A Dynamic Flow Control Algorithm for LTE-Advanced Relay Networks

Ping-Chen Lin, Ray-Guang Cheng, Senior Member, IEEE, and Yu-Jen Chang

Abstract—Relay technology is a candidate for extending the coverage or enhancing the throughput of next-generation cellular systems. In the downlink of an LTE-A relay network, a relay node (RN) normally reserves a small buffer for each user equipment (UE) such that the RN can minimize the number of forwarding packets during UE handover. The small buffer and a mismatch of the data rates between the access link and the relay link may result in buffer-overflow for some UEs while buffer-underflow for the other UEs served by the same RN. This paper presents a simple analytical model to illustrate the buffer-overflow and buffer-underflow problems in the downlink of an LTE-A relay network. A dynamic flow control algorithm (DFCA) is then proposed to minimize the buffer-overflow and buffer-underflow probabilities. The proposed DFCA is designed to minimize the feedback signaling overhead by dynamically adjusting the window size and the feedback frequency based on the relevant measures obtained from individual UE. Simulation results showed that most of the flow control schemes can effectively prevent the buffer-overflow problem. However, only DFCA can provide lower buffer-underflow probabilities and higher throughput for most of UEs than the other flow control schemes do.

Index Terms—LTE-A relay networks, dynamic flow control algorithm (DFCA), buffer-underflow/-overflow probability.

I. INTRODUCTION

RELAY TECHNOLOGY is a promising technique proposed to achieve coverage and throughput enhancements for next-generation cellular systems [1]–[4]. In 3GPP Long Term Evolution-Advanced (LTE-A) relay networks, each relay node (RN) is responsible for relaying packets from the base station (called the donor eNode-B or DeNB in LTE-A) to its user equipment (UE) in LTE-A. The wireless link connecting the DeNB and the RN is referred to as a relay link. The wireless link between the DeNB and UE or between the RN and UE is called an access link. Wireless links can use either in-band relaying or out-of-band relaying [5]. For in-band relaying, the relay link and access links share the same radio frequency (RF) channel. Therefore, the RN cannot transmit data to UE while receiving data from the DeNB. In out-of-band relaying, the relay link and access links use different RF channels, allowing the RN to communicate with UE and the DeNB simultaneously. The 3GPP standard defines two types of RNs in LTE-A networks [6]–[8]. A Type-I RN behaves like a base station and transmits its own physical cell identity (ID), common reference signal, and scheduling information. UE reports their channel quality information (CQI) to the Type-I RN and receive the scheduling information and hybrid automatic repeat request (HARQ) feedback directly from the RN. A Type-II RN acts as a repeater. The UE cannot detect the presence of a Type-II RN because the RN does not have a cell ID and does not transmit a common reference signal.

In the downlink of LTE-A relay networks, the DeNB can transmit packets to multiple UEs through the RN. At the RN, per-UE buffer is allocated and per-UE flow control is required to throttle the data rates among the DeNB and UEs. In the uplink, the RN aggregates packets from multiple UEs and forwards to the DeNB. A single buffer is used by all UEs and flow control between the RN and the DeNB is not required because the DeNB always adjusts its uplink data rate based on the total buffer occupancy reported by the RN. Therefore, we will focus on the downlink transmission in the rest of this paper. In LTE-A relay networks, the data rate of each UE served by an RN may depend on the bandwidth allocated to the UE at the relay link and the attainable data rate of the UE at the access link. All UEs served by the same RN share the fix bandwidth of a relay link. The DeNB allocates the bandwidth of the relay link to each UE based on a predetermined scheduling policy. The attainable data rate of each UE at the access link, however, is not fixed and depends on the signal to noise ratio (SNR) of the UE. Bandwidth is typically allocated based on the CQI reported by each UE. In LTE-A relay networks, however, UE reports its CQI to the RN, instead of the DeNB. Hence, the DeNB cannot properly allocate bandwidth to each UE due to a lack of CQI message. As a result, some of the buffers reserved for UEs in the RN will overflow due to the poor signal qualities of UEs [9], which is referred to as a buffer-overflow problem. In contrast, some of the buffers reserved for UEs in the RN will run out of packets if the RN does not receive enough packets from the DeNB in time [10], which is referred to as a buffer-underflow problem. The mismatch of the data rate between the relay link and the access link leads to buffer-overflow and buffer-underflow problems at the RN. As a result, some UEs may experience excess packet loss while some UEs may suffer from unexpected service interruption. Both of these problems should be prevented as much as possible.

Researchers have proposed several flow control or bandwidth allocation policies [11]–[19] to deal with the buffer-overflow and buffer-underflow problems. Both of policies rely on the statistics (e.g., congestion status, attainable data rate, CQI, or available buffer size of each UE) reported by the RN.
The RN may use explicit signaling of a single bit to indicate its congestion, a credit to specify the number of packets that it can accommodate, or a rate to identify its maximum sustainable data rate. However, the quantity and quality of reported statistics and the incurred signaling overhead may affect the performance of these policies. One study [11] proposes on/off flow control and gradual flow control policies. The DeNB may use the on/off flow control to activate or deactivate a flow at the relay link based on the flow congestion status indicated by the RN. DeNB may also use the gradual flow control (e.g., window-based flow control) to allocate the bandwidth of the relay link to UE based on the available UE buffer sizes reported by the RN. However, the gradual flow control approach lacks a general rule for identifying the congestion status of a flow and the right timing for RN to send the reported information is unknown. Several approaches can be used to allocate the bandwidth for the gradual flow control. Krishnan et al. [12] proposed a bandwidth allocation policy to optimize the bandwidth and power usage under constraints on required rate for relaying cellular systems. Ma et al. [13] proposed a bandwidth allocation policy for DeNB and the Type-I RNs to ensure global fairness of flows. In both policies, the RN should periodically feedback the information (e.g., bandwidth usage, power usage, or queue state) of each associated UE to the DeNB. Wang et al. [14] proposed a bandwidth allocation policy for IEEE 802.16j WiMAX network. In their policy, the relay station periodically determines an expected ingress rate based on the demanded traffic rates of all mobile stations and feedbacks the ingress rate to the base station in order to stabilize the buffer occupancy of the relay station. They took the signaling delay between relay stations and base station into design consideration but assumed mobile stations have similar characters to share a common buffer. Hence, the relay node may still suffer from the buffer-overflow and buffer-underflow problems. Jagannathan et al. [15] proposed threshold-based flow control policies to relieve network congestion. They presented an analytical model to analyze the performance of the threshold-based flow control policies. They assumed that the input data rates of each flow in normal state and congestion state are deterministic and the information is known by the transmitter and the receiver a priori. However, the input data rates of each flow at the relay link dynamically changes in LTE-A relay networks since the bandwidth of the relay link is shared by all UE and the signal quality of each UE may vary from subframe to subframe. In [16] and [17], the authors proposed a closed loop credit-based flow control (CBFC) scheme to avoid buffer-overflow at the receiver for multi-hop wireless networks. In CBFC, a receiver takes relevant measures to sender to determine the number of packets that can be forwarded by the sender. The granted packet volume, which is called credit, is periodically feedback to the sender as a capacity allocation message via the uplink signaling channels. To minimize the feedback overhead, CBFC utilizes two counters (i.e., \( N2 \) and \( N4 \)) to decrease the feedback frequency and the amount of feedback messages [16]. However, improper setting of the two counters in CBFC may result in the buffer-underflow problem, especially for the RN that reserves small buffer for each UE. In this case, the DeNB may wait for a long time to receive new credits but UEs with good signal quality can run out of their buffer very quickly. Weerawardane et al. [18], [19] proposed a congestion control and flow control (CCFC) scheme to avoid congestion in HSDPA networks. CCFC was developed based on the credit-based flow control scheme. Similar to CBFC, the receiver periodically feedbacks a capacity allocation message carrying the credits and reporting interval to control the data rate of the sender. Different to CBFC, CCFC utilizes two reporting intervals (i.e., flow control cycle time and congestion control cycle time [18]) to report the capacity allocation message. The authors suggested using short reporting interval for flows in flow control state and long reporting interval for flows in congestion state, which reduces the amount of feedback messages.

Existing studies rarely addressed the issue of buffer-underflow problem in designing their flow control schemes because the authors normally assumed that the receiver always reserves a large buffer for each flow. This assumption is valid for general transport network and some multi-hop wireless networks but is not applicable to LTE-A relay networks. In LTE-A relay network, a UE may frequently move among RNs. A serving RN should forward the buffered packets to DeNB and re-directs to a target RN during handover due to the lack of direct link between RNs. The packet forwarding occupies the bandwidth of the uplink relay links of the serving RN and the downlink relay links of the target RN. It also introduces extra packet forwarding delay. The bandwidth consumption and the forwarding delay increases as the buffer size grows. Therefore, the RN normally reserves a small buffer for each flow. The small buffer at the RN places a fundamental limitation in the designing of the flow control algorithm since it increases the risk of high buffer-overflow probabilities at the RN. Under the small buffer constraint, the RN can reduce its buffer-overflow probability by reducing the window size but at the risk of increasing its buffer-underflow probability. Therefore, the challenge of the flow control algorithm in LTE-A relay network is to simultaneously minimize the buffer-overflow and the buffer-underflow probabilities while maintaining a low signaling overhead.

This paper presents a simple analytical model to illustrate the buffer-overflow/-underflow problems of an LTE-A relay network. A dynamic flow control algorithm (DFCA) is then proposed to solve the aforementioned problems. The rest of the paper is organized as follows. Section II presents the system model for LTE-A relay network. An analytical model is presented to estimate the buffer-overflow/-underflow probabilities of flows in the RN if flow control is not applied. Section III describes the design concept and implementation details of the proposed DFCA. Section IV presents simulation results. Finally, Section V draws conclusions.

II. SYSTEM MODEL

This paper considers the downlink transmission of an LTE-A relay network consisting of a DeNB, \( N \) Type-I RNs, and \( M \) UEs as illustrated in Fig. 1. In this figure, \( R_N \) denotes the RNs served by the DeNB (\( 1 \leq i \leq N \)); \( U_{E_{i,j}} \) is the
This study uses the in-band relaying technique to allocate the RF channel for the relay links and access links. Fig. 2 shows the frame structure [20] and the time division duplex (TDD) subframe configurations [21] defined by LTE-A. In LTE-A relay networks, time is divided into fixed-length frames. Each frame consists of ns subframes (e.g., ns = 10 in Fig. 2) and each subframe contains multiple physical resource blocks (PRBs) [5]. The PRB is the minimum resource allocation unit. The number of bits that can be carried by a PRB depends on the chosen modulation and coding scheme (MCS). The DeNB may assign subframes to the relay links connecting DeNB and RNs, the access links connecting RNs and their serving UEs, and the access links connecting DeNB and UEs on a frame-by-frame basis. Let nR,i and nA,i be the number of subframe reserved by DeNB for the relay links and access links of RNi, respectively. The terms nR,i and nA,i can be determined based on the subframe configuration of DeNB. The upper part of Fig. 2 illustrates TDD subframe configuration number 4 and number 18. In the subframe configuration shown in this figure, DA, UA, DR, and UR represent the subframes reserved for downlink access link, uplink access link, downlink relay link, and uplink relay link, respectively. The portion of frame reserved by DeNB for the relay link and access links of RNi, denoted as αR,i and αA,i, respectively, can be derived from ns,nR,i, and nA,i, and is given by

\[ \alpha_{R,i} = \frac{n_{R,i}}{n_s}, \]

and

\[ \alpha_{A,i} = \frac{n_{A,i}}{n_s}. \]

For example, nR,i = 2 and nA,i = 1 in TDD subframe configuration number 4 and 18, which gives αR,i = 2/10 = 0.2 and αA,i = 1/10 = 0.1 respectively. Note that αA,i = 0.2 in both configurations due to nA,i = 2.

Let βi,j[k] be the portion of bandwidth in the relay link reserved by the DeNB for UEi,j at the kth subframe, where 0 ≤ βi,j[k] ≤ 1 and \[ \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_i,j[k] = 1 \] That is, the transmission rate of UEi,j allocated by DeNB is βi,j[k]li,j[k]. Note that βi,j[k] can be determined by either the DeNB or the RN and the value of βi,j[k] can be static or dynamic. The channel condition of the relay link between DeNB and RNi is normally stationary and can be assumed to be a constant [7]. In contrast, the channel condition of the access link between RNi and UEi,j changes dynamically due to the movement of UE. In LTE-A relay networks, DeNB does not know the statistics of UEi,j a priori. Therefore, the DeNB may assign a constant (same) value of βi,j[k] based on the limited information negotiated obtained at the connection establishment (e.g., the minimum bit rate of UEi,j). DeNB may also dynamically adjust βi,j[k] upon receiving the up-to-date statistics of UEi,j from RNi in a periodic or event-triggered manner. However, the price paid is the extra signaling overhead required to carry the statistics of UEi,j.

The following discussion presents an analytical model to investigate the performance of an LTE-A relay network adopting a fixed bandwidth allocation policy. That is, the DeNB always allocates the same portion of bandwidth to UEi,j at each subframe (βi,j[k] = βi,j). The analytical model is developed based on the general procedure proposed in [9]. In [9], the time-average dropping probability of a node with finite-length buffer is estimated by the ensemble-average dropping probability. Let Ai,j[k] and Di,j[k] be the number of arrival and departure packets at RNi for UEi,j at the kth subframe (unit: packet), which are given by

\[ A_{i,j}[k] = \beta_{i,j} l_{i,j}[k] (1 - PLR), \]

and

\[ D_{i,j}[k] = l_{i,j}[k] (1 - PLR), \]

where PLR is the packet loss rate [25]. For a packet of size L bits, PLR = 1 - (1 - BER)^L, where BER is the target bit error rate of the MCS [25].

Let \( P_{i,j}^{(O)} \) be the buffer-overflow probability at RNi for UEi,j. \( P_{i,j}^{(O)} \) is the percentage of arriving packets at RNi dropped because the buffer is full in an observation interval of T subframes. Consequently, \( P_{i,j}^{(O)} \) can be estimated by

\[ P_{i,j}^{(O)} = \lim_{T \to \infty} \max_{k} \left( 0, \frac{T}{\sum_{k=1}^{M} \frac{\alpha_{R,i} A_{i,j}[k]}{\alpha_{R,i} A_{i,j}[k]}} - \frac{K_{i,j}}{\sum_{k=1}^{M} \frac{\alpha_{R,i} A_{i,j}[k]}{\alpha_{R,i} A_{i,j}[k]}} \right) \]

\[ \geq \max \left( 0, \frac{\alpha_{R,i} E\{A_{i,j}\} - \alpha_{A,i} E\{D_{i,j}\}}{\alpha_{R,i} E\{A_{i,j}\}} \right), \]

where \( E\{A_{i,j}\} \) and \( E\{D_{i,j}\} \) are the average number of arrival and departure packets at RNi for UEi,j, respectively. In the steady state, we can have [9, Eq. (39)]

\[ E\{A_{i,j}\} = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} A_{i,j}[k] \]

\[ = \beta_{i,j} (1 - PLR) E\{l_{i,j}[k]\}, \]

and

\[ E\{D_{i,j}\} = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} D_{i,j}[k] \]

\[ = (1 - PLR) E\{l_{i,j}[k]\}, \]

where \( E\{l_{i,j}[k]\} \) is the average transmission rate of the relay link between DeNB and RNi, and \( E\{l_{i,j}[k]\} \) is the average transmission rate of the access link between RNi and UEi,j.

In LTE-A relay networks, the adaptive modulation and coding (AMC) scheme is employed [9], [22]. Let \( n_{\text{max}} \) be the maximum level of the MCS; Rn be the transmission rate of MCS level \( n \) (0 ≤ n ≤ \( n_{\text{max}} \)); and \( \gamma_n \) be the minimum SNR.
thresholds of the MCS level $n$ required to support a target bit error rate. Assume that $R_i$ can obtain immediate CQI feedback, $\gamma_{ij}$, from $UE_{ij}$ and uses MCS level $n$ to transmit data to $UE_{ij}$ at the next subframe if $\gamma_n \leq \gamma_{ij} \leq \gamma_{n+1}$. In [23, Eq. (2)], the authors showed that the probability density function (pdf) of the SNR $\gamma_{ij}$ for a constant-power variable multi-level quadrature amplitude modulation (QAM) scheme over Nakagami multipath fading channel model is given by

$$p_{\gamma}(\gamma_{ij}) = \frac{m^m \gamma_{ij}^{m-1}}{\Gamma(m,0)} \exp\left(-\frac{m\gamma_{ij}}{\gamma_{ij}}\right),$$

(8)

where $m$ is the Nakagami fading parameter ($m \geq 1/2$); $\gamma_{ij}$ is the average received SNR at $UE_{ij}$. The terms $\Gamma(m,x)$ is the complementary incomplete Gamma function, and is given by

$$\Gamma(m,x) = \int_x^\infty \tau^{m-1}e^{-\tau}d\tau.$$

(9)

The value of $E\{l_{ij}[k]\}$ in Eq. (7) can be derived by summing over the transmission rate of MCS level $n$ weighted by its probability [9]. That is,

$$E\{l_{ij}[k]\} = \sum_{n=0}^{n_{\text{max}}} R_n \int_{\gamma_n}^{\gamma_{n+1}} p_{\gamma}(\gamma_{ij}) d\gamma_{ij}$$

$$= \sum_{n=0}^{n_{\text{max}}} R_n \frac{\Gamma(m,m\gamma_{ij}/\gamma_{ij}) - \Gamma(m,m\gamma_{n+1}/\gamma_{ij})}{\Gamma(m,0)}.$$

(10)

The buffer-underflow probability at $R_i$ for $UE_{ij}$, $P_{ij}(U)$, is the percentage of packets that cannot be transmitted through the access link due to the lack of buffered packets of $UE_{ij}$ at $R_i$, during an observation interval of $T$ subframes. Therefore, $P_{ij}(U)$ is given by

$$P_{ij}(U) = \lim_{T \to \infty} \max_0 \left( \sum_{k=1}^{T} (\alpha_{A_i}D_{ij}[k] - \alpha_{R_i}A_{ij}[k]) \right)$$

$$= \max_0 \left( \frac{\alpha_{A_i}E_D[ij] - \alpha_{R_i}E_A[ij]}{\alpha_{A_i}E_D[ij]} \right).$$

(11)

Let $\mu_{ij}$ be the average throughput from the DeNB to $UE_{ij}$. $\mu_{ij}$ is the minimum of $UE_{ij}$'s transmission rates at the relay link and at the access link. That is,

$$\mu_{ij} = \min(\alpha_{R_i}E_A[ij], \alpha_{A_i}E_D[ij]).$$

(12)

Note that the effect of cell capacity and number of active users in the cell is modeled in terms of $l_{ij}[k]$ and $\beta_{ij}$, respectively. A higher cell capacity implies a higher value of $l_{ij}[k]$. Less active users in the cell implies a higher value of $\beta_{ij}$. It can be found from Eq. (3) that increasing either $\beta_{ij}$ or $l_{ij}[k]$ increases the number of arrival packets at $R_i$ for $UE_{ij}$ at the $k$th subframe, $A_{ij}[k]$. Additionally, based on Eqs. (5) and (11), we can also find that increasing either $\beta_{ij}$ or $l_{ij}[k]$ increases the buffer-overflow probability $P_{ij}(O)$ and decreases the buffer-underflow probability $P_{ij}(U)$.

The buffer-overflow problem may result in re-transmission of lost packets while the buffer-underflow problem may lead to a low utilization of $R_i$’s access link. The DeNB or the RN can adopt proper flow control policies to reduce the buffer-overflow and buffer-underflow probabilities and increase the throughput of the access link at the cost of extra signaling overhead.

A. Simulation Studies

Computer simulations were conducted on a C-based platform to verify the accuracy of the analysis. In the following figures, each point represented the mean value of 100 samples. Each sample was obtained by averaging data collected within 1000 seconds (or 100000 radio frames). A 95% confidence interval was shown on the figures and they fell in a quite narrow region. The simulation considered the downlink transmission of a two-hop Type-I RN ($N = 1$) serving 20 UEs ($M = 20$) in a 20MHz bandwidth LTE-A relay network under a Rayleigh fading channel ($m = 1$). It was assumed that the RN reserved equal buffer size for each UE and $K_{ij} = 30$ packets. The TDD subframe configuration numbers 4 (i.e., $\alpha_{A_i} = \alpha_{R_i} = 0.2$) and 18 (i.e., $\alpha_{A_i} = 0.2, \alpha_{R_i} = 0.1$) were investigated [21]. The signaling overhead required for carrying physical downlink common control channel (PDCCH) in the relay link was ignored.

The full-buffer traffic model defined in 3GPP TR 36.814 was used in the DeNB to evaluate the performance of the proposed flow control algorithm with continuous traffic and non-varying interference [5]. In full-buffer traffic model, packets are always available at the DeNB pending for transmission to all UEs served by the RN. Therefore, the performance of the flow control protocols can be fairly compared and may not be affected by the traffic model. The UE’s mobility model was not considered. We used stationary UEs with linearly increased average SNRs in order to demonstrate the rate-mismatch problem. The average SNR of $UE_{ij}$ was uniformly distributed within the range from 11 dB to 30 dB. That is, the average SNR of $UE_{ij}$, $\gamma_{ij}$, was set to $(10 + j)$ dB, $1 \leq j \leq 20$. The instantaneous SNR of $UE_{ij}$ varies due to Rayleigh fading. HARQ and AMC schemes were implemented in both relay link and access link. For simplicity, we used ‘packet’ to represent a transport block used in LTE-A networks. With HARQ, each transport block [5] is immediately re-transmitted up to five times if the transport block is not correctly received. With AMC, the DeNB and the RN adjust the MCS level based on the CQI reported by the receiver. The number of transport block that can be carried in a subframe depends on the chosen MCS level. A maximum MCS level of six (i.e., $n_{\text{max}} = 6$) was considered herein. The transmission rate of MCS level $n$, $R_n$, and the minimum SNR thresholds of the MCS level $n$, $\gamma_n$, to support a target bit error rate $BER = 10^{-6}$ was summarized in Table I [24].

For demonstration, the transport block size is chosen as the number of bits carried by 10 PRBs using MCS level 1 in a subframe. In LTE-A, one PRB contains 12 subcarriers multiplied by 7 OFDM symbols. A PRB using MCS level 1 can carry 84 bits (i.e., $7 \times 12$ symbols/PRB $\times 1$ bit/symbol),
which gives the transport block size (or, packet size $L$) of 840 bits, or 105 bytes. With 20 MHz bandwidth, the available capacities of the relay link and the access link were both equal to 200 PRBs/subframe [5]. It was assumed that the RN equally shares the capacity of the access link to all of the 20 UEs. Therefore, each UE has 10 PRBs/subframe at the access link. The highest MCS level was used at the relay link due to its excellent channel quality, which resulted in a constant data rate of $l_i[k] = E\{l_i[k]\} = 200$ PRBs/subframe $\times 84$ symbols/PRB $\times 4.5$ bits/symbol $= 9450$ bytes/subframe. The attainable data rate at the access link depends on the SNR of $UE_{i;j}$. That is, $l_{i;j}[k] = 10$ PRBs/subframe $\times 84$ symbols/PRB $\times R_n$ bits/symbol.

Figs. 3 to 5 show the simulation and analytical results for 20 UEs with average SNR ranging from 11 dB to 30 dB at the access link. Fig. 3 shows the buffer-overflow probability for UEs with different average SNR. Increases $\alpha_{R;i}$ resulted in a higher buffer-overflow probability due to the increased arrival data rate at the relay link. The buffer-overflow probability decreased for UEs with higher average SNR due to the increased departure rate at the access link. Fig. 4 shows the buffer-underflow probability for UEs with different average SNR. Increases in $\alpha_{R;i}$ may result in a lower buffer-underflow probability because it increases the arrival data rate at the relay link. The buffer-underflow probability is increased for UEs with higher average SNR due to the increased departure rate at the access link. Note that the high buffer-underflow probability occurs only for some UEs which had good signal qualities in the case of $\alpha_{R;i} = 0.1$. The high value of 0.5 in Fig. 4 resulted from the improper bandwidth allocation. For $\alpha_{R;i} = 0.1$, the bandwidth was not properly allocated since the bandwidth reserved for the relay link was only half of that reserved for the access link. Fig. 5 shows the throughput for UEs with different average SNR. Increases $\alpha_{R;i}$ resulted in a higher throughput because the arrival data rate at the relay link is increased. The throughput increased for UEs with higher average SNR due to the decreased buffer-overflow probability. Note that the arrival rate of $UE_{i;j}$ injected at the relay link exceeds the service rate of $UE_{i;j}$ at the access link for $\alpha_{R;i} = 0.2$. Hence, UEs with higher average SNR will have higher throughput. For $\alpha_{R;i} = 0.1$, the arrival rate of $UE_{i;j}$ injected at the relay link is lower than the service rate of $UE_{i;j}$ at the access link for $j \geq 6$. As a result, the throughput of these UEs (i.e., $UE_{i;j}$ for $j \geq 6$) are identical and does not increase as the average SNR increases. All simulation results agree with the analysis in Eqs. (5), (11), and (12), confirming the accuracy of the analysis.

III. DYNAMIC FLOW CONTROL ALGORITHM (DFCA)

The following discussion describes a dynamic flow control algorithm (DFCA) proposed for LTE-A relay networks. This study also investigates the performance indices of buffer-overflow probability, buffer-underflow probability, throughput, and signaling overhead of the proposed DFCA algorithm. The signaling overhead of the flow control algorithm is defined as the average number of signaling messages transmitted from the RN for each UE in one second. Each signaling message is carried by one packet.

The DFCA adopts the concept of the window-based flow control scheme to limit the number of packets transmitted from the DeNB to $RN_i$ for $UE_{i;j}$ at the relay link. That is, the sender (DeNB) can transmit a number of packets to a particular UE up to a UE-specific window size determined by the receiver ($RN_i$) before receiving an acknowledgment (ACK). Let $W_{i;j}[k]$ be the window size of $UE_{i;j}$ at the relay link determined at the $k$th subframe. $W_{i;j}[k]$ can be derived from the maximum number of packets designed to $UE_{i;j}$ that can be transmitted from the DeNB to the $RN_i$. In DFCA, $W_{i;j}[k]$ is dynamically adjusted and the $RN_i$ will send an ACK message carrying the new window size, $W_{i;j}[k]$, to the DeNB whenever it has forwarded the same amount of packets to $UE_{i;j}$ through the access link. In the implementation, the $RN_i$ maintains two counters $Q_{i;j}[k]$ and $S_{i;j}[k]$ similar to that used in CBFC [16] to count the number of buffered packets pending for transmission and the number of successfully transmitted packets to $UE_{i;j}$ at the $k$th subframe, respectively. In the following, we propose an analytical formula to determine $W_{i;j}[k]$.

The total buffer size reserved for $UE_{i;j}$ at $RN_i$ is generally equal to the number of buffered packets plus the number of successfully transmitted packets (i.e., $K_{i;j} = Q_{i;j}[k] + S_{i;j}[k]$). The buffer-overflow problem can be prevented if $Q_{i;j}[k] \leq K_{i;j}$. The buffer-underflow problem can be prevented if the $RN_i$ can response the ACK before running out of its buffer. In other words, the time required by $RN_i$ to transmit all of the buffered packets to $UE_{i;j}$ is no less than the uplink signaling delay plus the downlink packet transmission delay. That is,

$$Q_{i;j}[k] \leq t_u + t_d,$$

where $t_u$ is the uplink signaling delay (unit: subframe) from the $RN_i$ to the DeNB and $t_d$ is the downlink packet transmission delay (unit: subframe). The downlink packet transmission delay is the time required by the DeNB to transmit $W_{i;j}[k]$ packets of $UE_{i;j}$ to $RN_i$, which gives $t_d = W_{i;j}[k]/(\beta_{i;j}[k]l_i[k])$. Substitute $K_{i;j} = Q_{i;j}[k] + S_{i;j}[k]$, $t_d = W_{i;j}[k]/(\beta_{i;j}[k]l_i[k])$, and $S_{i;j}[k] = W_{i;j}[k]$, into Eq. (13), we can have

$$K_{i;j} - W_{i;j}[k]/l_i[k] - W_{i;j}[k]/(\beta_{i;j}[k]l_i[k]) \geq t_u.$$

The equation can be rewritten as

$$0 \leq W_{i;j}[k] \leq \frac{t_u l_i[k]}{K_{i;j} - W_{i;j}[k]/l_i[k] - W_{i;j}[k]/(\beta_{i;j}[k]l_i[k])}.$$

Hence, the maximum value of the window size in DFCA is given by

$$W_{i;j}[k] = \frac{l_i[k](K_{i;j} - t_u l_i[k]/l_i[k])}{l_i[k] + W_{i;j}[k]/(\beta_{i;j}[k]l_i[k])}.$$

In Eq. (16), $K_{i;j}$ and $l_i[k]$ are the buffer size and the data rate of $UE_{i;j}$, respectively; $l_i[k]$ is the data rate of the relay link; $t_u$ is the uplink signaling delay; $\beta_{i;j}[k]$ can be obtained directly from the observing the portion of packets transmitted by the DeNB at the relay link. Hence, all of these parameters in Eq. (16) are known by the RN. Note that the window size $W_{i;j}[k]$ obtained in Eq. (16) depends on the channel quality...
of UE$_{i,j}$. A higher $l_{i,j}[k]$ results in a smaller threshold. The maximum value of $W_{i,j}[k]$ occurs when $l_{i,j}[k] = 0$. In this case, $W_{i,j}[k] = K_{i,j}$ and no more packets can be forwarded form the DeNB to the RN unless all of the buffered packets are transmitted. Note that UE$_{i,j}$ may have a higher chance to handover to the other RN (or DeNB) under such bad channel quality. Hence, a higher window size at RN also helps to minimize the unnecessary packet forwarding and the signaling overhead. In contrast, a smaller window size is used if the channel quality is very good. In this case, the DFCA gives RN$_i$ a higher chance to request more packets from the DeNB before running out of its buffer. The operation of proposed DFCA is as follows:

Step 1 At the establishment of a new connection to UE$_{i,j}$, the DeNB sends $K_{i,j}$ packets designated to UE$_{i,j}$ to the RN$_i$. The RN$_i$ shall set $Q_{i,j}[0] = K_{i,j}$ and $S_{i,j}[0] = 0$.

Step 2 At RN$_i$, $Q_{i,j}[k]$ and $S_{i,j}[k]$ are dynamically updated. $Q_{i,j}[k]$ is increased by one whenever RN$_i$ receives a packets from DeNB. $S_{i,j}[k]$ is increased by one and $Q_{i,j}[k]$ is decreased by one whenever RN$_i$ successfully transmits a packet to UE$_{i,j}$.

Step 3 At the $k$th subframe, RN$_i$ calculate the window size $W_{i,j}[k]$ according to Eq. (16).

Step 4 RN$_i$ transmits an ACK to DeNB carrying $W_{i,j}[k]$ if $S_{i,j}[k - 1]$ exceeds $W_{i,j}[k]$ (i.e., $S_{i,j}[k - 1] \geq W_{i,j}[k]$). RN$_i$ resets $S_{i,j}[k]$ (i.e., let $S_{i,j}[k] = S_{i,j}[k - 1] - W_{i,j}[k]$) after transmitting the ACK.

Step 5 Upon receiving the ACK, DeNB can send extra $W_{i,j}[k]$ packets to UE$_{i,j}$.

Repeat Steps 2 to 5 until the UE leaves the coverage area of RN$_i$.

Note that the design concept of DFCA is quite different than that of CBFC although DFCA adopts a similar counter-based implementation approach as CBFC does. However, CBFC adopts a constant feedback frequency and the receiver periodically calculates and feedbacks the credits to the sender [16]. In addition, CBFC lacks a general rule to optimize the setting of the two counters (i.e., $N2$ and $N4$ [16]). In contrast, DFCA dynamically adjusts the feedback frequency and the window size based on the buffer size, the ingress rate and the egress rate of RN, and the uplink signaling delay and downlink packet transmission delay between RN and DeNB. DFCA also provides different reporting intervals for UEs with different signal qualities to further reduce the amount of feedback messages.

IV. SIMULATION RESULTS

In the following, the effectiveness of proposed DFCA was investigated. The same simulation scenarios specified in the first three paragraphs of Sec. II-A were considered. The performance metrics of signaling overhead, buffer-overflow probability, buffer-underflow probability, and throughput were investigated. A queue-based flow control algorithm (QFCA), a rate-based flow control algorithm (RFCA), a credit-based flow control (CBFC) [16], and a congestion control and flow control (CCFC) [18] are chosen as benchmark flow control algorithms for evaluating the effectiveness of the proposed DFCA. In all simulations, two reporting interval of $T_R = 1$ and 8 frames were considered.

In the simulations, it was assumed that DeNB utilized a weighted-fair-queueing (WFQ) algorithm to determine the value of $\beta_{i,j}[k]$ for scheduling the packet transmission at the relay link. Let $U_{i,j}[k]$ be the number of packets pending for transmission (i.e., the window size minus the number of packets transmitted after receiving the ACK) at DeNB to UE$_{i,j}$ at the $k$th subframe. Note that $U_{i,j}[k] = 0$ if the DeNB has transmitted packets up to the window size but has not received ACK from the RN. If the DeNB receives a new window size $W_{i,j}[k]$ before transmitting all of the pending packets (i.e., $U_{i,j}[k] \neq 0$), $U_{i,j}[k]$ will be updated by

$$U_{i,j}[k] = U_{i,j}[k - 1] + W_{i,j}[k].$$ (17)

In DFCA, $\beta_{i,j}[k]$ is assigned in proportion to $U_{i,j}[k]$ in each subframe with duration $T_s$. That is,

$$\beta_{i,j}[k] = \begin{cases} \frac{U_{i,j}[k]}{M} & \text{if } \sum_{j=1}^{M} U_{i,j}[k] > l_i[k]T_s, \\ \sum_{i=1}^{M} \frac{U_{i,j}[k]}{U_{i,j}[k]} & \text{if } \sum_{j=1}^{M} U_{i,j}[k] \leq l_i[k]T_s. \end{cases}$$ (18)

In QFCA, $\beta_{i,j}[k]$ is assigned in proportion to the available queue size of UE$_{i,j}$, $B_{i,j}[k]$, if the available queue size of all UEs in RN$_i$ is higher than the number of packets carried by the relay link in a subframe. Otherwise, $\beta_{i,j}[k]$ is also assigned based on available queue size of UE$_{i,j}$. Hence, $\beta_{i,j}[k]$ is given by

$$\beta_{i,j}[k] = \begin{cases} \frac{B_{i,j}[k]}{M} & \text{if } \sum_{j=1}^{M} B_{i,j}[k] > l_i[k]T_s, \\ \sum_{i=1}^{M} \frac{B_{i,j}[k]}{B_{i,j}[k]} & \text{if } \sum_{j=1}^{M} B_{i,j}[k] \leq l_i[k]T_s. \end{cases}$$ (19)

where $B_{i,j}[k] = K_{i,j} - Q_{i,j}[k]$ is the available queue size for UE$_{i,j}$ at subframe $k$.

In RFCA, $\beta_{i,j}[k]$ is assigned in proportion to the service rate of the access link (i.e., $l_{i,j}[k]$) if the aggregative service rates of all UEs in RN$_i$ is higher than the maximum rate of the relay link. Otherwise, $\beta_{i,j}[k]$ is assigned based on the service rate of the access link. Hence, $\beta_{i,j}[k]$ is given by

$$\beta_{i,j}[k] = \begin{cases} \frac{l_{i,j}[k]}{M} & \text{if } \sum_{j=1}^{M} l_{i,j}[k] > l_i[k], \\ \sum_{i=1}^{M} \frac{l_{i,j}[k]}{l_{i,j}[k]} & \text{if } \sum_{j=1}^{M} l_{i,j}[k] \leq l_i[k]. \end{cases}$$ (20)

In CBFC, $\beta_{i,j}[k]$ is assigned in proportion to the number of forwarded packets for UE$_{i,j}$ (i.e., $f_{i,j}$ in [16]) if the number of forwarded packets for all UEs in RN$_i$ is higher than the number of packets carried by the relay link in a subframe. Otherwise, $\beta_{i,j}[k]$ is also assigned based on the number of
forwarded packets for \( UE_{i,j} \).

\[
\beta_{i,j}[k] = \begin{cases} 
    \frac{f_{i,j}[k]}{\sum_{j=1}^{M} f_{i,j}[k]}, & \text{if } \sum_{j=1}^{M} f_{i,j}[k] > l_{i}[k]T, \\
    \frac{f_{i,j}[k]}{l_{i}[k]T}, & \text{otherwise},
\end{cases} \tag{21}
\]

where \( f_{i,j}[k] \) is the number of forwarded packets from \( RN_{i} \) to \( UE_{i,j} \) at subframe \( k \).

In CCFC, the data rate was additive-increased and multiplicative-decreased based on the predefined rule, which cannot be changed by the DeNB.

For RFCA and QFCA, the reporting interval was \( T_R \). For CBFC, \( N_2 = T_R \) and \( N_4 = 1 \) were considered. For CCFC, the flow control cycle time was 1 frame and the congestion control cycle time was \( T_R \) frames. Note that the DeNB can obtain up-to-date statistics if \( T_R = 1 \) but at the cost of higher signaling overhead. For DFCA, the RN reports its statistics only if \( S_{i,j}[k-1] \geq W_{i,j}[k] \) and is independent of \( T_R \).

Note that the RN transmits one signaling message (unit: packet) in each reporting interval in QFCA and RFCA and transmits one signaling message whenever the event-trigger condition is met in DFCA, CBFC, and CCFC.

Figs. 6 to 12 show the performance of five flow control algorithms. For comparison, the LTE-A relay network without flow control mechanism was only illustrated in Fig. 6. Figs. 6, 7, 8, and 9 show the buffer-overflow probabilities, the buffer-underflow probabilities, the throughput from the DeNB to the UE, and the signaling overhead of the five flow control algorithms for \( \alpha_{R,i} = 0.2 \) and UEs with different SNRs, respectively. Fig. 6 shows that RFCA and QFCA resulted in non-zero buffer-overflow probabilities if a long reporting interval \( (T_R = 8) \) was used. Therefore, the two cases were not illustrated in the following figures. RFCA with \( T_R = 1 \), QFCA with \( T_R = 1 \), DFCA, CBFC, and CCFC achieved zero buffer-overflow probability and thus, solved the buffer-overflow problem. Fig. 7 shows that the buffer-underflow probability of CBFC with \( T_R = 8 \) increased for UEs with higher average SNRs. It is because the data rates of access link reserved for these UEs were all higher than the credit replenishment rate of the RN due to a long reporting interval. RFCA, QFCA, and CBFC with \( T_R = 1 \), and DFCA had zero buffer-underflow probability. Fig. 8 shows the throughput from the DeNB to the UE was increased as the average SNR increased for all flow control algorithms. However, CBFC with \( T_R = 8 \) saturated for UEs with higher SNRs. CCFC had lower throughput because they had higher buffer-underflow probabilities. The throughput of CCFC increased as the reporting interval decreased. Note that increments of throughput for CCFC with \( T_R = 1 \) and \( T_R = 8 \) were similar because they used the same setting of additive-increase/multiplicative-decrease as in [18]. CBFC with \( T_R = 8 \) saturated faster than the others and was bounded by the credit replenishment rate of the RN. DFCA, RFCA, and QFCA achieved the highest throughput for all UEs. Fig. 9 shows the signaling overhead of RFCA, QFCA, CBFC, and CCFC with \( T_R = 1 \) was constant and depended only on \( T_R \). The signaling overhead of RFCA, QFCA, CBFC, and CCFC with \( T_R = 1 \) could be five to twenty times than that of DFCA. The signaling overhead of CCFC with \( T_R = 8 \) increased as the average SNR of UE increased. It is because UEs with higher average SNR experience less congestion than the other UEs. The signaling overhead of CCFC with \( T_R = 8 \) for UEs with lower average SNRs was lower and was almost a constant because a long reporting interval was used during congestion. As expected, the signaling overhead of DFCA increased as the average SNR of UE increased. Note that the signaling overhead of DFCA was slightly higher than that of CBFC with \( T_R = 8 \). However, the throughput CBFC with \( T_R = 8 \) was much lower than that of DFCA.

Figs. 10, 11, and 12 show the buffer-underflow probabilities, the throughput from the DeNB to the UE, and the signaling overhead of the five flow control algorithms for \( \alpha_{R,i} = 0.1 \) and UEs with different SNRs, respectively. The buffer-overflow probabilities of the five flow control algorithms were not demonstrated because they were all zero. For \( \alpha_{R,i} = 0.1 \), the aggregated service rates of all UEs in \( RN_k \) is higher than the maximum arrival rate of the relay link. Hence, the buffer-underflow probabilities depended on the bandwidth allocation policy. In CBFC and CCFC, the DeNB equally shared the bandwidth for flows. Fig. 10 show RFCA provided UEs with similar buffer-underflow probabilities since the DeNB allocated the bandwidth according to the attainable data rate of each UE. The buffer-underflow probabilities of QFCA increased as the average SNRs increased because the bandwidth was assigned in proportion to the available queue size. CBFC had very low buffer-underflow probabilities for UEs with low average SNRs but had very high buffer-underflow probabilities for UEs with high average SNRs. Fig. 11 show the throughput of all flow control algorithms increased as the average SNRs increased. However, CBFC and CCFC got saturated for UEs with high average SNR. Note that the total throughput of all UEs in DFCA, QFCA, and RFCA was similar, which was higher than that of CBFC and CCFC. It is because that CBFC and CCFC equally allocated bandwidth to all UEs and thus, UE with the highest average SNR spent the longest time in idle. Fig. 12 show CCFC with \( T_R = 8 \) and DFCA had the lowest signaling overhead. However, CCFC with \( T_R = 8 \) achieved a lower throughput than DFCA did. The signaling overhead of CCFC with \( T_R = 8 \) decreased for UEs with SNR ranging from 11 to 15 dB. For these UEs, their data rates were lower and thus, resulted in congestion at the RN. The RN adopted a longer reporting interval during congestion state, which decreased the signaling overhead. The signaling overhead of CCFC with \( T_R = 8 \) increased as the average SNR increased because congestion relieved as the data rate of UE increased.

V. Conclusion

This study investigates the buffer-overflow and buffer-underflow problems results from a mismatch of the data rates of the access link and relay link in downlink of LTE-A relay networks. First, we presented an analytical model to estimate the buffer-overflow and buffer-underflow probabilities for the
LTE-A relay networks without flow control functionality. We further proposed a DFCA to solve the buffer-overflow and buffer-underflow problems. Computer simulations were conducted to verify the accuracy of the analysis and the effectiveness of the proposed DFCA algorithm. Simulation results show that the proposed analytical model can accurately estimate the buffer-overflow and buffer-underflow probabilities. The results also show that DFCA can solve both the buffer-overflow and buffer-underflow problems simultaneously with reduced signaling overhead. In addition, DFCA can still work well even if the bandwidth is not properly allocated.

REFERENCES


TABLE I

<table>
<thead>
<tr>
<th>level</th>
<th>MCS</th>
<th>$R_{th}$ (bits/symbol)</th>
<th>$\gamma_n$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Silent</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>QPSK(1/2)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>QPSK(3/4)</td>
<td>1.5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>16QAM(1/2)</td>
<td>2</td>
<td>11.5</td>
</tr>
<tr>
<td>4</td>
<td>16QAM(3/4)</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>64QAM(2/3)</td>
<td>4</td>
<td>18.5</td>
</tr>
<tr>
<td>6</td>
<td>64QAM(3/4)</td>
<td>4.5</td>
<td>21</td>
</tr>
</tbody>
</table>

Ping-Chen Lin (S’09) was born in Tainan, Taiwan, in 1984. He received the B.E. and M.E. degrees in electronic and computer engineering, in 2006 and 2008, respectively, from National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan, where he is currently working toward the Ph.D. degree with the department of electronic and computer engineering. His current research interests include multi-hop wireless networks and machine-to-machine communications.

Yu-Jen Chang was born in Chunghua, Taiwan, in 1987. He received the B.S. degrees in electrical engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1991, 1993, and 1996, respectively.

He is currently a Professor with the Department of Electronic and Computer Engineering, NTUST, Taiwan. His research interests include multi-hop wireless networks and machine-to-machine communications.
Fig. 1. System model for LTE-Advanced relay networks.

Fig. 2. TDD frame structure for LTE-Advanced relay networks [20].

Fig. 3. The buffer-overflow probability of no flow control.

Fig. 4. The buffer-underflow probability of no flow control.

Fig. 5. The throughput of no flow control.

Fig. 6. The buffer-overflow probability of flow control algorithms ($\alpha_{R,t} = 0.2$).
Fig. 7. The buffer-underflow probability of flow control algorithms ($\alpha_{R,i} = 0.2$).

Fig. 8. The throughput of flow control algorithms ($\alpha_{R,i} = 0.2$).

Fig. 9. The signaling overhead of flow control algorithms ($\alpha_{R,i} = 0.2$).

Fig. 10. The buffer-underflow probability of flow control algorithms ($\alpha_{R,i} = 0.1$).

Fig. 11. The throughput of flow control algorithms ($\alpha_{R,i} = 0.1$).

Fig. 12. The signaling overhead of flow control algorithms ($\alpha_{R,i} = 0.1$).