A Request Control Scheme for Data Recovery in DVB-IPDC Systems with Spatial and Temporal Packet Loss

Wen-Hsin Yang, You-Chiun Wang, Yu-Chee Tseng, and Bao-Shuh Paul Lin

Abstract—Recently, *DVB-H (digital video broadcasting–handheld)* and *DVB-IPDC (IP datacast over DVB-H)* have been developed to support broadcasting services. DVB-H is designed to support digital video broadcast for handheld devices, while DVB-IPDC can integrate with an IP-relay network to complement the data loss problem in DVB-H. Assuming that WiMAX networks are adopted to support DVB-IPDC, this paper points out two critical problems: *group packet loss (GPL)* and *broadcast data handover (BDH)*. GPL occurs when there is a burst of retransmission requests for the same pieces of data with high spatial or temporal correlation. BDH happens when some devices that made the above requests handover to new serving cells. To solve these problems, we propose a *lazy wait* and a *group acknowledgement* schemes to alleviate duplicate requests by exploiting their spatial and temporal correlation. This not only reduces the requests submitted by neighboring devices in both space and time domains, but also avoids handovering devices from sending duplicate requests in new cells. Through mathematical analysis, we show how to adaptively adjust the timers of lazy wait and group acknowledgement based on channel quality. Simulation results prove that our schemes can efficiently reduce retransmission requests and retransmission packets, thus alleviating congestion in the IP-relay network.

Index Terms—broadcast, data recovery, DVB, DVB-H, WiMAX, wireless network.

1 Introduction

Digital video broadcasting–handheld (DVB-H) [9], [22] is developed from the successful *DVB-T (terrestrial)* system with special designs for handheld, battery-powered devices. DVB-H adopts a time-slicing mechanism to reduce the energy consumption of *mobile devices (MDs)*. Since MDs are vulnerable to packet loss in a wireless environment [5], [24], a feedback mechanism to request retransmissions is necessary. There are two possible solutions. One is to use the DVB-H return channel to request the lost packets. However, not only the return channel is quite small, but also the DVB-H server can be easily overloaded [32]. The other solution is to adopt *DVB-IPDC (IP datacast over DVB-H)* [6], [15], [17], which relies on cooperating with a separate wireless network, such as an IP-relay wireless network, to allow MDs to submit requests and receive lost packets.

This paper supports DVB-IPDC by adopting broadband WiMAX networks [20], [33] as an example to serve as the datacast channel. We consider MDs roaming inside the coverage of both DVB-H and WiMAX networks, as shown in Fig. 1. The DVB-H server continuously broadcasts digital videos to MDs. When a MD detects any packet[1] loss, it requests a nearby WiMAX *relay station (RS)* for retransmission. Under this architecture, we point out two critical problems: *group packet loss (GPL)* and *broadcast data handover (BDH)*. GPL occurs when there are bursty requests for retransmissions of the same pieces of data with high spatial or temporal correlation. Thus, some RSs may be seriously congested by incoming requests and outgoing retransmissions. On the other hand, BDH occurs when some MDs handover to neighboring RS cells after submitting their requests. Thus, duplicate requests and retransmissions could further congest the WiMAX network.

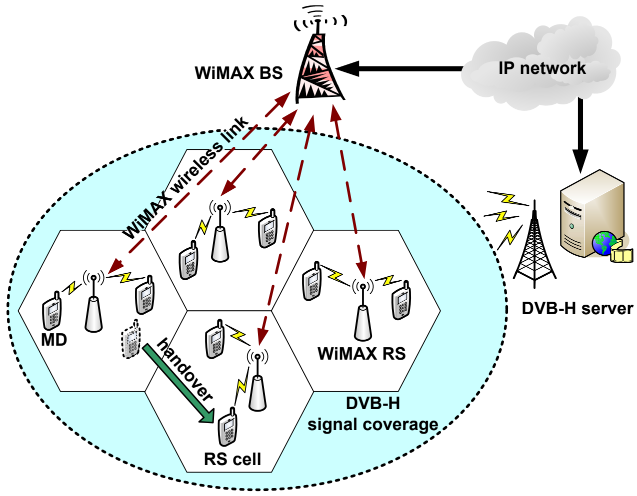


Fig. 1: The DVB-IPDC architecture with WiMAX networks as the datacast channel.

To solve these problems, we should prevent MDs from sending duplicate requests and deliver lost packets effectively. In fact, many requests could exhibit *spatial* or *temporal* correlation. In the space domain, neighboring MDs may send similar requests because they are interfered by the same noise or obstacles. Moreover, the handover behavior of MDs may extend this spatial correlation to neighboring cells. In the time domain, some MDs may lose a sequence of packets since the interference sources often exist for a spell. If redundant requests and retransmissions can be merged, the network efficiency can be greatly improved.

This paper proposes a *bulk recovery with lazy wait (BR-LW)* scheme to deal with the GPL and BDH problems by exploiting

The authors are with the Department of Computer Science, National Chiao-Tung University, Hsin-Chu, 30010, Taiwan.
E-mail: {wenhsin, wangyc, yctseng}@cs.nctu.edu.tw; bplin@mail.nctu.edu.tw

1. In this paper, a "packet" means a DVB-H packet.

the spatial and temporal correlation of retransmission requests. The idea is to adopt the *lazy wait* and *group acknowledgement* mechanisms. Specifically, when a MD detects any packet loss, it "lazily" waits a random period and only sends the request to its associated RS if needed. Then, each RS accumulates the requests from its MDs and broadcasts a group acknowledgement g_{ack} to its cell to announce the requests it has received. On receiving g_{ack}, each MD cancels those requests that appear in g_{ack} but have not been sent out yet. Also, the RS packs its accumulated requests in bulk and sends it to the WiMAX *base station (BS)*. The BS queries the DVB-H server for the lost packets and then sends these packets to MDs through the corresponding RSs.

In BR-LW, we adopt three special designs to reduce duplicate requests and improve network efficiency:

- With lazy wait, each MD can pack many small requests together and sends one bulk request to the RS. Thus, not only is the amount of message transmissions reduced but network contention is also alleviated. The waiting period should be adaptively adjusted based on the DVB-H channel condition. When the MD experiences a bad channel, a shorter period is adopted to quickly recover the lost packets; otherwise, a longer period is adopted to accumulate more requests.
- The group acknowledgement g_{ack} confirms not only the receipt of requests by the RS but also prevents MDs from sending duplicate requests. The timing to send g_{ack} should consider the spatial and temporal correlation of requests. When most requests exhibit high spatial or temporal correlation, g_{ack} is sent immediately to avoid further duplication. Otherwise, the RS can defer g_{ack} to collect more requests.
- The BS also takes advantage of the spatial and temporal correlation of retransmission requests to improve the transmission efficiency. In the space domain, we divide RSs into several multicast groups according to their positions, so that the BS can multicast the requested packets to adjacent RSs altogether to reduce the amount of transmissions. In the time domain, the BS proactively fetches correlated packets from the DVB-H server in advance to reduce the downloading latency.

Major contributions of this paper are three-fold. First, we point out the new GPL and BDH problems under the DVB-IPDC environment. Second, we propose the lazy wait and group acknowledgement schemes to solve these problems by exploiting the spatial and temporal correlation of requests. Simulation results show that our schemes can significantly reduce the amounts of retransmission requests and retransmission packets, thereby alleviating network congestion. Third, with mathematical analysis, we discuss how to adaptively adjust the timers of lazy wait and g_{ack} according to the channel conditions.

The rest of this paper is organized as follows. Section 2 gives the problem statements and surveys related work. Section 3 presents the BR-LW scheme while Section 4 discusses how to determine its system parameters. Simulation results are shown in Section 5. Conclusions are drawn in Section 6.

2 PRELIMINARIES

2.1 Problem Statements

We consider a DVB-H system with WiMAX networks as the datacast channel (refer to Fig. 1). Each MD is equipped with a DVB-H data receiver and a WiMAX wireless interface. The WiMAX networks consist of multiple BSs connected to the DVB-H server through an IP-based network. Each BS supports multiple RSs. MDs continuously receive the DVB-H broadcast data and ask their associated RSs to retransmit the lost packets. Each packet has a deadline and a packet missing its deadline will be dropped. We assume that each MD maintains a small playback buffer and there is a upper-layer protocol to manage the buffer to guarantee some video QoS (quality of service) requirements, such as jitter and PSNR (peak signal-to-noise ratio). Therefore, in this paper we only focus on the retransmission of lost packets to avoid missing their deadlines.

For each BS, it collects the requests from its RSs, packs and forwards these requests to the DVB-H server, and then sends the requested packets to MDs through the corresponding RSs. In the above architecture, the GPL problem occurs when MDs send duplicate requests for the same packets (that is, the requests exhibit spatial correlation) or multiple requests for consecutive packets (that is, the requests exhibit temporal correlation). The BDH problem occurs when a MD sends its requests to a RS, handovers to another RS, and then resends the same requests to this new associated RS again. Our goal is to reduce the above redundancy to improve retransmission efficiency and avoid potential congestion in the WiMAX networks.

2.2 Related Work

DVB-H is developed from the DVB-T system, which relies on uni-directional broadcast. Prior studies thus focus on integrating DVB-T with other wireless networks to provide two-way communications. For example, [31] suggests adopting GSM networks for this purpose. In [26], a GPRS network is adopted to support near-media-on-demand service such as MP3 delivery. References [18], [19] address the integration of 3G, WLAN, and DVB-T networks, where the objective is to manage the composite network resources. A seamless, cross-layer interworking scheme between DVB-T and IEEE 802.11 WLANs is proposed in [4] to provide interactive mobile TV services, where a gateway is deployed in each access network to conduct media adaptations. Similarly, [27] proposes a testbed for mobile interactive television, where DVB-H streams are accessed through WiFi networks. In [28], the concept of mobile social television is proposed, where the DVB-H content is shared among mobile users through peer-to-peer interactivity. Automatic configuration of an IP-based interworking model among WLAN, 3G, and DVB-H is addressed in [23].

Following the DVB-IPDC architecture, several research efforts consider recovering DVB-H data through an IP-relay wireless network. Reference [1] suggests multicasting a DVB-H stream to multiple receivers through a WiFi access point. This work relies on maintaining multicast group membership to avoid CSMA/CA contention; it differs from our work in that we use the datacast channel only for retransmissions. The study of [10] adopts a cellular network to transmit repair data for the receivers temporarily affected by noise, interference, or fading. Similarly, [11] integrates DVB-H with a 3G network to transmit additional parity data for the recovery purpose. Reference [12] proposes an application-layer FEC (forward error correction) scheme and a multi-burst coding scheme to deliver DVB-H data to improve transmission robustness. However, none of the above work addresses the GPL and BDH problems. References [7], [8] propose a *content delivery*

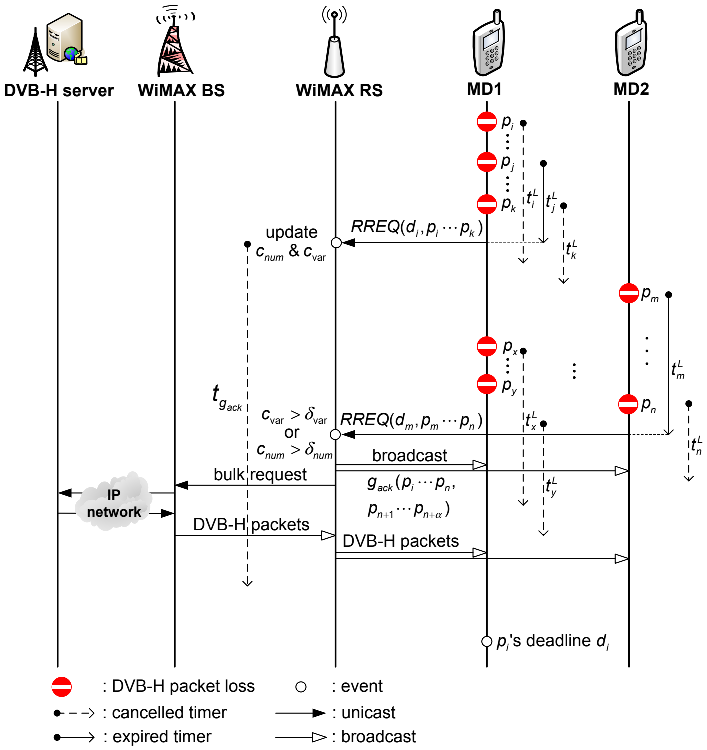


Fig. 2: The flowchart of BR-LW.

protocol (CDP) for the DVB-IPDC architecture to recover lost or corrupted packets, and [13] evaluates the cost of CDP's repair mechanism. Note that CDP relies on a point-to-point mechanism, while ours adopts a more efficient point-to-multipoint mechanism. In [29], MDs are organized as an ad hoc network and each MD encountering packet loss will query the lost packets from other MDs. [16] proposes a recovery mechanism based on a peer-to-peer basis connecting Mobile TV clients via UMTS to recover packets that have been lost during the Mobile TV DVB-H transmission. Different from [16], [29], our work follows the DVB-IPDC architecture to retransmit through an infrastructured network.

3 The BR-LW Scheme

In this section, we describe our BR-LW scheme. How to adaptively adjust its system parameters will be discussed in Section 4. Table 1 summarizes our notations. Fig. 2 illustrates the flowchart of BR-LW. We assume that each MD has a small buffer to store the received DVB-H packets before playback, so a MD can tolerate a reasonable delay d_i for each packet p_i. Our goal is to reduce the amount of message transmissions between MDs and RSs and that between RSs and BSs to preserve wireless bandwidth and avoid network congestion. Below, we present the operations conducted by MDs, RSs, and BSs.

3.1 Operations of MDs

Each MD continuously receives the packets sent from the DVB-H server, and maintains a table T_{MD} to record the information of the lost packets. Specifically, for each lost packet p_i, the MD inserts an entry (p_i, d_i, t_i^L) into its T_{MD}, where d_i is the deadline and t_i^L is the *lazy-wait timer* of packet p_i. Initially, the entry (p_i, d_i, t_i^L) is set as *"unqueried"*. The value of timer t_i^L is randomly selected from $(0, t_{\text{lazy}}]$, where t_{lazy} is a system parameter to be selected according to the current DVB-H channel condition. Intuitively, when the MD experiences a bad channel, a smaller t_{lazy} is set to enforce the MD to quickly react to packet loss; otherwise, a larger t_{lazy} is set to allow the MD to accumulate more requests. (We will discuss how to adaptively adjust t_{lazy} in Section 4.1.) Note that the order that these lazy-wait timers expired is not necessarily the same as the order that the corresponding packets were lost. Fig. 2 gives an example. Although packet p_i is lost earlier than packet p_j, timer t_i^L expires later than timer t_j^L.

When any lazy-wait timer expires, the MD will send to its associated RS a *recovery request* $RREQ(d_i, p_i, \cdots, p_k)$ containing the earliest packet deadline d_i and the indices of all lost packets $p_i, \cdots, p_k$ whose entries are "unqueried" in its T_{MD}. We assume that the transmissions of RREQs are more *reliable* (guaranteed by an acknowledgement mechanism). Then, the MD marks all entries in its T_{MD} as *"queried"* and cancels all active lazy-wait timers. For example, in Fig. 2, since timer t_j^L expires first, MD_1 will send a RREQ and cancel timers t_i^L and t_k^L.

Each MD continuously repeats the above actions (that is, inserts entries for lost packets into T_{MD} and marks entries as "queried" by sending RREQs). However, when the MD receives a *group acknowledgement* g_{ack} from its associated RS while its T_{MD} is non-empty, it will mark those entries in its T_{MD} that have the same packet indices in g_{ack} as "queried" and then cancel the corresponding lazy-wait timers. Fig. 2 shows an example. Supposing that MD_1 receives g_{ack} that contains the packet indices p_x and p_y, MD_1 will mark the entries (p_x, d_x, t_x^L) and (p_y, d_y, t_y^L) as "queried" and cancel the active timers t_x^L and t_y^L. When either deadline d_i expires or the MD receives packet p_i, the entry (p_i, d_i, t_i^L) is removed from T_{MD}.

To deal with the BDH problem, when a MD handovers to a new RS cell, two cases will be considered:

1) If all entries of the MD's T_{MD} are "unqueried", which means that there are no *pending* RREQs in its old RS, this handovering MD behaves the same as a non-handovering MD.
2) If there are some entries marked as "queried" in T_{MD}, which means that the handovering MD did not receive the requested packets from the *old* RS,[2] the MD will set all entries in its T_{MD} as "unqueried" and initialize a lazy-wait timer t_{HO}^L. The value of t_{HO}^L is also randomly selected from $(0, t_{\text{lazy}}]$. This lazy-wait timer is not associated with any packet. Its purpose is to ensure that this handovering MD still has a chance to send a RREQ to let the new RS know its (previously) lost packets if the MD does not experience any packet loss after handovering to the new RS cell.

When any of the lazy-wait timers expires, this handovering MD will send a RREQ to the new RS and cancel all active lazy-wait timers (including t_{HO}^L).

The above lazy wait design has three advantages. First, we can prevent MDs from sending a large number of *small* RREQs. Instead, a MD can combine all requests together and send a bulk RREQ. Thus, not only the overhead of packet headers is reduced but the network contention is also alleviated. Second, the *random* waiting periods help differentiate RREQs' transmission time, and thus reduce potential packet collisions. Third,

2. That is, the MD sends RREQs to the RS but the RS has not returned the requested packets yet.

TABLE 1: Summary of notations.

notation	definition
T_{MD}	the table stored in each MD to record the information of the lost packets
p_i	a lost packet with index i
d_i	the deadline of packet p_i
t_i^L	the lazy-wait timer of packet p_i
t_{lazy}	the threshold value to determine the lazy-wait timers
T_{RS}	the table stored in each RS to record the information of the received RREQs
n_i	the number of RREQs that request packet p_i
g_{ack}	the group acknowledgement
c_{num}	the number of entries in T_{RS}
c_{var}	the coefficient of variation of all n_is in T_{RS}
δ_{num}	the threshold value of c_{num} for the RS to trigger a g_{ack}
δ_{var}	the threshold value of c_{var} for the RS to trigger a g_{ack}
ζ	the threshold value of the sum of all n_is in T_{RS} for the RS to trigger a g_{ack}
μ_E	the expected number of successive lost packets by a MD
$\mathcal{R}$	the set of lost packets requested by RREQs during the $t_{g_{\text{ack}}}$ period
$\mathcal{N}_R$	the size of $\mathcal{R}$
ε_i	the minimum value of n_i for the RS to trigger a g_{ack}
$\mathcal{T}$	the period of simulation time
$\mathcal{F}$	the distribution of packet loss in the simulations
$\mathcal{L}$	the distribution of correlation feature of RREQs in the simulations

by deferring RREQs, the probability to receive g_{ack} before sending RREQs is increased, thus avoiding MDs from sending duplicate requests to the RS. Since a handovering MD will also conduct lazy wait in the new RS cell, some of its duplicate requests could be eliminated if packet loss in both new and old RS cells exhibit spatial correlation.

3.2 Operations of RSs

Each RS should keep the RREQs received from its associated MDs in a table T_{RS}. Each entry (p_i, n_i) in T_{RS} contains the index of a lost packet p_i and the number n_i of MDs that request packet p_i. Every time when T_{RS} changes from empty to non-empty, a timer $t_{g_{\text{ack}}}$ is initiated to determine the deadline to broadcast g_{ack}. The expiration of $t_{g_{\text{ack}}}$ should be earlier than the earliest packet deadline in each received RREQ. For example, in Fig. 2, $t_{g_{\text{ack}}}$ should expire earlier than the earliest packet deadline d_i.

In addition, the RS should check two parameters c_{num} and c_{var} whenever the content of its T_{RS} changes, where c_{num} is the total number of entries in T_{RS} and c_{var} is the *coefficient of variation* of all n_is in T_{RS} which is defined as the ratio of the standard deviation of all n_is to their mean. The RS will broadcast a g_{ack} immediately to its associated MDs if any of the following conditions is true:

1) $c_{\text{num}} \geq \delta_{\text{num}}$: Since the RS will clear c_{num} and T_{RS} whenever it sends out g_{ack} (this will be discussed later), the functionality of T_{RS} can be viewed as a "sliding window" to record which packets are lost among a *small* set of packets broadcasted from the DVB-H server in each round. Therefore, when the number of entries in T_{RS} (that is, c_{num}) exceeds a threshold δ_{num}, there is a high probability that consecutive or semi-consecutive packets are lost. In other words, these RREQs could exhibit temporal correlation.
2) $c_{\text{var}} \geq \delta_{\text{var}}$: When the coefficient of variation c_{var} exceeds a threshold δ_{var}, it means that *some* of the packets may be lost by a large number of MDs. In this case, these RREQs could exhibit spatial correlation.
3) $\sum_{(p_i,n_i)\in T_{\text{RS}}} n_i \geq \zeta$: When both $c_{\text{num}} < \delta_{\text{num}}$ and $c_{\text{var}} < \delta_{\text{var}}$, but the RS has collected a sufficient number ζ of RREQs, this also implies some sort of spatial correlation of received RREQs.
4) Timer $t_{g_{\text{ack}}}$ expires: If none of the above three conditions is met, it means that most RREQs are not correlated. Thus, the RS defers g_{ack} until timer $t_{g_{\text{ack}}}$ expires to collect more RREQs.

Parameters δ_{num}, δ_{var}, and ζ are thresholds to determine how quickly the RS should broadcast g_{ack} before timer $t_{g_{\text{ack}}}$ expires. The threshold ζ can be calculated by past statistics. We will discuss how to adaptively adjust δ_{num} and δ_{var} in Section 4. If any condition above is met, the RS broadcasts a g_{ack} containing all packet indices in its T_{RS}.

As to the content of the g_{ack}, our design will try to exploit both spatial and temporal correlation of RREQs to inhibit potential future RREQs. There are two rules.

1) In the space domain, g_{ack} will contain all packet indices that have received so far. This prevents nearby MDs from sending duplicate requests. Specifically, when a MD sends the request of a lost packet p_i to the RS, other neighboring MDs that also lose packet p_i but have not sent their requests yet will cancel their submissions after receiving g_{ack}. In Fig. 2, suppose that MD_2's RREQ contains packet indices p_x and p_y. If the g_{ack} from the RS also contains p_x and p_y, MD_1 will remove (p_x, d_x, t_x^L) and (p_y, d_y, t_y^L) from its T_{MD}.
2) In the time domain, the RS tries to *predict* future lost packets. Specifically, if it finds a sequence of packet indices $p_{n-\beta}$, $p_{n-\beta+1}$, $\cdots$, p_n in its T_{RS}, the g_{ack} will include α more packet indices p_{n+1}, p_{n+2}, $\cdots$, $p_{n+\alpha}$. So as to exploit temporal correlation of RREQs. Here α and β are system parameters.

After sending g_{ack}, the RS cancels timer $t_{g_{\text{ack}}}$ if it has not expired yet, resets its c_{num} and c_{var} to zeros, sends a bulk request containing all of the packet indices in g_{ack} to the BS, and clears its T_{RS}. Note that to reduce the overhead of packet headers and alleviate network contention, each RS also packs small requests together and then sends one bulk request to the BS.

For the BDH problem, we suggest dividing RSs into multiple *multicast groups* according to their relative positions in deployment, where adjacent RSs will be grouped together [14], [30]. Since MDs usually handover to adjacent RS cells, this behavior could make the received RREQs exhibit spatial correlation in each multicast group of RSs. Therefore, the BS can take advantage of multicast to alleviate BDH by simultaneously sending the requested packets to the RSs in the same multicast group. This issue will be further discussed later.

After obtaining the packets from the BS, the RS will send these packets to all its MDs through broadcasting. Using broadcasting has two advantages over using unicasting. First, the RS does not need to record which packets are lost by which MD, thus reducing storage and maintenance costs (especially when there are many handovering MDs). Second, the RS can significantly alleviate the amount of transmissions if most MDs lose similar packets. In addition, even when there are many handovering MDs, broadcasting can help reduce the cost of sending *orphan* packets (that is, those packets no longer requested by any MD).

3.3 Operations of BSs

Recall that RSs will be divided into multiple multicast groups according to their positions. For each multicast group of RSs, the BS will collect (and pack) their requests and send a bulk request to the DVB-H server to ask for retransmissions of lost packets. Then, the BS will multicast the requested packets to the member RSs in the multicast group. The above multicast takes advantage of spatial correlation of requests since MDs in some adjacent RS cells may lose similar packets. Besides, the handovering behavior of MDs extends this spatial correlation to neighboring RS cells. To take advantage of temporal correlation of requests, the BS can prefetch a small number of future correlated packets from the DVB-H server. In this way, the downloading latency can be reduced.

Note that the goal of this paper is to reduce redundant requests to avoid bandwidth waste and network congestion. How to guarantee video QoS requirements, such as jitter and PSNR, is out of the scope of this paper. Therefore, BR-LW adopts c_{num} and c_{var} to help RSs decide when to broadcast g_{ack} to alleviate redundant requests. For the delay concern, since packets have deadlines and MDs will drop those out-of-date packets, we consider a packet with *no delay* if the packet can be received by the MD before passing its deadline. Therefore, BR-LW makes MDs to notify the RS of packet deadlines so that these lost packets can be sent to MDs before they become out of date.

4 DETERMINING SYSTEM PARAMETERS OF BR-LW

In this section, we discuss how to adaptively adjust the system parameters of BR-LW according to the network situation.

4.1 Adjustment of lazy-wait timer t_{lazy}

The system parameter t_{lazy} determines how long a MD *waits* to send its RREQ after it incurs any packet loss. Intuitively, a shorter length of t_{lazy} is adopted if the MD incurs a bad DVB-H channel. In this case, since the MD is expected to lose a large number of packets, the MD has to send its RREQ immediately to recover the lost packets. Otherwise, a longer length of t_{lazy} is adopted to make the MD accumulate more requests into its RREQ. Therefore, we adjust the length of t_{lazy} in a *binary exponential manner* as follows:

$$t_{\text{lazy}} = (d_i - t_{\text{current}}) \times 2^{-\mu_E}, \tag{1}$$

where t_{current} is the current time and μ_E is the expected number of *successive* lost packets by the MD.

To calculate μ_E, we adopt a four-state *aggregated Markov process (AMP)*, which can approximate the behavior of a DVB-H channel (even when the MD is moving) [25]. AMP considers a finite-state channel error model, where the states are divided into two groups $\mathcal{S}_C$ and $\mathcal{S}_E$ that correspond to the *correct* and *erroneous* reception of packets, respectively. Each group is associated with an output symbol that emits at each state transition and ends at the given group. In particular, let $\mathcal{X} = \{\mathcal{X}_0, \mathcal{X}_1, \cdots\}$ be a time-homogenous Markov chain with a four-state space $\mathcal{S} : \{1,2,3,4\}$. We can divide $\mathcal{S}$ into two groups $\mathcal{S}_C : \{1,2\}$ and $\mathcal{S}_E : \{3,4\}$. Let $\Phi : \mathcal{S} \to \{c,e\}$ be an *emission function*, where $\Phi(i) = c$ if $i \in \mathcal{S}_C$ and $\Phi(i) = e$ if $i \in \mathcal{S}_E$. If we denote the output process of $\mathcal{X}$ as $\mathcal{Y} = \{\mathcal{Y}_0, \mathcal{Y}_1, \cdots\}$ with state space (c,e), then $\mathcal{Y}$ is an AMP.

The transition probability matrix of $\mathcal{X}$ is defined as

$$\mathcal{M}_{\mathcal{X}} = [P_{ij}]_{4\times 4} = \begin{bmatrix} \rho_1 & 0 & (1-\rho_1)\omega_3 & (1-\rho_1)\omega_4 \\ 0 & \rho_2 & (1-\rho_2)\omega_3 & (1-\rho_2)\omega_4 \\ (1-\rho_3)\omega_1 & (1-\rho_3)\omega_2 & \rho_3 & 0 \\ (1-\rho_4)\omega_1 & (1-\rho_4)\omega_2 & 0 & \rho_4 \end{bmatrix},$$

where P_{ij} is the transition probability from state i to state j, $1 \le i,j \le 4$; the condition $0 \le \rho_i \le 1$ holds for $i = 1,2,3,4$; ω_is, $i = 1,2,3,4$, are the weights such that $\omega_1 + \omega_2 = 1$ and $\omega_3 + \omega_4 = 1$.

Let $\pi^{(n)} = \left[\pi_1^{(n)}\ \pi_2^{(n)}\ \pi_3^{(n)}\ \pi_4^{(n)}\right]$ be the vector of state probability distribution, where each $\pi_i^{(n)}$, $i = 1,2,3,4$, is the unconditional probability of being in state i at time n. Let $\pi = \lim_{n\to\infty} \pi^{(n)} = [\pi_1\ \pi_2\ \pi_3\ \pi_4]$. If the limit exists, π is a stationary distribution. Thus, we can calculate π by $\pi \cdot \mathcal{M}_{\mathcal{X}} = \pi$ and $\pi \cdot \mathbf{1} = 1$. Therefore, we can obtain that

$$\pi_i = \frac{\omega_i}{(1-\rho_i)\sum_{j=1}^{4} \frac{\omega_j}{1-\rho_j}}, \quad \text{for } i = 1,2,3,4.$$

With π, we can calculate the probability that a MD incurs erroneous reception of packets by

$$\begin{aligned} Prob(\mathcal{Y} = e) &= \sum_{i\in\mathcal{S}_E} \pi_i \\ &= \pi_3 + \pi_4 \\ &= \frac{\omega_3}{(1-\rho_3)\sum_{j=1}^{4} \frac{\omega_j}{1-\rho_j}} + \frac{1-\omega_3}{(1-\rho_4)\sum_{j=1}^{4} \frac{\omega_j}{1-\rho_j}}. \end{aligned} \tag{2}$$

On the other hand, the value of $Prob(\mathcal{Y} = e)$ can be measured by the average *packet error rate* of the MD during a fixed observation window. Therefore, the moving behavior of the MD is also considered because it will impose an effect on the received signal strength. By assigning weights ω_1, ω_2, ω_3, and

ω_4, we can adopt the Levenberg-Marquardt algorithm [21] to calculate the variables ρ_1, ρ_2, ρ_3, and ρ_4 in Eq. (2).

Now, Let $\mathcal{D}_E$ be the discrete random variable of the dwell time in state group $\mathcal{S}_E$. The probability mass function of $\mathcal{D}_E$ is defined by [25]

$$\begin{aligned} P_{\mathcal{D}}(n) &\triangleq Prob(\mathcal{D}_E = n) \\ &= Prob(\mathcal{X}_{j+2} \in \mathcal{S}_E, \cdots, \mathcal{X}_{j+n} \in \mathcal{S}_E, \\ &\quad \mathcal{X}_{j+n+1} \in \mathcal{S}_C \mid \mathcal{X}_j \in \mathcal{S}_C, \mathcal{X}_{j+1} \in \mathcal{S}_E) \\ &= \sum_{i=3}^{4} \omega_i \rho_i^{n-1}(1-\rho_i), \quad \text{for } n = 1, 2, \cdots. \end{aligned}$$

Then, the probability-generating function of $\mathcal{D}_E$ is calculated by

$$\begin{aligned} G_{\mathcal{D}}(z) &= \sum_{n=1}^{\infty} P_{\mathcal{D}}(n) z^n \\ &= \sum_{n=1}^{\infty} \left(\sum_{i=3}^{4} \omega_i \rho_i^{n-1}(1-\rho_i) \right) \cdot z^n \\ &= \sum_{n=1}^{\infty} \left(\sum_{i=3}^{4} \frac{\omega_i(1-\rho_i)}{\rho_i} \right) \cdot \rho_i^n z^n \\ &= \sum_{i=3}^{4} \frac{\omega_i(1-\rho_i)}{\rho_i} \cdot \sum_{n=1}^{\infty} (\rho_i z)^n \\ &= \sum_{i=3}^{4} \frac{\omega_i(1-\rho_i)}{\rho_i} \cdot \frac{\rho_i z}{1-\rho_i z} \end{aligned} \tag{3}$$

Let us define $f_i(z) = \frac{\rho_i z}{1-\rho_i z}$. Then, Eq. (3) is written as

$$G_{\mathcal{D}}(z) = \sum_{i=3}^{4} \frac{\omega_i(1-\rho_i)}{\rho_i} \cdot f_i(z).$$

The kth derivative of $f_i(z)$ is calculated by

$$\begin{aligned} f_i^{(k)}(z) &= \frac{d^{k-1}}{dz^{k-1}} \left(f_i'(z) \right) \\ &= \frac{d^{k-1}}{dz^{k-1}} \left(\frac{\rho_i(1-\rho_i z) - \rho_i z(-\rho_i)}{(1-\rho_i z)^2} \right) \\ &= \rho_i \cdot \frac{d^{k-1}}{dz^{k-1}} \left(\frac{1}{(1-\rho_i z)^2} \right) \\ &= \frac{k! \rho_i^k}{(1-\rho_i z)^{k+1}} \end{aligned}$$

Therefore, the kth derivative of $G_{\mathcal{D}}(z)$ is calculated by

$$G_{\mathcal{D}}^{(k)}(z) = \sum_{i=3}^{4} \frac{k! \rho_i^{k-1} \omega_i (1-\rho_i)}{(1-\rho_i z)^{k+1}}, \quad \text{for } k = 1, 2, \cdots. \tag{4}$$

By Eq. (4), we can calculate the expectation of $\mathcal{D}_E$ as

$$\mu_E = G_{\mathcal{D}}^{(1)}(1) = \frac{\omega_3}{1-\rho_3} + \frac{1-\omega_3}{1-\rho_4}. \tag{5}$$

By substituting Eq. (5) into Eq. (1), we can obtain t_{lazy}.

4.2 Adjustment of threshold δ_{num}

The threshold δ_{num} determines how quickly the RS will broadcast g_{ack}. When $c_{\text{num}} \geq \delta_{\text{num}}$, the RS sends a g_{ack} because the received RREQs may exhibit temporal correlation. Let $\mathcal{R} = \{p_1, p_2, \cdots, p_{|\mathcal{R}|}\}$ be the set of the lost packets requested by RREQs that the RS collects during the $t_{g_{\text{ack}}}$ period. The size of $\mathcal{R}$ is

$$\mathcal{N}_R \leq \gamma \times e_{MD} \times t_{g_{\text{ack}}},$$

where γ is the packet transmission rate of the DVB-H server and e_{MD} is the average packet error rate of a MD. Then, we adopt the *Zipf's law* [34] to calculate δ_{num}. According to the Zipf's law, many types of data studied in the physical and social sciences such as on-demand broadcasting and video-on-demand services [2] can be approximated by *Zipfian distributions*. Thus, we model the arrival of RREQs by a Zipfian distribution. Besides, due to the lazy wait scheme, we assume that a continuous or near-continuous sequence of packet indices are put together in a single RREQ if the packet loss exhibits temporal correlation.

Suppose that the packets in $\mathcal{R}$ are ranked *decreasingly* by the number of RREQs that query them. The Zipf's law predicts that out of these $\mathcal{N}_R$ packets, the frequency of a packet with rank k that appears in RREQs is

$$f(p_k) = \frac{\frac{1}{k^\theta}}{\sum_{i=1}^{\mathcal{N}_R} \frac{1}{i^\theta}},$$

where θ is an exponent characterizing the Zipfian distribution. For instance, supposing that $\mathcal{N}_R = 5$ and $\theta = 0.6$, the frequencies of packets p_1, p_2, p_3, p_4, and p_5 appearing in RREQs are 0.33, 0.22, 0.17. 0.15, and 0.13, respectively. In other words, about 33%, 22%, 17%, 15%, and 13% RREQs request packets p_1, p_2, p_3, p_4, and p_5, respectively. Therefore, we calculate δ_{num} by

$$\sum_{i=1}^{\delta_{\text{num}}} f(p_i) \geq F_{th},$$

where F_{th} is the expected cumulative frequency of the first δ_{num} packets that most frequently appear in RREQs. Clearly, we have $F_{th} > 0.5$ (that is, the first δ_{num} packets are requested by more than half of all RREQs). For instance, we can set $F_{th} = 0.7$ and thus calculate $\delta_{\text{num}} = 3$ in the previous example. In this case, since the first three packets are requested by at least 70% RREQs, the RS can broadcast g_{ack} without waiting the remaining 30% RREQs.

4.3 Adjustment of threshold δ_{var}

The threshold δ_{var} also determines how quickly the RS will broadcast g_{ack}. When $c_{\text{var}} \geq \delta_{\text{var}}$, the RS sends a g_{ack} because the received RREQs may exhibit spatial correlation.

Let ε_1, ε_2, $\cdots$, and $\varepsilon_{\mathcal{N}_R}$ be the minimum values of n_1, n_2, $\cdots$, and $n_{\mathcal{N}_R}$ that trigger the RS to broadcast a g_{ack}, respectively, where $\varepsilon_i > 0$ for all $i = 1..\mathcal{N}_R$. In other words, if $n_i \geq \varepsilon_i$ for *any* $i = 1..\mathcal{N}_R$, the RS sends a g_{ack}. Then, δ_{var} is calculated by the coefficient of variation of all ε_is, $i = 1..\mathcal{N}_R$:

$$\delta_{\text{var}} = \frac{\sqrt{\frac{1}{\mathcal{N}_R} \sum_{i=1}^{\mathcal{N}_R} (\varepsilon_i - \varepsilon_{\text{avg}})^2}}{\varepsilon_{\text{avg}}}, \tag{6}$$

where $\varepsilon_{avg} = \frac{1}{\mathcal{N}_R} \sum_{i=1}^{\mathcal{N}_R} \varepsilon_i$. Note that the sum of all ε_is, $i = 1..\mathcal{N}_R$, is smaller than ζ (otherwise, by case 3 in Section 3.2, the RS sends a g_{ack} due to $\sum_{i=1}^{\mathcal{N}_R} n_i \geq \zeta$). Thus, we have

$$\sum_{i=1}^{\mathcal{N}_R} \varepsilon_i \leq \sum_{i=1}^{\mathcal{N}_R} n_i < \zeta. \tag{7}$$

The values of ε_1, ε_2, $\cdots$, and $\varepsilon_{\mathcal{N}_R}$ can be calculated by an optimization problem whose objective is to minimize the total communication cost between the RS and its MDs. In particular, let $\Theta(\zeta)$ be the number of messages exchanged between the RS and its MDs during the $t_{g_{\rm ack}}$ period, under the constraint in Eq. (7). Therefore, we have

$$\Theta(\zeta) = \mathcal{N}_{RREQ} + \mathcal{N}_{g_{\rm ack}},$$

where $\mathcal{N}_{RREQ}$ is the number of RREQs sent by MDs and $\mathcal{N}_{g_{\rm ack}}$ is the number of $g_{\rm ack}$ sent by the RS. Since the RS sends only one $g_{\rm ack}$ during the $t_{g_{\rm ack}}$ period, we have $\mathcal{N}_{g_{\rm ack}} = 1$. Thus, the above equation is written as

$$\Theta(\zeta) = \mathcal{N}_{RREQ} + 1. \quad (8)$$

Our goal is to find the values of ε_1, ε_2, $\cdots$, and $\varepsilon_{\mathcal{N}_R}$ such that $\Theta(\zeta)$ is minimized.

Since $\sum_{i=1}^{\mathcal{N}_R} n_i \geq \zeta$ will trigger the RS to send a $g_{\rm ack}$, the maximum number of RREQs received by the RS before sending $g_{\rm ack}$ is ζ. Therefore, before the condition $c_{\rm var} \geq \delta_{\rm var}$ is satisfied, the expectation of $\mathcal{N}_{RREQ}$ is

$$\begin{aligned} & E[\mathcal{N}_{RREQ}] \\ &= \zeta \times Prob((\text{no } g_{\rm ack} \text{ is sent due to } c_{\rm var} < \delta_{\rm var}) \mid (\text{no } g_{\rm ack} \\ &\quad \text{is sent due to } \textstyle\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta)) \\ &= \zeta \times Prob((n_i < \varepsilon_i, \forall p_i \in \mathcal{R}) \mid (\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta)) \\ &= \zeta \times \frac{Prob((n_i < \varepsilon_i, \forall p_i \in \mathcal{R}) \cap (\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta))}{Prob(\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta)}. \end{aligned} \quad (9)$$

We assume that each MD sends its RREQs *independently* with other MDs. By Eq. (7), we obtain that

$$\begin{aligned} & Prob((n_i < \varepsilon_i, \forall p_i \in \mathcal{R}) \cap (\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta)) \\ &= Prob(n_i < \varepsilon_i, \forall p_i \in \mathcal{R}) \times Prob(\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta) \\ &= \prod_{i=1}^{\mathcal{N}_R} Prob(n_i < \varepsilon_i) \times Prob(\sum_{i=1}^{\mathcal{N}_R} n_i < \zeta). \end{aligned} \quad (10)$$

By substituting Eq. (10) into Eq. (9), we have

$$E[\mathcal{N}_{RREQ}] = \zeta \times \prod_{i=1}^{\mathcal{N}_R} Prob(n_i < \varepsilon_i). \quad (11)$$

Thus, by Eq. (11), Eq. (8) is written as

$$\Theta(\zeta) = \zeta \times \prod_{i=1}^{\mathcal{N}_R} Prob(n_i < \varepsilon_i) + 1. \quad (12)$$

Since we have $\varepsilon_i > 0$ for each $i = 1..\mathcal{N}_R$, we can obtain that $Prob(n_i < \varepsilon_i) > 0$. Besides, since ζ is a given constant, the only way to minimize $\Theta(\zeta)$ in Eq. (12) is to minimize the value of $Prob(n_i < \varepsilon_i)$ for each $i = 1..\mathcal{N}_R$.

Recall that each n_i, $i = 1..\mathcal{N}_R$, is the number of RREQs received by the RS. Thus, a combination of all n_is reflects the statistical quantity of RREQs that should follow some distribution (*e.g.*, Zipfian distribution). Based on the definition, ε_i is the minimum value of n_i, $i = 1..\mathcal{N}_R$, that triggers the RS to broadcast a $g_{\rm ack}$. Thus, the combination of all ε_is also follows the same distribution. Therefore, the value of $Prob(n_i < \varepsilon_i)$ is expected to increase when n_i is close to ε_i for each $i = 1..\mathcal{N}_R$. Based on the above observation, we define a *distance function* of n_i and ε_i as follows:

$$H(n_i, \varepsilon_i) = \begin{cases} \frac{\max(n_i,\varepsilon_i)}{\min(n_i,\varepsilon_i)} & \text{if } n_i < \varepsilon_i \\ 1 & \text{otherwise.} \end{cases}$$

When $n_i < \varepsilon_i$, we have $H(n_i, \varepsilon_i) > 1$ and its value decreases when n_i is more close to ε_i. On the other hand, when $n_i \geq \varepsilon_i$, we define the value of $H(n_i, \varepsilon_i)$ to be one. By transforming the probability $Prob(n_i < \varepsilon_i)$ to the distance function $H(n_i, \varepsilon_i)$, our objective (to minimize $\Theta(\zeta)$ in Eq. (12)) becomes

$$\text{minimize} \prod_{i=1}^{\mathcal{N}_R} H(n_i, \varepsilon_i). \quad (13)$$

By taking logarithm in Eq. (13), we calculate the optimal values of $\varepsilon_1, \varepsilon_2, \cdots$, and $\varepsilon_{\mathcal{N}_R}$ to minimize $\Theta(\zeta)$ by the following optimization equations:

$$\begin{aligned} & \text{minimize} \\ & \qquad \sum_{i=1}^{\mathcal{N}_R} \log H(n_i, \varepsilon_i), \\ & \text{subject to} \\ & \qquad n_i > 0 \quad \text{for } i = 1..\mathcal{N}_R, \\ & \qquad \varepsilon_i > 0 \quad \text{for } i = 1..\mathcal{N}_R, \\ & \qquad 0 < \sum_{i=1}^{\mathcal{N}_R} \varepsilon_i \leq \zeta. \end{aligned} \quad (14)$$

We can use the Karush-Kuhn-Tucker (KKT) algorithm [3] to solve Eq. (14). Then, by applying the values of all ε_is, $i = 1..\mathcal{N}_R$, into Eq. (6), we can calculate $\delta_{\rm var}$.

5 Performance Evaluation

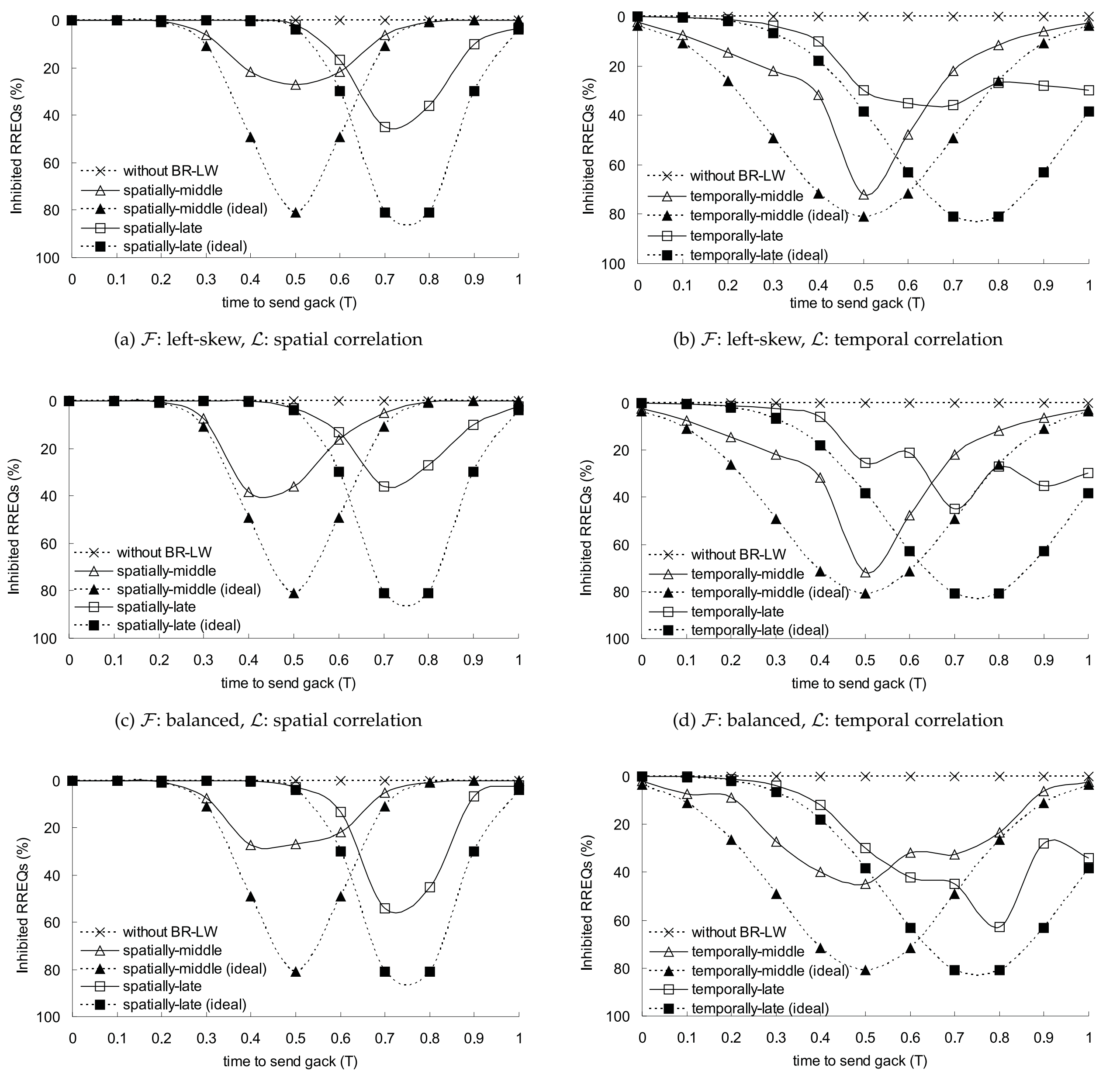
In this section, we verify the effectiveness of BR-LW by simulations. In our simulations, there is one BS supporting four RSs, as shown in Fig. 1. These four RSs belong to the same multicast group, and each RS serves ten MDs initially. We measure the amounts of RREQ submissions from the MDs and packet retransmissions from the RS under different distributions of *DVB-H packet loss* $\mathcal{F}$ and *correlation feature of RREQs* $\mathcal{L}$. The two distributions $\mathcal{F}$ and $\mathcal{L}$ are normal distributions and independent with each other. For distribution $\mathcal{F}$, we consider three scenarios:

- **Left-skew:** The peak of $\mathcal{F}$ appears in the left-hand side. $\mathcal{F}(\mathcal{X}_\mathcal{F} = 0 \sim \mathcal{T}, \mu_\mathcal{F} = \frac{1}{4}\mathcal{T}, \sigma_\mathcal{F} = \frac{1}{5}\mathcal{T})$.
- **Balanced:** The peak of $\mathcal{F}$ appears in the middle. $\mathcal{F}(\mathcal{X}_\mathcal{F} = 0 \sim \mathcal{T}, \mu_\mathcal{F} = \frac{1}{2}\mathcal{T}, \sigma_\mathcal{F} = \frac{1}{5}\mathcal{T})$.
- **Right-skew:** The peak of $\mathcal{F}$ appears in the right-hand side. $\mathcal{F}(\mathcal{X}_\mathcal{F} = 0 \sim \mathcal{T}, \mu_\mathcal{F} = \frac{3}{4}\mathcal{T}, \sigma_\mathcal{F} = \frac{1}{5}\mathcal{T})$.

Here, $\mathcal{X}_\mathcal{F}$, $\mu_\mathcal{F}$, and $\sigma_\mathcal{F}$ are the random variable, mean value, and standard deviation of $\mathcal{F}$, respectively, and $\mathcal{T}$ is the period of simulation time. Distribution $\mathcal{F}$ models the needs of sending RREQs by MDs. The three scenarios, namely *left-skew*, *balanced*, and *right-skew*, indicate that a mass of packet loss occur at time around $0.25\mathcal{T}$, $0.5\mathcal{T}$, and $0.75\mathcal{T}$, respectively.

For distribution $\mathcal{L}$, we consider four scenarios:

- **Spatially-middle:** Spatially-correlated RREQs are generated in the middle of the simulation. $\mathcal{L}(\mathcal{X}_\mathcal{L} = 0 \sim \mathcal{T}, \mu_\mathcal{L} = \frac{1}{2}\mathcal{T}, \sigma_\mathcal{L} = \frac{1}{10}\mathcal{T})$.



(a) $\mathcal{F}$: left-skew, $\mathcal{L}$: spatial correlation

(b) $\mathcal{F}$: left-skew, $\mathcal{L}$: temporal correlation

(c) $\mathcal{F}$: balanced, $\mathcal{L}$: spatial correlation

(d) $\mathcal{F}$: balanced, $\mathcal{L}$: temporal correlation

(e) $\mathcal{F}$: right-skew, $\mathcal{L}$: spatial correlation

(f) $\mathcal{F}$: right-skew, $\mathcal{L}$: temporal correlation

Fig. 3: The ratio of inhibited RREQs by BR-LW under different distributions of $\mathcal{F}$ and $\mathcal{L}$.

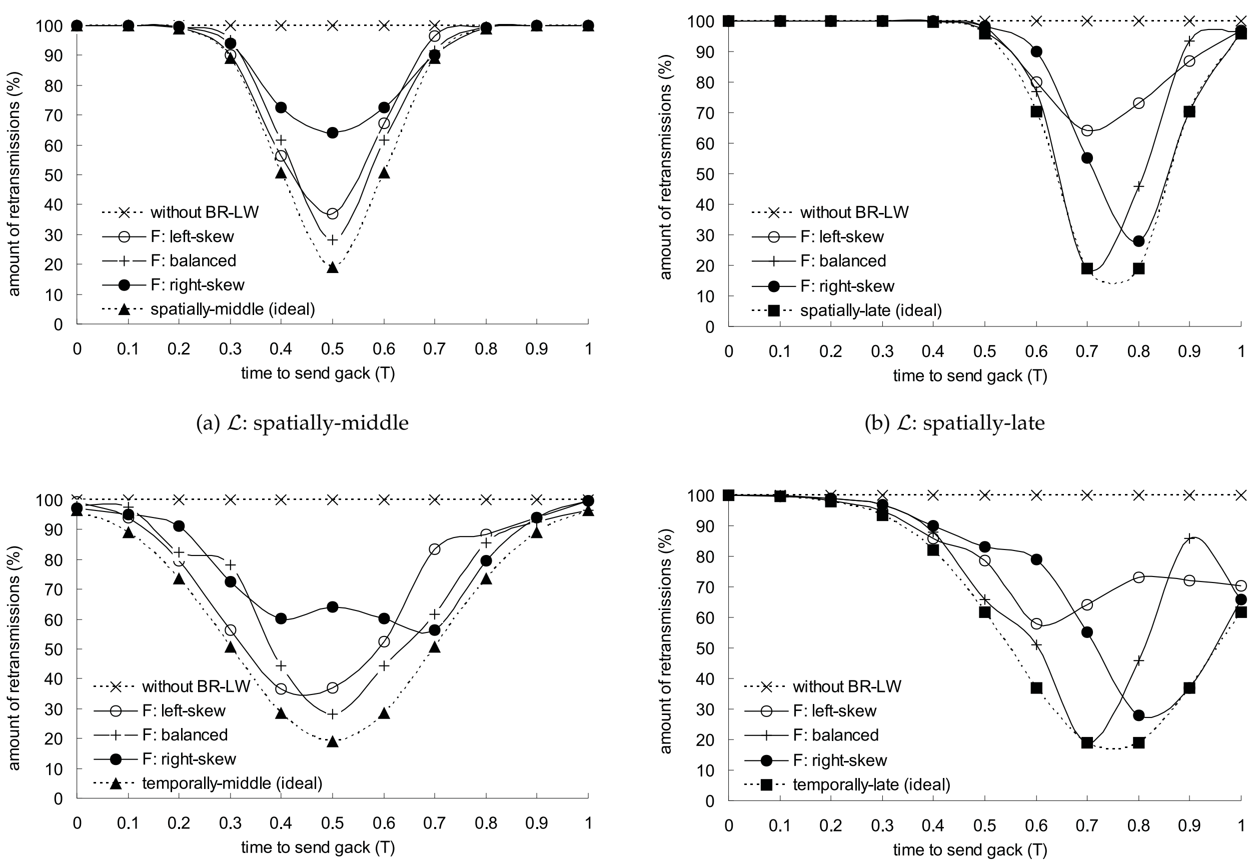
- **Spatially-late:** Spatially-correlated RREQs are generated in the late portion of the simulation. $\mathcal{L}(\mathcal{X}_\mathcal{L} = 0 \sim \mathcal{T}, \mu_\mathcal{L} = \frac{3}{4}\mathcal{T}, \sigma_\mathcal{L} = \frac{1}{10}\mathcal{T})$.
- **Temporally-middle:** Temporally-correlated RREQs are generated in the middle of the simulation. $\mathcal{L}(\mathcal{X}_\mathcal{L} = 0 \sim \mathcal{T}, \mu_\mathcal{L} = \frac{1}{2}\mathcal{T}, \sigma_\mathcal{L} = \frac{1}{5}\mathcal{T})$.
- **Temporally-late:** Temporally-correlated RREQs are generated in the late portion of the simulation. $\mathcal{L}(\mathcal{X}_\mathcal{L} = 0 \sim \mathcal{T}, \mu_\mathcal{L} = \frac{3}{4}\mathcal{T}, \sigma_\mathcal{L} = \frac{1}{5}\mathcal{T})$.

Here, $\mathcal{X}_\mathcal{L}$, $\mu_\mathcal{L}$, and $\sigma_\mathcal{L}$ are the random variable, mean value, and standard deviation of $\mathcal{L}$, respectively. Distribution $\mathcal{L}$ models the inhibited submissions of RREQs after the RS broadcasts a g_{ack}. To model the spatial correlation of RREQs, the two scenarios, namely *spatially-middle* and *spatially-late*, indicate that a mass of correlated RREQs are generated at time around $0.5\mathcal{T}$ and $0.75\mathcal{T}$, respectively. The similar scenarios (that is, *temporally-middle* and *temporally-late*) are also applied to model the temporal correlation of RREQs.

5.1 Ratio of Inhibited RREQs

We first evaluate the ratio of inhibited RREQs by BR-LW. To show the effect of g_{ack} timing, we do not apply the adjustment mechanisms in Sections 4.2 and 4.3. Instead, we let the RS send a g_{ack} once at *fixed* time $0.1\mathcal{T}$, $0.2\mathcal{T}$, $\cdots$, and $\mathcal{T}$. Note that the RS sends only one g_{ack} and MDs still submit their RREQs after receiving a g_{ack}.

Fig. 3 shows the ratio of inhibited RREQs under different distributions of $\mathcal{F}$ and $\mathcal{L}$. For comparison, we consider an *ideal* case where only one MD sends RREQs and other MDs are silent before the RS sending g_{ack}. As can be seen, g_{ack} timing significantly affects the ratio of inhibited RREQs. The peak (that is, the maximum ratio of inhibited RREQs) of the ideal case in the spatially-middle (respectively, temporally-middle) scenario of $\mathcal{L}$ locates at time $0.5\mathcal{T}$ because a mass



(a) $\mathcal{L}$: spatially-middle

(b) $\mathcal{L}$: spatially-late

(c) $\mathcal{L}$: temporally-middle

(d) $\mathcal{L}$: temporally-late

Fig. 4: The amount of packet retransmissions by BR-LW under different distributions of $\mathcal{F}$ and $\mathcal{L}$.

of spatially (respectively, temporally) correlated RREQs are generated at time $0.5\mathcal{T}$. In this case, most MDs can find similar requests from the $g_{\rm ack}$ and thus avoid sending duplicate RREQs. Similarly, the peak of the ideal case in the spatially-late (respectively, temporally-late) scenario of $\mathcal{L}$ appears at time between $0.7\mathcal{T}$ and $0.8\mathcal{T}$ since a mass of spatially (respectively, temporally) correlated RREQs are generated at time $0.75\mathcal{T}$. However, since the RS will put its predicted future requests in the $g_{\rm ack}$ if RREQs exhibit temporal correlation, the effect of $g_{\rm ack}$ timing is less significant in the temporally-middle and temporally-late scenarios compared to the spatially-middle and spatially-late scenarios of $\mathcal{L}$.

We then discuss the effect of $\mathcal{F}$ on inhibited RREQs. From Fig. 3(a), (c), and (e), we observe that the peak of spatially-middle (respectively, spatially-late) scenario of $\mathcal{L}$ locates at time between $0.4\mathcal{T}$ and $0.5\mathcal{T}$ (respectively, $0.7\mathcal{T}$ and $0.8\mathcal{T}$), which is close to that of the ideal case. In the spatially-middle scenario of $\mathcal{L}$, the balanced scenario of $\mathcal{F}$ has the highest ratio of inhibited RREQs (refer to Fig. 3(c)). The reason is that most MDs lose a mass of packets at time around $0.5\mathcal{T}$ and these packets exhibit high spatial correlation. On the other hand, in the spatially-late scenario of $\mathcal{L}$, the right-skew scenario of $\mathcal{F}$ has the highest ratio (refer to Fig. 3(e)). Similarly, since most MDs incur serious packet loss at time around $0.75\mathcal{T}$ and these packets also exhibit high spatial correlation, sending a $g_{\rm ack}$ at time $0.75\mathcal{T}$ can reduce the most duplicate RREQs.

From Fig. 3(b), (d), and (f), we observe that the peak of temporally-middle scenario of $\mathcal{L}$ appears at time $0.5\mathcal{T}$, which is the same as that of the ideal case. The peak of temporally-late scenario of $\mathcal{L}$ appears at time $0.7\mathcal{T}$ and $0.8\mathcal{T}$, which is close to that of the ideal case. The ratio of inhibited RREQs in the temporally-late scenario of $\mathcal{L}$ oscillates because a mass of temporally-correlated RREQs are generated late (at time $0.75\mathcal{T}$) and thus the RS may not predict many future requests. Comparing Fig. 3(b), (d), and (f), in the temporally-middle scenario of $\mathcal{L}$, the right-skew scenario of $\mathcal{F}$ has the lowest ratio. The reason is that most MDs lose their packets in the late portion of the simulation and these packets may be uncorrelated (since a mass of temporally-correlated RREQs are generated in the middle of the simulation). However, when most MDs lose their packets in the late portion of the simulation (that is, the right-skew scenario of $\mathcal{F}$ in Fig. 3(f)), the temporally-late scenario of $\mathcal{L}$ has the highest ratio because the peaks of distributions $\mathcal{F}$ and $\mathcal{L}$ overlap.

5.2 Reduction of Packet Retransmissions

Using the same setting in the previous section, we then measure the reduced amount of packet retransmissions by BR-LW. Recall that the RS will broadcast the packets after sending $g_{\rm ack}$. Thus, the $g_{\rm ack}$ timing can be approximately viewed as the timing to broadcast packets. To show the effect of such broadcasting timing, after the RS broadcasts the packets (at the time of sending $g_{\rm ack}$), we allow MDs to receive unicast packets and submit RREQs simultaneously.

Fig. 4 shows the amount of packet retransmissions under different distributions $\mathcal{F}$ and $\mathcal{L}$. Without BR-LW, the RS always unicasts the requested packets to MDs after receiving RREQs. In the *ideal* case, we assume that the RS can know the lost packets by *all* MDs before sending g_{ack}. From Fig. 4, g_{ack} timing significantly affects the amount of packet retransmissions. The peak (that is, the minimum amount of packet retransmissions) of the ideal case in the spatially-middle (respectively, temporally-middle) scenario of $\mathcal{L}$ locates at time $0.5\mathcal{T}$ because a mass of spatially (respectively, temporally) correlated RREQs are generated at time $0.5\mathcal{T}$. In this case, most of the retransmitted packets can meet the requirements of most MDs. Similarly, the peak of the ideal case in the spatially-late (respectively, temporally-late) scenario of $\mathcal{L}$ appears at time between $0.7\mathcal{T}$ and $0.8\mathcal{T}$ since a mass of spatially (respectively, temporally) correlated RREQs are generated at time $0.75\mathcal{T}$. Again, the effect of g_{ack} timing is less significant in the temporally-middle and temporally-late scenarios compared to the spatially-middle and spatially-late scenarios of $\mathcal{L}$, since the RS predicts the future lost packets if RREQs exhibit temporal correlation.

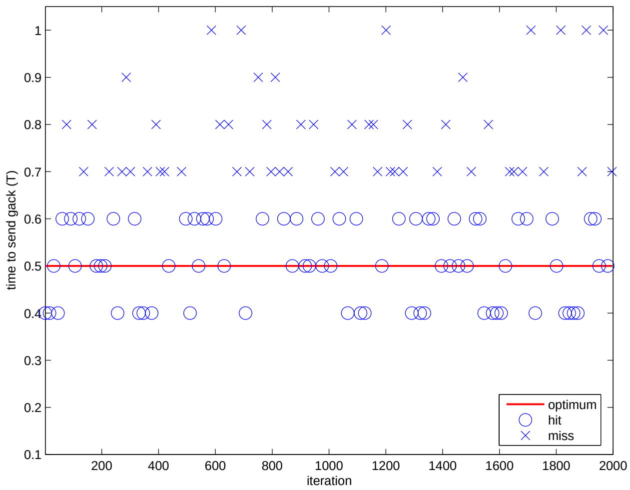
Next, we discuss the effect of $\mathcal{F}$ on packet retransmissions. In Fig. 4(a), the peaks of the three scenarios of $\mathcal{F}$ appear at time $0.5\mathcal{T}$, which are the same as that of the ideal case. The right-skew scenario of $\mathcal{F}$ has the smallest peak, because most MDs experience serious packet loss at time $0.75\mathcal{T}$ but most RREQs exhibit high spatial correlation at time $0.5\mathcal{T}$. In this case, a g_{ack} can only reflect the spatial correlation of a small number of RREQs, as it is sent early at time $0.5\mathcal{T}$. On the other hand, in Fig. 4(b), the peaks of the three scenarios of $\mathcal{F}$ locate at time $0.7\mathcal{T}$ or $0.8\mathcal{T}$, which are close to that of the ideal case. The left-skew scenario of $\mathcal{F}$ has the smallest peak. The reason is that most MDs experience serious packet loss at early time $0.25\mathcal{T}$ but most spatially-correlated RREQs are generated at late time $0.75\mathcal{T}$.

For the temporally-middle and temporally-late scenarios of $\mathcal{L}$ (that is, Fig. 4(c) and (d)), the peaks of different scenarios of $\mathcal{F}$ may not necessary be the same. The reason is that the RS will put its prediction of future lost packets in a g_{ack}. However, similar to Fig. 4(a) and (b), the right-skew (respectively, left-skew) scenario of $\mathcal{F}$ in Fig. 4(c) (respectively, Fig. 4(d)) has the smallest peak, because g_{ack} can only reflect the temporal correlation of a small number of RREQs.

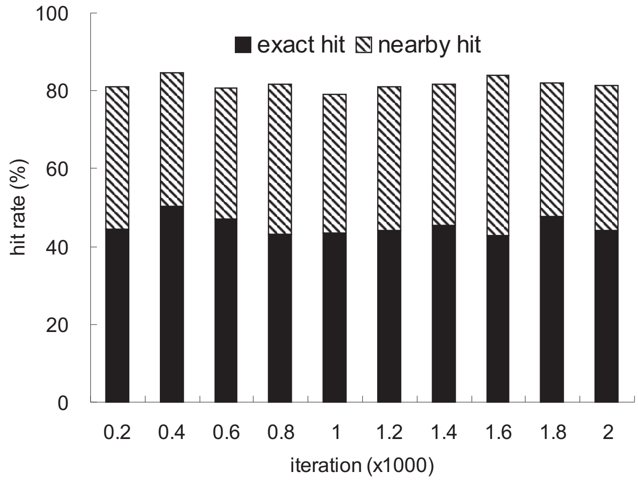
5.3 Adjustment of δ_{var}

Recall that the RS will send a g_{ack} when $c_{\text{var}} \geq \delta_{\text{var}}$, which means that the received RREQs may exhibit high spatial correlation. We also analyze how to adjust δ_{var} in Section 4.3. In this section, we verify the correctness of our analysis on δ_{var}. The simulation consists of 2000 iterations, each with a period of $\mathcal{T}$. In an iteration, we divide the time into 10 segments and check when the condition $c_{\text{var}} \geq \delta_{\text{var}}$ occurs. We adopt the spatially-middle and spatially-late scenarios of $\mathcal{L}$.

Fig. 5(a) shows when the condition $c_{\text{var}} \geq \delta_{\text{var}}$ occurs in the spatially-middle scenario of $\mathcal{L}$. From Figs. 3 and 4, we find that the optimal time to send a g_{ack} is $0.5\mathcal{T}$. We define that the calculation of δ_{var} *hits* the optimal value if the time when $c_{\text{var}} \geq \delta_{\text{var}}$ occurs locates between time $0.4\mathcal{T}$ and $0.6\mathcal{T}$ (that is, we tolerate an error of $0.1\mathcal{T}$). Fig. 5(b) shows the hit rate of δ_{var}, where the *exact hit* means that $c_{\text{var}} \geq \delta_{\text{var}}$ occurs exactly at time $0.5\mathcal{T}$ and the *nearby hit* means that $c_{\text{var}} \geq \delta_{\text{var}}$ occurs at time either $0.4\mathcal{T}$ or $0.6\mathcal{T}$. As can be seen, in the spatially-middle scenario of $\mathcal{L}$, the overall hit rate (that is, the sum of exact hits and nearby hits) is always higher than 80%. Besides, the exact hit rate is always higher than 40%. The above result shows the correctness of our analysis on δ_{var} in the spatially-middle scenario of $\mathcal{L}$.



(a) time to send g_{ack} (due to $c_{\text{var}} \geq \delta_{\text{var}}$)



(b) hit rate of δ_{var}

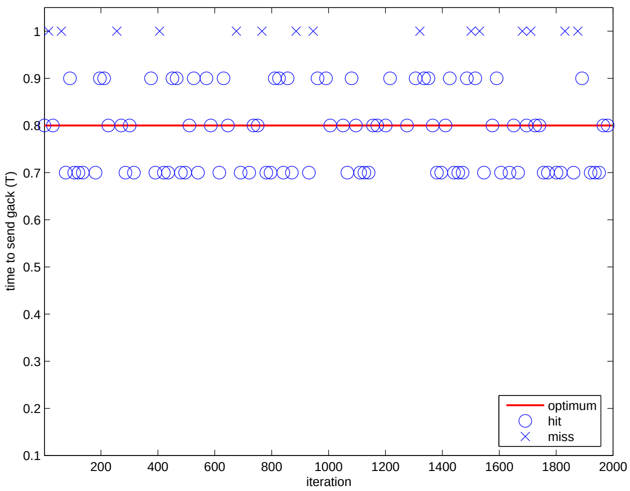
Fig. 5: The time when $c_{\text{var}} \geq \delta_{\text{var}}$ occurs and the hit rate of δ_{var} in the spatially-middle scenario of $\mathcal{L}$.

Fig. 6(a) shows when the condition $c_{\text{var}} \geq \delta_{\text{var}}$ occurs in the spatially-late scenario of $\mathcal{L}$. Again, from Figs. 3 and 4, we find that the optimal time to send a g_{ack} is $0.8\mathcal{T}^3$. The calculation of δ_{var} is said to *hit* the optimal value if the time when $c_{\text{var}} \geq \delta_{\text{var}}$ occurs locates between time $0.7\mathcal{T}$ and $0.9\mathcal{T}$. Fig. 6(b) shows the hit rate of δ_{var}, where the *exact hit* means that $c_{\text{var}} \geq \delta_{\text{var}}$ occurs exactly at time $0.8\mathcal{T}$ and the *nearby hit* means that $c_{\text{var}} \geq \delta_{\text{var}}$ occurs at time either $0.7\mathcal{T}$ or $0.9\mathcal{T}$. As can be seen, in the spatially-late scenario of $\mathcal{L}$, the overall hit rate is always higher than 85%. Besides, the exact hit rate is always higher than 35%. The above result shows the correctness of our analysis on δ_{var} in the spatially-late scenario of $\mathcal{L}$.

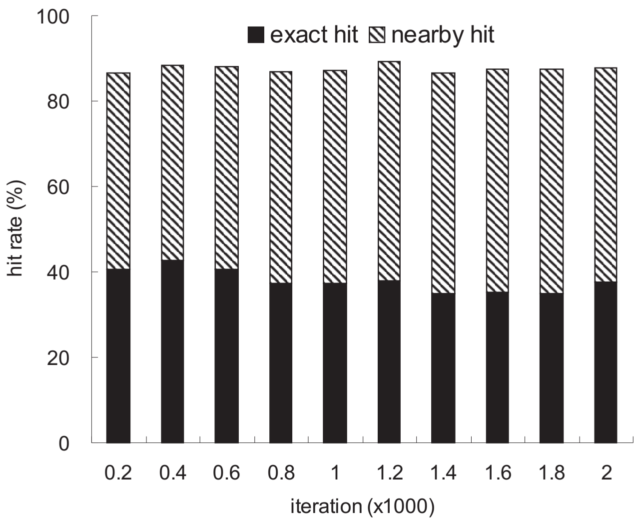
5.4 Effect of Handovers

The previous experiments consider the behavior only inside a single RS cell. In this section, we evaluate the effect of handovers. The mobility of each MD follows the random waypoint model, and thus the occurrence of handovers can be simulated by a uniform distribution. In addition, to measure the effect of handover on the timing of sending g_{ack}, we allow

3. The optimal value in fact locates at time between $0.7\mathcal{T}$ and $0.8\mathcal{T}$. For convenience, we take $0.8\mathcal{T}$ as the optimal value.

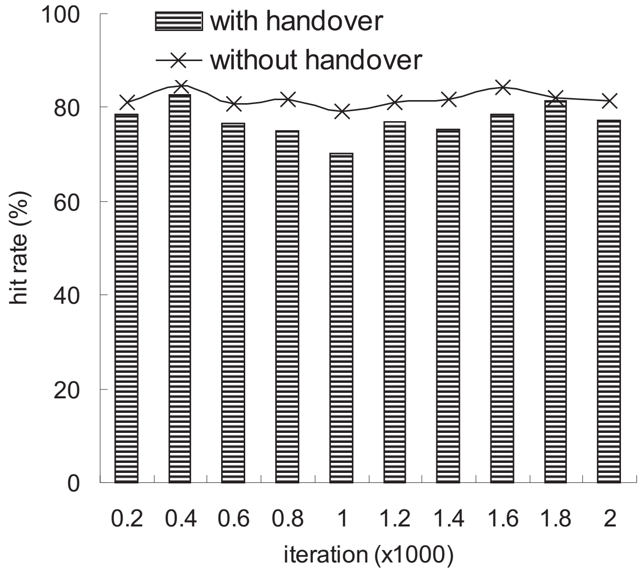


(a) time to send g_{ack} (due to $c_{\text{var}} \geq \delta_{\text{var}}$)

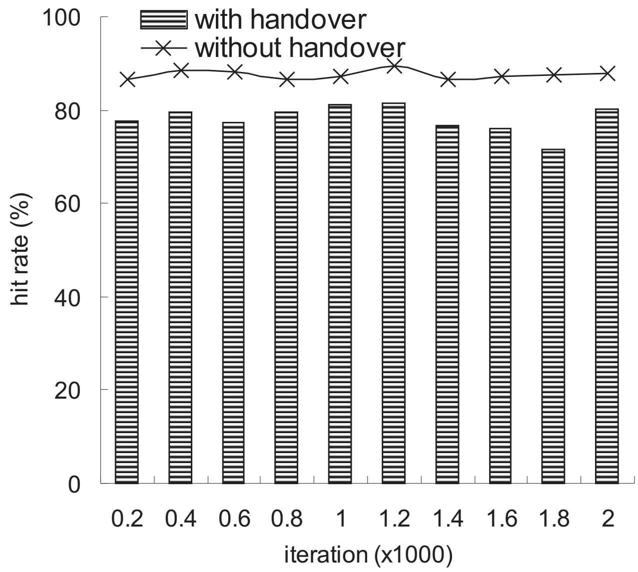


(b) hit rate of δ_{var}

Fig. 6: The time when $c_{\text{var}} \geq \delta_{\text{var}}$ occurs and the hit rate of δ_{var} in the spatially-late scenario of $\mathcal{L}$.



(a) hit rate of δ_{var} in the spatially-middle scenario of $\mathcal{L}$



(b) hit rate of δ_{var} in the spatially-late scenario of $\mathcal{L}$

Fig. 7: The hit rate of sending g_{ack} (due to $c_{\text{var}} \geq \delta_{\text{var}}$) when there are handovers.

each MD to send out all of its "unqueried" RREQs to the new associated RS after handovering.

Fig. 7(a) and (b) show the hit rates of sending g_{ack} due to $c_{\text{var}} \geq \delta_{\text{var}}$ (including both exact and nearby hits) in the spatially-middle and spatially-late scenarios of $\mathcal{L}$, respectively. For comparison, we also observe the hit rate when there are no handovers. Because the handovering behavior of MDs extends the spatial correlation of RREQs to neighboring RS cells, our lazy-wait operation can help reduce the redundant requests sent from handovering MDs. Therefore, there is only a slight decrease of hit rate caused by handovering MDs. From Fig. 7, the decreases of hit rates in the spatially-middle and spatially-late scenarios of $\mathcal{L}$ are about 4.5% and 9.5%, respectively. On average, the hit rate always exceeds 75% even when there are handovers, which shows the effectiveness of our BR-LW scheme.

6 Conclusions

DVB-H supports the broadcast service of digital video to handheld devices in a mobile environment, but it may suffer from serious packet loss due to unreliable wireless transmissions. In this paper, we follow the DVB-IPDC architecture by integrating a DVB-H system with a broadband WiMAX network to support the packet recovery mechanism. We have addressed two critical GPL and BDH problems and developed a novel BR-LW scheme to solve these problems. BR-LW exploits the spatial and temporal correlation of recovery requests, and efficiently reduces duplicate request submissions while merging retransmissions of lost packets. Simulation results have verified that BR-LW significantly reduces the amount of message transmissions inside a WiMAX RS cell, thereby alleviating network congestion while improving communication efficiency. Furthermore, with mathematical analysis, we have discussed how to adaptively determine the timing for each MD to send its requests based on the channel condition, as well as the timing for each RS to broadcast a group acknowledgement according to the spatial and temporal correlation of received requests.

For the future work, we will further investigate how to improve video quality of MDs by considering some QoS requirements such as delay, jitter, and PSNR in the proposed scheme.

REFERENCES

[1] R. Akester, "Reducing multicast collision loss for digital TV over 802.11 wireless networks," *Proc. WSEAS Int'l Conf. Multimedia, Internet and Video Technologies*, 2004.

[2] S.A. Azad and M. Murshed, "An efficient transmission scheme for minimizing user waiting time in video-on-demand systems," *IEEE Comm. Letters*, vol. 11, no. 3, pp. 285–287, 2007.

[3] E.K.P. Chong and S.H. Zak, *An Introduction to Optimization*, John Wiley & Sons Inc., 2008.

[4] I. Djama and T. Ahmed, "A cross-layer interworking of DVB-T and WLAN for mobile IPTV service delivery," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 382–390, 2007.

[5] O. Eerenberg, A. Koppelaar, A.M. Stuivenwold, and P.H.N. de With, "IP-recovery in the DVB-H link layer for TV on mobile," *Proc. IEEE Int'l Conf. Consumer Electronics*, 2006, pp. 411–412.

[6] ETSI TR 102 469, "Digital video broadcasting (DVB); IP datacast over DVB-H: architecture," *European Telecomm. Standards Institute*, 2006.

[7] ETSI TS 102 472, "Digital video broadcasting (DVB); IP datacast over DVB-H: content delivery protocols," *European Telecomm. Standards Institute*, 2006.

[8] ETSI TS 102 591, "Digital video broadcasting (DVB); IP datacast over DVB-H: content delivery protocols (CDP) implementation guidelines," *European Telecomm. Standards Institute*, 2007.

[9] G. Faria, J.A. Henriksson, E. Stare, and P. Talmola, "DVB-H: Digital broadcast services to handheld devices," *Proc. IEEE*, vol. 94, no. 1, pp. 194–209, 2006.

[10] D. Gomez-Barquero and A. Bria, "Repair mechanisms for broadcast transmissions in hybrid cellular & DVB-H systems," *Proc. IEEE Int'l Symp. Wireless Comm. Systems*, 2006, pp. 398–402.

[11] D. Gomez-Barquero, N. Cardona, A. Bria, and J. Zander, "Affordable mobile TV services in hybrid cellular and DVB-H systems," *IEEE Network*, vol. 21, no. 2, pp. 34–40, 2007.

[12] D. Gomez-Barquero, D. Gozalvez, and N. Cardona, "Application layer FEC for mobile TV delivery in IP datacast over DVB-H systems," *IEEE Trans. Broadcasting*, vol. 55, no. 2, pp. 396–406, 2009.

[13] B. Hechenleitner, "Repair cost of the IPDC/DVB-H file repair mechanism," *Proc. IEEE Wireless Telecomm. Symp.*, 2008, pp. 137–144.

[14] A. Helmy, M. Jaseemuddin, and G. Bhaskara, "Efficient micro-mobility using intra-domain multicast-based mechanisms (M&M)," *ACM SIGCOMM Computer Comm. Review*, vol. 32, no. 5, pp. 61–72, 2002.

[15] C. Heuck, "An analytical approach for performance evaluation of hybrid (broadcast/mobile) networks," *IEEE Trans. Broadcasting*, vol. 56, no. 1, pp. 9–18, 2010.

[16] P. Hummelbrunner, S. Buchinger, W. Robitza, D. Selig, M. Nezveda, and H. Hlavacs, "Peer to peer mobile TV recovery system," *Proc. Int'l Interactive Conf. Interactive TV & Video*, 2010, pp. 263–272.

[17] M. Kornfeld and G. May, "DVB-H and IP datacast-broadcast to handheld devices," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 161–170, 2007.

[18] D. Kouis, P. Demestichas, G. Koundourakis, and M.E. Theologou, "Resource management of IP-enabled DVB-T networks in the context of wireless B3G systems," *Wireless Networks*, vol. 13, no. 2, pp. 165–175, 2007.

[19] D. Kouis, D. Loukatos, K. Kontovasilis, G. Kormentzas, and C. Skianis, "On the effectiveness of DVB-T for the support of IP-based services in heterogeneous wireless networks," *Computer Networks*, vol. 48, no. 1, pp. 57–73, 2005.

[20] K. Lu, Y. Qian, H.H. Chen, and S. Fu, "WiMAX networks: from access to service platform," *IEEE Network*, vol. 22, no. 3, pp. 38–45, 2008.

[21] D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[22] G. Mastorakis, G. Kormentzas, and E. Pallis, "A fusion IP/DVB networking environment for providing always-on connectivity and triple-play services to urban and rural areas," *IEEE Network*, vol. 21, no. 2, pp. 21–27, 2007.

[23] P. Pangalos, J.M.D.L.T. Velver, M. Dashti, A. Dashti, and H. Aghvami, "Confirming connectivity in interworked broadcast and mobile networks," *IEEE Network*, vol. 21, no. 2, pp. 13–20, 2007.

[24] D. Plets, W. Joseph, L. Verloock, E. Tanghe, L. Martens, E. Deventer, and H. Gauderis, "Influence of reception condition, MPE-FEC rate and modulation scheme on performance of DVB-H," *IEEE Trans. Broadcasting*, vol. 54, no. 3, pp. 590–598, 2008.

[25] J. Poikonen, J. Paavola, and V. Ipatov, "Aggregated renewal Markov process with applications in simulating mobile broadcast systems," *IEEE Trans. Vehicular Technology*, vol. 58, no. 1, pp. 21–31, 2009.

[26] C. Rauch, W. Kellerer, and P. Sties, "Hybrid mobile interactive services combining DVB-T and GPRS," *Proc. European Personal Mobile Comm. Conf.*, 2001.

[27] R. Schatz, N. Jordan, and S. Wagner, "Beyond broadcast–a hybrid testbed for mobile TV 2.0 services," *Proc. Int'l Conf. Networking*, 2007.

[28] R. Schatz, S. Wagner, and N. Jordan, "Mobile social TV: extending DVB-H services with P2P-interaction," *Proc. IEEE Int'l Conf. Digital Telecomm.*, 2007.

[29] K. Sinkar, A. Jagirdar, T. Korakis, H. Liu, S. Mathur, and S. Panwar, "Cooperative recovery in heterogeneous mobile networks," *Proc. IEEE Int'l Conf. Sensor, Mesh and Ad Hoc Comm. and Networks*, 2008, pp. 395–403.

[30] A. Stephane, A. Mihailovic, and A.H. Aghvami, "Mechanisms and hierarchical topology for fast handover in wireless IP networks," *IEEE Comm. Magazine*, vol. 38, no. 11, pp. 112–115, 2000.

[31] P. Sties and W. Kellerer, "Radio broadcast networks enable broadband Internet access for mobile users," *Proc. EUNICE*, 1999.

[32] N.H. Vaidya and S. Hameed, "Scheduling data broadcast in asymmetric communication environments," *Wireless Networks*, vol. 5, no. 3, pp. 171–182, 1999.

[33] W.H. Yang, Y.C. Wang, Y.C. Tseng, and B.S.P. Lin, "Energy-efficient network selection with mobility pattern awareness in an integrated WiMAX and WiFi network," *Int'l J. Comm. Systems*, vol. 23, no. 2, pp. 213–230, 2010.

[34] G.K. Zipf, *Human behaviour and the principle of least efffort: an introduction to human ecology*, Addison-Wesley, 1949.