A Multipath Bandwidth Aggregation Method with parallel D/M/1 Queueing System Assumption

*Yuncong Hong*

Abstract — The arise of millimeter wave communication and VLC (Visible Light Communication) is making us considering more about the possibility of combining the LOS-preferred communication with current 802.11 protocol for large and robust indoor throughput enhancement. This article discusses about the two-node multi-link situation, with queueing process as D/M/1 for each data link. Under this simple assumption, a simple iterative method for solving the optimization problem is put forward, and a simulation implemented to show its optimization.

Index Terms — Bandwidth aggregation, queueing system, D/M/1, minimax problem.

1. Introduction

With the increasing demand for indoor large throughput transmission, which is always stressed in downlink demand design, the original communication methods like VLC and developing millimeter wave communication are proposed as supplements to indoor network connection. Although the wireless PHY protocol are designed to overcome the complex multipath environment, there still exists imbalance among different communication protocol under complex wireless environment, which prevent us to regard them together as one single virtual datalink.

For millimeter wave communication and visual light communication which occupy sparse bandwidth could provide large throughput, but encounter severe attenuation when the LOS channel loses. While compared with traditional 802.11b transmission, which function currently on crowded bandwidth, having more robustness for its physical layer characteristics and good design for frame transmission. To make up for both the shortness, we could combine the data stream initially on each link to multi-link transmission, thus share the advantages among the aggregated data link diversity: robustness provided by 802.11, throughput provided by sparse bandwidth with VLC or millimeter wave communication. An algorithm, or protocol should be put forward for the design, reallocating the data stream into different data link according to their performance. The performance metric should be linear, with bandwidth, propagation delay, protocol differences into consideration. In this article, a queueing system model assumption is proposed for each data link to estimate the end-to-end delay, and adapt load distributing policy based on that.

The rest of the paper is organized as follows: Section II formulate the problem into detail and elaborate the queueing system model; in Section III, we focus on the minimax optimization problem in section II and give an approximated solution; in Section IV some simulation and general results are presented; related work will be given in section V and draw the conclusion in section VI.

1. problem formulation

At first, Kendall’s notation for queueing system should be introduced. A queueing process is generally combined by two stochastic process, called *arrival proces*s and *service process*. For Kendall’s notation, A/B/s, respectively taken as Arrival Process, Service Process interval time random variable, and the service counter number for the queue. The acronym, ‘M’ means Exponential distribution, ‘D’ means deterministic distribution (non-stochastic).

The queueing theory is used in approximate estimation of the performance of each data link, for the reason that the policy adapted is carried out over the protocol implementation layer, which means we could not grasp the exact information about how busy the link is. Queueing theory give us a good estimate on the average queueing time for each ‘income’, thus we could adapt the policy in the right.

1. Proposed model

The overall model is given in Fig.1. In the figure, the whole transmission process is combined by two part of queueing system. The first part, A/D/1, represents the *source controller*, which aggregate the data together and taking the arrival interval as a general distribution process. The second part is constituted by a series of paralleled D/M/1 system which function as the estimation of the data link, simply assuming the service departure process as exponential distribution, without memory.

The well-marked ‘D’, deterministic process, is placed near the first joint. That means when the load distribution policy is made at that joint, the income to it and the outcome form it, is determined, and that decrease the randomness of the queueing system. But with the randomness of the paralleled exponential service interval, the rearranged arrival to the *resembler* would have reordering problem.

The consideration of reordering, will introduce an instinct thought that the end-to-end delay, which mainly affect the throughput, could be regarded as the time until last segment coming into the reassembler. And this leads to a simple optimization representation:

(1)

Where represent the allocated portion for i-th link.

The simple thought that *maximum* of the average waiting time

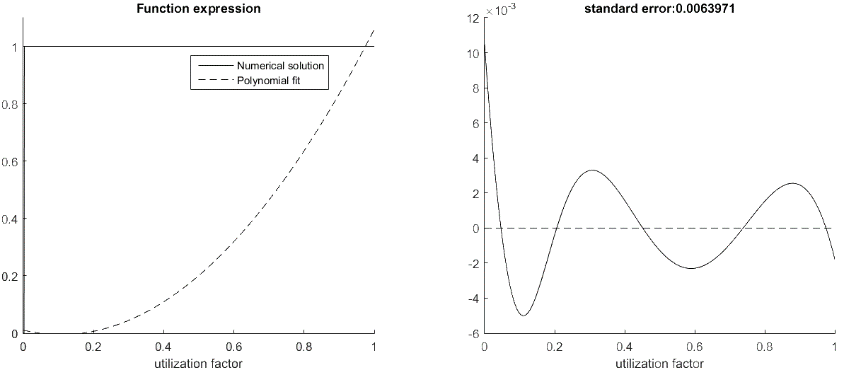
in queue representation, implies that we have to segment the data flow of packets into subtle packets, and thus we are possible to estimate the arrival of a certain data segments.

Fig. 2. Polynomial fit for transcendental equation (left) and standard error analysis (right)

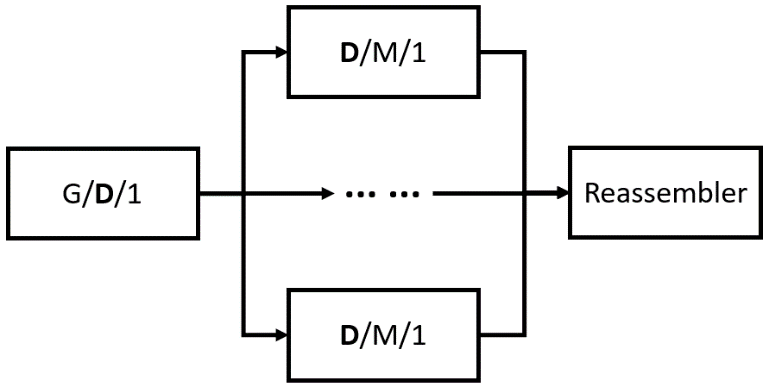
In summary, for a reasonable method to solve the bandwidth aggregation problem, we chop each packet into sub-packets [2], and estimate each packet arrival with the last arrival sub-packet, via the D/M/1 queueing system assumption.

Fig. 1. Overview of the queueing system modeling

1. Equations elaboration

For a certain data link, the overall delay could be expressed in the general form as the network definition:

(2)

where they respectively represent: processing delay, queueing delay, transmission delay, propagation delay. To make it theoretical, the first three part could be seen as *system time* in queueing system, and the last component is out of the queueing and need extra expression, thus formed:

(3)

according to [1], the average *in system time* is:

(4)

where means the *waiting time*, means *service time*. For D/M/1 system, the average waiting time could be expressed in detail:

(5)

where represents the departure rate for service process, and means the probability that the link is busy at the arrival of the packet [3]. And plus the service time to form the total system time:

(6)

The detailed deduction will be given in appendix. And is described by:

(7)

with as utilization factor for the queueing system.

After the theoretical modeling, we are going to put it into practice and align the unit representation; for i-th data link:

(8)

where PK means the input packet size (unit, bytes), is the packet splitting portion for i-th link, represents propagation delay (unit, seconds), is the bandwidth of i-th link (unit, bps) which is the physical meaning of [4]. And we could check that the unit of is ‘seconds’.

And for the utilization factor related with :

(9)

where means average arrival rate, and means average service (departure) rate; and we calculate the deterministic component v with its arithmetic mean:

(10)

where v(t) times means the allocated bytes for one certain datalink.

With the previous work, finally, we again formulate the cost function as:

s.t. (11)

1. Description complement
2. Solution of transcendental equation for .

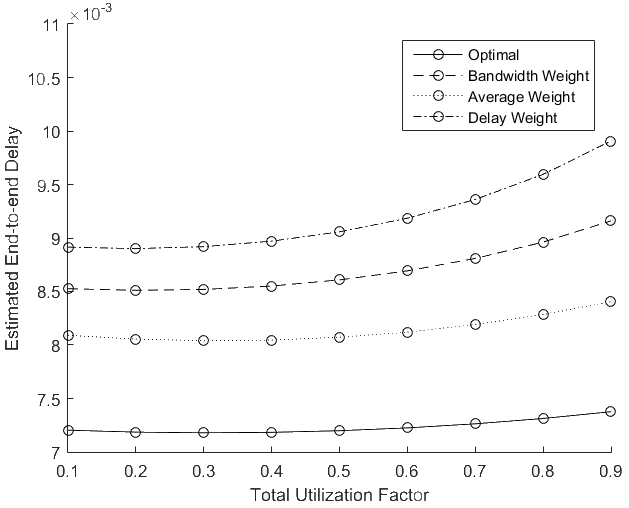
The numerical solution is shown in figure 2 with concrete line. And we notice that it violates our instinct that the system couldn’t transform to busy state immediately when the utilization factor increases. It’s because the delta function assumption for ‘deterministic process’ introduce the weird condition. To recover to the realistic case, we do not solve the equation straightly here, but choose a second order polynomial fit here with:

(12)

the fitting curve is shown in left of figure 2 with dash line. Although it seems like a quite difference between the numerical and fitting curves, the actual error is as small as 0.006 shown in right of figure 2. What causes this is that also on the exponent component in equation (7). The fitting not only relax the singularity brought by delta function, but also better describe the tendency between and .

1. Deduction for equation (6)

Deduction for equation (6) is elaborated in the Appendix. It is a simple supplement for the deduction of average waiting time for D/M/1 in [1].

1. minimax optimization problem

This section will introduce a method for solving the minimax optimization problem introduced by (11).

For simulation need, we could use Matlab toolbox *fminimax* function to solve the problem directly, with line search method. But when look closely inside the equation (11), we could find it in the linear form like:

(13)

where A is constant independent of , based on the solution proposed by [3], a successive approximation method. It is described in the following step but need some fix:

1. Firstly, determine with an initial policy, like bandwidth weight policy [3], or a simple average weight policy, and calculate the cost according to (8);
2. Find the *worst link i* and the *best link k* according to cost function, then find a to minimize the difference between the i-th and k-th link.
3. Repeat step 2 until must be negative, or the difference for the current turn is tolerant or equals to zero.

With the ideal situation, a balance would achieve:

(14)

While we always have to stop the process due to negative situation, the ideal situation for minimax problem solution could not always be reached.

1. Simulation and Results

In this simulation part, we first determine the situation under which we could achieve a better close to the reality result.

TABLE I

configuration of datalink parameters

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Bandwidth** | **Propagation Delay** | **Header size** |
| 1 | 4 Mbps | 0/5 ms | 224 bit |
| 2 | 2.67 Mbps | 0/3 ms |
| 3 | 1.33 Mbps | 0/1 ms |

The parameter setup is illustrated in Table 1. For the total 3 datalinks, the bandwidth BW and propagation delay F is considered. The header size is given according to [4] in practical reason to perform as overhead part, which is not covered in our discussion.

A simple simulation is carried out running with Matlab. In this program, the Matlab toolbox *fminimax* function is used for its simplicity solving the constrained linear minimax problem, and the optimization results are obtained with finite line search steps to the closet solution which is under tolerance.

The simulation results are shown in Figure 3. Four sets of weight policy are selected including the optimal one, and the results show that it’s really optimal. The utilization factor is chosen as dependent variable, over the range which we take interest in.

Fig. 3. The simulation results with different weight assumption.

1. Citing Previous Work

The similar work is done by [3][4]. With their proposed SPMLD model, this article could develop in a quite fast way and master the basic conception of the function and deduction of the G/M/1 queueing system. In this article, several fixup and detail elaboration additional are shown for a more precise expression. The ‘sub-Packet’ statement was cited, and move on to a simple expression as ‘segments on data link’ for more formal express.

1. Conclusion and future work

In this article, a D/M/1 queueing system assumption is established for a single datalink. Based on the assumption, a method of end-to-end delay estimation is proposed as the criterion for the splitting policy for bandwidth aggregation, with sub-packet splitting. The solving of the minimax optimization proves that it outperforms other splitting policy and the potential to improve the aggregated throughput.

As future work, a practical system is to be implemented for evaluating the actual performance; Compared work is planned to compare with other related work.

1. Appendix

This part will give the deduction for equation (6).

For original system expression:

and for D/M/1 system, the corresponding waiting time and service time expression:

Therefore, the sum-up time expression:

Inverse transform to pdf representation:

Giving the cdf conditioned described in Chapter 6 in [1]:

After the calculation:

Compute the expectation out, the equation (6) obtained:

Acknowledgement

The author would like to express his sincere gratitude for the presentation audience and reviewers who provide the suggestions to improve the paper quality.

References

[1] Kleinrock, Leonard. *Queuing Systems*. 1st ed. New York, N.Y.: John Wiley & Sons, 1975. Print.

[2] Wu, J., Wu, X., & Chen, J. (2013). SPMLD: Sub-Packet Based Multipath Load Distribution for Real-Time Multimedia Traffic.

[3] Wu, J., Yang, J., Shang, Y., Cheng, B., & Chen, J. (2014). SPMLD: Sub-packet based multipath load distribution for real-time multimedia traffic.

[4] S. Prabhavat et al., “Effective delay-controlled load distribution over multipath networks,” IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 10, 1730–1741, 2011.