Performance Analysis and Enhancement of Cooperative Retransmission Strategy for Delay-Sensitive Real-Time Services

Wei Song
Faculty of Computer Science
University of New Brunswick, Canada
wsong@unb.ca

Weihua Zhuang
Electrical and Computer Engineering
University of Waterloo, Canada
wzhuang@bcr.uwaterloo.ca

Abstract—As a very promising technique, multi-hop relay has been considered in many wireless networks. It can take advantage of the inherent broadcasting nature of wireless transmission and facilitate cooperative communications. In this paper, we develop an effective analytical framework to study the delay performance of cooperative retransmission strategies. All neighbor nodes overhearing the in-progress transmission cooperate in a distributed manner and contribute to retransmissions. In particular, we focus on the application of cooperative retransmission for delay-sensitive real-time services. Based on the proposed analytical framework, the cumulative distribution function of packet transfer delay can be numerically evaluated. Accordingly, we investigate the delay outage probability (i.e., the probability of violating the maximum delay bound), which is an essential statistical quality-of-service (QoS) metric for real-time services. Further, an enhancement approach is proposed to reduce unnecessary power consumption on retransmissions. It dynamically adapts the transmission probabilities of all participating nodes, depending on current retransmission count. As shown in the numerical results, the adaptive cooperative strategy can achieve a better trade-off between satisfying delay constraint and minimizing total power consumption.

Index Terms—Cooperative wireless communications, quality-of-service, performance analysis, delay-sensitive services.

I. INTRODUCTION

Traditionally, wireless networks are developed based on a hierarchical infrastructure with a single-hop wireless transmission as the last-mile access option. However, the single-hop infrastructured wireless networks have exposed some weakness after decades of operation. Multi-hop relay becomes a very promising technique to overcome obstacle blocking for radio transmissions and alleviate power limitation. As such, extended coverage and enhanced system capacity can be achieved with a low cost. For example, a mobile multi-hop relay mode (MMR) is specified in IEEE 802.16j for wireless metropolitan networks (WirelessMAN), which is also termed as WiMAX for worldwide interoperability for microwave access by The WiMAX Forum [1]. Also, the IEEE 802.11s working group is amending IEEE 802.11 wireless local area networks (WLAN) with mesh networking, so that the physical layer and medium access control (MAC) layer can support both broadcast/multicast and unicast delivery over self-configuring multi-hop topologies [2].

Taking advantage of multi-hop relay, cooperative communications can exploit the spatial diversity similar to the multiple input and multiple output (MIMO) technique. At the same time, it can relax the deployment and cost constraints of MIMO systems [3]. In the literature, there has been extensive research on cooperative transmission technologies. In [4], Lin et al. study the optimal placement of relay stations and relay timeslot allocation for IEEE 802.16j MMR networks. The non-uniform traffic distribution can be satisfied with an optimal selection of relay stations. In [3], a node-cooperative automatic retransmission request (ARQ) scheme is proposed for wireless ad hoc networks. It takes into account a Gilbert-Elliot wireless fading channel [5,6] and assumes simultaneous retransmissions are enabled with a centralized multiple access mechanism.

However, the selection of cooperative partners may involve an excessive overhead and even a centralized controller for resource allocation. As a result, extra access latency may be intolerable for delay-sensitive services such as real-time multimedia applications. On the other hand, the next-generation wireless networks have been envisioned to be heterogeneous and may differ from each other significantly. There is an increasing level of network control decentralization to distribute the intelligence of a central resource controller among mobile nodes, which can cooperate in a mesh mode. Hence, cost-effective distributed strategies are preferable to cooperative transmissions. In [7], Xiong et al. propose two cooperative retransmission strategies enabling distributed implementation. Each cooperative node helps with packet retransmissions based on independent transmission probabilities. A heuristic greedy algorithm is developed to numerically approach the optimal solution. Also, the authors analyze the performance of the proposed strategies in terms of transmission success probabilities. As observed, the optimal transmission probabilities vary with the number of relay neighbors and also the channel conditions.

In this paper, we propose a new analytical framework for the second cooperative retransmission strategy, which achieves superior performance. Here, we focus on the performance analysis in terms of delay outage probability. This is because delay characteristics are more critical for quality-of-service (QoS)
provisioning to real-time services [8]. Based on a discrete-time Markov chain, we can obtain the optimal transmission probabilities to bound delay outage probability. Further, using the proposed analytical framework, the cooperative strategy is enhanced by adapting transmission probabilities to the number of retransmission attempts. The enhancement approach can achieve a better trade-off between delay constraint and limitation on power consumption.

The remainder of this paper is organized as follows. In Section II, we introduce the previous cooperative retransmission strategy under study and adapt the corresponding system model toward delay-sensitive real-time services. The performance analytical framework is presented in Section III, followed by numerical results in Section IV. Section V gives conclusion and discusses future work.

II. COOPERATIVE RETRANSMISSIONS FOR DELAY-SENSITIVE SERVICES

Taking advantage of a multi-hop networking infrastructure, cooperative retransmissions can exploit redundant neighboring relays to improve ARQ efficiency and delay performance. Similar to [7], we consider a network model as shown in Fig. 1. In addition to the source node $S$ and destination node $D$, there are $K$ relay nodes in the neighborhood. The wireless links between all these nodes are assumed to be memoryless packet erasure channel [9]. In particular, the success probability of packet transfer over the direct channel from node $S$ to node $D$ is denoted by $P_{sd}$, while that of the relay channel from the source node $S$ to the $K$ neighboring relay nodes and then to the destination node $D$ is denoted by $P_{sn}$ and $P_{nd}$, respectively. Moreover, a packet transmission from a given node among the $K$ neighbors can be successfully overheard by any of the other $(K-1)$ nodes with a probability $P_{nn}$. The wireless links between neighboring relay nodes are omitted in Fig. 1 for presentation clarity. As seen in this network model, we differentiate the packet transmission success probabilities for three groups of links, i.e., the source-relay links, relay-destination links, and inter-relay links. In practice, the transmission success probabilities may not be equal for each group of links. However, they can be at the same level if the relay nodes are selected properly. Hence, to simplify analysis, we do not differentiate the transmission success probability for each individual link.

In the literature, many cooperative schemes have been proposed to select a best forwarding node from a candidate neighbor list [10]. Dynamic link information is necessary to make an optimal selection, which induces a large signaling overhead. There are also some studies that consider a clustered structure for cooperative communications [11]. A virtual cluster is established for the source node and relay nodes. The transmission performance can be further improved through the node coordination within the cluster. Nonetheless, there may be extra overhead to set up and maintain relay clusters. Moreover, some previous schemes assume online channel assignment to enable multiple simultaneous transmissions of relay nodes [3]. However, the resource allocation by a central controller may involve a long latency [12], which can be intolerable for delay-sensitive services. In contrast, the cooperative strategies proposed in [7] consider uncoordinated retransmissions by all available neighbor nodes. As a result, the above signaling and control overhead can be avoided. Although multiple simultaneous transmissions result in a collision, the probabilistic relaying manner can bound the collision probability.

Particularly, we are interested in the second cooperative strategy in [7] and develop an effective analytical framework for it. Moreover, the strategy is enhanced to further reduce retransmission power consumption. The original Xiong’s cooperative strategy works as follows. The source node ($S$) is transmitting to the destination node ($D$) packet by packet. The duration of each transmission attempt including acknowledgment is assumed to be a fixed timeslot. First, node $S$ sends a packet to node $D$, which is successful with a probability $P_{sd}$ and correctly overheard by a relay node with a probability $P_{sn}$. If the first transmission fails, node $S$ retransmits the packet in the next timeslot with a probability $\tau_s$, while all relay nodes with a correct copy simultaneously resend with a probability $\tau_n$. A final successful reception by node $D$ is only possible without collisions from multiple simultaneous retransmissions. Otherwise, node $S$ and more relay nodes, which receive a packet copy from node $S$ or other relay nodes, will continue the retransmission process as in the preceding timeslots.

It is known that real-time services such as multimedia applications are subject to stringent delay constraints but relatively loss-tolerant. However, for wireless networks, due to inherent high channel variations, it is very challenging to provide deterministic QoS guarantee [13]. That is, in terms of packet transfer delay $D$, due to wireless channel randomness, it is hard to ensure $D \leq D_{\text{max}}$, where $D_{\text{max}}$ is the upper bound of packet transfer delay. It is more practical to provision statistical QoS guarantee, which allows for a small probability ($\varepsilon$) of QoS violation, i.e., to achieve

$$\Pr\{D \geq D_{\text{max}}\} < \varepsilon. \tag{1}$$

With respect to the preceding cooperative retransmission strategy, suppose the packet transfer delay bound $D_{\text{max}} = m$ ($m \geq 1$) timeslots. Then, the statistical delay constraint is
defined by a bounded delay outage probability as [14]

\[ P_{out}^{D} = \Pr\{D \geq m\} < \varepsilon. \] (2)

In the next section, we introduce our analytical framework to evaluate the delay statistics of the cooperative retransmission strategy.

III. PERFORMANCE ANALYSIS AND ENHANCEMENT

A. Analytical Framework

After the first transmission from only the source node (S), the number of neighboring relay nodes receiving a correct copy follows a binomial distribution \( \pi^{(0)} \) [7], given by

\[ \pi^{(0)}(k) = C^k_K P_{sn}^k (1 - P_{sn})^{K-k}, \quad k = 0, 1, ..., K. \] (3)

For the subsequent retransmissions involving both the source node and relay nodes, a heuristic greedy algorithm is proposed in [7] to minimize packet loss probability. The local optimal transmission probabilities of the source node and relay nodes \( \tau_s^* \) and \( \tau_{n,i}^* \) respectively are numerically solved for each retransmission attempt without considering the delay bound.

Actually, since the packet error of each retransmission is independent for a memoryless erasure channel, the retransmission process can be modeled by a discrete-time Markov chain as shown in Fig. 2. The state of the Markov chain defines the number of relay nodes having a correct packet copy after one retransmission timeslot. According to the cooperative strategy, the state transition probability matrix takes the form of an upper triangular matrix, given by

\[
Q = \begin{bmatrix}
    p_{0,0} & p_{0,1} & p_{0,2} & \cdots & p_{0,(K-1)} & p_{0,K} \\
    0 & p_{1,1} & p_{1,2} & \cdots & p_{1,(K-1)} & p_{1,K} \\
    0 & 0 & p_{2,2} & \cdots & p_{2,(K-1)} & p_{2,K} \\
    0 & 0 & 0 & \ddots & \vdots & \vdots \\
    0 & 0 & 0 & 0 & p_{(K-1),(K-1)} & p_{(K-1),K} \\
    0 & 0 & 0 & 0 & 0 & p_{K,K}
\end{bmatrix}
\] (4)

The state transition probability \( p_{0,0} \) is either because the source node chooses not to retransmit or the new retransmission fails to be received by any of the \( K \) relay nodes. Actually, if only the source node retransmits, the number of relay nodes with error-free reception follows a binomial distribution as that after the initial transmission. Hence, we have the state transition probability \( p_{0,j} \) for \( j \geq 1 \) as shown in (5).

\[ p_{0,j} = \begin{cases} 
    (1 - \tau_s) + \tau_s (1 - P_{sn})^K, & \text{if } j = 0 \\
    \tau_s C^j_K P_{sn}^j (1 - P_{sn})^{K-j}, & \text{if } 1 \leq j \leq K.
\end{cases} \] (5)

That is, the state transition probability \( p_{0,0} \) is either because the source node chooses not to retransmit or the new retransmission fails to be received by any of the \( K \) relay nodes. Actually, if only the source node retransmits, the number of relay nodes with error-free reception follows a binomial distribution as that after the initial transmission. Hence, we have the state transition probability \( p_{0,j} \) for \( j \geq 1 \) as shown in (5).

In the second case, for \( 1 \leq i \leq K \), i.e., when at least one relay node has a correct packet copy, collisions may occur if the source node and the relay node(s) retransmit simultaneously. Thus, the state transition \( (i \rightarrow i) \) for \( 1 \leq i \leq K \) implies one of the following three situations: 1) neither the source node nor the relay node(s) retransmit; 2) only one node retransmits but fails due to channel errors; and 3) more than one node retransmits and causes collisions. Hence, the state transition probability \( p_{i,i} \) for \( 1 \leq i \leq K \) consists of three corresponding terms, given by

\[ p_{i,i} = P_{NT}(i) + P_{E}(i) + P_{C}(i), \quad i = 1, ..., K \] (6)

where

\[
P_{NT}(i) = (1 - \tau_s)^i (1 - \tau_n)^i
\] (7)

\[
P_{E}(i) = \tau_s (1 - \tau_n)^i (1 - P_{sn})^{K-i} + (1 - \tau_s) C^i_{K} \tau_n (1 - \tau_n)^{i-1} (1 - P_{nn})^{K-i}
\] (8)

\[
P_{C}(i) = 1 - (1 - \tau_s)^i (1 - \tau_n)^i - \tau_s (1 - \tau_n)^i
\] (9)

The state transition \( (i \rightarrow j) \) for \( j \geq i + 1 \) indicates that either the source node or only one relay node has been transmitting error-free without collisions. Thus, we have

\[
p_{i,j} = \tau_s (1 - \tau_n)^i C^j_{K-i} (1 - P_{sn})^{K-j} P_{sn}^{j-i} + (1 - \tau_s) C^j_{K-i} \tau_n (1 - \tau_n)^{j-1} \cdot C^{j-i}_{K-n-i} (1 - P_{nn})^{K-j} P_{nn}^{j-i}.
\] (10)

Then, according to the state transition probability matrix \( Q \) in (4), we can obtain the probability distribution of the...
number of relay nodes having a correct packet copy after the \( l \)th transmission timeslot as [15]

\[
\bar{\pi}^{(l)} = \bar{\pi}^{(0)} \cdot Q^l, \quad l = 1, 2, \ldots
\]  

(11)

where \( \bar{\pi}^{(0)} \) is given by (3).

For delay-sensitive services, a statistical QoS constraint needs to be satisfied, as defined in (2). Actually, the delay outage probability \( P_{out} \) is equivalent to the probability that the original transmission from the source node and \( m - 2 \) retransmission attempts by the source node and neighboring relay nodes all fail. That is,

\[
P_{out}^D = (1 - P_{sd}) \prod_{l=1}^{m-2} \left[ \sum_{k=0}^{K} \pi^{(l-1)}(k) \cdot (1 - P_S(k)) \right]
\]

(12)

where \( P_S(k) \) is the probability of a successful retransmission from the source node and \( k \) relay nodes, given by

\[
P_S(k) = \begin{cases} 
\tau_s P_{sd}, & \text{if } k = 0 \\
(1 - \tau_s)C_k \tau_n (1 - \tau_n)^{k-1} P_{sd} + \tau_s (1 - \tau_n)^k P_{sd}, & \text{if } 1 \leq k \leq K.
\end{cases}
\]

(13)

According to (12), the cooperative retransmission strategy can be evaluated with respect to the statistical delay constraint defined in (2). Further, the delay statistics can be obtained from (12). In particular, we evaluate the mean and variance of packet transfer delay as follows [16]

\[
D = \sum_{d=1}^{d_{max}} \frac{\Pr\{D \geq d\}}{1 - \Pr\{D \geq d_{max} + 1\}}
\]

(14)

\[
\sigma_D^2 = \sum_{d=1}^{d_{max}} 2d \frac{\Pr\{D \geq d\}}{1 - \Pr\{D \geq d_{max} + 1\}} - D^2
\]

(15)

where \( d_{max} \) is the maximum number of retransmission limit before the packet is proactively dropped and \( \Pr\{D \geq d\} \) can be derived similar to (12) as

\[
\Pr\{D \geq d\} = (1 - P_{sd}) \prod_{l=1}^{d-1} \left[ \sum_{k=0}^{K} \pi^{(l)}(k) \cdot (1 - P_S(k)) \right].
\]

(16)

B. Enhancement Approach

As proposed in [7], the transmission probabilities of the source node and relay nodes (\( \tau_s \) and \( \tau_n \), respectively) for each retransmission attempt are independent of current retransmission state. Actually, extra power consumption is involved due to unnecessary transmission collisions among the source node and relay nodes. Initially, the transmission success probability increases because more relay nodes on average hold a correct packet copy and help with the retransmission. Later on, when even more nodes still retransmit based on the same transmission probabilities, more collisions will happen and the effectiveness of cooperative packet delivery may not be improved further.

To address such limitation, we enhance the static cooperative strategy by dynamically adapting nodes’ transmission probabilities depending on the retransmission state. It is true that a large overhead is possible to collect updated information of all cooperative nodes after each retransmission attempt. Nonetheless, current retransmission count can be easily obtained from the original timestamp of the packet. Hence, we consider state-dependent transmission probabilities for the source node and relay nodes. Initially, based on the effective analytical framework in Section III-A, the nodes’ transmission probabilities is searched to minimize the delay outage probability \( P_{out}^D \). Here, we can use the optimization toolbox of MATLAB [17] to numerically obtain the optimal solutions. The algorithm implemented in MATLAB is a sequential quadratic programming (SQP) method. First, cubic quadratic interpolation techniques are used to estimate peak values in the semi-infinite constraints. Then, the peak values form a set of constraints that are supplied to the SQP method.

Given the numerical solutions to minimizing \( P_{out}^D \), we select the minimum \( \tau_s^* \) and \( \tau_n^* \) for the first retransmission attempt. The subsequent transmission probabilities are then dynamically adapted to further reduce total power consumption for cooperative retransmissions. We use the average number of retransmission attempts of both the source node and relay nodes, denoted by \( N_T \), as a quantitative indication for power consumption. Thus, we have

\[
N_T = \sum_{l=1}^{d_{max}} \left[ \tau_s(l) + \tau_n(l)N_R^{(l-1)} \right]
\]

(17)

where \( \tau_s(l) \) and \( \tau_n(l) \) are the transmission probabilities of the source node and relay nodes, respectively, for the \( l \)th retransmission attempt, and \( N_R^{(l-1)} \) is the average number of relay nodes available for \( l \)th retransmission, given by

\[
N_R^{(l-1)} = \sum_{k=0}^{K} k \cdot \pi^{(l-1)}(k), \quad l = 1, 2, \ldots, d_{max}.
\]

(18)

Here, the probability distribution \( \pi^{(l-1)} \) before the \( l \)th transmission timeslot can be obtained by extending (11) as follows

\[
\bar{\pi}^{(l)} = \bar{\pi}^{(0)} \prod_{r=1}^{l} Q(\tau_s(r), \tau_n(r)), \quad l = 1, 2, \ldots, d_{max}
\]

(19)

where the transition probability matrix \( Q(\tau_s(r), \tau_n(r)) \) for the \( r \)th retransmission attempt depends on the adaptive transmission probabilities \( \tau_s(r) \) and \( \tau_n(r) \). To reduce \( N_T \) without violating the delay constraint, we adapt \( \tau_s(\cdot) \) and \( \tau_n(\cdot) \) based on the expected numbers of relay nodes available, which can be evaluated using (18), i.e.,
\[ \tau_s(r + 1) = \tau_s(r) \frac{N_{R}^{(r-1)} + 1}{N_{R}^{(r)} + 1}, \quad \tau_n(r + 1) = \tau_n(r) \frac{N_{R}^{(r-1)} + 1}{N_{R}^{(r)} + 1} \]

\[ \tau_s(1) = \tau_s^*, \quad \tau_n(1) = \tau_n^*, \quad r = 1, 2, ..., \max_d. \] 

(20)

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, numerical results are presented demonstrating the effectiveness of our analytical framework and enhancement approach discussed in Section III. Also, we investigate the statistics of packet transfer delay with the adaptive cooperative retransmission strategy. Given in Table I are the analysis parameters.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{sd} )</td>
<td>0.01 \sim 1.0</td>
<td>Packet transmission success probability of direct source-to-destination channel</td>
</tr>
<tr>
<td>( P_{sn} )</td>
<td>0.7</td>
<td>Packet transmission success probability of source-to-neighbor channel</td>
</tr>
<tr>
<td>( P_{nd} )</td>
<td>0.7</td>
<td>Packet transmission success probability of neighbor-to-destination channel</td>
</tr>
<tr>
<td>( P_{nn} )</td>
<td>0.7</td>
<td>Packet transmission success probability of neighbor-to-neighbor channel</td>
</tr>
<tr>
<td>( K )</td>
<td>5</td>
<td>Number of neighboring relay nodes</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>0.1</td>
<td>Statistical constraint for transfer delay</td>
</tr>
<tr>
<td>( m )</td>
<td>7</td>
<td>Upper bound for delay outage</td>
</tr>
<tr>
<td>( \max_d )</td>
<td>20</td>
<td>Retransmission limit before proactive packet dropping</td>
</tr>
</tbody>
</table>

![Fig. 3. Performance of the adaptive cooperative retransmission strategy in terms of statistical delay constraint defined in (2).](image)

Fig. 3 shows the performance of the adaptive cooperative retransmission strategy in terms of delay outage probability \( P_{out}^{D} \). Given in (2), a statistical delay constraint is usually defined as a bounded \( P_{out}^{D} \). Our proposed analytical framework can effectively evaluate whether or not this performance constraint can be satisfied. It can be seen in Fig. 3 that \( P_{out}^{D} \) first decreases almost linearly with \( P_{sd} \), which is the transmission success probability of the direct channel from the source node to the destination node. When \( P_{sd} \) is sufficiently large, the direct channel can be even better than the two-hop relay channel involving \( K \) neighbor nodes. As a result, the source node can contribute more to the retransmission and there is an evident discontinuity in the curve of Fig. 3.

In Fig. 4, we compare the performance of the static and adaptive cooperative strategies with respect to the average retransmission attempts \( \bar{N}_T \) in (17). As discussed in Section III-B, the adaptive cooperative strategy can further lower down the power consumption on unnecessary retransmissions. When the direct channel is extremely poor, e.g., \( P_{sd} \leq 0.43 \), it is hard to satisfy a low bound for the delay outage probability \( P_{out}^{D} \). In such cases, the gap between the static and adaptive strategies in terms of \( P_{out}^{D} \) is as small as 0.91%. Nonetheless, the adaptive strategy can reduce the unnecessary retransmissions by more than 23%. On the other hand, when the direct channel is sufficiently good, the adaptive strategy can satisfy the statistical delay constraint given in (2). At the same time, the average number of retransmission attempts \( \bar{N}_T \) is also significantly reduced.

Moreover, based on our analytical framework in Section III-A, we investigate the statistic properties of packet transfer delay \( D \). Fig. 5 shows the average and variance of \( D \), which are evaluated using (14) and (15), respectively. Similar to Fig. 3, the discontinuity is due to the relative conditions between the direct channel and two-hop relay channels. Also, it is observed that the squared coefficient of variation \( C_v^2 = \frac{\sigma_D^2}{D^2} \) is
CDF of packet transfer delay

Fig. 5. Statistics of packet transfer delay in terms of average $\bar{D}$ and variance $\sigma^2_D$, which are evaluated using (14) and (15), respectively.

within the range of $(0.86, 1.24)$. Approximately, we can model the packet transfer delay with an exponential distribution. As illustrated in Fig. 6, the cumulative distribution function (CDF) of $D$ derived from (16) can closely fit to an exponential distribution.

V. CONCLUSION AND FUTURE WORK

In this paper, we study the performance of distributed cooperative retransmission strategies [7] for delay-sensitive real-time services. Based on a discrete-time Markov chain, an analytical framework is developed to effectively evaluate the delay outage probability, by which a statistical delay constraint is usually defined. Further, an enhancement approach is proposed to dynamically adapt the transmission probabilities of the source node and neighboring relay nodes. As shown in the numerical results, the adaptive strategy outperforms the static strategy with respect to reducing the total power consumption. Also, both cooperative strategies present approximately the same level of performance in terms of satisfying the statistical delay constraint. Moreover, we investigate the statistic properties of packet transfer delay when adaptive cooperative retransmission is applied. It is observed that the squared coefficient of variation is around 1.0 for the given settings under study. Hence, the packet transfer delay can be approximated by an exponential distribution.

In this study, we only consider a single traffic flow from a given source node to the destination node. All the available neighbor nodes cooperate in a distributed manner for retransmissions. In the future, we are interested in extending this study and considering random traffic arrivals. By using effective scheduling, the neighbor nodes can be shared by multiple flows and cooperate to relay the random traffic. The transmission probabilities of all participating nodes need to be determined properly to minimize the impact of collisions. As such, we expect the overall performance can be further improved by exploiting the multiplexing gain.

REFERENCES