Performance Analysis of Frame Aggregation under Unsaturated Conditions

Li Zhijie, Fang Xuming
Key Laboratory of Information Coding & Transmission, Southwest Jiaotong University, Chengdu 610031, Sichuan Province, P.R.China

Abstract: This paper proposes an analysis model of frame aggregation in error-free channel with unsaturated traffic and fixed aggregation size. Integrated with model of channel access, calculation of MAC (Media Access Control) average service time and queue model of frame aggregation, our model can get the stable result with a recursive algorithm, and it further derive the throughput and latency of frame aggregation in steady state. As the impact of traffic, frame length, collision probability, buffer size, aggregation size and interactive effects are taken into consideration, the effect of every parameter could be evaluated and the major factor which degrades the performance of frame aggregation can be determined in different situation with this model. By the simulation and numerical analysis, this model confirmed its accuracy. The proposed model can be used in the design, optimization and deployment of WLAN (Wireless Local Area Network) and WMN (Wireless Mesh Network) widely.

I. INTRODUCTION

IEEE 802.11 technology has made great progress with the wide deployment of wireless LAN and wireless mesh network, 802.11n PHY (Physical) layer data rate has reached 600Mbps recently [1-2]. With the increase of data rate, the overload, which is not reduced due to the increase of data rate such as channel access delay and PHY preamble, becomes the main factors affecting the MAC (Media Access Control) layer throughput efficiency. Research shows the MAC throughput efficiency is only about 10% when the data rate reaches 532Mbps [3-4]. With the adopting of frame aggregation technology which joins multiple data frames into a single aggregation frame to transmit, the average load per sub-frame could be well reduced, and thereby the efficiency of MAC layer is greatly improved.

Impact of frame aggregation on the throughput and delay of MAC layer are dependent on the network, traffic and other parameters. Under saturated traffic and error-free channel, MAC throughput increases a-
long with the increment of aggregation size, i.e. sub-frames number in the aggregation frame [3-4]. Compared with non-aggregation mode, frame aggregation mode starts transmitting only when enough frames are aggregated. Thus for unsaturated traffic, besides channel conditions and contention access mechanism, the effects of the aggregation size, buffers number and traffic characteristics shall be considered. The queuing effect in MAC layer buffers influences not only queuing delay and blocking probability, but also the waiting time for aggregation. Therefore, the performance of frame aggregation under unsaturated traffic must take all the effects into consideration, including network parameters such as channel access mechanism, number of nodes and collision probability, and node parameters such as service delay and blocking probability of buffering queue, aggregation waiting time, as well as interaction between network parameters and node parameters. Meanwhile the performance analysis must cover throughput and delay. It will be difficult to realize an accurate and comprehensive analysis of frame aggregation performance under unsaturated traffic if any of the above-mentioned parameters is over-simplified or only the throughput is considered.

Bianchi model in Ref. [5] describes correctly the relationship between channel contention and throughput, and unsaturated performance analysis is mostly based on this model or its extension. To present the situation of no frame available in MAC layer, Ref. [6] introduces post-backoff states while Ref. [7] only uses an idle state. However these two models all simplify the traffic arrival and queuing service in MAC into a single probability parameter, which is only related to the arrival rate in contention access process. Therefore they can not fully reflect the impacts of traffic arrival and queuing on the service process of MAC frame. They can only be used to analyze non-aggregated throughput, and the delay performance obtained is inaccurate. In Ref. [8] a buffer with empty state indicating is used, and probability of this state is obtained by directly using the queue occupation rate of M/M/1/K model, without considering the practical queuing effect. In addition, the simulation shows the probability of frame service time does not entirely obey Poisson distribution [9,11]. These factors have affected the accuracy of the model. Using the transfer function in transform domain of Bianchi model, Ref. [9] analyzes the result of the queue to obtain the probability generation function of frame service time. This approach considers impact of channel access on the queue service, but the assumption saying that every node would be involved in channel contention at any time, which is used in service time calculation, is inaccurate. Thus it is not applied to frame aggregation. In short, these methods are not taken into account the impact of frame aggregation, especially of aggregation waiting time.

Ref. [10], based on Bianchi model, analyzes the influence of channel on the frame aggregation throughput, which indicates that optimal aggregation size exists in case where the channel error is. In Ref. [11], frame aggregation is modeled as a batch service queue model, which analyzes the effect of aggregation size on delay and channel occupation. However the probability distribution of frame service time used is an approximate result based on finite simulation. To make it worse, queue analysis only takes into account the impact of arrival rate and frame length, while ignores the effect of channel access collision and its interactions with arrival rate on performance. These greatly influence the accuracy and application scope of this model. Analysis of frame aggregation performance in Ref. [12] is based on batch service model with a fixed Transmission Opportunity (TXOP) and the maximum aggregation size. It is only suit for specific scenes because of these limitations.

In this paper, under the conditions of ideal channel, unsaturated traffic and limited buffer size, considering the effect of channel access collision and node queue status, we propose a performance analysis model of fixed-sized frame aggregation and analyze the resulted throughput and delay. The rest of this paper is organized as follows. Section II establishes a complete performance analysis model. Section III provides analysis and simulation
II. PERFORMANCE ANALYSIS MODEL

2.1 Analysis model for aggregation frame transmission probability

Assume a WLAN with N nodes which are in the same collision area, and the channel between any two nodes is error-free, that means no hidden terminal is existed and frame transmission failure is caused only by the conflict. Arrival rate of frame with length of L at each node is $\lambda$. The buffer can store up to K frames. Frame aggregation size is m, that means when the frames number in the buffer $X(t)$ is less than m, transmission can not be started until enough frames arrived.

A new MAC state of E, which indicates no enough frames is available in buffer to satisfy the requirements for aggregation size, is added into the two-dimensional discrete Markov chain of [13], as shown in Figure 1. Here, n is the maximum backoff level, i.e. the maximum number of retransmission; q denotes the probability of contention access process when enough frames exists in buffer right after the last time of aggregation frame transmission. Transmission completion contains 2 cases - either frame transmitted successfully or frame dropped when the number of retransmission reaches the limitation. $p_a$ is the transition probability from state E to state $(0,k)$. $p$ is the collision probability of one node which is a constant independent of current times of retransmission [13]. $CW_{min}$ is the maximum Contention Window (CW) size in initial stage, $CW_{max}$ is the maximum CW allowed. $n'$ is defined as the backoff levels from $CW_{min}$ to $CW_{max}$, i.e. $2^{n'} \cdot (CW_{min} + 1) = CW_{max} + 1$. In general $n' \leq n$.

Denote the maximum CW in backoff stage i by $CW_i$, and $W_i = CW_i + 1$, $W_0 = CW_{min} + 1$, then,

$$W_i = \begin{cases} 2^i \cdot W_0 & 0 < i \leq n' \\ 2^{n'} \cdot W_0 & n' < i \leq n \end{cases}$$

(1)
b_{i,k} denote the steady-state probability of backoff state (i,k). \(i \in \{0,n\}\) is backoff level. \(k \in \{0,W_i-1\}\) is CW size in stage i. \(b_i\) is the steady-state probability of state E, i.e. \(b_i = \lim_{t \to \infty} P\{X(t) < m\}\). From it, there are

\[b_{i,0} = p^i \cdot b_{0,0}\]

(2)

\[b_i \cdot p_a = (\sum_{i=0}^{n-1} (1-p) \cdot p_{i,0} + p_{a,0}) \cdot (1-q)\]

(3)

\[b_{0,k} = \left[ b_i \cdot p_a + \left( \sum_{i=0}^{n-1} (1-p) p_{i,0} + p_{a,0}\right) \cdot q \right] \cdot \frac{(W_0 - k)}{W_0}\]

(4)

\[= b_{0,0} \cdot \frac{(W_0 - k)}{W_0}\]

Therefore, there is

\[b_{i,k} = p \cdot b_{i-1,0} \cdot \frac{(W_i - k)}{W_i} = p^i \cdot b_{0,0} \cdot \frac{(W_i - k)}{W_i} \quad 0 < i \leq n\]

(5)

Because \(\sum_{i=0}^{n} \sum_{k=0}^{n-1} b_{i,k} = 1\), then

\[b_{0,k} = \begin{cases} \frac{W_0 (1 - (2p)^{n+1}) (1-p) + (1-2p)(1-p^{n+1})}{2(1-b_i)(1-p)} & n \leq n \vspace{2mm} \\ \frac{W_0 (1 - (2p)^{n+1}) (1-p) + (1-2p)(1-p^{n+1}) + W_0 2^{n} \cdot p^{n+1} (1-2p)(1-p^{n+1})}{2(1-b_i)(1-p)} & n > n \end{cases}\]

(7)

\(\tau\) denote the probability of aggregation frame transmission in any time slot, then

\[\{\tau = \sum_{i=0}^{n} b_{i,0} = b_{0,0} \cdot (1-p^{n+1})/(1-p)\]

\[p = 1 - (1-\tau)^{n+1}\]

(8)

Because N is determined through calculating the nonlinear equations of (7) and (8), collision probability \(p\) of a single node and transmitting probability \(\tau\) can be obtained when \(b_i\) is determined. When \(b_i = 0\), the result is consistence with the model of saturated traffic without state of E in Ref. [13].

2.2 Analysis of MAC average service time

MAC layer service time \(T\) is defined as the duration from the start of channel contention to the completion of aggregation frame transmission, and it is a random variable.

Every transmission is divided into two phases—channel access and channel transmission. Let \(T_s\) denote average time for channel access, \(T_x\) denote the average time for channel transmission, then the average time for one transmission is

\[T_x = T_s + T_x\]

(9)

As the effective payload length \(m \cdot L\) of aggregation frame is relatively long, it is assumed that channel access always uses RTS/CTS access mechanism. Then the channel occupancy time \(t_s\) and the conflict time \(t_c\) of one transmission can be expressed as [5, 10]

\[t_s = RTS + SIFS + ACK + DIFS\]

\[t_c = RTS + CTS + DATA + ACK + 3 \cdot SIFS + DIFS\]

(10)

where, RTS/CTS/ACK are the channel occupancy time of different control frames respectively. DATA is the occupancy time of aggregated data frame. Thus

\[T_x = (1-p) \cdot t_s + p \cdot t_c\]

(11)

Because \(p\) is the collision probability obtained at a single node, let \(p_{e,N-1}\) and \(p_{s,N-1}\) denote the probabilities of collision and successful transmission of other \(N-1\) nodes respectively. There are

\[p_{e,N-1} = (N-1) \cdot \tau \cdot (1-\tau)^{N-2}\]

\[p_{s,N-1} = p - p_{e,N-1}\]

(12)

Average time for one step of backoff is

\[t_{b_{i,k}} = (1-p) \cdot \sigma + p_{d,N-1} \cdot t_s + p_{d,N-1} \cdot t_c\]

(13)

where, \(\sigma\) is the time for physical time slot. Because the average CW size at the stage i is \(W_i/2\), and the backoff process goes on to the next stage with the
2.3 Queuing model and analysis of aggregation

Frames in buffer at the time of \( t_n \). The state space is \( \theta = \{0, 1, \cdots, K\} \), and the state transition matrix is

\[
P = \begin{pmatrix}
0 & \cdots & \cdots & \cdots & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & \cdots & \cdots & \cdots \\
1 & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}
\]

By given the average service time \( T \), we define the traffic density as \( \rho = \lambda \cdot T \). Let \( a_k \) be the probability of \( k \) frames arrived in \( T \), then

\[
a_k = e^{-\rho} \cdot \frac{\rho^k}{k!} \quad k = 0, 1, \cdots
\]

For frame aggregation transmission, \( m \) subframes are sent out in every \( T \). From (17), \( P \) can be expressed as

\[
P = \begin{pmatrix}
a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots & a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots \\
a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots & 1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}
\]

Define the vector of steady-state probability distribution \( \Pi = \left[ \pi_0, \pi_1, \cdots, \pi_K \right] \), then

\[
\begin{cases}
\Pi = \Pi \cdot P \\
\sum_{i=0}^{K} \pi_i = 1
\end{cases}
\]

From above-listed equation, the steady-state probability \( \pi_i \) can be obtained.

Let \( P_i (i \in \{0, K\}) \) be the probability, which the number of frames is \( i \) at the arrival time of frames. Due to the PASTA property of Poisson process \([14]\), \( P \) is also the probability, which the number of frames is \( i \) at any time \( t \). So, with the quantity of frames is less than \( m \) at any time \( t \), the probability is

\[
b_i = \sum_{i=0}^{m-1} P_i
\]

Let \( p_b \) be the probability that newly arrived frames are rejected at any time \( t \), i.e., blocking probability, there

\[
p_b = 1 - \sum_{i=0}^{K} P_i
\]

For \( P_i \), there is \([14]\):

\[
P_i = P\{X = i\} = \pi_i \cdot (1 - p_b)
\]

When the system is stable, the arrival rate of

\[
T_a = \left( \sum_{i=0}^{n} p_i \cdot W_i / 2 \right) / \left( \sum_{i=0}^{n} p_i \right) \cdot t_{bo}
\]

In one transmission of aggregation frame, the probability of successful transmission is \( 1 - p \). The probability of collision is \( p \). Frame is dropped when the count of collision reaches the maximum number of retransmission. Thus the probability of frame loss is

\[
p_d = \frac{p \cdot b_{\|\|}}{\left( \sum_{i=0}^{n} b_{\|\|} \right)} = \left( 1 - p \right)^2 \cdot p \cdot \frac{1}{(1 - p + p_d)} \cdot \tau
\]

Therefore, \( m \) data subframes are consumed with the probability of \( 1 - p + p_d \) in one transmission of aggregation frame. The average service time of aggregation frame is

\[
T = T_a / (1 - p + p_d)
\]

2.3 Queuing model and analysis of aggregation frame

Now analyze the queuing model of buffer. Assume the frame arrival process obeys Poisson distribution, and the change of frame quantity in the system only occurs when aggregation frame service is finished \([14]\). Let \( t_1 , t_2 , \cdots \) denote serial of time values when aggregation frame services have been finished. \( n \) is the number of frames arriving at the number \( n \) service time \( (t_n+1 - t_n) \), then the number of frames in buffer can be expressed as

\[
X(t_n+1) = \begin{cases}
X(t_n) + \alpha & X(t_n) < m \\
X(t_n) & X(t_n) \geq m
\end{cases}
\]

Because the service time for current aggregation frame is independent of the previous service times, the service time is also independent of the number of frames in buffer, and furthermore because of the memoryless nature of Poisson process, \( X(t_{n+1}) \) only depends on the length of \( t_{n+1} - t_n \) and \( X(t_n) \) and has nothing to do with the value of \( t_n \) and the number of frames in buffer at time before \( t_n \). The discrete time random process \( X = (X_0 = 0, X_n = X(t_n), n \in \{1, 2, \cdots\}) \) is a Markov chain, which is consisted of quantity of frames in buffer at the time when services are finished. Their states are the number of

\[
\pi_i \in \{0, 1, \cdots, K\}
\]

probability of \( p \), thus

\[
T_a = \left( \sum_{i=0}^{n} p_i \cdot W_i / 2 \right) / \left( \sum_{i=0}^{n} p_i \right) \cdot t_{bo}
\]
frame being accepted equals to the service departure rate of frames, so
\[
\lambda \cdot (1 - p_e) = (1 - b_e) / T = (1 - \sum_{i=0}^{m-1} P_i) / T
\]  
\(24\)

Eqs. (22)-(24) can be transformed to
\[
\begin{cases}
P_i = \pi_i / (\rho + \sum_{i=0}^{m-1} \pi_i) & \text{0} \leq i \leq K \\
p_e = 1 - 1 / (\rho + \sum_{i=0}^{m-1} \pi_i)
\end{cases}
\]  
\(25\)

2.4 Modeling algorithm
Under saturated condition, frames are always available to be transmitted, that means \(b_e = 0\). In this situation, \(p\), \(T\) and \(P\) can be calculated, then \(p_e\) and other results of a stable system can be obtained easily. However, under unsaturated condition, \(b_e\) is unknown, and it determines \(p\) and \(T\). However \(b_e\) also depends on \(p\) and \(T\), so the stable result can not be obtained with such simple calculation.

If conditions of the network and traffic have been determined, parameters of \(N\), \(\lambda\), \(L\), \(K\), \(m\) will have constant values. After a long run, the system can always reach a stable state, \(b_e\), \(p\) and \(T\) get stable results respectively. Therefore when parameters are determined, let \(b_e\) has larger value, \(p\) and \(T\) be smaller, then the newly obtained \(b_e\) is smaller when using the new \(T\). Let \(b_e\) be the smaller value while \(p\) and \(T\) be the larger ones, then the newly obtained \(b_e\) will be the larger. This trend for \(b_e\), \(p\) and \(T\) makes it possible that a convergent, stable result can always be obtained after a serial of recursive calculations. The convergence process of recursive calculation also reflects the stability process of the system. Therefore, after parameters are determined, the above-mader analysis can always result in a stable and unique result with the recursive calculation as follows

1) Initialization. Set the stable threshold of \(b_e\):
\[\xi < 1;\] index of calculation step: \(i = 0\), \(b_e(i) = 0.9999\), which means the initial queue is empty;

2) Solve equation (7) and (8) with \(b_e(i)\), \(p\) and \(\tau\) are obtained;

3) Solve equation (9 ~ 16) to obtain the average service time \(T\);

4) And solve equation (20) to obtain \(I\). Calculate the new \(b_e\), i.e. \(b_e(i + 1)\), by using (21);

5) If \(|b_e(i + 1) - b_e(i)| \leq \xi\), the process of recursive calculation ends with a stable results, i.e., \(b_e = b_e(i + 1)\); otherwise, \(i = i + 1\), and return to 2).

The threshold \(\xi\) can be set to a smaller value (such as 1e-6), the stable and accurate result of \(p\), \(b_e\) and others can be obtained with the above-made recursive calculation.

2.5 Performance analysis
Frame loss can be caused when the buffer is full, or the retransmission reaches the limited times. Effective payload of aggregation frame is \(m \cdot L\). Effective service rate of queue is \((1 - b_e) / T\), and it accurately reflects the frame drop caused by the limited buffer size. For the frames leaving the buffer, the probability of dropping frame because of retransmission is \(p_d / (1 - p + p_d)\). So the throughput of node \(i\) is expressed as
\[Th_i = (1 - b_e) \cdot m \cdot L \cdot [1 - p_d / (1 - p + p_d)] / T\]  
\(26\)

Based on PHY bandwidth \(BW\), the normalized system throughput \(\text{Th}\) is
\[\text{Th} = N \cdot Th_i / BW\]
\(27\)

Let \(Q\) denote the average queue length of node, i.e. the average number of frames in buffers is
\[\bar{Q} = \sum_{i=0}^{m} i \cdot P_i\]  
\(27\)

According to Little formula [15], the average waiting time for aggregation, i.e. the average delay \(\bar{D}\) is
\[\bar{D} = \bar{Q} / \lambda \cdot (1 - p_\lambda)\]  
\(28\)

If \(m = 1\), the obtained results are the performances of non-aggregation.

III. VALIDATION AND PERFORMANCE EVALUATION

Firstly, the model is validated with the network simulator NS2 [16]. Let \(BW = 4\) Mbps, \(W_o =
16, n = 4, n = 6, and other parameters used in the analysis and simulation can be found in Ref. [2]. Nodes are distributed in the 10 × 10m area randomly with transmitting power that is set to cover the transmission range of 250m, and static routing is used. So all the nodes belong to the same collision domain and no hidden terminal being existed. Frames arrive as Variable Bit Rate (VBR) traffic. Simulation time is 50 seconds, each simulation run 3 times with different random seed respectively.

When L = 1024 bytes and K = 50, the throughput and delay obtained via analysis and simulation are shown in Figures 2-3 respectively. In these figures, lines represent analysis result, and the markers are the corresponding simulation results. It shows that the results of analysis and simulation are coincident. And when λ is small, the delay caused by waiting for aggregation of enough frames will be increased along with the λ decreasing. It indicates that this model obtains the throughput and delay of frame aggregation under different traffic load accurately, and Poisson assumption of frame arrival does not affect the results of the steady state performance.

Figures 4-5 are the analysis results of p and bₑ respectively. Set the minimum arrival rate be the saturated arrival rate λₛ when bₑ reaches 0. When λ ≥ λₛ, bₑ = 0, p get a constant and saturated value of pₑₑ. The system throughput also reaches a constant and stable saturated value of Thₛ, no matter what λ is. With the fixed N, the greater the m is, the greater the Thₛ will be. It indicates that throughput is mainly determined by pₑₑ and m in this scope of λ. On the other hand, when λ < λₛ, throughput is mainly determined by traffic load, and Th is increased approximately proportional to N with the λ increasing.
The larger the $p_{\text{max}}$ is, the smaller the $T_h$ will be. $p_{\text{max}}$ is determined by $N$. The bigger the $N$ is, the bigger the $p_{\text{max}}$ and $T$ will be. So the $T_h$ obtained is smaller. The greater the $\lambda_s$ is, the greater scope of the linear throughput will be, so is the $T_h$. Value of $\lambda_s$ is dependent of $N$ and $m$. With the fixed $m$, the larger the $N$ is, the larger $p$ and $T$ will be. So the $\lambda_s$ and $T_h$ are smaller too. With the fixed $N$, the larger the $m$ is, the smaller the $p$ and $T$ will be. Therefore $\lambda_s$ and $T_h$ are greater, and also shows a greater $\lambda_s$ is obtained for frame aggregation. Additionally, it also indicates that with the saturated traffic, the throughput is enhanced by reducing collision probability.

Denote the saturated arrival rate at $m = 1$ by $\lambda_{s,1}$. When $\lambda > \lambda_{s,1}$, the frame aggregation can always improve the system throughput. When $\lambda < \lambda_{s,1}$, the system throughput can not be improved. From the collision probability curve in Figure 5, it shows that when $m > 1$ and $\lambda < \lambda_{s,1}$, $p$ of frame aggregation is smaller than that of non-aggregated. However the throughput is not improved by reducing $p$. This is because that the $T$ decreased by reducing $p$ is elimated by the prolonged waiting time caused by reducing $\lambda$. All these show that collision probability is the main factor of throughput efficiency at the PHY data rates used in this paper.

From the delay result, it can be seen, when $\lambda > \lambda_s$, a constant maximum delay is obtained. The greater the $N$ is, the greater the value of the maximum delay will be. The greater the $m$ is, the smaller this maximum delay will be. Because $\lambda > \lambda_s$, the node queue is saturated approximately, i.e. $Q = K$ and $b_e = 0$. Meanwhile, the more nodes there are, the larger the $p_{\text{max}}$ and $T$ will be, and the longer the waiting time in the queue of aggregation frame will be. So the greater the $\bar{D}$ will be. With the fixed $N$, the greater the $m$ is, the smaller the $p_b$ will be, and the smaller the $\bar{D}$ will be also. When $\lambda < \lambda_s$, $\bar{D}$ is smaller and independent of $N$, which is mainly determined by $m$. The greater the $m$ is, the greater the $\bar{D}$ will be. When $m > 1$ and $\lambda < \lambda_s/2$, $\bar{D}$ decreases along with the $\lambda$ increasing. When $\lambda \in [\lambda_s/2, \lambda_s]$ or $m = 1$, the waiting time for aggregation is small. The delay is mainly determined by transmission time, so delay is small. From the average queue length in Figure 6, it can be seen that $\bar{Q} = 0$ when $m = 1$. But when $m > 1$, $\bar{Q}$ has stable value which is independent of $\lambda$, and the greater the $m$ is, the greater the $\bar{Q}$ will be. When $\lambda < \lambda_s/2$, $p_b = 0$, as $m > 1$, the smaller the $\lambda$ is, the greater the $\bar{D}$ will be; the greater the $m$ is, the more remarkably $\bar{D}$ will be, which is increased with the $\lambda$ decreasing. This is because the delay of frame aggregation is mainly composed of waiting time for aggregation. The smaller the $\lambda$ is, the longer the waiting time will be when $\lambda$ is very small.

When $\lambda$ is very small, although the delay of frame aggregation increases, the system throughput is still proportional to $\lambda$. This is because although waiting for aggregation will result in performance degradation in a single node, $b_e$ and $p$ varies slowly, $p$ and $T$ are nearly constant as the change of $\lambda$ when $\lambda$ is very small. So the frame aggregation will not cause the waste of channel bandwidth in network with $N$ nodes.
When \( \lambda < \lambda_s \), \( b_e \) decreases and \( p \) increases with the \( \lambda \) increasing. When \( \lambda \) is small, \( b_e \) and \( p \) varies slowly with the \( \lambda \) changing. When \( \lambda \) is approaching to \( \lambda_s \), they change drastically, and approach to 0 and \( p_{\text{max}} \) respectively, which means small change of \( \lambda \) will result in a huge change of \( b_e \) and \( p \). So the system is quickly saturated. The more the nodes there are, the more significantly the \( b_e \) and \( p \) will change, and the faster the system will be saturated. Because of the trend of rapid saturation of \( b_e \) and \( p \), a peak throughput \( T_{\text{peak}} \) is obtained before the system is saturated. Denote the arrival rate where \( T_{\text{peak}} \) is reached by \( \lambda_{\text{opt}} \), and define the increased ratio of peak throughput as \( R = (T_{\text{peak}} - T_s)/T_s \).

From the results of \( T_{\text{peak}}, T_s \) and \( R \) are shown in Figures 7-8. It shows that whether \( T_{\text{peak}} \) exists or not, it depends on \( N \). When \( N < 10 \), the rapid saturation trends of \( b_e \) and \( p \) are not significant. \( T_{\text{peak}} \) does not exist. Meanwhile when \( N \geq 10 \), the greater the \( N \) is, the greater the \( R \) will be. With the fixed \( N \), \( R \) decreases slightly with \( m \) increasing. Value of \( T_{\text{peak}} \) is mainly determined by \( N \) and \( m \). With the fixed \( m \), \( T_{\text{peak}} \) and \( T_s \) are decreased with \( N \) increasing, and \( T_{\text{peak}} \) is decreased more slowly. \( T_{\text{peak}} \) decreases very slowly and tends to stabilize with \( N \) increasing when \( N \geq 30 \), which indicates that the stable value of \( T_{\text{peak}} \)'s independent of \( N \) can be obtained when \( N \) is large. From the curve of \( p \) in Figure 5, it is evident that \( p \) at \( \lambda_{\text{opt}} \) is smaller than \( p_{\text{max}} \). The throughput at \( \lambda_{\text{opt}} \) mainly depends on the arrival rate rather than number of nodes, therefore number of nodes has little effect on \( T_{\text{peak}} \).

Figure 9 is a delay results for \( T_{\text{peak}} \) and \( T_s \). It can be seen that, with the same \( N \) and \( m \), \( D \) at \( \lambda_{\text{opt}} \) is smaller than that at \( \lambda_s \). But \( Q \) at these 2 points have nearly the same value because \( \lambda_{\text{opt}} \) is only slightly smaller than \( \lambda_s \). And from Figures 4-5, it also can be seen, \( b_e \) at \( \lambda_{\text{opt}} \) is much greater than that at \( \lambda_s \), and its \( p \) and \( T \) are much smaller than that at \( \lambda_s \). So the delay in queue at \( \lambda_{\text{opt}} \) is less than that at \( \lambda_s \). With the fixed \( N \), \( b_e \) becomes smaller. But \( T_{\text{peak}} \) becomes bigger with the \( m \) increasing. This is because, under the saturated traffic, \( Q = K \), i.e.
queue length has nothing to do with m. But \( p_b \) becomes smaller with the m increasing, so \( D \) at \( \lambda_1 \) is larger, as shown in (27). At \( \lambda_{opt} \), \( p \) and T are small, \( p_b \approx 0 \); and \( \lambda_{opt} \) increases with the m increasing. The larger \( \lambda_{opt} \) means larger \( Q \), so the larger m means larger delay. But the difference between different \( Q \) at different m is relatively small, so the increase of m at \( \lambda_{opt} \) has less influence on the delay.

From the above-made analysis, it can be seen that when the traffic load is \( \lambda_{opt} \), not only the optimized stable throughput which is nearly independent of nodes number can be obtained, but also the less delay which is nearly not affected by aggregation size can be obtained.

IV. CONCLUSIONS

Model of performance analysis on frame aggregation under unsaturated traffic is established in this paper. Throughput and delay under different traffic arrival rates, buffer sizes, aggregation sizes and access collisions can be obtained using this model. Analysis and simulation show that model is accurate and reliable.

From the results of simulation and analysis, it indicates that, when the traffic is not saturated and the PHY data rate is low, the frame aggregation with fixed aggregation size can not improve the system throughput; the delay performance gets worse under unsaturated traffic, furthermore the deterioration of delay is serious when the traffic load is very low. When frame aggregation works with a traffic load that peak throughput occurs, throughput can be improved and delay is decreased greatly. Meanwhile the decreased delay is less affected by aggregation size. In particular, when the number of nodes is large, the performance improvement at the peak throughput is more evident. This shows the frame aggregation can effectively promote the MAC efficiency.

This model can analyze and evaluate the impact on frame aggregation of different parameters accurately and can determine the main factors affecting system performance. Therefore it can be used in network design, optimization and deployment widely. A further research on the optimal design for frame aggregation under different conditions, especially on how to dynamically select optimal aggregation size under different traffic load will be carried out.

Acknowledgements

This work was partially supported by National Natural Science Foundation of China under Grant No.60772085, 61071108; Sino–Finland Joint Project under Grant No.2010DFB10570 and China Fundamental Research Funds for the Central Universities under Grant No.SWJTU 09ZT14.

References


Biographies

Li Zhijie, currently is a Ph.D. candidate at Southwest Jiaotong University, China. His research interests include wireless resource management and wireless mesh network. Email: sclslzj @163.com

Fang Xuming, is a Ph.D., professor, Ph.D. advisor at Southwest Jiaotong University, China. Prof. Fang is currently engaged in the research on wireless resource management, mobile ad hoc network, wireless mesh network and wireless sensor network. Email: xmfang@swjtu.edu.cn