

$$\overline{W} = \mathbf{E}[\mathbf{E}[W|K]] = \mathbf{E}[q'K] = q'\overline{K} \tag{8}$$

$$\begin{aligned}\overline{W^2} &= \mathbf{E}\left[\mathbf{E}[W^2|K]\right] = \mathbf{E}\left[Kq'(Kq' + 1 - q')\right]\\ &= (q')^2\overline{K^2} + q'(1-q')\overline{K}\end{aligned} \tag{9}$$

By $\overline{W}$ given in (8), $\overline{W^2}$ given in (9), and the well-known formulas of the mean and the second moment of a negatively binomial distribution, we can derive $\overline{N_R}$ and $\overline{{N_R}^2}$ as:

$$\overline{N_R} = \mathbf{E}[\mathbf{E}[N_R|W]] = \mathbf{E}[W/P_{RD}] = \frac{\overline{W}}{P_{RD}} = \frac{q'}{P_{RD}}\overline{K} \tag{10}$$

$$\begin{aligned}\overline{{N_R}^2} &= \mathbf{E}\left[\mathbf{E}[{N_R}^2|W]\right] = \mathbf{E}\left[\frac{W^2 + W(1-P_{RD})}{{P_{RD}}^2}\right]\\ &= \frac{1}{{P_{RD}}^2}\overline{W^2} + \frac{1-P_{RD}}{{P_{RD}}^2}\overline{W}\\ &= \frac{(q')^2}{{P_{RD}}^2}\overline{K^2} + \frac{q'(2-q'-P_{RD})}{{P_{RD}}^2}\overline{K}\end{aligned} \tag{11}$$

Next, we derive $\overline{N}$ and $\overline{N^2}$ based on the above observations.

Theorem 3: Given the first and second moments of segment size ($\overline{K}$ and $\overline{K^2}$), $\overline{N}$ and $\overline{N^2}$ for the LS protocol are

$$\begin{aligned}\overline{N} &= \left(\frac{1}{q} + \frac{q'}{P_{RD}}\right)\overline{K}\\ \overline{N^2} &= \left(\frac{1}{q^2} + \frac{(q')^2}{{P_{RD}}^2}\right)\overline{K^2}\\ &\quad + \left(\frac{1-q}{q^2} + \frac{q'(2-q'-P_{RD})}{{P_{RD}}^2}\right)\overline{K} + \frac{2q'}{qP_{RD}}(\overline{K})^2\end{aligned}$$

where $q = P_{SD}+P_{SR}-P_{SD}P_{SR}$ is the probability of helpful transmissions and $q' = P_{SR}(1-P_{SD})/q$ is the probability of helpful transmissions failing to reach the destination.

Proof: By definition, $N = N_S + N_R$, which implies that $\overline{N} = \overline{N_S} + \overline{N_R}$ and $\overline{N^2} = \overline{{N_S}^2} + \overline{{N_R}^2} + 2\overline{N_S N_R}$. Using (6) and (10), it is straightforward to derive $\overline{N}$ as:

$$\overline{N} = \overline{N_S} + \overline{N_R} = \left(\frac{1}{q} + \frac{q'}{P_{RD}}\right)\overline{K}$$

To derive $\overline{N^2} = \overline{{N_S}^2} + \overline{{N_R}^2} + 2\overline{N_S N_R}$, what is left is to derive $\overline{N_S N_R}$, since $\overline{{N_S}^2}$ and $\overline{{N_R}^2}$ are given in (7) and (11) respectively. In Appendix B, we prove $\overline{N_S N_R} = \overline{N_S}\cdot\overline{N_R}$. So $\overline{N_S N_R}$ can be obtained by multiplying (6) and (10):

$$\overline{N_S N_R} = \overline{N_S}\cdot\overline{N_R} = \frac{q'}{qP_{RD}}(\overline{K})^2 \tag{12}$$

Substituting (7), (11), and (12) into $\overline{N^2} = \overline{{N_S}^2} + \overline{{N_R}^2} + 2\overline{N_S N_R}$, we get

$$\begin{aligned}\overline{N^2} &= \left(\frac{1}{q^2} + \frac{(q')^2}{{P_{RD}}^2}\right)\overline{K^2}\\ &\quad + \left(\frac{1-q}{q^2} + \frac{q'(2-q'-P_{RD})}{{P_{RD}}^2}\right)\overline{K} + \frac{2q'}{qP_{RD}}(\overline{K})^2\end{aligned}$$

Therefore, we have proven this theorem. ■

Here we give an interpretation of how $\overline{N}$ for the LS protocol comes from. The source keeps sending blocks until there are K helpful transmissions. So the source takes $\overline{K}/q$ transmissions on average. Helpful transmissions fail to reach the destination with probability q' and it is the relay's responsibility to deliver these missing blocks to the destination over link $\overline{RD}$. On average, there are $\overline{K}q'$ missing blocks. Because a transmission over link $\overline{RD}$ reaches the destination with probability P_{RD}, the relay takes $\overline{K}q'/P_{RD}$ transmissions on average. Summing up the two numbers of transmissions, we get that the total number of transmissions the source and relay nodes take is $\overline{N} = (\frac{1}{q} + \frac{q'}{P_{RD}})\overline{K}$. This interpretation can be regarded as a verification of the above proof.

With Theorem 3, it is easy to obtain the saturation throughput, average segment delay, and buffer occupancy for the LS protocol, simply by substituting $\overline{N}$ and $\overline{N^2}$ given in Theorem 3 into the formulas in Theorem 1.

C. The HP Protocol

This subsection aims to analyze the saturation throughput, average segment delay, and buffer occupancy for the HP protocol. With the commonly used formulas listed in Theorem 1, what we need to do is to derive $\overline{N}$ and $\overline{N^2}$ for HP.

For a contained presentation, $\overline{N}$ and $\overline{N^2}$ for HP are derived in this subsection, assuming all segments are of a constant size

k. Note that these results under this assumption can be easily generalized to the case of having random-sized segments, by unconditioning $E[N|K=k]$ and $E[N^2|K=k]$ over all possible segment sizes as:

$$\overline{N} = \sum_{k=1}^{\infty} \Pr(K=k)\,\mathbf{E}[N|K=k]$$
$$\overline{N^2} = \sum_{k=1}^{\infty} \Pr(K=k)\,\mathbf{E}[N^2|K=k]$$

Starting here, we assume that all segments are of equal size k. For a tractable analysis, we make two simplifications:

- In the first stage, the source sends h/P_{SR} coded blocks and the destination receives $\lfloor h/P_{SR}\cdot P_{SD}\rfloor$ coded blocks.
- In the second stage, the destination needs to receive $k' = k-\lfloor h/P_{SR}\cdot P_{SD}\rfloor = \lceil k-h/P_{SR}\cdot P_{SD}\rceil$ blocks to retrieve a complete segment.

The above simplifications implies $N = h/P_{SR}+N_2$, where the first term h/P_{SR} is a constant and the second term N_2 is the number of blocks sent by either source or relay in the second stage until the destination receives k' blocks. To derive $\overline{N}$ and $\overline{N^2}$, what we need to do is to derive $\overline{N_2}$ and $\overline{{N_2}^2}$. Because the destination receives k' blocks in the second stage, we know $N_2 = N_2(k')$, where $N_2(x)$ is defined as the random variable representing the number of blocks sent (either by source or relay) for the destination to receives x blocks in the second stage. It is defined that $N_2(-1) = N_2(0) = 0$ because no block would be sent out if the destination does not have to receive any block. $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$ are defined as the mean and the second moment of $N_2(x)$, respectively. By this definition, $\overline{N_2} = \overline{N_2(k')}$ and $\overline{{N_2}^2} = \overline{{N_2}^2(k')}$.

To analyze $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$, we observe the second stage for the three-node (source-relay-destination) network. The second stage is split into the first *round* and the remaining rounds. The first round finishes right before source transmits the second coded block in the second stage; the remaining rounds starts upon source's second transmission and finishes at the end of the second stage.

The first round has six possible outcomes. Each represents a combination that *i)* the coded block sent by the source may or may not reach the destination, *ii)* the relay may or may not overhear the coded block, and *iii)* if the relay overhears the coded block, the recoded block sent by the relay may or may not reach the destination. Considering the six outcomes and using the law of total expectation, $\overline{N_2(x)}$ can be formulated into the following recursive expression for $x \geq 1$:

$$\begin{aligned}\overline{N_2(x)} = \quad & P_{SD}(1-P_{SR})\,\overline{1+N_2(x-1)}\\ &+ P_{SD}P_{SR}P_{RD}\,\overline{2+N_2(x-2)}\\ &+ P_{SD}P_{SR}(1-P_{RD})\,\overline{2+N_2(x-1)}\\ &+ (1-P_{SD})(1-P_{SR})\,\overline{1+N_2(x)}\\ &+ (1-P_{SD})P_{SR}P_{RD}\,\overline{2+N_2(x-1)}\\ &+ (1-P_{SD})P_{SR}(1-P_{RD})\,\overline{2+N_2(x)} \qquad (13)\end{aligned}$$

where each term at the right-hand side corresponds to one of the six outcomes. Take the first term as an example. The first term corresponds to the outcome when the coded block sent (by source) in the first round reaches the destination but not the relay. The probability of this outcome is $P_{SD}(1-P_{SR})$. In this case, the total number of blocks sent in the first round is 1, because a relay sends a recoded block only if it overhears a coded block. The destination, which needs to receive x blocks in total and has already received one block in the first round, will receive $x-1$ blocks in the remaining rounds. This is how we get the first term of $\overline{N_2(x)}$'s recursive expression; the other five terms can be derived in a similar way.

Similarly, $\overline{{N_2}^2(x)}$ can be formulated into the following recursive expression for $x \geq 1$:

$$\begin{aligned}\overline{{N_2}^2(x)} = \quad & P_{SD}(1-P_{SR})\,\overline{(1+N_2(x-1))^2}\\ &+ P_{SD}P_{SR}P_{RD}\,\overline{(2+N_2(x-2))^2}\\ &+ P_{SD}P_{SR}(1-P_{RD})\,\overline{(2+N_2(x-1))^2}\\ &+ (1-P_{SD})(1-P_{SR})\,\overline{(1+N_2(x))^2}\\ &+ (1-P_{SD})P_{SR}P_{RD}\,\overline{(2+N_2(x-1))^2}\\ &+ (1-P_{SD})P_{SR}(1-P_{RD})\,\overline{(2+N_2(x))^2} \quad (14)\end{aligned}$$

Based on the above observations, we can exploit a recurrence solving technique called annihilation to derive $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$. The derivation is shown in Theorem 4. With $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$ as well as the formula of $N = h/P_{SR} + N_2(k')$, it is easily to obtain $\overline{N} = h/P_{SR} + \overline{N_2(k')}$ and $\overline{N^2} = (h/P_{SR})^2 + 2(h/P_{SR})\overline{N_2(k')} + \overline{{N_2}^2(k')}$.

Theorem 4: If all segments are of an equal size k, $\overline{N_2(x)}$, $\overline{{N_2}^2(x)}$, $\overline{N}$, and $\overline{N^2}$ for the HP protocol are

$$\begin{aligned}\overline{N_2(x)} &= \frac{-r'(1-r)}{(2-r)^2}(r-1)^x + \frac{r'(1-r)}{(2-r)^2} + \frac{r'}{2-r}x\\ \overline{{N_2}^2(x)} &= \alpha(r-1)^x + \beta x(r-1)^x + \gamma + \delta x + \epsilon x^2\\ \overline{N} &= h/P_{SR} + \overline{N_2(k')}\\ \overline{N^2} &= (h/P_{SR})^2 + 2(h/P_{SR})\overline{N_2(k')} + \overline{{N_2}^2(k')}\end{aligned}$$

where the constants are

$$\begin{aligned}k' &= \lceil k - h/P_{SR}\cdot P_{SD}\rceil\\ r &= \frac{P_{SD}+P_{SR}P_{RD}-2P_{SD}P_{SR}P_{RD}}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}}\\ r' &= \frac{1+P_{SR}}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}}\end{aligned}$$

$$\begin{bmatrix}\alpha\\ \beta\\ \gamma\\ \delta\\ \epsilon\end{bmatrix} = \begin{bmatrix}(r-1)^{-1} & -1(r-1)^{-1} & 1 & -1 & (-1)^2\\ (r-1)^0 & 0(r-1)^0 & 1 & 0 & 0^2\\ (r-1)^1 & 1(r-1)^1 & 1 & 1 & 1^2\\ (r-1)^2 & 2(r-1)^2 & 1 & 2 & 2^2\\ (r-1)^3 & 3(r-1)^3 & 1 & 3 & 3^2\end{bmatrix}^{-1}\begin{bmatrix}0\\ 0\\ \overline{{N_2}^2(1)}\\ \overline{{N_2}^2(2)}\\ \overline{{N_2}^2(3)}\end{bmatrix}$$

and $\overline{{N_2}^2(1)}$, $\overline{{N_2}^2(2)}$, and $\overline{{N_2}^2(3)}$ can be computed by (16).

Proof: What we need to do is to derive the closed forms of $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$. The two closed forms can be obtained by annihilating the recursive expressions of $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$ shown in (13) and (14), respectively.

Step 1: Derivation of the closed form of $\overline{N_2(x)}$.

After some algebraic manipulations on (13), we get

$$\begin{aligned}\overline{N_2(x)} = & \frac{1+P_{SR}}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}} \\ & + \frac{P_{SD}+P_{SR}P_{RD}-2P_{SD}P_{SR}P_{RD}}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}}\overline{N_2(x-1)} \\ & + \frac{P_{SD}P_{SR}P_{RD}}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}}\overline{N_2(x-2)}\end{aligned}$$

The above equation is equivalent to the below recurrence for $x \geq -1$.

$$\overline{N_2(x+2)} - r\overline{N_2(x+1)} - (1-r)\overline{N_2(x)} = r' \tag{15}$$

Denote the shift operator by $\boldsymbol{E}$.[5] Regarding $\overline{N_2(x)}$ as a function of x, the recurrence (15) can be rewritten as

$$\big(\boldsymbol{E}-(r-1)\big)\big(\boldsymbol{E}-1\big)\overline{N_2(x)} = r'$$

Observing the above equation, becasue $(\boldsymbol{E}-1)$ annihilates the constant r', we know $(\boldsymbol{E}-(r-1))(\boldsymbol{E}-1)^2$ annihilates $\overline{N_2(x)}$. So we get the generic solution for $x \geq -1$:

$$\overline{N_2(x)} = a\cdot(r-1)^x + b\cdot 1^x + c\cdot x\,1^x = a(r-1)^x + b + cx$$

where a, b, c are constants that satisfy the system of equations corresponding to the boundary conditions at $x = -1, 0, 1$:

$$\begin{cases}\overline{N_2(-1)} = 0 = \frac{1}{r-1}a + b - c \\ \overline{N_2(0)} = 0 = a + b \\ \overline{N_2(1)} = r' = (r-1)a + b + c\end{cases}$$

$\overline{N_2(-1)}$ and $\overline{N_2(0)}$ both equal zero, because $N_2(-1) = N_2(0) = 0$ by definition. $\overline{N_2(1)} = r'$ can be easily obtained by substituting x by -1 into (15).

Solving the above system of equations, we get the constants

$$a = \frac{-r'(1-r)}{(2-r)^2}, b = \frac{r'(1-r)}{(2-r)^2}, c = \frac{r'}{2-r}$$

and hence

$$\overline{N_2(x)} = \frac{-r'(1-r)}{(2-r)^2}(r-1)^x + \frac{r'(1-r)}{(2-r)^2} + \frac{r'}{2-r}x$$

Step 2: Derivation of the closed form of $\overline{{N_2}^2(x)}$.

After some algebraic manipulations on (14), we get

$$\begin{aligned}\overline{{N_2}^2(x)} = & \frac{1}{P_{SD}+(1-P_{SD})P_{SR}P_{RD}}\Big\{1+3P_{SR} \\ & + 2(1-P_{SD})(1+P_{SR}-2P_{SR}P_{RD})\overline{N_2(x)} \\ & + 2(P_{SD}+P_{SD}P_{SR}+2P_{SR}P_{RD}-4P_{SD}P_{SR}P_{RD})\overline{N_2(x-1)} \\ & + 4P_{SD}P_{SR}P_{RD}\overline{N_2(x-2)} \\ & + (P_{SD}+P_{SR}P_{RD}-2P_{SD}P_{SR}P_{RD})\overline{{N_2}^2(x-1)} \\ & + P_{SD}P_{SR}P_{RD}\overline{{N_2}^2(x-2)}\Big\}\end{aligned} \tag{16}$$

The above equation is equivalent to the below recurrence for $x \geq -1$.

$$\begin{aligned}&\overline{{N_2}^2(x+2)} - r\overline{{N_2}^2(x+1)} - (1-r)\overline{{N_2}^2(x)} \\ &\quad = c' + c_2\overline{N_2(x+2)} + c_1\overline{N_2(x+1)} + c_0\overline{N_2(x)}\end{aligned} \tag{17}$$

[5]The (linear) shift operator $\boldsymbol{E}$ is, for any function f, $(\boldsymbol{E}\,f)(x) = f(x+1)$ for all x. More generally, for any positive integer n, the operator $\boldsymbol{E}^n$ shifts its argument n times. That is, $\boldsymbol{E}^n\,f(x) = f(x+n)$. The compound operator $\boldsymbol{E}-1$ is defined by setting $(\boldsymbol{E}-1)f = \boldsymbol{E}\,f + (-1)f$ for any function f. $(\boldsymbol{E}-1)^2$ is shorthand for $(\boldsymbol{E}-1)(\boldsymbol{E}-1)$, which applies $(\boldsymbol{E}-1)$ twice.

where c', c_2, c_1, c_0 are constants.

Regarding $\overline{N_2(x)}$ and $\overline{{N_2}^2(x)}$ as functions of x, the recurrence (17) can be rewritten as

$$\begin{aligned}&\big(\boldsymbol{E}-(r-1)\big)\big(\boldsymbol{E}-1\big)\overline{{N_2}^2(x)} \\ &\quad = c' + \big(c_2\,\boldsymbol{E}^2 + c_1\,\boldsymbol{E} + c_0\big)\overline{N_2(x)}\end{aligned} \tag{18}$$

In Step 1, we know $(\boldsymbol{E}-1)$ annihilates any constant and $(\boldsymbol{E}-(r-1))(\boldsymbol{E}-1)^2$ annihilates $\overline{N_2(x)}$. Therefore, $(\boldsymbol{E}-(r-1))(\boldsymbol{E}-1)^2$ annihilates the right-hand side of the above equation, which implies the left-hand side is also annihilated. Thus, $\overline{{N_2}^2(x)}$ is annihilated by $[(\boldsymbol{E}-(r-1))(\boldsymbol{E}-1)^2]\cdot[(\boldsymbol{E}-(r-1))(\boldsymbol{E}-1)] = (\boldsymbol{E}-(r-1))^2(\boldsymbol{E}-1)^3$.

So we get the generic solution for $x \geq -1$:

$$\overline{{N_2}^2(x)} = \alpha(r-1)^x + \beta x(r-1)^x + \gamma + \delta x + \epsilon x^2$$

where $\alpha, \beta, \gamma, \delta, \epsilon$ are constants that satisfy the system of equations representing the boundary conditions at $x = -1, 0, 1, 2, 3$:

$$\begin{cases}\overline{{N_2}^2(\text{-}1)} = (r-1)^{\text{-}1}\alpha + (\text{-}1)(r-1)^{\text{-}1}\beta + \gamma + (\text{-}1)\delta + (\text{-}1)^2\epsilon \\ \overline{{N_2}^2(0)} = (r-1)^0\alpha + 0(r-1)^0\beta + \gamma + 0\delta + 0^2\epsilon \\ \overline{{N_2}^2(1)} = (r-1)^1\alpha + 1(r-1)^1\beta + \gamma + 1\delta + 1^2\epsilon \\ \overline{{N_2}^2(2)} = (r-1)^2\alpha + 2(r-1)^2\beta + \gamma + 2\delta + 2^2\epsilon \\ \overline{{N_2}^2(3)} = (r-1)^3\alpha + 3(r-1)^3\beta + \gamma + 3\delta + 3^2\epsilon\end{cases}$$

$\overline{{N_2}^2(-1)} = \overline{{N_2}^2(0)} = 0$ because $N_2(-1) = N_2(0) = 0$ by definition. The values of $\overline{{N_2}^2(1)}$, $\overline{{N_2}^2(2)}$, and $\overline{{N_2}^2(3)}$ can be computed using (16).

Solving the above system of equations, we get the constants α, β, γ, δ, and ϵ. So far, we have proven this theorem. ■

VI. Discussions on HP

It is desirable that the HP protocol performs comparably well to the LS protocol. In particular, we hope that HP has the same saturation throughput as LS has. Because saturation throughput mainly depends on $\overline{N}$ (cf. Theorem 1), we attempt to make the value of $\overline{N}$ of the HP protocol equal that of the LS protocol. A key to this goal is to choose an appropriate value for h.

Under the LS protocol, given the segment $K = k$, the average numbers of blocks that node S and node R send out are respectively

$$N_S^{(\text{LS})} = \frac{1}{q}k = \frac{1}{P_{SD}+P_{SR}-P_{SD}P_{SR}}k$$

and

$$N_R^{(\text{LS})} = \frac{q'}{P_{RD}}k = \frac{P_{SR}(1-P_{SD})}{P_{RD}(P_{SD}+P_{SR}-P_{SD}P_{SR})}k$$

Under the HP protocol, a relay node does not start to send out recoded blocks until it has received h coded blocks. To make HP achieve a performance comparable to LS, we hope that in the next time slot after node S sends the $N_S^{(\text{LS})}$-th coded block, node R sends out the $N_R^{(\text{LS})}$-th recoded block and the segment transmission completes. Because the probability of node R overhearing a block sent by node S is P_{SR}, node R should hold quiescent until it has overheard

$$h = N_S^{(\mathrm{LS})} P_{SR} - N_R^{(\mathrm{LS})} = \frac{P_{SR}(P_{SD} + P_{RD} - 1)}{P_{RD}(P_{SD} + P_{SR} - P_{SD}P_{SR})} k \tag{19}$$

coded blocks. After that, it sends out a recoded block each time it overhears a coded block.[6]

We note that by observing (19), a challenge occurs when $P_{SD} + P_{RD} < 1$. In this situation, a negative value of h implies that a relay node should not only start to help from the very first but also help the source node more aggressively than the HP protocol could. That is, instead of sending out a recoded block each time when overhearing a coded block, the relay node probably should send out more than one recoded block. To this end, we develop an advanced RANC ARQ protocol, called the work-based opportunistic RANC protocol. The preliminary work about the work-based protocol is published in [37] and is beyond the scope of this paper.

VII. Performance Evaluation

In this section, we evaluate the RANC protocols—LS and HP—and the C-ARQ protocol—OF—in terms of saturation throughput and average segment delay. The LS protocol can be regarded as an ideal scheme, offering a performance bound. The OF protocol is representative of C-ARQ protocols. The performance of LS and HP is compared with OF to see performance gains brought by a RANC ARQ protocol over a C-ARQ protocol. The performance of HP is compared with LS to see how much room left for a RANC ARQ protocol without the need of per-packet ACKs to be further improved. This section presents both theoretical results and simulation results. The simulation results are obtained using the in-house simulator we develop.

For a specific RANC ARQ protocol, the throughput gain is defined as its saturation throughput (η) divided by OF's saturation throughput (η_{OF}):

$$G_\eta = \frac{\eta}{\eta_{\mathrm{OF}}}$$

Throughput gain is the greater the better. A value greater than one implies outperformance against the OF protocol.

While related work has shown that C-ARQ protocols outperform conventional ARQ protocols, this section intends to show that RANC protocols can even perform better than C-ARQ protocols. The reasons that the two RANC ARQ protocols, LS and HP, are expected to outperform the C-ARQ protocol, OF, are two-fold. First, both LS and HP indeed take link reception probabilities into consideration (explicitly or implicitly) and therefore, they transmit packets more efficiently than OF does. Second, the use of network coding minimizes unneeded retransmissions caused by ACK loss.

In the following subsections, the RANC ARQ protocols and the C-ARQ protocol are extensively evaluated and compared with each other, by theoretical results and through simulations under various environments. Comparing theoretical results and

[6]The number of received blocks cannot be negative. So, for both analysis and simulations, h is set to 0 if (19) turns out to be a negative number.

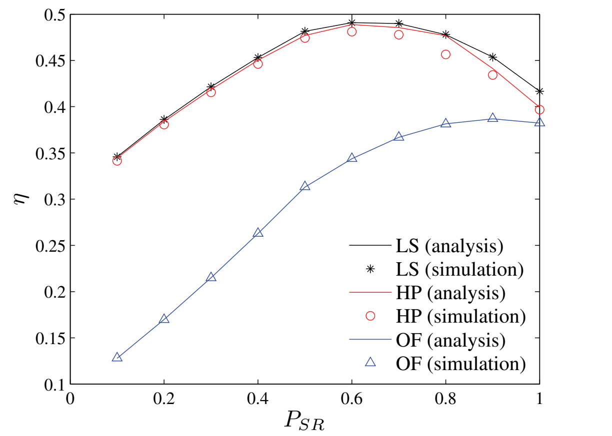

Fig. 6. Theoretical/simulated saturation throughput. $P_{SD} = 0.3$. $P_{RD} = \min(1.5 - P_{SR}, 1)$. $K = 100$.

simulation counterparts helps to ensure the analysis's validity. In addition to the performance evaluation under generic environments presented in sections VII-A, we also present a case study in Section VII-B.

A. Generic Environments

In the first analysis/simulation setup, we show theoretical results and simulation counterparts in the case where link probabilities P_{SR} and P_{RD} vary in such a way that a relay node near the source node has a high P_{SR} but a low P_{RD}, while a relay node close to the destination node has a low P_{SR} but a high P_{RD}. More precisely, P_{RD} is assumed to be either $1.5 - P_{SR}$ or one, whichever smaller. F is set to eight, that is, there are eight flows (source-destination pairs) in this simulation setup.

In this setup, saturation throughput (η) and throughput gain (G_η) are shown in Fig. 6 and Fig. 7, respectively. As one can observe, the theoretical results of all protocols match extremely well to their simulation counterparts, except HP. For HP, due to a few simplifications mentioned in Section V-C, there is a small difference between the theoretical results and simulation counterparts. The difference is less than 1% mostly and less than 5% for the largest one. This helps ensure the validity of the performance analysis derived in Section V.

Let us rank the protocols according to their saturation throughput. As observed in Fig. 6 and Fig. 7, LS performs best, which is expected. The second place is HP; HP's performance is very close to LS's performance. OF performs worst among these protocols. As observed, our proposed RANC protocols, LS and HP, significantly outperform the C-ARQ protocol, OF, in terms of saturation throughput from both absolute and relative perspectives. This matches our aforementioned expectation and explanation.

Fig. 6 and Fig. 7 also show that the performance of HP is quite close to the performance of LS. This makes HP attractive. In other words, HP is a simple yet efficient RANC protocol.

A phenomenon of turnaround is observed in Fig. 6: Along the x-axis, the saturation throughput first increases and then decreases after a certain point. This turnaround phenomenon is relevant to how the location of the relay node affects the

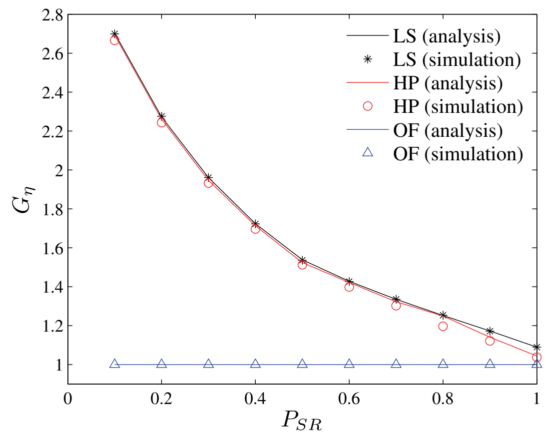

Fig. 7. Theoretical/simulated throughput gain. $P_{SD} = 0.3$. $P_{RD} = \min(1.5 - P_{SR}, 1)$. $K = 100$.

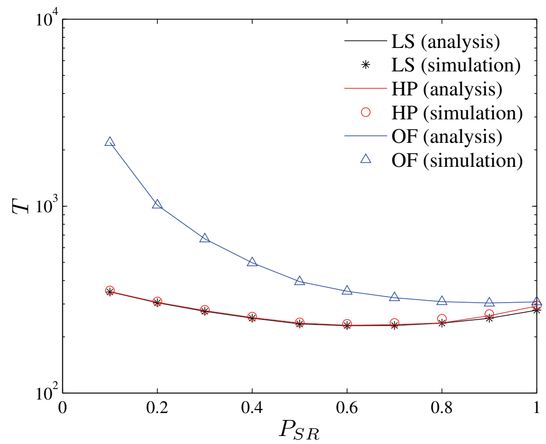

Fig. 8. Segment delay vs. P_{SR}. $P_{SD} = 0.3$. $P_{RD} = \min(1.5 - P_{SR}, 1)$. $K = 100$. $F = 8$. $T_F = 1$. The average packet arrival rate is set to 0.1.

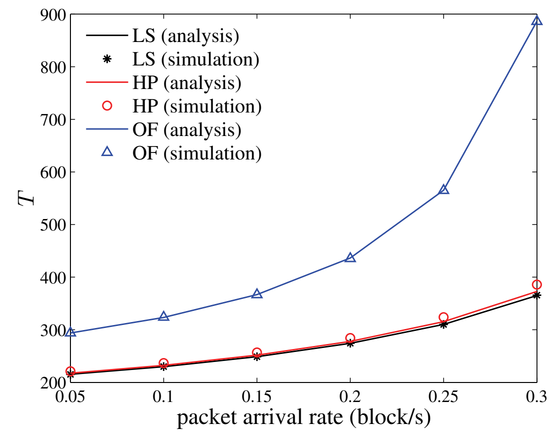

Fig. 9. Segment delay vs. packet arrival rate. $P_{SD} = 0.3$. $P_{SR} = 0.7$. $P_{RD} = 0.8$. $K = 100$. $F = 8$. $T_F = 1$.

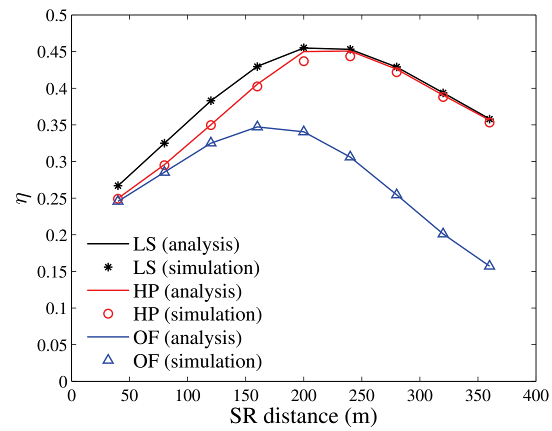

Fig. 10. Saturation throughput vs. source-relay distance in the case study.

performance. If the relay node is chosen to be close to the source but far away from the destination, P_{RD} is small; such a relay is not a good choice, because the packets it sends out are not very likely to reach the destination. If the relay node is chosen to be close to the destination but far away from the source, P_{SR} is small; such a relay does not help much, either, because the relay seldom overhears the packets sent by the source. A good relay node would lie close to the midpoint of the source and destination. This explains the turnaround phenomenon.

The same comparison conclusions drawn in terms of saturation throughput also apply in terms of segment delay. As shown in Fig. 8, all the theoretical results match well to the simulation counterparts; even HP only has a difference less than 5%. The ranking order is LS best, HP second place, and OF worst. LS and HP outperform OF drastically; meanwhile, HP has a performance almost comparable to LS.

In the second analysis/simulation setup, we evaluate the impact of packet arrival rate on T. The packet arrival rate is equal to the segment arrival rate (λ_s) times the average segment size ($\overline{K}$). As one can observe in Fig. 9, LS performs best again, in terms of segment delay. The second place, which tightly follows the champion, is still HP. And OF remains the worst one among these protocols. As the packet arrival rate increases, T increases with an increasing slope for all the protocols.

B. Case Study

In this subsection, a case study is presented to evaluate, in a WiMAX system, the performance gains of our proposed RANC ARQ protocols compared to the representative C-ARQ protocol, OF. Consider a base station (i.e., source) in the WiMAX system communicating to a subscriber (i.e., destination) or vice versa, with the help of a relay station (i.e., relay). Suppose that path loss and fading affect communications between source and destination, between source and relay, and between relay and destination. The channel parameters as well as modulation and coding are set according to pages 75-76 in [40]. In this setup, the probability of successful transmission over a link of length d_{ij} (which connects nodes i and j) is

$$p_{ij} = Q\left(\frac{24.7 + 30\ log_{10} d_{ij}}{6}\right)$$

where $Q(\cdot)$ is the Gaussian Q-function.

In this simulation setup, a relay station is located somewhere on the line segment connecting the base station and the subscriber. The distance between the base station and the subscriber is set to 400 meters.

Fig. 10, Fig. 11, and Fig. 12 show saturation throughput, throughput gain, and segment delay in this simulation setup,

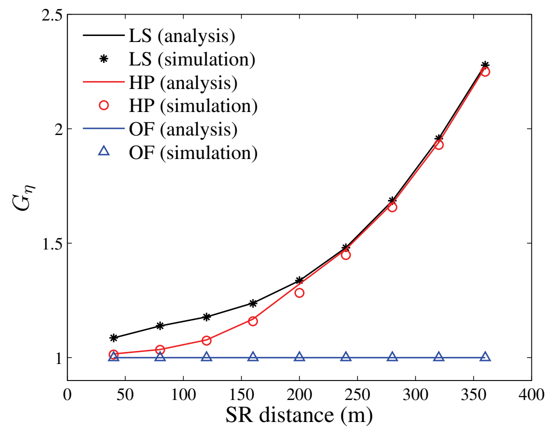

Fig. 11. Throughput gain vs. source-relay distance in the case study.

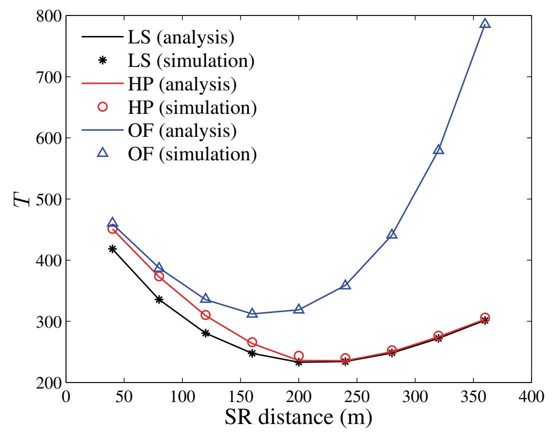

Fig. 12. Segment delay vs. source-relay distance in the case study.

respectively. As one can observe, the theoretical results and the simulation counterparts match extremely well to each other, except HP. For HP, the difference between theoretical results and simulation counterparts is small—less than 3% in saturation throughput and less than 4% in segment delay. In this case study, it is observed that the performance ranking is the same as previous subsection. LS performs best, HP is the second place, and OF performs worst. It is also observed that as long as the relay station is not far away from the subscriber (i.e., destination), the performance of HP is almost comparable to that of LS.

VIII. Conclusion

In this paper, the framework of the relay-assisted network-coding (RANC) ARQ protocols, which leverage both network coding and opportunistic retransmission, has been addressed. In particular, two RANC ARQ protocols—the LS protocol which offers a fundamental limit to any single-relay RANC ARQ protocol and the HP protocol which is simple yet efficient—have been proposed.

Moreover, an analysis framework for single-relay RANC ARQ protocols in TDMA-based networks has been presented, which essentially behave as a M/G/1 queues with vacation. We have derived saturation throughput and average segment delay for LS and HP, as well as for the OF protocol which is representative of C-ARQ protocols. Extensive analysis and simulations show that the two proposed RANC ARQ protocols bring significant performance gains over the representative C-ARQ protocol.

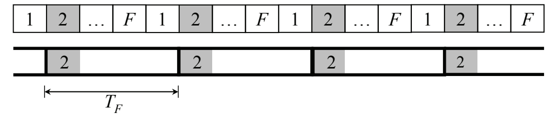

Fig. 13. The rectangles with thick edges represent the service rounds for flow 2, which are shifted time frames.

Appendix A
Proof of Theorem 1

It is known in textbooks that a conventional TDMA (and TDM) system, in which a source node sends each of its Poisson arrival packets directly to its destination node once, can be modeled as an M/G/1 queue with fixed-length vacation: When the queue is empty, the server takes a vacation of length T_F. If the queue is still empty after a vacation, the server takes another vacation of equal length.

Different from the conventional TDMA system, the TDMA system described in Section III allows relay nodes to help and also allows source and relay nodes to send the same packets (or some version of the same packets) multiple times. In order to connect such a TDMA system to the M/G/1 model with vacation, let us observe one flow (i.e., one source-relay-destination triple) at a time. In particular, we observe service rounds for a certain flow. Take flow 2 as an example. The rectangles with thick edges in Fig. 13 represent service rounds for flow 2, where the service rounds for a certain flow begin at the starts of time slots allocated to that flow. Each of the service rounds for a certain flow is essentially a shifted time frame. So the length of a service round is T_F.

One can regard the source and relay for a certain flow as a single entity, since the relay helps the source only in the flow's time slots. For a certain flow in the TDMA system, if the source and relay have no block to send at the start of a service round for the flow, they cannot send any block in that service round, even if a new block may arrive during the service round. This time period of length T_F is a "vacation", in which the source and relay cannot send any block. Focusing on the service rounds for a certain flow, the TDMA system we consider essentially behaves as an M/G/1 queue with vacation.

In an M/G/1 queue with vacation, the average waiting time (excluding the service time), denoted by W_Q, can be computed:

$$W_Q = \frac{\lambda_s \mathbf{E}[X^2]}{2(1-\lambda_s \mathbf{E}[X])} + \frac{\mathbf{E}[V^2]}{2\,\mathbf{E}[V]} \tag{20}$$

where X and V are random variables representing the service time and the vacation length, respectively.

In the TDMA system we consider, X depends on the actual number of service rounds, which is equal to N. More precisely, $X = NT_F$. The vacation length is $V = T_F$. Substituting these values into (20), we get

$$W_Q = \frac{\lambda_s \cdot \overline{N^2} \cdot {T_F}^2}{2(1-\lambda_s \overline{N} T_F)} + \frac{T_F}{2} \tag{21}$$

The segment transmission finishes earlier than the last service round ends. Observing Fig. 13, the segment transmission is done upon the end of the first time slot in the last service round. So we know that the segment delay is the length of the first $N-1$ service rounds plus the length of the first time slot in the last service round. Using (21), the average segment delay is

$$\begin{aligned} T &= W_Q + \left((\overline{N} - 1) + \frac{1}{F} \right) T_F \\ &= \frac{\lambda_s \cdot \overline{N^2} \cdot {T_F}^2}{2(1 - \lambda_s \overline{N} T_F)} + \left(\overline{N} - \frac{F-1}{F} + \frac{1}{2} \right) T_F \end{aligned}$$

Using Little's formula, we can derive the buffer occupancy

$$L = \lambda_s T = \frac{{\lambda_s}^2 \cdot \overline{N^2} \cdot {T_F}^2}{2(1 - \lambda_s \overline{N} T_F)} + \lambda_s \left(\overline{N} - \frac{F-1}{F} + \frac{1}{2} \right) T_F$$

The saturation throughput is the maximum departure rate which is equal to the average service rate. Since the average service rate is $1/(\overline{N} T_F)$ segments per second and a segment consists of $\overline{K}$ blocks on average, we know

$$\eta = \overline{K} / (\overline{N} T_F)$$

So far we have proven Theorem 1.

APPENDIX B
PROOF OF $\overline{N_S N_R} = \overline{N_S} \cdot \overline{N_R}$ FOR THE LS PROTOCOL.

We prove $\overline{N_S N_R} = \overline{N_S} \cdot \overline{N_R}$ by showing that N_S and N_R are independent to each other.

Let us observe the three-node (source-relay-destination) network only at the moments when helpful transmissions are taking place. The source in total sends K helpful transmissions, all in the first stage. Hence there are K time epochs observed. Denote the K time epochs by $t_1, t_2, \ldots, t_K$. Observing the network at these K time epochs, we get:

$$\begin{aligned} W &= w_1 + w_2 + \ldots + w_K \\ N_S &= n_1 + n_2 + \ldots + n_K \end{aligned}$$

where n_i is the number of coded blocks sent (by source) in the time interval $(t_{i-1}, t_i]$, and w_i is a binary random variable indicating whether the helpful transmission at t_i fails to reach the destination. $w_i = 1$ if the helpful transmission at t_i reaches the relay, not the destination. $w_i = 0$ otherwise.

It is straightforward that n_is are iid geometrically-distributed random variables with parameter $q = P_{SD} + P_{SR} - P_{SD} P_{SR}$. This is because a coded block reaches either relay or destination with probability q.

Now let us pay attention to the random variable w_i. By definition, w_i depends only on the helpful transmission at the time moment t_i and thus is independent of n_i. Regardless of the value of n_i, w_i is irrelevant to the first $n_i - 1$ transmissions in the time interval $(t_{i-1}, t_i]$ at all; it depends only on the last transmission. In fact, w_is are iid Bernoulli trials, each with a successful probability $q' = P_{SR}(1 - P_{SD})/q$.

The above argument implies that all w_is and n_js, $1 \le i, j \le K$, are independent of each other, which implies that W and N_S are independent of each other. Because N_R depends on W only and is independent of any of the values of n_is, N_R and N_S are independent of each other. Therefore, $\overline{N_S N_R} = \overline{N_S} \cdot \overline{N_R}$.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers and the editor for their many helpful suggestions. This research was supported in part by National Science Council, Taiwan, under grants 100-2219-E-007-003 and 101-2221-E-007-017.

REFERENCES

[1] J. Jin, B. Li, and T. Kong, "Is random network coding helpful in WiMAX?" in *Proc. 2008 IEEE INFOCOM*.

[2] Z. Li, Q. Luo, and W. Featherstone, "N-in-1 retransmission with network coding," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2689–2694, Sept. 2010.

[3] M.-H. Lu, P. Steenkiste, and T. Chen, "Design, implementation and evaluation of an efficient opportunistic retransmission protocol," in *Proc. 2009 ACM MobiCom*.

[4] I. Cerutti, A. Fumagalli, and P. Gupta, "Delay models of single-source single-relay cooperative ARQ protocols in slotted radio networks with Poisson frame arrivals," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 371–382, Apr. 2008.

[5] D. Murray, T. Koziniec, and M. Dixon, "Solving ACK inefficiencies in 802.11 networks," in *Proc. 2009 IEEE Intl. Conf. Internet Multimedia Syst. Architecture Applications*.

[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

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

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

[9] S. Bhadra, S. Shakkottai, and P. Gupta, "Min-cost selfish multicast with network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5077–5087, Nov. 2006.

[10] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," in *Proc. 2008 IEEE INFOCOM*.

[11] X.-B. Liang, "Matrix games in the multicast networks: maximum information flows with network switching," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2433–2466, June 2006.

[12] Q.-T. Vien, L.-N. Tran, and H. X. Nguyen, "Network coding-based ARQ retransmission strategies for two-way wireless relay networks," in *Proc. 2010 Intl. Conf. Software Telecommun. Comput. Netw.*

[13] S. Fu, K. Lu, Y. Qian, and M. Varanasi, "Cooperative network coding for wireless ad-hoc networks," in *Proc. 2007 IEEE GLOBECOM*.

[14] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: practical wireless network coding," in *Proc. 2006 ACM SIGCOMM*.

[15] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

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

[17] E. Zimmermann, P. Herhold, and G. Fettweis, "On the performance of cooperative relaying protocols in wireless networks," *Eur. Trans. Telecommun.*, vol. 16, no. 1, pp. 5–16, Jan.–Feb. 2005.

[18] B. Rong and A. Ephremides, "Cooperation above the physical layer: the case of a simple network," in *Proc. 2009 IEEE International Symp. Inf. Theory*.

[19] E. Zimmermann, P. Herhold, and G. Fettweis, "The impact of cooperation on diversity-exploiting protocols," in *Proc. 2004 IEEE VTC – Spring*.

[20] A. K. Sadek, K. J. R. Liu, and A. Ephremides, "Cognitive multiple access via cooperation: protocol design and performance analysis," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3677–3696, Oct. 2007.

[21] J. Si, Z. Li, and Z. Liu, "Threshold based relay selection protocol for wireless relay networks with interference," in *Proc. 2010 IEEE ICC*.

[22] B. Maham, A. Hjorungnes, and M. Debbah, "Outage probability analysis of multi-relay delay-limited hybrid-ARQ channels," in *Proc. 2010 IEEE VTC – Fall*.

[23] B. Maham, A. Behnad, and M. Debbah, "Analysis of outage probability and throughput for half-duplex hybrid-ARQ relay channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3061–3070, Sept. 2012.

[24] X. Wang and C. Yang, "A MAC protocol supporting cooperative diversity for distributed wireless ad hoc networks," in *Proc. 2005 IEEE PIMRC*.

[25] J. Alonso-Zarate, E. Kartsakli, C. Verikoukis, and L. Alonso, "Persistent RCSMA: a MAC protocol for a distributed cooperative ARQ scheme in wireless networks," *EURASIP J. Advanced Signal Process.*, Dec. 2008.

[26] S. S. N, C.-T. Chou, and M. Ghosh, "Cooperative communication MAC (CMAC)—a new MAC protocol for next generation wireless LANs," in *Proc. 2005 Intl. Conf. Wireless Netw., Commun. Mobile Comput.*.

[27] N. Agarwal, N. Agarwal, L. N. Kannan, M. Tacca, and A. Fumagalli, "IEEE 802.11b cooperative protocols: a performance study," in *Proc. 2007 IFIP/TX6 NETWORKIG 2007*.

[28] J. Alonso-Zarate, E. Kartsakli, C. Verikoukis, and L. Alonso, "A novel near-optimum medium access control protocol for a distributed cooperative ARQ scheme in wireless networks," in *Proc. 2008 IEEE PIMRC*.

[29] J. Alonso-Zarate, L. Alonso, C. Skianis, and C. Verikoukis, "Analysis of a distributed queuing medium access control protocol for cooperative ARQ," in *Proc. 2010 IEEE GLOBECOM*.

[30] A. Antonopoulos, C. Skianis, and C. Verikoukis, "Network coding-based medium access control protocol for cooperative wireless networks," in *Proc. 2012 IEEE ICC*.

[31] K. Tan, Z. Wan, H. Zhu, and J. Andrian, "CODE: cooperative medium access for multirate wireless ad hoc network," in *Proc. 2007 IEEE SECON*.

[32] A. Antonopoulos and C. Verikoukis, "Network coding-based cooperative ARQ scheme," in *Proc. 2011 IEEE ICC*.

[33] P. Fan, Z. Chen, W. Chen, and K. B. Letaief, "Reliable relay assisted wireless multicast using network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 749–762, June 2009.

[34] A. Fanous and A. Ephremides, "Network-level cooperative protocols for wireless multicasting: stable throughput analysis and use of network coding," in *Proc. 2010 IEEE Information Theory Workshop*.

[35] Q. Song, Y. Li, Z. He, and J. Lin, "On reliable multicast with network coding-ARQ for relay cooperation cells," in *Proc. 2012 IEEE VTC – Spring*.

[36] B. Shrader and T. C. Royster, "Cooperative multicast strategies under heterogeneous link loss rates," in *Proc. 2011 IEEE GLOBECOM*.

[37] J.-C. Kao, F.-W. Chen, M.-H. Yang, and T.-L. Wang, "Coverage enhancement to IEEE 802.11p using work-based opportunistic relay-assisted network-coding ARQ," in *Proc. 2011 IEEE Asia Pacific Wireless Commun. Symp.*

[38] N. Abuzainab and A. Ephremides, "Energy efficiency of cooperative relaying over a wireless link," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2076–2083, June 2012.

[39] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 52–59, Apr. 2009.

[40] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall, 2007.

Jung-Chun Kao (S'04, M'09) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, in 1999; the M.S. degree in electrical engineering from University of Southern California, Los Angeles, in 2003; and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, in 2008. He joined the faculty of the Department of Computer Science, National Tsing Hua University, Taiwan, in August 2008. His research interests include wireless ad-hoc and sensor networks, cooperative communications, network coding, and design and analysis of communication protocols. He is a member of the IEEE.

Fu-Wen Chen (S'09) received the BS degree from the Department of Computer Science, National Tsing Hua University (NTHU), Hsinchu, Taiwan, R.O.C., in 2008. One year after his master's study in the same department, he transferred to the Ph.D. program due to academic excellence. He is currently a Ph.D. candidate. His research interests include wireless ad-hoc and sensor networks, cooperative communications, network coding, distributed algorithms, and performance evaluation. He is a student member of the IEEE.