

Optimization of ARQ Distribution for HARQ Strategies in Delay-Bounded Networks

Jaya Goel and J. Harshan
Bharti School of Telecom Technology and Management,
Indian Institute of Technology Delhi, India

Abstract—Inspired by several delay-bounded mission-critical applications, optimizing the end-to-end reliability of multi-hop networks is an important problem subject to end-to-end delay constraints on the packets. Towards that direction, Automatic Repeat Request (ARQ) based strategies have been recently proposed wherein the problem statement is to distribute a certain total number of ARQs (that capture end-to-end delay) across the nodes such that the end-to-end reliability is optimized. Although such strategies provide a fine control to trade end-to-end delay with end-to-end reliability, their performance degrades in slowly-varying channel conditions. Pointing at this drawback, in this work, we propose a Chase Combining Hybrid ARQ (CC-HARQ) based multi-hop network addressing a similar problem statement of how to distribute a certain total number of ARQs such that the end-to-end reliability is optimized. Towards solving the problem, first, we identify that the objective function of the optimization problem is intractable due to the presence of Marcum-Q functions in it. As a result, we propose an approximation on the objective function and then prove a set of necessary and sufficient conditions on the near-optimal ARQ distribution. Subsequently, we propose a low-complexity algorithm to solve the problem for any network size. We show that CC-HARQ based strategies are particularly appealing in slow-fading channels wherein the existing ARQ strategies fail.

Index Terms—Multi-hop networks, High-reliability, Bounded delay, Chase Combining HARQ

I. INTRODUCTION

In the fifth generation (5G) wireless networks and beyond, Internet-of-Things (IoTs), massive Machine-type communications (mMTC) etc., [1] are pitched to play a vital role in deploying mission-critical applications. Examples include reliable and timely delivery of status updates in vehicular networks, facilitating real-time industrial automation tasks by deploying massive low-power IoT devices in factory settings, and so on. Be it vehicular networks or industrial IoT settings, one of the main challenges is the problem of designing signalling schemes over large-scale devices and understanding their fundamental limits in achieving ultra-reliable communication with bounded delay constraints on the packets. Furthermore, generalizing these mission-critical problem statements to multi-hop wireless network settings is also important owing to its proven expansion in the coverage area [2], especially when involving low-powered devices. Thus, in this work, we address the challenges in envisioning delay-bounded mission-critical applications over multi-hop wireless networks [3], [4].

Ongoing research in wireless multi-hop networks has already shown the challenges in achieving high end-to-end reliability with bounded delay-constraint on the packets. [5],

[6] recently proposed a decode-and-forward (DF) based relaying strategy along with retransmission protocols at the intermediate nodes. In such a scheme, reliability within a hop was taken care of by the number of retransmissions allotted to a given node, and the end-to-end delay on the packets was dictated by the total number of ARQs allotted to all the nodes in the multi-hop network. By using the information on processing delays at each node, a certain total number of retransmissions was estimated corresponding to the deadline on the packets.

In order to understand the relation between the bounded delay constraints and the total number of retransmissions, we first explain the *worst-case deadline* approach of computing the maximum number of retransmissions by assuming that the delay overheads from ACK/NACK in the reverse channel are sufficiently small compared to the payload. Suppose that the processing time at each hop is τ_p seconds (which includes packet encoding and decoding time), the delay incurred for packet transmission at each hop is τ_d seconds (which includes the propagation delay and the time-frame of the packet), and the delay incurred because of NACK overhead is τ_{NACK} (which is the time taken for the transmitter to receive the NACK). Given the stochastic nature of the wireless channel at each link, the total number of packet retransmissions before the packet reaches the destination is a random variable, denoted by n , and as a result, the end-to-end delay between the source and the destination is upper bounded by $n \times (\tau_p + \tau_d + \tau_{NACK})$ seconds. In particular, when $\tau_{NACK} \ll \tau_p + \tau_d$, the end-to-end delay can be approximated as $n \times (\tau_p + \tau_d)$ seconds. Thus, when the packet size and the decoding protocol at each node are established, and when the deadline on end-to-end delay (denoted by τ_{total}) is known, we may impose an upper bound on n , provided by $q_{sum} = \lfloor \frac{\tau_{total}}{\tau_p + \tau_d} \rfloor$. The values of τ_p and τ_d at each relay node are assumed to be identical in order to obtain a relation between the end-to-end delay τ_{total} and the total number of ARQs q_{sum} . While identical τ_d holds in practice as the packet length is the same at each hop, the propagation delay is usually negligible. However, if τ_p at each relay are not identical, then an upper bound on the total number of ARQs can still be obtained by considering the maximum of the processing delays offered by all the relays in the chain. Overall, this approach implies that q_{sum} captures the maximum number of retransmissions that can be tolerated over the multi-hop network in order to respect the deadline on the delay.

Once q_{sum} is obtained for a given bounded-delay constraint, several questions were posed in [6] on how to distribute these retransmissions across different nodes to maximize end-to-end reliability. In particular, Type-1 Automatic Repeat Request (ARQ) protocol was employed, wherein the receiver node at each hop transmits an ACK or NACK depending on the success in decoding. Subsequently, on every retransmission, the receiver node discards the previous version of the packet and only uses the latest packet to decode the information. We note that when these solutions are used in the scenario where the devices are static and their surrounding does not change over time, e.g., factory settings, Type-1 ARQs are not applicable for two apparent reasons: Firstly, in Type-1 ARQ, the previous packet is discarded; therefore, it implies the wastage of resources, and secondly, as the channel may not vary with time, there is no benefit in discarding the previous packets and then decoding the new packet that is re-transmitted under the same channel conditions. In such cases, advanced ARQ strategies, such as Type-2 ARQs [8] is beneficial, wherein on every retransmission, the receiver node combines the latest packet along with its previous copies to decode the information. Although an optimization problem on ARQ distribution has been studied in [5], [6] for Type-1 ARQs, there is no such effort on Type-2 ARQs to achieve high reliability in delay-bounded scenarios. Among many variants of Type-2 ARQs schemes that are often used in practical systems, a popular scheme is Chase Combining Hybrid ARQ (CC-HARQ), wherein each retransmission block is identical to the original code block [7].

Motivated by the drawbacks of using Type-1 ARQs in wireless networks with slowly varying channels, in this work, we propose a CC-HARQ based DF strategy for achieving high reliability under delay-bounded scenarios. In particular, we consider a multi-hop network dominated by slowly varying wireless channels with arbitrary line-of-sight (LOS) components. Along the similar lines of [5], we impose an upper bound on the total number of ARQs required in the network following the CC-HARQ strategy based on the end-to-end delay requirements of the application. Subsequently, we formulate an optimization problem of allocating the optimal ARQ distribution to each intermediate link such that the packets reach the destination with high reliability.¹ Due to the intractability of the objective function in the optimization problem, first, we propose an approximation for the objective function at a high signal-to-noise-ratio (SNR). After that, we obtain the necessary and sufficient conditions on the near-optimal ARQ distribution followed by a low-complexity algorithm to solve the optimization problem. We show that our CC-HARQ based framework outperforms the Type-1 ARQ based strategy in [5] when channels are slowly varying.

In terms of novelty, this is the first CC-HARQ based strategy

¹Although it appears that allocating q_{sum} ARQs to the first node of the network seems optimal, such an approach would require each node to explicitly communicate the residual ARQs in the packet. This, in turn, leads to additional communication-overhead in the packet, and such a provision may not be allowed in certain applications.

that optimizes the reliability aspects of multi-hop networks with bounded-delay constraints. Moreover, the analytical results of this work cannot be viewed as a straightforward extension of prior works since the objective function used for optimization is unique to the CC-HARQ strategy, which was not dealt with bounded-delay applications hitherto.

For visibility purposes, this manuscript is also made available on arXiv [10] with the same title. Therefore, [10] should not be treated as prior-art when evaluating the novelty of this submission.

II. CC-HARQ BASED MULTI-HOP NETWORK MODEL

Consider a network with N hops, as shown in Fig. 1, consisting of a source node (S), a set of $N - 1$ relays R_1, R_2, \dots, R_{N-1} and a destination node (D). By aggregating the information bits in the form of packets, we communicate these packets from S to D by using the $N - 1$ intermediate relays. We assume that the channel between any two successive nodes is characterized by Rician fading that varies slowly over time. We model the complex baseband channel of the k -th link, for $1 \leq k \leq N$, as

$$h_k = \sqrt{\frac{c_k}{2}}(1 + \iota) + \sqrt{\frac{(1 - c_k)}{2}}g_k,$$

where $\iota = \sqrt{-1}$, $0 \leq c_k \leq 1$ is the LOS component, $1 - c_k$ is the Non-LOS component, and g_k is a Gaussian random variable with distribution $\mathcal{CN}(0, 1)$. In this channel model, c_k is a deterministic quantity, which characterizes different degrees of Rician fading channels, and also makes sure that $\mathbb{E}[|h_k|^2] = 1$ holds for any c_k . At the extreme ends, it is well known that $c_k = 0$ gives us the Rayleigh fading channels and $c_k = 1$ provides the Gaussian channels.

By assuming that the intermediate relays are sufficiently far apart from each other, throughout the paper, we use the vector $\mathbf{c} = [c_1, c_2, \dots, c_N]$ to denote the LOS components of the N -hop network. Let $\mathcal{C} \subset \mathbb{C}^L$ denote the channel code used at the

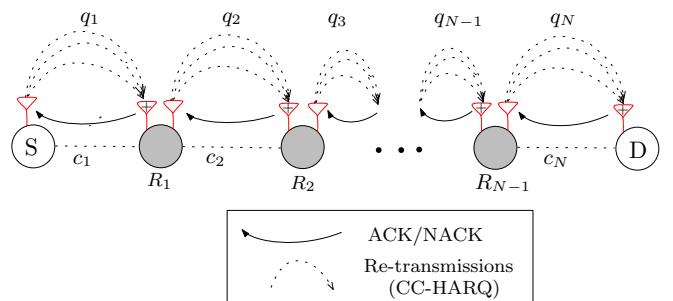


Fig. 1: Illustration of an N -hop network with a source node (S), the relay nodes R_1, \dots, R_{N-1} and the destination node (D) follow the CC-HARQ protocol at each intermediate link. Also, each intermediate link can be characterized by an LOS component $c_k \forall k \in [N]$.

source of rate R bits per channel use, i.e.,

$$R = \frac{1}{L} \log(|\mathcal{C}|).$$

Let $\mathbf{x} \in \mathcal{C}$ denote a packet (which is a codeword in the code) transmitted by the source node such that its average energy per channel use is unity. When \mathbf{x} is sent over the k -th link, for $1 \leq k \leq N$, the corresponding baseband symbols gathered at the receiver over L channel uses is $\mathbf{y}_k = h_k \mathbf{x} + \mathbf{w}_k \in \mathbb{C}^L$, where \mathbf{w}_k is the additive white Gaussian noise (AWGN) at the receiver of the k -th link, distributed as $\mathcal{CN}(0, \sigma^2 \mathbf{I}_L)$. We assume that h_k is known at the receiver of the k -th link owing to channel estimation; however, the transmitter of the k -th link does not know h_k . Since h_k is sampled from an underlying distribution and the realization remains constant for L channel uses, the transmission rate R may not be less than the instantaneous mutual information of the k -th link. Therefore, in such cases, the corresponding relay node will fail to correctly decode the packet. The probability of such an event is well captured by

$$P_k = \Pr\left(R > \log_2(1 + |h_k|^2 \gamma)\right) = F_k\left(\frac{2^R - 1}{\gamma}\right), \quad (1)$$

where $\gamma = \frac{1}{\sigma^2}$ is the average signal-to-noise-ratio (SNR) of the k -th link, $F_k(x)$ is the cumulative distribution function of $|h_k|^2$, defined as

$$F_k\left(\frac{2^R - 1}{\gamma}\right) = 1 - Q_1\left(\sqrt{\frac{2c_k}{(1 - c_k)}}, \sqrt{\frac{2(2^R - 1)}{\gamma(1 - c_k)}}\right), \quad (2)$$

such that $Q_1(\cdot, \cdot)$ is the first-order Marcum-Q function [11]². To support the transmission rate, we follow the CC-HARQ strategy wherein a receiver node asks the transmitter node for retransmission of the packet and combines the received packet with the previous failed attempts to recover the information. To explain further, for a given N -hop network model, a transmitter node gets an ACK or NACK from the next node in the chain, indicating the success or failure of the transmission, respectively. Upon receiving a NACK, the transmitter retransmits the packet, and the receiver combines the current packet with previously received copies of the packet. Let q_k be the maximum number of attempts given to the transmitter of the k -th link to retransmit the packet on demand. Consolidating the number of attempts given to each link, the ARQ distribution of the multi-hop network is represented by the vector $\mathbf{q} = [q_1, q_2, \dots, q_N]$. Since we are addressing bounded delay applications, we impose the sum constraint $\sum_{i=1}^N q_i = q_{sum}$, for some $q_{sum} \in \mathbb{Z}_+$, which captures an upper bound on the end-to-end delay on the packets.

Note that the packet does not reach the destination if an intermediate node fails to deliver the packet to its next node despite using the allotted number of attempts. Since the packet can be dropped in any of the links, we use packet-drop-

²It is well known that the expression for P_k given in (1) captures the error expression in asymptotic block-length regimes. However, it has been shown in [9, Theorem 2] that (1) serves as a saddle-point approximation on the error-probability expressions in non-asymptotic block lengths, especially when the block-length K is in the order of a few hundreds.

probability (PDP) as our reliability metric of interest, given by

$$p_d = P_{1q_1} + \sum_{k=2}^N P_{kq_k} \left(\prod_{j=1}^{k-1} (1 - P_{jq_j}) \right), \quad (3)$$

where P_{kq_k} represents the outage event at k -th link after using the q_k attempts with the CC-HARQ protocol. In particular, we have $P_{kq_k} = \Pr\left(R > \log_2(1 + (\sum_{j=1}^{q_k} |h_{kj}|^2) \gamma)\right)$, where h_{kj} is the channel realization at the k -th link for the j -th attempt. The expression on P_{kq_k} is obtained due to the maximum ratio combining technique. Under the scenario, when the channel remains fixed over multiple attempts, we can rewrite the above equation as $\Pr\left(R > \log_2(1 + |h_{k1}|^2 q_k \gamma)\right)$ which is equal to $F_k\left(\frac{2^R - 1}{q_k \gamma}\right)$. It can be observed that on every retransmission, the packet gets added to its previous copies due to CC-HARQ protocol, which results in an increased effective SNR of the given link.

For the proposed CC-HARQ based multi-hop network, we are interested in minimizing the PDP given in expression (3) for a given sum constraint on the total number of ARQs given by $q_{sum} \in \mathbb{Z}_+$ such that $\sum_{i=1}^N q_i = q_{sum}$. Henceforth, we formulate our optimization problem in Problem 1 as shown below. Throughout this paper, we refer to the solution of Problem 1, as the optimal ARQ distribution.

Problem 1: For a given \mathbf{c} and $\gamma = \frac{1}{\sigma^2}$, solve

$$\begin{aligned} q_1^*, q_2^*, \dots, q_N^* &= \arg \min_{q_1, q_2, \dots, q_N} p_d \\ \text{subject to } q_k &\geq 1, q_k \in \mathbb{Z}_+ \forall k \in [N], \sum_{i=1}^N q_i = q_{sum}. \end{aligned}$$

III. ANALYSIS ON THE OPTIMAL ARQ DISTRIBUTION USING CC-HARQ

The PDP expression in (3) is dependent on the Marcum-Q function, and tackling the Marcum-Q function analytically is challenging because it contains a modified Bessel function of first kind [11]. Therefore, in this section, we first present an approximation of the Marcum-Q function at a high SNR and then use it to provide necessary and sufficient conditions on the optimal ARQ distribution when using the approximations of the Marcum-Q function.

A. Approximation on the Marcum-Q function

Theorem 1: For a given N -hop network and at high SNR regime, we can approximate the Marcum-Q function as $Q_1(a_k, b_k) \approx \tilde{Q}_1(a_k, b_k) \triangleq 1 - (b_k^2/2)e^{-a_k^2/2}$ for all $k \in [N]$.

Proof: In the context of the N -hop network discussed in the previous section, the Marcum-Q function of first-order associated with the k -th hop is given by

$$e^{-\frac{a_k^2}{2}} e^{-\frac{b_k^2}{2}} \sum_{i=0}^{\infty} \left(\frac{a_k^2}{2}\right)^i \left(\sum_{m=0}^{\infty} \left(\frac{a_k b_k}{2}\right)^{2m} \frac{1}{m!(m+i)!}\right),$$

where $a_k = \sqrt{(2c_k)/(1 - c_k)}$, $b_k = \sqrt{2(2^R - 1)/\gamma(1 - c_k)}$ such that $\gamma = 1/\sigma^2$. We highlight that a_k depends on the LOS

component and b_k is dependent on SNR. It can be observed that at high SNR, b_k is very small, and henceforth, we use the small values of b_k to derive the approximation. On expanding $e^{-b_k^2/2}$ using Taylor series, and neglecting the higher power terms, we can approximate $e^{-b_k^2/2}$ by $(1 - b_k^2/2)$. Similarly, by expanding the internal summation starting with index m , we can neglect the higher power terms of b_k and approximate the summation terms as

$$\sum_{m=0}^{\infty} \left(\frac{a_k b_k}{2}\right)^{2m} \frac{1}{m! \Gamma(m+i+1)} \approx \frac{1}{i!} + \frac{1}{(1+i)!} \left(\frac{a_k^2}{2}\right) b_k^2.$$

Therefore, we can rewrite $Q_1(a_k, b_k)$ as

$$Q_1(a_k, b_k) \approx e^{-\frac{a_k^2}{2}} \left(1 - \frac{b_k^2}{2}\right) \sum_{i=0}^{\infty} \left(\frac{a_k^2}{2}\right)^i \left(\frac{1}{i!} + \frac{1}{(1+i)!}\right) \left(\frac{a_k^2}{4}\right) b_k^2.$$

Now, by using $\sum_{n=0}^{\infty} x^n/n! = e^x$ and re-arranging the terms in above equation, we can write

$$Q_1(a_k, b_k) \approx e^{-\frac{a_k^2}{2}} \left(1 - \frac{b_k^2}{2}\right) \left[e^{\frac{a_k^2}{2}} + \frac{b_k^2}{2} \left(e^{\frac{a_k^2}{2}} - 1\right)\right].$$

On further solving the above equation and by neglecting the higher power terms of b_k , we obtain

$$Q_1(a_k, b_k) \approx \tilde{Q}_1(a_k, b_k) \triangleq 1 - \frac{b_k^2}{2} e^{-\frac{a_k^2}{2}}. \quad (4)$$

This completes the proof. \blacksquare

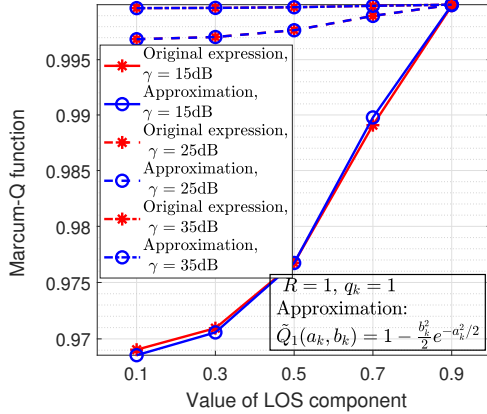


Fig. 2: Comparison of original expression of Marcum-Q function with its proposed approximation at different values of SNR.

To validate the approximation given in (4) for different values of SNR and the LOS components, we present the simulation results in Fig. 2 where the original expression of Marcum-Q function is compared with its approximated expression given in (4). It can be observed from Fig. 2 that the approximation on the Marcum-Q function almost coincides with the original expression.

By using the approximation in (4), we can approximate (1) as

$$\tilde{P}_k = 1 - \tilde{Q}_1(a_k, b_k) \triangleq \frac{b_k^2}{2} e^{-\frac{a_k^2}{2}} = \frac{\phi}{(1-c_k)} e^{-\frac{a_k^2}{2}},$$

where $\phi = \frac{2^R-1}{\gamma}$, and \tilde{P}_k denotes the approximation on P_k . Similarly, for a given q_k retransmissions, we can approximate P_{kq_k} as \tilde{P}_{kq_k} , given by

$$\tilde{P}_{kq_k} \triangleq \frac{\phi}{q_k(1-c_k)} e^{-\frac{a_k^2}{2}}. \quad (5)$$

Finally, by using (5), we can write the approximate version of (3) as

$$\tilde{p}_d = \tilde{P}_{1q_1} + \sum_{k=2}^N \tilde{P}_{kq_k} \left(\prod_{j=1}^{k-1} (1 - \tilde{P}_{jq_j}) \right), \quad (6)$$

where \tilde{p}_d denotes the approximation on p_d . Henceforth, using the approximated expression on p_d as above, we formulate an optimization problem in Problem 2 as shown below. Throughout this paper, we refer to the solution of Problem 2, as the near-optimal ARQ distribution since the objective function in Problem 2 is an approximation on the objective function in Problem 1.

Problem 2: For a given \mathbf{c} and a high SNR $\gamma = \frac{1}{\sigma^2}$, solve

$$q_1^*, q_2^*, \dots, q_N^* = \arg \min_{q_1, q_2, \dots, q_N} \tilde{p}_d$$

$$\text{subject to } q_k \geq 1, q_k \in \mathbb{Z}_+ \forall k, \sum_{i=1}^N q_i = q_{\text{sum}}.$$

B. Sufficient and Necessary Conditions on the Near-Optimal ARQ Distribution in CC-HARQ Strategy

Before we obtain the necessary and sufficient conditions on the optimal ARQ distribution, we show that a link with a higher LOS component must not be given more ARQs than the link with a lower LOS component.

Theorem 2: For a given LOS vector \mathbf{c} and a high SNR regime in a CC-HARQ based multi-hop network, the solution to Problem 2 satisfies the property that whenever $c_i \geq c_j$, we have $q_i \leq q_j \forall i, j \in [N]$.

Proof: To highlight the location of c_i and c_j , we write \mathbf{c} as $[c_1, c_2, \dots, c_i, \dots, c_j, \dots, c_{N-1}, c_N]$ such that $j > i$. Assume that $c_j > c_i$, and q_i and q_j denote the number of ARQs allotted to the i -th and j -th link, respectively. Furthermore, let us suppose that $q_i = q_j = q$ and we have one ARQ with us. Now, the problem is deciding whether to allocate that additional ARQ to the i -th link or the j -link that results in lower PDP. Towards solving this problem, we consider an equivalent multi-hop network with LOS vector $\mathbf{c}' = [c_1, c_2, \dots, c_{N-1}, \dots, c_N, \dots, c_i, c_j]$, wherein \mathbf{c}' is obtained from \mathbf{c} by swapping c_i with c_{N-1} and c_j with c_N . By using a similar approach given in [6, Theorem 1], we can

show that the PDP of the multi-hop networks with the LOS vectors \mathbf{c} and \mathbf{c}' are identical. Furthermore, the PDP of the N -hop network with LOS vector \mathbf{c}' , is written as

$$\begin{aligned} \tilde{p}_d &= \tilde{P}_{1q_1} + \tilde{P}_{2q_2}(1 - \tilde{P}_{1q_1}) + \dots \\ &+ \left(\tilde{P}_{iq_i} + \tilde{P}_{jq_j}(1 - \tilde{P}_{iq_i}) \right) \prod_{k \in [N] \setminus \{i, j\}} (1 - \tilde{P}_{kq_k}). \end{aligned}$$

Note that \tilde{P}_{iq_i} and \tilde{P}_{jq_j} appear only in the last term of the above expression. Since the question of allocating the additional ARQ is dependent only on the expression $\tilde{P}_{iq_i} + \tilde{P}_{jq_j}(1 - \tilde{P}_{iq_i})$, we henceforth do not use the entire expression for PDP. Additionally, since $q_i = q_j = q$, we obtain one of the following expressions when allocating the additional ARQ,

$$\begin{aligned} A &= \tilde{P}_{i(q+1)} + \tilde{P}_{jq}(1 - \tilde{P}_{i(q+1)}), \\ B &= \tilde{P}_{iq} + \tilde{P}_{j(q+1)}(1 - \tilde{P}_{iq}). \end{aligned}$$

Since $c_i < c_j$, we know that $\tilde{P}_{iq} > \tilde{P}_{jq}$ for $q = 1$. To prove the statement of the theorem, we use the approximation given in (5) to show that $A < B$. Therefore, using (5), we write

$$\begin{aligned} A &= \frac{\phi e^{-\frac{a_i^2}{2}}}{(q+1)(1-c_i)} + \frac{\phi e^{-\frac{a_j^2}{2}}}{q(1-c_j)} \left(1 - \frac{\phi e^{-\frac{a_i^2}{2}}}{(q+1)(1-c_i)} \right), \\ B &= \frac{\phi e^{-\frac{a_i^2}{2}}}{q(1-c_i)} + \frac{\phi e^{-\frac{a_j^2}{2}}}{(q+1)(1-c_j)} \left(1 - \frac{\phi e^{-\frac{a_i^2}{2}}}{q(1-c_i)} \right). \end{aligned}$$

On solving the difference $A - B$, we get

$$\phi \left(\frac{e^{-a_j^2/2} - e^{-a_i^2/2} + c_j e^{-a_i^2/2} - c_i e^{-a_j^2/2}}{(1-c_i)(1-c_j)} \right) \underbrace{\left(\frac{1}{q} - \frac{1}{q+1} \right)}.$$

In the above equation, note that the second product term in the bracket is positive. Therefore, to prove that $A - B < 0$, first, we show that $e^{-a_j^2/2} - e^{-a_i^2/2} + c_j e^{-a_i^2/2} - c_i e^{-a_j^2/2} \leq 0$ for any $i, j \in [N]$. To proceed further, we start by assuming that the above inequality is true, and by rewriting the above equation, we get $e^{-a_j^2/2}(1-c_i) \leq e^{-a_i^2/2}(1-c_j)$ which in turns equal to $\frac{(1-c_i)}{(1-c_j)} \leq \frac{e^{-a_i^2/2}}{e^{-a_j^2/2}}$. Now, by taking the logarithm on both sides and on expanding both a_i and a_j , we obtain the inequality $\log \left(\frac{1-c_i}{1-c_j} \right) \leq \frac{c_j - c_i}{(1-c_i)(1-c_j)}$. Let $x = \frac{1-c_i}{1-c_j}$, and therefore, $x - 1 = \frac{1-c_i}{1-c_j} - 1 = \frac{c_j - c_i}{(1-c_j)}$. Furthermore, by using x , we can rewrite the above inequality as $\log x \leq \frac{x-1}{1-c_i}$, where $\frac{1}{1-c_i} \geq 1$ because $c_i \leq 1$. Moreover, by using the standard inequality i.e. $\log x \leq (x-1)$, we can prove that the inequality $\log x \leq \frac{x-1}{(1-c_i)}$ is true for any $i, j \in [N]$. Hence, this completes the proof that $A - B < 0$. ■

Henceforth, we use the set $\mathbb{S} = \{\mathbf{q} \in \mathbb{Z}_+^N \mid \sum_{j=1}^N q_j = q_{sum} \ \& \ q_j \geq 1 \ \forall j\}$ to define the search space for the optimal ARQ distribution. Furthermore, for a given $\mathbf{q} \in \mathbb{S}$, its neighbors are defined as below.

Definition 1: The set of neighbors for a given $\mathbf{q} \in \mathbb{S}$ is defined as $\mathcal{H}(\mathbf{q}) = \{\bar{\mathbf{q}} \in \mathbb{S} \mid d(\mathbf{q}, \bar{\mathbf{q}}) = 2\}$, where $d(\mathbf{q}, \bar{\mathbf{q}})$ denotes the Hamming distance between \mathbf{q} and $\bar{\mathbf{q}}$.

In the following definition, we present a local minima of the space \mathbb{S} by evaluating the PDP of the CC-HARQ based multi-hop network over the vectors in \mathbb{S} .

Definition 2: To be a local minima of \mathbb{S} , an ARQ distribution $\mathbf{q}^* \in \mathbb{S}$ must satisfy the condition $p_d(\mathbf{q}^*) \leq p_d(\mathbf{q})$, for every $\mathbf{q} \in \mathcal{H}(\mathbf{q}^*)$, such that $p_d(\mathbf{q}^*)$ and $p_d(\mathbf{q})$ represent the PDP evaluated at the distributions \mathbf{q}^* and \mathbf{q} , respectively.

Using the above definition, we derive a set of necessary and sufficient conditions on the local minima in the following theorem.

Theorem 3: For a given N -hop network with LOS vector \mathbf{c} , the ARQ distribution $\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_N^*]$ is said to be a local minima if and only if q_i^* and q_j^* for $i \neq j$ satisfy the following bounds

$$q_j^{*2} K_i - q_j^*(K_i + K_j K_i) - C_1 \leq 0, \quad (7)$$

$$q_j^{*2} K_i + q_j^*(K_i - K_j K_i) - C_2 \geq 0, \quad (8)$$

where $C_1 = -K_j K_i + q_i^{*2} K_j + q_i^*(K_j - K_j K_i)$, $C_2 = K_j K_i + q_i^{*2} K_j - q_i^*(K_j + K_j K_i)$ with K_t for $t \in \{i, j\}$ given by

$$K_t = \frac{\phi}{1-c_t} e^{\left(\frac{-c_t}{1-c_t} \right)}. \quad (9)$$

Proof: According to Definition 1, it can be observed that a neighbor of \mathbf{q}^* in the search space \mathbb{S} differs in two positions with respect to \mathbf{q}^* . Let these neighbors are of the form $\hat{\mathbf{q}}_+ = [q_1^*, q_2^*, \dots, q_i^* + 1, \dots, q_j^* - 1, \dots, q_N^*]$ and $\hat{\mathbf{q}}_- = [q_1^*, q_2^*, \dots, q_i^* - 1, \dots, q_j^* + 1, \dots, q_N^*]$ that differs in two positions at i and j provided $q_i^* - 1 \geq 1$ and $q_j^* - 1 \geq 1$. Because of the type of search space and the expression of PDP, we invoke the results from [6, Theorem 1] that the PDP remains identical after swapping intermediate links. Therefore, instead of considering the multi-hop network with LOS vector $\mathbf{c} = [c_1, c_2, \dots, c_i, \dots, c_j, \dots, c_{N-1}, c_N]$, we consider its permuted version with the LOS vector $\mathbf{c} = [c_1, c_2, \dots, c_{N-1}, \dots, c_N, \dots, c_i, c_j]$, wherein the i -th link is swapped with $(N-1)$ -th link, and the j -th link is swapped with N -th link. Consequently, the local minima and its two neighbors are respectively of the form $\mathbf{q}^* = [q_1^*, q_2^*, \dots, q_{N-1}^*, \dots, q_N^*, \dots, q_i^*, q_j^*]$, $\hat{\mathbf{q}}_+ = [q_1^*, q_2^*, \dots, q_{N-1}^*, \dots, q_N^*, \dots, q_i^* + 1, q_j^* - 1]$ and $\hat{\mathbf{q}}_- = [q_1^*, q_2^*, \dots, q_{N-1}^*, \dots, q_N^*, \dots, q_i^* - 1, q_j^* + 1]$. By using the definition of local minima, we have the inequalities

$$\tilde{p}_d(\mathbf{q}^*) \leq \tilde{p}_d(\hat{\mathbf{q}}_+), \text{ and } \tilde{p}_d(\mathbf{q}^*) \leq \tilde{p}_d(\hat{\mathbf{q}}_-), \quad (10)$$

where $\tilde{p}_d(\mathbf{q}^*)$, $\tilde{p}_d(\hat{\mathbf{q}}_+)$ and $\tilde{p}_d(\hat{\mathbf{q}}_-)$ represent the PDP evaluated at the distributions \mathbf{q}^* , $\hat{\mathbf{q}}_+$, and $\hat{\mathbf{q}}_-$, respectively. Because of the fact that $\hat{\mathbf{q}}_+$ and $\hat{\mathbf{q}}_-$ differ only in the last two positions

and the structure of the PDP, it is possible to demonstrate that the inequalities in (10) are equivalent to

$$\tilde{P}_{iq_i^*} + \tilde{P}_{jq_j^*} \left(1 - \tilde{P}_{iq_i^*}\right) \leq \tilde{P}_{i(q_i^*+1)} + \tilde{P}_{j(q_j^*-1)} \left(1 - \tilde{P}_{i(q_i^*+1)}\right), \quad (11)$$

$$\tilde{P}_{iq_i^*} + \tilde{P}_{jq_j^*} \left(1 - \tilde{P}_{iq_j^*}\right) \leq \tilde{P}_{i(q_i^*-1)} + \tilde{P}_{j(q_j^*+1)} \left(1 - \tilde{P}_{i(q_i^*-1)}\right), \quad (12)$$

respectively. First, let us proceed with (11) to derive a necessary and sufficient conditions on q_i^* and q_j^* . On expanding \tilde{P}_{iq} and \tilde{P}_{jq} , and by using the approximation on outage probability given in (5), the inequality in (11) can be rewritten as

$$\frac{K_i}{q_i^*} + \frac{K_j}{q_j^*} \left(1 - \frac{K_i}{q_i^*}\right) \leq \frac{K_i}{q_i^*+1} + \frac{K_j}{q_j^*-1} \left(1 - \frac{K_i}{q_i^*+1}\right),$$

where K_i and K_j can be obtained from (9). On solving the above equation, we obtain

$$\frac{K_i}{q_i^*} - \frac{K_i}{q_i^*+1} + \frac{K_j}{q_j^*} - \frac{K_j}{q_j^*-1} \leq \frac{K_i K_j}{q_i^* q_j^*} - \frac{K_i K_j}{(q_i^*+1)(q_j^*-1)}.$$

After further modifications, we can rewrite the above equation as

$$q_j^{*2} K_i - q_j^* (K_i + K_i K_j) \leq -K_i K_j + q_i^{*2} K_j + q_i^* (K_j - K_i K_j). \quad (13)$$

In the above equation, we can replace $(-K_i K_j + q_i^{*2} K_j + q_i^* (K_j - K_i K_j))$ by C_1 , and by rearranging the terms, we get (7). This completes the proof for the first necessary condition. The second necessary condition for (12) can be proved along the similar lines of the above proof to obtain (8). It can be observed that the two conditions of this theorem are also sufficient because the bounds are obtained by rearranging the terms in the condition on local minima. ■

IV. LOW-COMPLEXITY LIST-DECODING ALGORITHM

Using Theorem 3, we are ready to synthesize a low-complexity algorithm to solve Problem 2.

Proposition 1: If the ARQ distribution \mathbf{q} is chosen such that $\frac{q_j^*}{q_i^*} = \sqrt{\frac{K_j}{K_i}}$, for $i \neq j$, then \mathbf{q} is a local minima of the search space at a high SNR.

Proof: At a high SNR, ϕ is very small, and therefore, the product $K_i K_j$ is negligible. Therefore, we can rewrite (7) and (8) as

$$\begin{aligned} q_j^{*2} K_i - q_j^* K_i &\leq q_i^{*2} K_j + q_i^* K_j, \\ q_j^{*2} K_i + q_j^* K_i &\geq q_i^{*2} K_j - q_i^* K_j, \end{aligned}$$

respectively. On rearranging the above equations, we get

$$\begin{aligned} q_j^{*2} K_i - q_i^{*2} K_j &\leq q_j^* K_i + q_i^* K_j, \\ q_j^{*2} K_i - q_i^{*2} K_j &\geq -q_j^* K_i - q_i^* K_j, \end{aligned}$$

respectively. Now, by equating $q_j^* K_i + q_i^* K_j = \epsilon_{i,j}$, the above equations become

$$q_j^{*2} K_i - q_i^{*2} K_j \leq \epsilon_{i,j}, \quad (14)$$

$$q_j^{*2} K_i - q_i^{*2} K_j \geq -\epsilon_{i,j}, \quad (15)$$

where $\epsilon_{i,j}$ is a positive number dependent on K_i, K_j, q_i^*, q_j^* . If we choose q_i^* and q_j^* such that $q_j^{*2} K_i - q_i^{*2} K_j = 0$, for $i \neq j$, then it ensures that the inequalities in (14) and (15) are trivially satisfied. Thus, choosing

$$\frac{q_j^*}{q_i^*} = \sqrt{\frac{K_j}{K_i}}, \quad (16)$$

for every pair $i, j \in [N]$ such that $i \neq j$, satisfies the sufficient conditions given in (14) and (15) at high SNR values. This completes the proof. ■

Based on the results in Proposition 1, we formulate Problem 3, as given below, as a means of solving Problem 2 at a high SNR.

Problem 3: For a given $\{K_1, K_2, \dots, K_N\}$ and q_{sum} , find $q_1^*, q_2^*, \dots, q_N^*$ such that

$$\frac{q_j^*}{q_i^*} = \sqrt{\frac{K_j}{K_i}}, \quad \forall i, j \in [N] \text{ where } i \neq j,$$

$$q_i^* \geq 1, q_i^* \in \mathbb{Z}_+, \quad \forall i, \quad \sum_{i=1}^N q_i^* = q_{sum}.$$

However, from Problem 3, it is observed that a solution is not guaranteed because the ratio $\sqrt{\frac{K_j}{K_i}}$, which is computed based on the LOS components and the SNR, need not be in \mathbb{Q} . Therefore, we first propose a method to solve Problem 3 without the integer constraints, i.e., to find an ARQ distribution $\mathbf{q} \in \mathbb{R}^N$. This type of problem can be solved by using the system of linear equations of the form $\mathbf{R}\mathbf{q} = \mathbf{s}$ to obtain $\mathbf{q}_{real} = \mathbf{R}^{-1}\mathbf{s}$, where $\mathbf{q} = [q_1, q_2, \dots, q_N]^T$, $\mathbf{s} = [0, 0, \dots, 0, q_{sum}]^T$ and $\mathbf{R} \in \mathbb{R}^{N \times N}$ such that $\mathbf{R}(j, j) = 1$ for $1 \leq j \leq N$, $\mathbf{R}(N, j) = 1$, for $1 \leq j \leq N$, $\mathbf{R}(j, j+1) = -r_{j,j+1}$, for $r_{j,j+1} \triangleq \sqrt{\frac{K_{j+1}}{K_j}}$ and $1 \leq j \leq N-1$, and the rest of the entries of \mathbf{R} are zeros. Subsequently, using the solution \mathbf{q}_{real} in \mathbb{R}^N , we propose to obtain a solution in the true search space \mathbb{S} by generating a list as given in Algorithm 1 [5]. Finally, once the list is generated, the ARQ distribution that minimizes the PDP would be the solution of the proposed algorithm.

Algorithm 1 List Creation Based Algorithm

Input: $\mathbf{R}, \mathbf{s}, q_{sum}, \mathbf{c} = [c_1, c_2, \dots, c_N]$

Output: $\mathcal{L} \subset \mathbb{S}$ - List of ARQ distributions in search space \mathbb{S} .

- 1: Compute $\mathbf{q}_{real} = \mathbf{R}^{-1}\mathbf{s}$, $\tilde{\mathbf{q}} = \lceil \mathbf{q}_{real} \rceil$.
 - 2: **for** $i = 1 : N$ **do**
 - 3: **if** $\tilde{q}_i = 0$ **then**
 - 4: $\tilde{q}_i = \tilde{q}_i + 1$
 - 5: **end if**
 - 6: **end for**
 - 7: Compute $E = \left(\sum_{i=1}^N \tilde{q}_i\right) - q_{sum}$
 - 8: $\mathcal{L} = \{\mathbf{q} \in \mathbb{S} \mid d(\mathbf{q}, \tilde{\mathbf{q}}) = E, q_j \neq q_i \text{ for } c_i < c_j\}$.
-

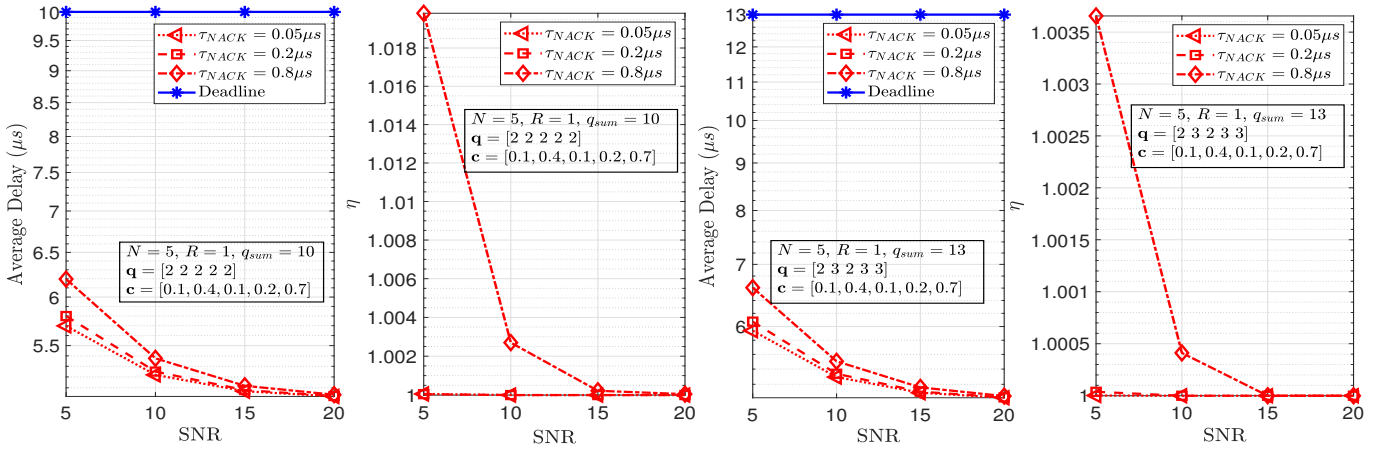


Fig. 3: Illustration on the average delay on packets and deadline violation parameter (η) for several values of $\tau_{NACK} = \{0.05, 0.2, 0.8\}$ microseconds while implementing CC-HARQ strategy.

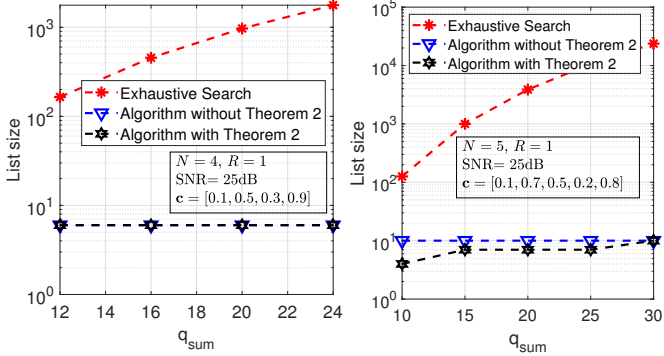


Fig. 4: Comparison of list size between exhaustive search and the proposed low-complexity algorithm with and without Theorem 2 .

V. SIMULATION RESULTS

In the first part of this section, we present simulation results on the delay profiles of the packets under the CC-HARQ protocol. We show that when the ACK/NACK delay overheads are sufficiently small as compared to the payload, there is a high probability with which the packets arrive at the destination within the given deadline. To obtain the results, first, we take $q_{sum} = \lfloor \frac{\tau_{total}}{\tau_p + \tau_d} \rfloor$, by assuming $\tau_{NACK} = 0$ and $\tau_p + \tau_d = 1$ microsecond where $\tau_{total}, \tau_{NACK}, \tau_p$ and τ_d (with μs unit) are defined in Section I. Then, we consider non-zero values of $\tau_{NACK} = \{0.05, 0.2, 0.8\} \mu s$ during the packet flow, and present the results for the following metrics: (i) the number of packets that did not reach the destination due to lack of ARQs present at the intermediate nodes (defined by P_{drop}), (ii) the number of packets that reach the destination after the given deadline (defined by $P_{deadline}$), and (iii) the average delay on the packets. The plots are as shown in Fig. 3 for a given N and a LOS vector \mathbf{c} with different q_{sum} and various SNR values. It can be observed that when τ_{NACK} is very small, specifically at values $\tau_{NACK} = \{0.05, 0.2\}$ microseconds, the average delay

is sufficiently lower than the q_{sum} (the deadline). Also, the deadline violation parameter $\eta = \frac{P_{drop} + P_{deadline}}{P_{drop}}$ is nearly or equal (in the case of $\tau_{NACK} = 0.05 \mu s$) to one. However, when $\tau_{NACK} = 0.8 \mu s$, the average delay slightly moves towards the deadline and also, η exceeds one, showing the deadline violation of the packets.

In the rest of this section, we present the simulation results to validate our theoretical analysis and to showcase the benefits of using the CC-HARQ protocol over a non-HARQ based multi-hop model [6]. On the one hand, the computational complexity for solving Problem 1 and Problem 2 using an exhaustive search is $\binom{q_{sum}-1}{N-1}$. On the other hand, the computational complexity of our method is determined by the complexity of computing the inverse of a matrix and that of Algorithm 1. Furthermore, while the complexity for computing the inverse is $O(N^3)$, the number of computations for generating the list is bounded by $\binom{N}{E}$, where E is the leftover number of ARQs after the ceiling operation. To demonstrate the benefits of using our low-complexity method, we plot the curves for the list size of an exhaustive search and the proposed algorithm in Fig. 4 for several values of q_{sum} with $N = 4$ and $N = 5$. Specifically, we plot the list size both with and without incorporating the results of Theorem 2. When using Theorem 2, by subtracting one ARQ from all possible $\binom{N}{E}$ positions from $\tilde{\mathbf{q}}$, we discard those ARQ distributions which follow the rule $q_i > q_j$ whenever $c_i > c_j$. As a result, we see that the list size shortens after incorporating the rule of Theorem 2 for $N = 5$. Furthermore, based on the simulation results, we observe that the ARQ distribution, which minimizes the PDP from the list \mathcal{L} matches the result of exhaustive search, thereby confirming that our list contains the optimal ARQ distribution. Although, we used high SNR results of Proposition 1 to synthesize the list-decoding method, we observe that the size of the list reduces significantly at low and medium-range of SNR values.

Finally, to showcase the benefits of using CC-HARQ in a multi-hop network, we plot the minimum PDP against several

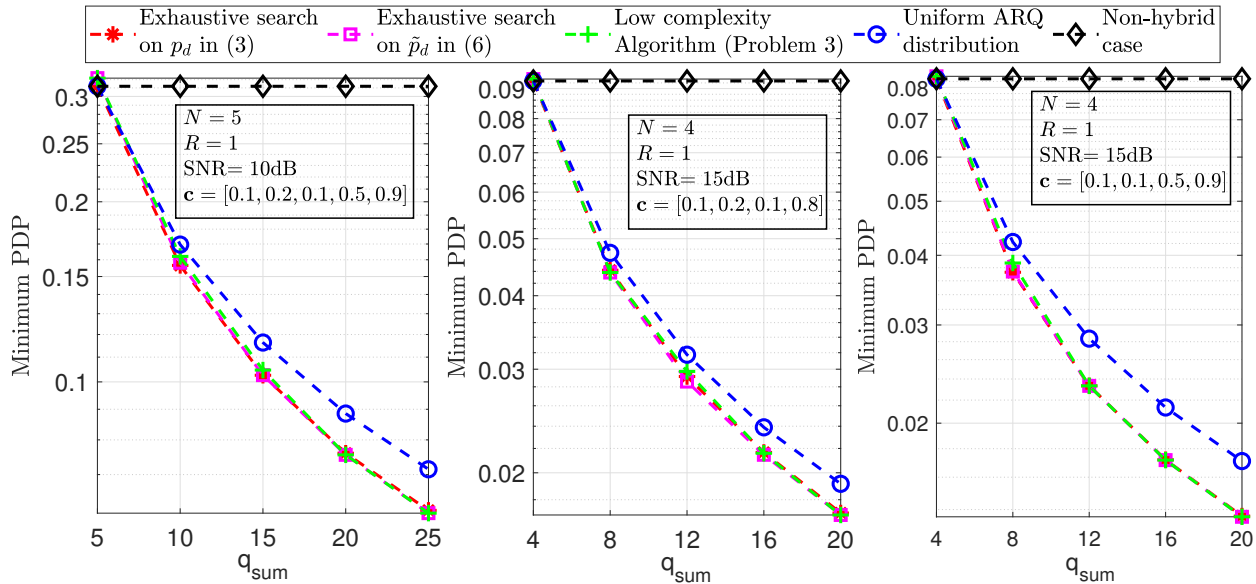


Fig. 5: PDP comparison between (i) an exhaustive search using original PDP expression in CC-HARQ strategy, (ii) an exhaustive search using approximated PDP expression in CC-HARQ strategy, (iii) the proposed low-complexity algorithm, (iv) uniform ARQ distribution in CC-HARQ based strategy, and (v) ARQ distribution in non-HARQ strategy [6].

values of q_{sum} in Fig. 5 by using (i) an exhaustive search on p_d (as given in (3)), (ii) an exhaustive search on \tilde{p}_d (as given in (6)), (iii) the proposed low-complexity algorithm, (iv) uniform ARQ distribution in CC-HARQ based strategy, and (v) non-HARQ strategy. There are mainly two observations from the simulation results in Fig. 5: (a) the PDP improves in the case of optimal ARQ distribution (with the exhaustive search on both p_d and \tilde{p}_d) over uniform ARQ distribution and non-hybrid based ARQ distribution, and (b) our proposed approximation \tilde{p}_d produces reasonably accurate results.

VI. DISCUSSION

In this work, we have presented a novel CC-HARQ framework to achieve high reliability with bounded-delay constraints in a slow-fading multi-hop network. In particular, by using the CC-HARQ protocol at the intermediate relay, we have posed the problem of distributing a given number of ARQs such that the PDP is minimized. We have shown that the problem is non-tractable because the underlying PDP expression contains the Marcum-Q function. Towards solving the problem, (i) we have provided a novel approximation on the Marcum-Q function in the high SNR regime and have subsequently used the approximation by formulating a similar optimization problem, (ii) we have derived a set of necessary and sufficient conditions on the near-optimal ARQ distribution by using the approximated version of the optimization problem, and finally, (iii) we have proposed a low-complexity list-based algorithm to solve the optimization problem that yields near-optimal ARQ distribution. We have validated our approximation and the efficacy of our algorithm through extensive simulation results.

REFERENCES

- [1] Schulz et al., "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, Feb. 2017.
- [2] O. S. Badarneh and F. S. Almeshmadi, "Performance of multi-hop wireless networks in α - μ fading channels perturbed by an additive generalized Gaussian noise," *IEEE Communications Letters*, vol. 20, no. 5, pp. 986–989, May 2016.
- [3] F. S. Shaikh and R. Wismüller, "Routing in multi-hop cellular Device-to-Device (D2D) networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2622–2657, 2018.
- [4] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 124–130, June 2018.
- [5] Jaya Goel and J. Harshan, "On the Optimal ARQ Distribution for Low-Latency Communication over Line-of-Sight Dominated Multi-Hop Networks," *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, pp. 1–8, June 2020.
- [6] Jaya Goel and J. Harshan, "Listen to Others' Failures: Cooperative ARQ Schemes for Low-Latency Communication over Multi-Hop Networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 09, pp. 6049–6063, Sept. 2021.
- [7] T. V. Chaitanya, E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in iid Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 61, no. 05, pp. 1835–1846, May 2013.
- [8] C. Shen, and M. P. Fitz, "Hybrid ARQ in multiple-antenna slow fading channels: Performance limits and optimal linear dispersion code design," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5863–83, Sept. 2011.
- [9] P. Mary et al. "Finite Blocklength Information Theory: What is the Practical Impact on Wireless Communications?," *IEEE Globecom Workshops (GC Wkshps)*, Washington DC, USA, pp. 1–6, 2016.
- [10] Jaya Goel and J. Harshan, "Optimization of ARQ Distribution for HARQ Strategies in Delay-Bounded Networks," available online on *arXiv:2203.08381v1 [cs.IT]*, 2022.
- [11] G. E. Corazza and G. Ferrari, "New bounds for the Marcum Q-function," *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 3003–3008, Nov. 2002.