MULTI-CLASS TRAFFIC MANAGEMENT IN 4G NETWORK

Damodar Nayak and Rabindra Kumar Dale
Department of CSE, Silicon Institute of Technology, Bhubaneswar, India

ABSTRACT

The 4G network is a heterogeneous network, supports wireless access (WiFi, WiMAX, Bluetooth), cellular networks (GSM, CDMA, GPRS, UMTS), WSN, MANET and other earlier generation networks (PSTN, ISDN) etc. In 4G each network has different characteristics with multi traffic classes to provide QoS. To manage the multiple traffic classes 4G network will need to support QoS provisioning and full mobility. In this paper we propose a scheme using queueing model M/G/1 for multiple traffic class management for better QoS. The objective of this paper is to minimize the delay and maximize the throughput using M/G/1 queueing model. This model uses the network state like traffic arrival and service rates, compute the end-to-end QoS parameter statistics like delay and throughput of traffic flow through each available network. Multi-mode mobile Terminal (MT) manages the multi-class traffic by using multiple queues for multiple services. Each queue has packet arrival rate, service rate, utilization and delay as prioritized traffic high to low priorities. Higher priority will be scheduled first for better QoS in 4G.

KEYWORDS: QoS, MT, 4G, M/G/1, PQ

I. INTRODUCTION

Last few years there are many changes in mobile computing and different types of networks are implemented. Now a day’s mobile computing are one of the best research areas. In different network many devices are known as node. Here a mobile node or mobile terminal introduced in 4G network. 4G mobile node is a multimode mobile terminal (MT) in [2] can have the access of different network in 4G. If it is in ad-hoc network it can route the packets to its neighbor node. If it is in wireless sensor node it sends packets as sensor node until sink node not found. In case of cellular network, it works as subscriber station (SS) and so on. 4G as heterogeneous network MT can access all the networks. In a 4G network environment, there is a wide range of QoS parameters. As a rule of thumb, key parameters are packet delay or latency, bandwidth, packet jitter, and packet loss. However, how to provide QoS for the multiclass traffic in the 4G network environment is not as easy task. Mobile communication today has providing varying levels of coverage and QoS[12].

In this work the queueing model is used to manage the delay in 4G network. There is a classifier and scheduler for MT. Classifier keeps the packet in different queue as per packet type of traffic class. Different queue has different priorities in MT. The scheduler has to decide which packet should be served first. Higher priority packets are served first then its next lower priority packets are served and so on. Multi-mode terminal is the mobile user having a multi mode MT can access the Internet services through cellular base stations (BSs), WMAN BSs, or WLAN access points (APs), and will also be able to support peer-to-peer communication with others using MANET connectivity. The 4G architecture with overlapping networks in the hotspot areas using different wireless technologies, and these different wireless networks are connected to the common Internet Protocol (IP) backbone. In a 4G, the basic components include MTs, BSs/APs and their access networks, and the core IP network (IP backbone).
In this paper, an analytical model for implementing Priority Queueing (PQ) in a 4G node to calculate the queueing delay is presented. The model is based on M/D/1 queueing system (a special class of M/G/1 queueing systems). There are five different traffic classes are considered as supported in 4G network. The exact packet delay for corresponding classes is also calculated. Here it is proposed that different queues for the different types of traffic with classifier and scheduler. All classes can use the bandwidth each other. This is one of the ways to get for cost and end-to-end constraints. This work is focused on discovering a least cost, delay constrained path for real time data.

In this work five traffic classes are taken. These are listed below.

(1) Unsolicited Grant Service (UGS) as class C1
(2) Real-Time Packet Services (RTPS) as class C2
(3) Extended Real Time Packet Services (ERTPS) as class C3
(4) Non-Real Time Packet Services (NRTPS) as class C4
(5) Best Effort (BE) as class C5

II. RELATED WORK

Among research areas 4G network is one of the most best research area to make differentiate and valuable. Multiple classes of traffic that have different time constraints where no user can feel it. Due to advance in research a new paradigm in telecoms systems known as the Next Generation Network (NGN) is emerging. A clear and fixed definition of NGN is yet to be developed, but there’s universal consensus that it will comprise the following [3]:

(1) A layered architecture, consisting of functionally differentiated planes, namely the Transport plane, the Control plane, the Application plane, and the Management plane.
(2) Open and standardized interfaces to facilitate interworking of systems across the planes.

Below shown the 4G heterogeneous network where different networks with MT(multimode mobile terminal) are working in single framework. Here the performance of packet scheduler of queueing model for multi-class traffic in a multi-mode mobile terminal(MT) is considered. In a 4G network, the mobile terminal (MT) may be equipped with multiple wireless network interfaces so that the integrated heterogeneous environment allows a user to access a particular network, depending on the specific application needs and types of available radio access networks in [2].

The architecture of 4G heterogeneous network is shown in figure-1

![4G Heterogeneous networks architecture](image)

Figure 1. 4G Heterogeneous networks architecture

2.1 Plane Description

Some description in [3] about the different plane

2.1.1. Transport plane:

The Transport plane (or the Data/User plane) is responsible for the transfer of user traffic across the network. The router decides what to do with packets arriving on an inbound interface. It says roughly about the routing control plane.

2.1.2. Control plane:

The Control plane is responsible for the control of bearer capabilities in the data plane through signaling and also the facilitation of interworking between systems. It define certain packets to be
discarded, or certain packets for which a high QoS is defined by such mechanisms as differentiated services

2.1.3. Application plane:
In this plane many operation are handled as service creation and deployment. This plane provides an infrastructure for the provision and management of services, and defines standard interfaces to common functionality

2.1.4. Management plane:
The Management plane performs such tasks as QoS provisioning, security, and network management. The Management plane will be distributed (not localized) and cross layered. Management plane has ultimate control over all transport plane and control plane entities

2.1.5. Access:
The Transport plane moves data packets across the access, the edge, and the core networks. The access network will be a heterogeneous network with various wireless and wireline technologies operating in concert to provide service to the user. A wireless access network will typically consist of mobile terminals, a wireless access point, and an access router.

2.1.6. Edge:
The edge network will be a system of routers that manage ingress into the core network through admission control and QoS provisioning.

2.1.7. Core:
The core network will be a high-speed optical switching and transport system. The entire data plane will be packet switched.

The layered architecture of 4G network is shown in figure -2

![Layered architecture of 4G network](image)

**Figure 2.** Layered architecture of 4G network

III. **QUEUEING MODEL**

First In First Out (FIFO) is the traditional queueing algorithm used in several topologies that requires no configuration. FIFO queueing has no decision power about packet priority. FIFO queueing keeps packets and schedule them according to their arrival. The explosion sources can cause extended delays in scheduling real time traffic, and potentially to network control and signalling messages. FIFO queueing system was an effective network traffic controller before, but now advance networks require more sophisticated algorithms. In case of overflow situation the dropping packets; even be a higher-priority packet. FIFO cannot differentiate among from high-priority to low-priority packet. To the requirement of 4G we need priority queue implementation.
3.1. Priority Queueing

Here implementation of priority queueing model is required. To eradicate the demerits of traditional FIFO discipline, priority queueing (PQ) is suggested to meet the desired QoS for real time traffic. In this work, five queues in a 4G node are considered, class1, class2, class3, class4 and class5 for respective traffic class. All classes are arranged in higher to lower priority order. Here, strict priority logic is used in the scheduler, means first always serves highest-priority queue. If higher priority queue is empty, it comes to its next lower priority queue and it will serve the low-priority queue. Here the scheduler of the 4G node is serving different output queues simultaneously, and behaves like a multiple queues/single server system. Here we exploit M/G/1 queueing system to build a queueing model for a 4G node which is behaving like multiple queues/single server system. Before discussing about the different type of queueing notations some assumptions should be taken. The packets that are related to high-priority queue one (Q1) to low-priority queue five (Q5) are called Class-1 (C1) to Class-5 (C5) packets with the average length of L1 to L5, respectively. C1 to C5 packets are coming according to Poisson process with arrival rate 1 to 5 respectively. 4G network is a set of networks like MANET, Ad-hoc, WLAN, Satellite, Cellular etc. Let 5 traffic classes for 4G QoS of each network (VoIP, Streaming Audio/Video, VoIP (VoIP with Activity Detection, FTP, Data transfer, Web Browsing).For each network traffic there are 5 queues. Let Q1,Q2, …Q5 queues of class C1,C2, …,C5 respectively. Let L1,L2,…L5 are the length of 5 queue respectively. Let λ1, λ2, … λ5 are the arrival rate of packets. Let µ1,µ2 ,..., µ5 are the service rate of packets.

3.2. Priority Queue Q1 (Packets C1)

At the beginning first we start higher priority queue Q1; we assume that the average service requirement for traffic class C1 packet is E[S1]=1/µ1. The second moment of service requirement for traffic class C1 packet is E[S1^2]. Here, the main objective is to finding the queueing delay for all higher priority class C1 packets. A packet is selected randomly in [1] and explained its arrival time using the PASTA property of Poisson arrival streams. Here, the queueing delay is defined as the expected waiting time E[W1] in queue Q1 for traffic class C1 packet before its being serviced. According to priority scheduler logic, the expected waiting time of higher priority class C1 packet consists of two points

(1) T_R : Those packets already in service and needs remaining service time to get service.
(2) T_s : The time needs to serve all the packets with the same priority (C1) that are already present in the system at the arrival of this new randomly selected packet.

The expected waiting time can be written as form of equation:

\[ E[W_1] = E[T_R] + E[T_s] \]  

\[ (Eq-1) \]

E[T_R] is the remaining expected time for a packet those already in service at the time scheduler is busy. The probability of scheduler (server) is busy is known as ρ (rho) or utilization. A packet of C1 is in service with probability \( \rho_1 = \lambda_1E[S_1] \), is the utilization of class one(C1) packets. Though packet arrival time is random selection, the remaining service time can be viewed as that obtained for generic random variables S [8]. Thus, the remaining processing time of a C1 packet is equal to \( E[S_1^2]/2E[S_1] \). However, because at the arrival time of the randomly selected packet, the class (either C1 or C2) which is already being served is unknown, the final equation should be modified to:

\[ E[T_R] = \sum_{k=1}^{c} \rho_k \frac{E[S_1^2]}{2E[S_1]} \]

\[ (Eq-2) \]

The second term in equation (Eq-1), E[T_s] is the expected total time to serve all C1 packets that are already waiting in Q1 upon arrival of the randomly selected packet. Assume that the expected number of packets already waiting in queue one is E[N1]. Due to the PASTA property and little’s Law, on average there are E[N1]=λ1E[W1] class 1 (C1) packets upon arrival of this randomly selected packet [7]. Since the packets already waiting in Q1 each requires on average service time on, E[T_s] can be written as:

\[ E[T_s] = E[N_1]1/\mu_1 = \lambda_1E[W_1]/\mu_1 \]

\[ (Eq-3) \]

If we are substituting in Eq-1 for E[T_R] and E[T_s] then E[W1] can be calculated as follows A very similar expression has been given in [11] for M/G/1 with priority as well:
According to the deterministic probability service time of scheduler of a 4G node is applied. The scheduler in the 4G node needs time units to serve a $C_k$ packet with the transmission rate of $R$, and the average service requirement for a $C_k$ packet is, the second moment of service requirement of a $C_k$ packet can be expressed as follows:

$$E[S_k^2] = \text{Var}[S_k] + (E[S_k])^2 \tag{Eq-5}$$

In our approach, because the scheduler is having a fixed service time to serve its packets and hence functioning similar to a M/D/1 queueing system, the service time would be deterministic with zero variance, i.e., $\text{Var}[S_k]=0$ [1]. Thus, Eq-5 can be simplified as: $E[S_k^2] = (E[S_k])^2 = \left(\frac{L_k}{R}\right)^2$ Based on the above assumption, the remaining service time of a $C_1$ packet will be $L_1/2R$ on average with the probability of $\rho_1$ when a $C_1$ packet is in service. However, since upon arrival of the randomly selected packet, it is not clear which packet (either $C_1$ or $C_2$) is in service, Eq-4 is modified as:

$$E[W_1] = \frac{\sum_{k=1}^{2} \rho_k \frac{L_k}{2R}}{1 - \rho_1} = \frac{\rho_1 \frac{L_1}{2R} + \rho_2 \frac{L_2}{2R}}{1 - \rho_1} \tag{Eq-6}$$

Where $\rho_1$ and $\rho_2$ are the utilizations caused by $C_1$ or $C_2$ packets with average lengths of $L_1$ and $L_2$, respectively.

### 3.3. Priority Queue $Q_2$ (Packets $C_2$)

This is the priority queue which should be scheduled after the 1st higher traffic class $C_1$. We obtain the expected waiting time for a randomly selected $C_2$ packet arriving to the priority queue $Q_2$ by analyzing the events that constitute this delay. The total work of the system is defined as the (random) sum of all service times at any time that will be required by the packets in the system at that instant. The waiting time of a $C_2$ packet (which is the low priority queue) can be written as in [10],

$$E[W_2] = E[X_1] + E[X_2] + E[X_3] + ... \tag{Eq-7}$$

Where $E[X_1]$ is the expected total work seen by the arriving $C_2$ packet in $Q_1$ and $Q_2$, plus the rest work needed to finish the service of a packet, which is already in service (if any).

$E[X_1]$ can be written as

$$E[X_1] = E[T_R] + E[T_2] \tag{Eq-8}$$

$E[T_R]$ which can be calculated in the same way, as we have calculated for $Q_1$ and $E[T_2]$ is the time needed to serve all the packets of the higher priority class $C_1$, and equal priority class $C_2$ upon the arrival of the randomly selected $C_2$ packet is related to the number of packets per class in the both queues ($Q_1$ and $Q_2$) upon arrival of the $C_2$ packet. Referring to the PASTA property and the Little’s
law, there are $E[N_k] = \lambda_k E[W_k]$ of $C_k$ packets on average upon arrival of a new $C_2$ packet. Since each requires service time on average, $E[T_2]$ can be calculated as follows.

$$E[T_2] = \sum_{k=1}^{2} \frac{E[N_k]}{\mu_k} = \sum_{k=1}^{2} \frac{\lambda_k E[W_k]}{\mu_k} = \sum_{k=1}^{2} \rho_k E[W_k]$$

By putting the values of $E[T_k]$ and $E[T_2]$, we can write equation (8) as follows:

$$E[X_1] = \sum_{k=1}^{2} \rho_k \frac{L_k}{2R} \rho_1 E[W_1] + \rho_2 E[W_2]$$

Now, $E[X_3]$ is the expected amount of work associated with higher priority $C_1$ packets arriving during $E[X_1]$, $E[X_3]$ is the expected amount of work associated with $C_1$ packets arriving during $E[X_3]$ and so on. Figure-3 showing the waiting time of an arriving packet of $C_2$ is in fact given by the total workload building in front of it. In figure-3 arrows denote the arrival times of $C_1$ packets, and all the oblique lines shown with 45° angle with the time axis. In this figure-3 the expected waiting time is $E[W_2] = E[X_1] + E[S_1]E[M_1] + E[S_1]E[M_2] + \ldots$ Since the service times and the arrival process are independent. For a stationary packet arrival process, this can be $E[M_j] = E[E[M_j|X_j]] = c_1 E[X_j]$.

Due to mentioned in dependence, where $c_1 > 0$ is a constant particular to the arrival process. That is, expectation of the number of arrivals in any period of time is proportional to the length of that period because of stationarity in time and linearity of expectation. According to stationary poisson traffic input process, higher priority class $C_1$ is the expected number of arrivals per unit time of $C_1$ packet is $\rho_1 E[W_1] + \rho_2 E[W_2]$. So the expected waiting time can be reduced to

$$E[W_2] = E[X_1] + E[S_1]c_1 E[X_1] + E[S_1]c_1 E[X_2] + \ldots$$

$$= E[X_1] + \frac{c_1}{\mu_1} (E[X_1] + E[X_2] + \ldots)$$

$$= E[X_1] + \frac{c_1}{\mu_1} E[W_2]$$

In other words, during $E[W_2]$ time units, the $C_2$ packet has to wait $\lambda_1 E[W_2]$ packets of $C_1$ arrive on average, each requiring $1/\mu_1$ service time. Hence $c_1/\mu_1 E[W_2]$ can be written like

$$\frac{\lambda_1 E[W_2]}{\mu_1} = \rho_1 E[W_2]$$

This

$$\frac{\lambda_1 E[W_2]}{\mu_1} = \rho_1 E[W_2]$$

Substituting all the values in equation, $E[W_2]$ can be calculated as follows:

$$E[W_2] = \sum_{k=1}^{2} \rho_k \frac{L_k}{2R} + \rho_1 E[W_1] + \rho_2 E[W_2] + \rho_1 E[W_2]$$

Bringing $E[W_2]$ to one side and by simplifying, then it can be:

$$E[W_2] = \frac{\sum_{k=1}^{2} \rho_k \frac{L_k}{2R} + \rho_1 E[W_1]}{1 - \rho_1 - \rho_2}$$

....(Eq-9)

3.4. Other Queues ($Q_3$, $Q_4$ and $Q_5$)

In this way waiting time for class $C_3$, $C_4$ and $C_5$ can be calculated.
3.4.1. Class $C_3(Q_3)$

The expected waiting time for traffic class $C_3$ can be calculated as

$$E[W_3] = \sum_{k=1}^{2} \rho_k \frac{L_k}{2R} + \rho_1 E[W_1] + \rho_2 E[W_2]$$

$$1 - \rho_1 - \rho_2 - \rho_3$$

...Eq-10

3.4.2. Class $C_4(Q_4)$

The expected waiting time for traffic class $C_4$ can also be calculated as

$$E[W_4] = \sum_{k=1}^{2} \rho_k \frac{L_k}{2R} + \rho_1 E[W_1] + \rho_2 E[W_2] + \rho_3 E[W_3]$$

$$1 - \rho_1 - \rho_2 - \rho_3 - \rho_4$$

...Eq-11

3.4.3. Class $C_5(Q_5)$

The expected waiting time for traffic class $C_5$ can also be calculated as

$$E[W_5] = \sum_{k=1}^{2} \rho_k \frac{L_k}{2R} + \rho_1 E[W_1] + \rho_2 E[W_2] + \rho_3 E[W_3] + \rho_4 E[W_4]$$

$$1 - \rho_1 - \rho_2 - \rho_3 - \rho_4 - \rho_5$$

...Eq-12

IV. SIMULATION AND PERFORMANCE ANALYSIS

4.1. Network Setup

The Simulation for this work is done in matlab programming. Total there are 5 networks are considered as an example. Area of all networks are taken in rectangular shape with size [0,0,1000,1000]. As heterogeneous networks all different overlapped networks are built inside this area. Area of MANET is [300,300,500,500], WiMAX is [400,400,500,500], WiFi [380,10,520,500], LTE is [10,500,400,500] and WSN is [10,10,500,500]. The network setup figure shown in figure-4.

![Figure-4: Network Setup](image-url)
assumed with the transmission speed of different as per different network; with queue length size is 40 packet capacities as their buffer sizes.

4.2. Results

Simulation Parameters are arrival rate, service rate, queue length, rate of transmission, throughput and delay. In table-1 the simulation output for Utilization vs Delay is shown. Here when the utilization is 2.5% at that time expected delay will be for class C1=0.1047 micro second(ms), C2=0.1103 ms, C3=0.1164 ms, C4=0.123 ms and C5=0.1265 ms and when arrival rate is increased the utilization is 5% at that time expected delay will be 0.2127, 0.2366, 0.2648, 0.2983 and 0.3172 for C1, C2, C3, C4 and C5 respectively. Like this also when utilization is 22% at that time the respective delays are 1.1646 ms, 2.1214 ms, 5.0788 ms, 28.4516 ms and 46.9596 ms and so on. Here total traffics are classified and packets are distributed to different queues.

Table-1: Expected Delay for different traffic classes

<table>
<thead>
<tr>
<th>Throughput (%)</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>W5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.55</td>
<td>0.1047</td>
<td>0.1103</td>
<td>0.1164</td>
<td>0.123</td>
<td>0.1265</td>
</tr>
<tr>
<td>5.05</td>
<td>0.2127</td>
<td>0.2366</td>
<td>0.2648</td>
<td>0.2983</td>
<td>0.3172</td>
</tr>
<tr>
<td>7.55</td>
<td>0.3267</td>
<td>0.3848</td>
<td>0.4599</td>
<td>0.5594</td>
<td>0.6199</td>
</tr>
<tr>
<td>10.05</td>
<td>0.4469</td>
<td>0.5593</td>
<td>0.7203</td>
<td>0.9624</td>
<td>1.1241</td>
</tr>
<tr>
<td>12.55</td>
<td>0.574</td>
<td>0.7664</td>
<td>1.0749</td>
<td>1.6167</td>
<td>2.0242</td>
</tr>
<tr>
<td>15.05</td>
<td>0.7087</td>
<td>1.0138</td>
<td>1.5702</td>
<td>2.7576</td>
<td>3.8004</td>
</tr>
<tr>
<td>17.55</td>
<td>0.8514</td>
<td>1.3119</td>
<td>2.2844</td>
<td>4.9751</td>
<td>7.905</td>
</tr>
<tr>
<td>20.05</td>
<td>1.0031</td>
<td>1.6747</td>
<td>3.3598</td>
<td>10.1644</td>
<td>20.4571</td>
</tr>
<tr>
<td>22.55</td>
<td>1.1646</td>
<td>2.1214</td>
<td>5.0788</td>
<td>28.4516</td>
<td>46.9596</td>
</tr>
</tbody>
</table>

Figure-5: Utilization vs. Delay

4.3. Analysis

The Table1 shows the differences in delays for Q1 to Q5 for several utilizations of the 4G nodes. Here Q5 delays exceed those of Q4 in all position and Q4 exceeds those of Q3 delays and so on. As shown
in Figure 5, we can notice the characteristic of priority queue (PQ), as the utilization increases, there is a increase in the queueing delay of class 2 (C2) packets. The 1st horizontal axis in Figure 5.2 shows the variable of arrival rates in units of utilization, while other shows the range of the variable of delay in units of micro seconds. So in the figure -5 the graph shows the change in delays over the arrival rates of the packets. As per packet arrival the result of delay is changing. From utilization 2.5% to 22% the impact of priority queueing implementation in the 4G node increased sharply in the delay for queues at higher arrival rate for the numerical and simulation results.

V. CONCLUSION

The main objective of this paper to calculate expected delay for multiple traffic classes. This paper have presented the closed-form expressions of the expected queueing delay for multiple classes of traffic like C1 to C5 in a 4G node through the implementation of priority queueing based on M/G/1 queueing system. The results have been verified through simulation studies. Provided results are one of way to analyze the performance of priority queueing implementation in a 4G node. In this technique 5 classes of traffic have different expected waiting time and utilization is shown. Its result is finding out the most appropriate queueing scheme for implementation for 4G networks.

VI. FUTURE WORK

In our future work we will implement 4G network traffic condition with multiple servers for multi-class traffics. We will use different types of queueing model and try to find out which queueing model provide best QoS under heavy traffic.

REFERENCES


Authors Biography

Damodar Nayak has done M.Tech. in Computer Science and Engineering in Biju Patnaik University of Technology (BPUT), Odisha, India. Now he is working in department of Computer Science and Engineering of Silicon Institute of Technology (SIT, BBSR). He has also done MCA from IGNOU. His interested research area is mobile computing.
Rabindra Kumar Dalei has done his M.Tech (CS) from Utkal University, Bhubaneswar, Odisha, India. Also he has done my MCA from NIT Warangal, AP, India. Presently he is working as a Sr. Asst. Professor in Silicon Institute of Technology, Bhubaneswar, India. His research area is Wireless Sensor Network.