Packet Length Model for Multihop Transmission

Tawan Thongpook, Anjum Pervez† and Dobri Atanassov Batovski‡

† London South Bank University, ‡ Assumption University, Bangkok, Thailand

Abstract: The packet length in wireless ad hoc networks is a subject of conflicting requirements imposed by the data link layer and the network layer. The dynamically changing signal-to-noise (SNR) ratio in wireless channels requires adaptive selection of the short frame length and the transmission rate at the data link layer. Most frequently, maximum throughput and minimum delay at the network layer are achieved when the optimal frame length is shorter than the packet length. In this case a packet can be transmitted in several consecutive short frames. The frame length and the packet length should be determined independently because the frame length is dependent on the channel conditions, while the packet length depends on the characteristics of both routing and data link layers. The optimal packet length for minimum delay in multihop paths with deterministic linear servers is considered as a function of the node delays and the transmission delays. An analytical model is derived to determine the optimal packet length.

1. Introduction.

The routing in wireless ad hoc networks depends on the distance between the nodes as well as their mobility, which in turn results in variable attenuation and noise levels. The data transmission is established through packet exchange on a multihop basis. The input and output transmission rates of a node may differ from each other. Furthermore, the selection of a higher transmission rate does not always follow to the increase of the information data rate due to the increased bit error rate (BER) and the necessity of multiple retransmissions and variable amount of redundancy introduced by the error-control coding techniques. The data link layer should be dynamically adapted to the channel conditions and one of the important parameters is the optimal frame length required for minimum number of retransmissions per frame. Related issues have been studied systematically [1-3] for the existing wireless networks. The common approach is to encapsulate a packet into a frame for the delivery between two adjacent nodes. Figure 1 shows this scenario where the frame header schematically represents the total number of additional bytes (data link header, trailer, etc.) added on the Data Link Layer and the packet itself represents the encapsulated payload.

The alternative approach proposed here is based on the elimination of the direct dependence of the frame length from the packet length. If the packet length is greater than the optimal frame length, then the packet is fragmented into a number of shorter frames, which is followed by a corresponding reassembly of the frames at the receiver for the reconstruction of the packet. On the other hand, if the packet length is shorter than the optimal frame length, then the packet is encapsulated within a single frame. Figure 2 schematically shows this scenario where the frames are separated by Request-to-Send/Clear-to-Send (RTS/CTS) handshaking, backoff, and Acknowledgement (ACK) operations.

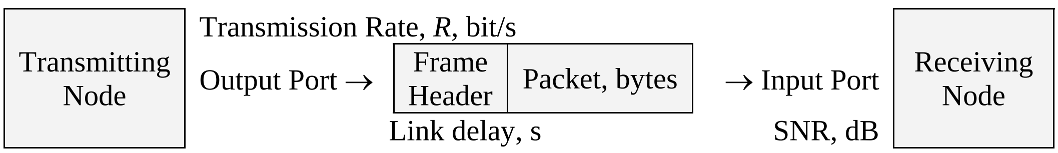

Figure 1. A transmission of a packet encapsulated in a single frame.

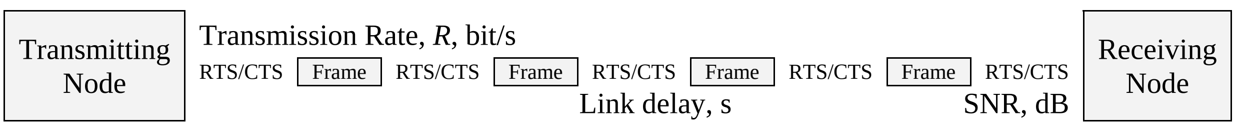

Figure 2. A transmission of a packet in a series of shorter frames.

2. Analytic Model.

The service rate at the output of a server will be considered as a linear function of the length of the digital content. The node design includes Input and Output Servers for the defragmentation/fragmentation of the packets on the Data Link Layer as well as a Routing Server on the Network Layer as shown in Figure 3.

Data Link Layer		Network Layer		Data Link Layer
Input Server (Defragmentation) $T_{\text{input}} = N_{\text{frames}}[A_1(H_{\text{frame}} + F) + B_1]$	→	Routing Server $T_{\text{routing}} = A_2(H_{\text{routing}} + P) + B_2$	→	Output Server (Fragmentation) $T_{\text{output}} = N_{\text{frames}}[A_3(H_{\text{frame}} + F) + B_3]$

Figure 3. A configuration of linear servers.

The processing times at the Input and Output Servers as well as at the Routing Server are given by the expressions

$$T_{\text{input}} = N_{\text{frames}}(1 + N_{\text{w, same port}}Q(P))\ (A_1(H_{\text{frame}} + F) + B_1), \tag{1}$$

$$T_{\text{routing}} = (1 + N_{\text{w}}\ Q(P))(A_2(H_{\text{routing}} + P) + B_2), \tag{2}$$

$$T_{\text{output}} = N_{\text{frames}}(1 + N_{\text{w, same port}}\ Q(P))(A_3(H_{\text{frame}} + F) + B_3), \tag{3}$$

where H_{frame} is the length of the frame header; F is the length of the fragment encapsulated into the frame; H_{routing} is the length of the packet header; P is the length of the packet payload; MTU is the length of the Maximum Transmission Unit at the network layer; A_i and B_i, for i = 1, 2, 3, are corresponding coefficients; and the number of frames N_{frames} equals

$$N_{\text{frames}} = \lceil (H_{\text{routing}} + P)/F \rceil, \tag{4}$$

where $\lceil\ \rceil$ represents a ceiling function. The term N_{w} represents the average number of other cross-traffic packets of the same length (equal to the length of the MTU, when $H_{\text{routing}} + P$ = MTU), waiting in the First-In-First-Out (FIFO) queues of the Routing Servers within the path consisting of L links. According to the all-port model, one can assume that an average number of cross-traffic packets per node, $N_{\text{w, same port}}$, will be transmitted through the same port and an average number of cross-traffic packets per node, $N_{\text{w, other ports}}$, will be transmitted through the remaining ports of the nodes,

$$N_{\text{w}} = N_{\text{w, same port}} + N_{\text{w, other ports}}. \tag{5}$$

Equations 1-3 include the average number of cross-traffic packets, $N_{\text{w}}Q(P)$ which depends on an unknown queuing function $Q(P)$. In general, the function $Q(P)$ indirectly represents the queuing policies that can be implemented in networks of queues for different congestion levels. This study is limited to the assumption of a linear function $Q(P) = P(C/(\text{MTU} - H_{\text{routing}})) + (1 - C)$ for a small number of waiting packets, where C is a small scaling parameter.

The total processing time of a packet in a node equals the sum of the individual processing times

$$T_{\text{server}} \approx T_{\text{input}} + T_{\text{routing}} + T_{\text{output}}, \tag{6}$$

For a selected transmission rate R and average RTS/CTS/backoff/ACK time $T_{\text{handshaking}}$, the transmission time can be obtained accordingly,

$$T_{\text{transmission}} \approx N_{\text{frames}}(1 + N_{\text{w, same port}}Q(P))(T_{\text{handshaking}} + (H_{\text{ftrame}} + F)/R), \tag{7}$$

The end-to-end transmission delay of a single packet over L equivalent links is given by

$$T_{\text{single packet}} \approx L(T_{\text{transmission}} + T_{\text{server}}). \tag{8}$$

For a given message of length M, the number of packets equals

$$N_{\text{packets}} = \lceil M/P \rceil \tag{9}$$

and the total time for the transmission of the entire message can be calculated as the sum of node processing delays and transmission delays,

$$T_{\text{end-to-end}} \approx T_{\text{single packet}} + (N_{\text{packets}} - 1)(T_{\text{transmission}} + T_{\text{server}}). \tag{10}$$

Here the number of frames, N_{frames}, the processing time at the Routing Server, $T_{routing}$, and the transmission time, $T_{transmission}$, are functions of the payload length, P. One can obtain the minimum total delay by plotting the graph $T_{total}(P)$ vs. P and visually observing the minima of the curve.

Also, one can eliminate the ceiling functions in the equations above and use the exact ratios instead. In first approximation, this does not change significantly the numerical value of the optimal payload length, P. Then, the first derivative of $T_{end\text{-}to\text{-}end}(P)$ with respect to P can be easily calculated. Setting it to zero, $d\,T_{end\text{-}to\text{-}end}(P) / dP = 0$, one can obtain an analytical expression for the optimal payload length. The resultant equation can be solved directly. Note, however, that there are cases when the solutions are lower than F or greater than the MTU.

A numerical illustration of the function $T_{end\text{-}to\text{-}end}(P)$ without the use of ceiling functions to obtain a smooth curve is shown in Figure 4. This result assumes: R = 11 Mbps, F = 256 bytes, $T_{handshaking} \cong 1$ ms, H_{frame} = 28 bytes, $H_{routing}$ = 40 bytes, $A_1 = A_2 = A_3 \cong 8/R$, $B_1 = B_2 = B_2 = 0$, message size M = 10 kB, number of links L = 4, $N_{w,\ same\ port} = 1$, $N_{w,\ other\ ports} = 1$, and $C = 0$. The optimal packet payload, P = 366 bytes. This payload can be split up into two frames - frame 1 to carry a fragment of 256 bytes and frame 2 to carry 110 bytes. The two frames can then be transmitted consecutively to deliver the whole payload over the noisy link. This will result in a minimum delay for the whole multihop path.

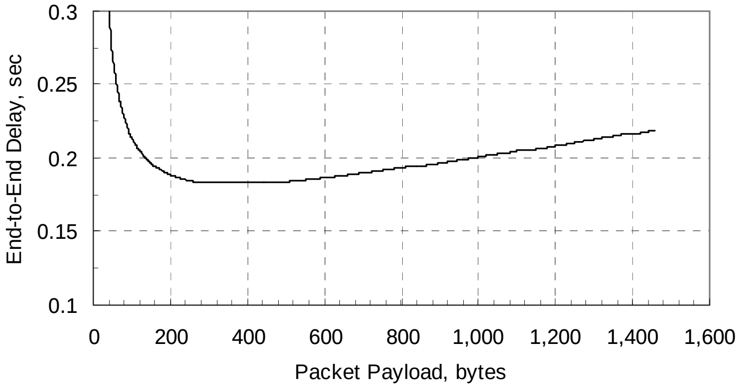

Figure 4. Sample determination of the optimal packet payload length, P.

Figure 5 shows a sample dependence of the optimal packet payload, P, versus the message length, M, when C = 0.1 using the same numerical values for the remaining terms.

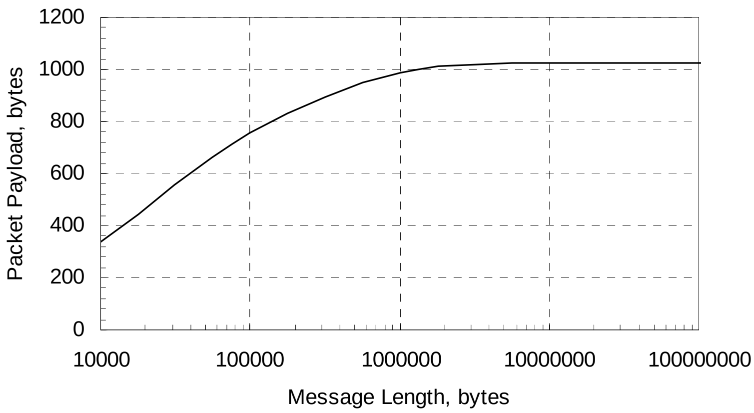

Figure 5. Sample dependence of the optimal packet payload P versus M for C = 0.1.

As it can be seen from Figure 5, there is a *saturation payload length* when the message size approaches infinity for a given value $C > 0$. The optimal payload length depends strongly on the value of the scaling parameter. A further study on the dependence $Q(P)$ of the number of waiting packets from the packet length in arbitrary networks of queues in the presence of cross-traffic might provide practical solutions related to the reduction of congestions in hot spots. One should be able to determine whether a hot spot topology is a subject to constant, decreasing (constructive), or increasing (catastrophic) behaviour of the average number of waiting packets for P < MTU in order to make a decision on the eventual use of packet length control.

3. Discussion.

Short messages that can be encapsulated within a single frame are mainly end-to-end Acknowledgement (ACK) packets, voice packets, etc. Long messages, related to the transmission of E-mail, real-time and non-real-time multimedia applications, etc., are usually transmitted as a sequence of packets of lengths equal to the MTU. This results in a bimodal-like packet length distribution, typically observed in wired Local Area Networks (LANs). The wireless networks suffer from decreased SNR and increased BER. For high values of the SNR, typically greater than 12 dB, the frame length is comparable to the MTU of the standard approach with no packet fragmentation. For values of the SNR, lower than 12 dB, the packets can be fragmented into several frames. The dependence of the packet length on the routing layer with respect to the queuing models and the channel conditions should be taken into consideration for the reliable cross-layer designs with maximum packet payload lengths limited to the saturation length. This would result in quite different packet length distributions in multihop wireless networks if compared to the wired ones. Furthermore, the packet length distributions would dynamically change with respect to the network topology, the congestion levels, the power constraints, and the mobility of the nodes.

The multihop networks also require novel node designs for the improved functionality of all-port routing configurations where several ports can be used simultaneously for the exchange of information among first neighbours. Separate ports might use different frame lengths on the data link layer for establishing optimal connections with the said neighbours. The packet lengths on the routing layer can be determined as a result of a periodic update of the network status. Load balancing schemes in this case can send packets of the same length to different ports. Then the said packets would be fragmented with various frame lengths according to the channel conditions.

4. Conclusions.

The obtained sample analytical expression for the end-to-end delay at the routing layer as a function of the packet length deals with the delivery of a single message (segment) being first fragmented into packets at the source node and further fragmented into frames during the transmissions among the adjacent intermediate nodes where homogeneous channel conditions and delays are assumed within the multihop path. The analytical model can be generalized for heterogeneous networks with different server and transmission parameters, which would require frequent exchange of Resource management (RM) packets to collect information about the status of the nodes and the channel conditions. Thus, the good knowledge of the congestion levels and the adaptive control of the frame length becomes an essential requirement for the implementation of this model in practice.

Acknowledgments.

The Monfort Brothers of St. Gabriel, Assumption University, are gratefully appreciated.

References.

[1] Y. Sankarasubramaniam, Ian F. Akyildiz, and S.W. McLaughlin, Energy Efficiency Based Packet Size Optimization in Wireless Sensor Networks, *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications (SNPA'03)*, in Conjunction with the 38th International Conference on Communications (ICC 2003), Anchorage, Alaska, USA, 11-15 May 2003, ISBN 0-7803-7879-2.

[2] Liljana Gavrilovska and Vladimir Atanasovski, Influence of Packet Length on IEEE 802.11b Throughput Performance in Noisy Channels, *Proceedings of the 1st International MAGNET Workshop D6.2.5*, Shanghai, China, 11-12 November 2004, ISBN 87-91696-36-4.

[3] Wei Liu, Xiang Chen, Yuguang Fang, and John M. Shea, Courtesy Piggybacking: Supporting Differentiated Services in Multihop Mobile Ad Hoc Networks, *IEEE Transactions on Mobile Computing*, Vol. 3, No. 4, October-December 2004.