An improved QoS technique to minimize delay for Multimedia Application in MANET

Lalit Johari, Rahul Kumar Mishra

***Abstract*— A mobile ad hoc network (MANET) is a self-arranging framework less system of mobile devices associated by wireless and it getting well known methods for information exchange nowadays. As more media applications are tried and sent, the need to give the essential Quality of Service (QoS) ability is expanding. In this paper, a proficient delay minimization system is proposed in view of the Multi server queuing model for application content. The application content is grouped into video, sound and data. To limit the delay, bandwidth capacity is utilized as limitation parameter. This framework is proposed to get the required QoS fulfilled courses for each kind of utilization content and to decrease the queuing delay.**

Index Terms—**MANET, Quality of Service functionalities, Queuing delay minimization**

I. Introduction

A mobile ad hoc network (MANET)[1] is a remote system that utilizations multi- hop shared routing rather than static system framework to give network connectivity. MANETs have applications in quickly conveyed and dynamic military and regular citizen frameworks. The system topology in a MANET ordinarily changes with time. The requirement for managing continuous applications for clients of MANET has turned out to be critical. The forthcoming content of this paper makes essential commitment on scheduling algorithms that classifies and organizes the real-time traffic with the goal of enhancing the execution of the real-time applications in light of content in MANET. This paper executes the Queuing model through which the change in execution in performance is demonstrated which is in terms of delay.

Mobile ad hoc network (MANET), with no settled frameworks, enable portable terminals to set up a temporary system for moment correspondence. The MANETs maintain huge applications in these situations, including fiasco recuperation, crisis help, portable conferencing, combat zone correspondence, electronic payments whenever and anyplace, dynamic database access, mobile workplaces, vehicular administrations and so on. Along these lines, the rise and the anticipating eventual fate of real-time and multimedia applications have blended the need of high caliber of Service bolster in remote and mobile systems administration condition. Be that as it may, giving Quality of Service (QoS) for Mobile Ad-Hoc Networks (MANETs) is a burdening assignment owing to the dynamic topology and constrained assets in MANET's.

Quality of Service (QoS) [3] is the performance level of an administration offered by the system to the client. The level of the administration depends on a few parameters or imperatives frequently known as accessible data transfer capacity, end-to-end delay, delay varieties or jitter, probability of packet loss and so on.

For real time applications, the information rate and delay are the key variables, while, in military utilize, security and dependability turn out to be more essential. If there should arise an occurrence of emergency circumstances, the key factor ought to be the accessibility. It demonstrates that the QoS factors differ from application to application.

As a result of the insufficient accessibility of transmission transfer speed in MANET's[5], QoS procedures need to advance the meager asset by giving greater need to the real-time streams over best-exertion streams so as to agree to the QoS prerequisite, such as delay bounds and throughput.

In this paper, we give Queuing modeling a calculation for real time applications. The objective of the framework is for picking the most significant way to fulfill the required QoS imperatives of a specific application. The rest of this paper is organized as follows: in Section II we focus a light on the term delay and its types. Section III discusses about the proposed queuing model in detail. In section IV, the results from comprehensive simulations are evaluated and Section VI draw conclusion and future work.

II. DELAY

Mathematically End to End delay is shown below:

dend-end= N [dtrans+dprop+dproc+ dqueue]

Where dend-end= end-to-end delay , dtrans= transmission delay , dprop= propagation delay ,
dproc= processing delay , N= number of links (Number of routers + 1) , dqueue = Queuing delay

A. Transmission delay: In a network normally, transmission delay (or store-and-forward delay) is the measure of time required to push the greater part of the data packets/bundle's bits into the wire. As such, this is the deferral caused by the data rate of the connection. Transmission delay is a component of the data packets / bundle's length and has nothing to do with the separation between the two nodes. This delay is corresponding to the parcel's length in bits, It is given by the following formula:

Transmission delay = number of bits/rate of transmission (bits every second)

B. Propagation delay: In computer networks, propagation delay is the measure of time it takes for the head of the signal to make a trip from the sender to the receiver. It can be

***Manuscript received August, 2017*.**

Lalit Johari, *School of computer Applications, IFTM University Moradabad, INDIA, 9997213060.*
Dr. Rahul Kumar Mishra, *School of computer Applications, IFTM University Moradabad, INDIA, 9639004077.*

processed as the proportion between the connection length and the spread speed over the particular medium. Propagation delay is equivalent to d/s where d is the distance and s is the wave proliferation speed. In wireless communication, s=c, i.e. the speed of light.

C. Processing delay: In a network, processing delay is the time which takes by switches to process the packet / bundle header. Processing delay is a key part in delay. Amid Processing of a bundle, switches may check for bit-level mistakes in the parcel that happened amid transmission and in addition figuring out where the bundle's next goal is. Processing delays in rapid switches are ordinarily on the request of microseconds or less. After this nodal preparing, the switch guides the parcel to the line where additionally deferral can happen (queuing delay).

D. Queuing Delay: The queuing postpone is the time that a bundle/ packet spends in a line at a hub while sitting tight for different parcels to be transmitted. On the off chance that the node is a fast switch at that point there is one queue for each friendly connection, so a parcel sits tight just for different bundles that are going over a similar connection.
The queuing delay is identified with the transmission delay dtrans by the accompanying surmised condition.

$$dqueue = dtrans * lqueue$$

Here, lqueue is the average length of the queue. The average queue length relies upon the load factor, which is the proportion of the endeavored interface transmission rate to the connection most extreme transmission rate. The average queue length is ordinarily under 1 for a load factor under 1/2. At the point when the load factor surpasses 1, the queue length develops without bound.

III. PROPOSED MODEL TO MINIMIZE QUEUING DELAY

In this proposed framework, right off the bat, the application layer decides the content type of application. The application content is characterized into three classes: Class 1 (Video), Class 2 (Audio) and Class 3 (Data) and sends the substance sort to arrange layer to choose the specific gathering of parameters.

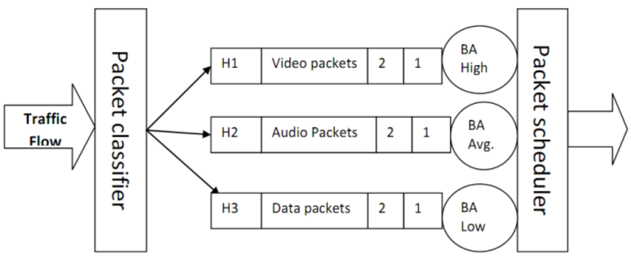


Figure 1: Proposed System Model

Hop count, end-to-end postpone and data transfer capacity are utilized for choosing the enhanced route. Among these parameters, the proper parameters for a specific application contents are grouped together to find the improved routes as appeared in figure 1.

In the above figure if we characterize the data transmission limitations according to the necessity of the substance so we can get the greatest throughput. In other words , to increase high throughput, route ought to have enough data transmission or should have a specific bandwidth as per the need of the content type. Thus, bandwidth parameter is utilized as imperative parameter to gather the bandwidth ensure routes. Means each route will ensure the adequate and pre determined data transfer capacity. The desire route and data transfer capacity will be chosen to get best throughput.

Data is characterized for the standard application sort. In light of the request of the best exertion conveyance of this kind of class, just hop count is utilized for course disclosure. Audio is proposed for the delay sensitive application. For this, hop count and end-to-end defer parameters are decided for choosing the route.

Video is characterized as the Class 3. For this kind of substance, parameters are utilized for requirement and in addition advancement. To increase high throughput, route should have enough transfer speed. Thus, data transfer capacity parameter is utilized as limitation parameter to gather the transmission capacity ensure route. This application content is likewise delay sensitive application, hop count and end-to-end postpone are utilized for picking the enhanced route among the bandwidth ensure routes.

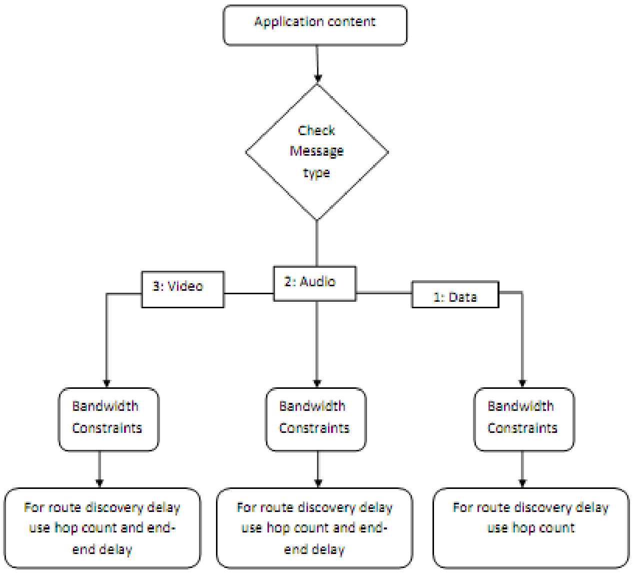


Figure 2: Flowchart for Proposed System Model

Here a strict-priority queue is policed, either by bandwidth or a percentage of the bandwidth as given in the bellow algorithm:

```
policy-map policy1
        class video
 priority X
         class audio
priority Y
         class text
bandwidth Z
         class class-default
 fair-queue
```

The previously mentioned setup indicates that there is just a single Priority Queue of size of 1 Mbps which is time shared between the two applications by the understood police. Video and sound classes of activity are set into the high need line and get strict need lining over information movement. The classes will be independently rate-constrained regardless of the possibility that they go into a similar line, for video movement X kbps and for sound activity Y Kbps will be rate restricted.

So X+Y+Z=1Mbps

A queuing framework comprises of at least one server that give administration or something to that affect to arriving data packets. Data packets who land to discover all servers occupied for the most part go along with at least one (lines) before the servers, henceforth the name queuing frameworks. There are a few regular illustrations that can be depicted as lining frameworks [7], for example, bank employee benefit, PC frameworks, fabricating frameworks, support frameworks, interchanges frameworks et cetera. Parts of a Queuing System: A lining framework is described by three segments: Arrival process, Service component & Queue discipline.

A. *Arrival Process*

Arrivals may begin from one or a few sources alluded to as the calling population. The calling Population can be restricted or boundless. A case of a restricted calling populace might be that of a settled number of machines that fail randomly. The arrival process comprises of depicting how data packets touch base to the framework. Suppose that Ai is the between entry time between the entries of the (i-1)th and ith packets, we should signify the mean (or expected) between landing time by E(A) and call it (λ); = 1/(E(A) the entry recurrence.

B. *Service component*

The service mechanism of a queuing system is specified by the number of servers (denoted by s), each server having its own queue or a common queue and the probability distribution of customer's service time.

C. *Queue Discipline*

Discipline of a queuing system means the rule that a server uses to choose the next customer from the queue (if any) when the server completes the service of the current customer.

IV. RESULTS ANALYSIS

In this model there are N numbers of identical independent parallel servers which receive data packets from a same source but in different parallel queues each one receiving packets at a rate of λ/ N . The following figure shows how a typical multiple single servers' model looks like.

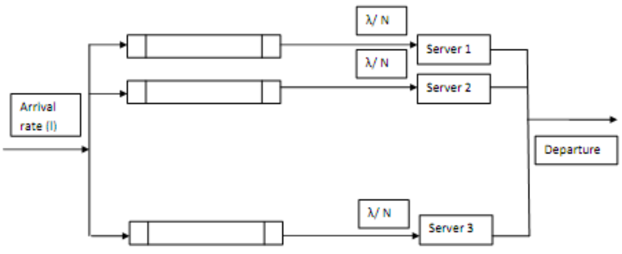


Figure 3: M/M/1 Queuing Model

In M/M/1 model there is only one server. The important results of this model are:

1. Average number of packets in the system = L = $\rho/1-\rho$
2. Average number of packets in the system = Lq = $\rho^2/1-\rho$
3. Expected waiting time in the system W = $L/\lambda = (1/\lambda)\ \lambda/\mu-\lambda$
$= 1/\mu-\lambda$
4. Expected waiting time in the queue Wq = Lq/λ
$= 1/\lambda \times \lambda 2\ /\mu(\mu-\lambda)$
$= \lambda/\mu(\mu-\lambda)$

In the following tables the values of μ and λ which are Arrival rate and Services rate respectively are taken from the secondary data sources and the value of L , Lq , W , Wq and ρ are calculated using the above formulas.

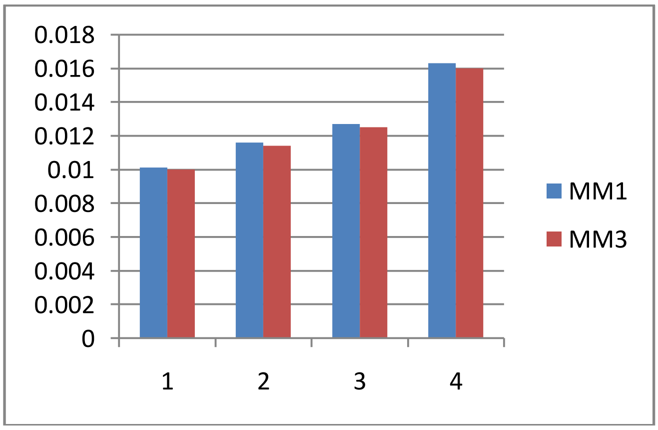
M/M/1							
μ	λ	C	L	Lq	W	Wq	ρ
Arrivals / Second	Services / Second	Number of Servers	Average No of packets in System	Average No of packets in Queue	Average Time Spent in System	Average Time Waiting in Line	Server Utilization
30	0.3	1	0.0101	0.0001	0.0337	0.0003	0.01
35	0.4	1	0.0116	0.0001	0.0289	0.0003	0.0114
40	0.5	1	0.0127	0.0002	0.0253	0.0003	0.0125
50	0.8	1	0.0163	0.0003	0.0203	0.0003	0.016

Table 1: Queuing Delay calculation using M/M/1 [9]

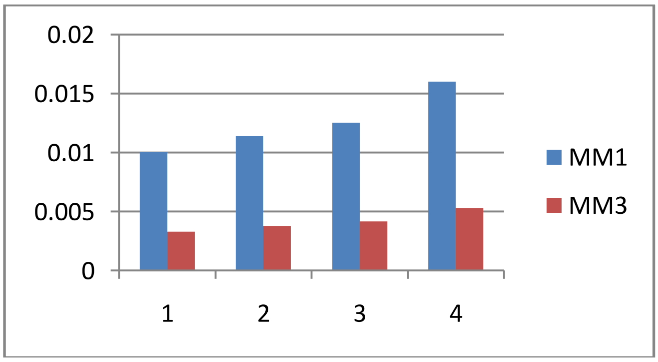
M/M/3							
μ	λ	C	L	Lq	W	Wq	ρ
Arrivals / Second	Services / Second	Number of Servers	Average No of packets in System	Average No of packets in Queue	Average Time Spent in System	Average Time Waiting in Line	Server Utilization
30	0.3	3	0.01	0	0.0333	0	0.0033
35	0.4	3	0.0114	0	0.0286	0	0.0038
40	0.5	3	0.0125	0	0.025	0	0.0042
50	0.8	3	0.016	0	0.02	0	0.0053

Table 2: Queuing Delay calculation using M/M/3 [9]

GRAPH GENERATION



Graph 1: Comparison between two scenarios in terms of Average Number of packets in System



Graph 2: Comparison between two scenarios in terms of Server Utilization

Above tables compare the single server model and multiple server model of the queuing model. Here M/M/1 represent that all the data packets are processed by the single server while the M/M/3 represents that all the data packets are processed by the three servers (as shown in above figure). Here μ and λ represent the arrival rate and service rate in seconds. Above tables data shows that when we increase the number of servers than the Queuing delay (Lq) decreases. Here each data packet spend less time in the queue (Wq). The server utilization becomes low in the case of M/M/3 means server is ready to process next data packet that are in ready queue.

V. CONCLUSION AND FUTURE WORK

With the above table's data and graphs we get the results that by using the multi-server model, queuing delay can be minimize. This whole system can be implement using the Riverbed Modeler Academic Edition 17.5 (OPNET), where the mathematical model can be tested on actual environment. In this way multi server queuing model can reduce queuing delay providing the bandwidth constraints on the basis of the content type.

REFERENCES

[1] Roja Kiran Basukala, Kyong-Heon Han, Dong-You Choi*, Young-Bai Kim, Seung-Jo Han," QoS Assurance of multimedia traffic in Residential Network with Hybrid Coordination Function and Queuing Disciplines" in 2009 Pacific-Asia Conference on Circuits, Communications and System IEEE.

[2] Sotirios-Angelos Lenas, Stylianos Dimitriou, Fani Tsapeli, Vassilis Tsaoussidis,2010. Queue Management Architecture for Delay Tolerant Networking.

[3] Tetsuji Hirayama,2010. Analysis Of Multi class Feedback Queues And Its Application To A Packet Scheduling Problem, Journalof Industrial And Management Optimization.

[4] Jui-Chi Chen.2012. Optimized Packet Scheduling Management: Maximizing Bandwidth Utilization for Next-Generation Mobile Multimedia Communications,Wireless Pers Communication.

[5] Jun Zou, Dongmei Zhao,2009.Real time CBR traffic scheduling in IEEE 802.16 based wireless mesh networks.

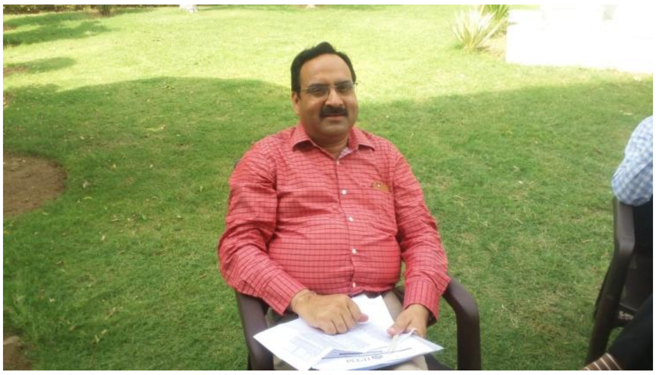
[6] Brunonas Dekeris, Tomas Adomkus, Aurelijus Budnikas, 2006.Analysis of QoS Assurance using Weighted Fair Queuing(WFQ) Scheduling.

[7] Discipline with Low Latency Queue(LLQ),International Conference on Information Technology Interfaces.

[8] Transportation Systems Engineering,,Ch. 45. Queuing Analysis. nptel.ac.in/courses/105101008/downloads/cete_45.pdf

[9] http://www.supositorio.com/rcalc/rcalclite.htm

Lalit Johari is pursuing his doctorate in computer science from IFTM University, Moradabad, India. His area of interest is Computer networking, Real time operating system amd Mobile Ad-Hoc network. Author has done MCA from U P Technical University , Lucknow , India from 2003-2006. Apart from this paper author has limited papers in the different fields of Computer Science.



Dr. Rahul Kumar Mishra presently working as a director in IFTM University, Moradabad , India. He has done his doctorate from Bhagwant University , Ajmer , India. He has vast experience on Programming Concepts , Design & Analysis of Algorithms , Database Management System and Distributed System. He has a profound interest in writing research papers and text books. He has authored many prominent research papers and books on Computer Science & Applications.