Design and Analysis of QoS-aware Scheduling Schemes for IoT Applications

Reema Sharma, NHCE, sharma80reema@gmail.com
M.Dhivya, NHCE, saidhivya@gmail.com
Sanjeev Sharma, NHCE

Abstract — *Internet of Things (IoT) designed using different technologies with numerous applications is becoming complex. In certain applications, the Quality of Service (QoS) needs to be a stringent requirement. To assure the necessities of these applications, it is crucial to define QoS models which can classify IoT applications and provide necessary QoS factors. Furthermore, providing QoS becomes more critical if resources available are inadequate. Addressing these issues, the paper proposes novel QoS-responsive models for providing priorities to delay and loss sensitive applications. These models allow efficient management of resources to provide superior treatment to real time applications without causing significant degradation to the performance of other network traffic. Traffic from Radio Frequency Identification (RFID), nodes equipped with IEEE 802.15.4 transceivers in sensors networks, security surveillance cameras in an intelligent home network and e-Health services etc. is collected for simulation. Furthermore, these models depict their feasibility through a range of IoT applications. The QoS-models are compared for allocated weights, packet lost ratio and waiting time in queue and validated with extensive simulation studies.*

Index Terms — *Average Queue Length; Buffer Management; Delay Sensitive Applications; Quality of Service; Internet of Things*

I. INTRODUCTION

INTERNET OF THINGS (IOT) is the network of sensor embedded physical objects for the purpose of connecting and exchanging data with other devices over the Internet. Most of the applications employ very small data streams but can't meet the expense of having unsuccessful connections at the time of congestion. Prioritization of traffic streams becomes more important when current IoT environment is expecting billion of devices to be connected. In IoT, devices with various capabilities are connected with Internet protocol (IP) and web services to make quick decisions and to transfer information without depending upon the human intervention [1]. Businesses seem to look forward for opportunities where real time and streaming data will generate new markets and improve existing services. For example, usage of IoT technologies in industry include: (i) Intelligent transport solutions to make traffic flows faster, prioritize schedules to repair vehicles and to reduce accidents [2], (ii) Remote health care monitoring to improve quality of health care unit and easy access to it [3], (iii) Sensors installed in airports, buildings and smart homes for security and information gathering [4], and (iv) Smart electric grids to improve system's reliability and to efficiently connect renewable resources [5]. The use of IoT technologies in such applications has stimulated the increase in real time data, which resulted in challenging issues of storing, accessing and providing QoSon the data. The most fundamental challenge of providing QoS in relatively scarce network is to allocate the available resources in an efficient manner to improve overall system performance.

After vigilant study and understanding, research communities and academic organizations have defined certain QoS architectures and QoS schemes based on IoT components, data classification, enabling technologies, application areas and relations between these modules. Optimized QoS can be achieved by executing different QoS schemes or by developing algorithms to optimize or improve one or more QoS parameters. The QoS in traditional networks is more steady as compared to QoS in IoT environment. Actually additional QoS attributes are required for IoT environment as IoT paradigm is fast growing, need to be more scalable and need to attain full correlation between connected devices.It also needs to provide them with intelligence and smartness by aiding their adaptation and independent conduct while assuring privacy, trust and security of the users and their data.

Therefore, the QoS issues in IoT need further improvement. Also, by improving hardware, routing and speed, etcto achieve QoS, it might impact cost of maintenance and service for the operators. Different existing QoS schemes such as call admission control scheme, dynamic allocation scheme using renegotiation, fault tolerant dynamic channel allocation can be applicable to a part of IoT network but not applicable to dynamic IoT environment. The knowledge of applications which need QoS, characteristics of different applications, packet size of the application and traffic classification is very important. We need to determine which applications are business-critical and analyze all the applications which are competing for the network resources. Furthermore, we need to understand the characteristics of the applications as some applications are sensitive to packet loss, some are sensitive to delay and some may steal lot of bandwidth such asbursty traffic.

TABLE I
QOS PARAMETERS AT DIFFERENT IOT LAYERS

Layers.	QoS Parameters			
Sensing	Accuracy& Precision	Energy Consumption	Coverage	Cost
Network	Bandwidth Allocation	Storage Capacity	Allocation Rate	Throughput optimization
Application	Service Cost	Service Time	Reliability	Load

The chief aim of this paper is to develop efficient service models to provide QoS factors required for different IoT applications. According to the proposed models, IoT resources can be effectively allocated to various services while satisfying QoS factors and maximizing system performance. We build an intellectual packet prioritizing scheme that adapts well to the dynamic requirements of IoT applications. Our main contribution includes the following:

- Modeling and analysis of service models based on categorization of IoT applications to provide certain QoS factors to satisfy the necessities of those services.
- Design and development of prediction based QoS model for packet scheduling to prioritize high priority data and simultaneously reduce packet loss of non-priority data.
- Comparison, simulation and performance studies of the developed models.
- Testing the scheduler efficiency for different application areas in IoT by considering different data rates, buffer sizes, packet sizes, etc.

The various sections in the paper are organized as follows. In Section II, we investigate related work, while design and analysis of QoS-Models are discussed in sections III with comparison of models on the basis of weights allocated to different services. Section IV provides the details of performance metrics and simulation results. Section V discusses about applications and proposed service models in IoT. Section VI concludes the paper.

II. RELATED WORK

Existing researches in IoT contributes towards energy efficient approaches, connectivity standards, sensing in complex environment, working towards significant challenges added due to the different type of applications and unacceptable QoS of delay sensitive information. QoS-alert model for providing guarantees using routing schemes and packet scheduling schemes are available in the literature, in which few models are specifically designed for IoT architecture. In the models designed for IoT architecture, basically the dynamic nature of IoT need to be considered otherwise some of the services related to different applications may fail. Very few models as discussed below are available to provide dynamic nature of IoT and it is observed that models with Markov decision may reduce the failure rate significantly.

A. QoS Awareness in IoT

In IoT scenario, attaining QoS to certain delay sensitive applications is extremely important. IoT technology also provides additional value to urgent operations in terms of obtaining efficient cooperation.

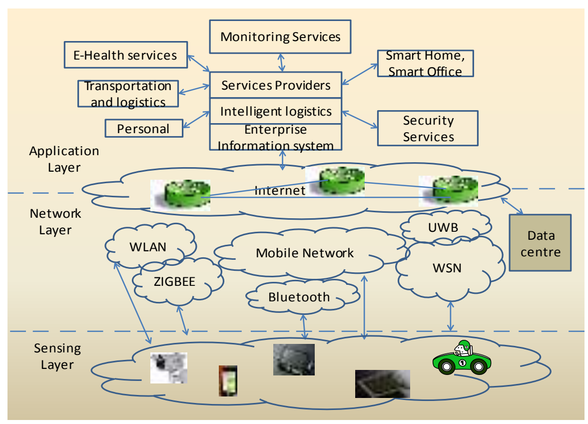

Fig. 1. Layers in IoT architecture [20]

The response time for these services and QoS demands need to be fulfilled. Klepec and Kos [6] offered a priority scheme for 2 queues and suggested packet transit time function for a delay sensitive application with minimum bandwidth constraint. The model is uncomplicated and straightforward; the effective use of more prior data was exhibited at the cost of high data loss for low priority data. Further, various models have been designed in order to inspect and observe energy efficiency [7], network topologies, issues connected with performance [8][9] and the attainability of bandwidth [10]. The explanation to emergency services is not effective as discussed in the above studies.So, to address the delay critical applications, an efficient packet scheduling scheme becomes necessary.

Numerous devices and varied networks limited to IoT makes it intricate to meet different QoS requirements [11]. Packet scheduling is most effective technique to decide which packets should be serviced or dropped. Packet scheduling can provide service differentiation and can help in separating critical data packets from non-urgent data packets. So, QoS alertness can be included in IoT by assigning traffic primacies and scheduling them with proper algorithm [12]. In [13], a QoS alert message scheduling procedure is projected in IoT environment. The author divided messages into 2 classes, critical messages and non-urgent data packets and anticipated QoS scheduling method. QoS requirement is one of the vital factors that is required to be mentioned for operative communication conferring to type of service. In [14], a multi-dimensional QoS, decision is anticipated on the basis of IoT approximation method in which every stakeholder is linked into a solitary value to create best formation of interactions. In [15], authors suggested an effective packet loss control technique by indicating retransmissions on maximum reliable link in the instance of packet dropped but with the norms that the nodes are immobile throughout their presence.

III. DESIGN AND ANALYSIS OF DYNAMIC SCHEDULING SCHEMES

Each QoS architecture introduces scenarios for different application areas in IoT. The difference in QoS architecture is based on the random number of active flows, traffic classification, priority constraints and amount of bandwidth allocated. Our model is loosely based on weighted round robin (WRR) scheme which does not consider packet size into account and thus the model needs to estimate proper weightvalues according to packet sizes. Hence, different classes are assigned packets of variable sizes to check that the bandwidth is allocated precisely.

The main difference between the existing approaches and our work is that for allocating the weights properly, we measure the increase in average queue length at different time slots and dynamically assign the weights to various services instead of usingan average queue length of different services to decide upon the weights allocated to them.Our approach greatly improves the performance in terms of packet lost and waiting time in queue.

TABLE II
HIGH, MEDIUM AND LOW PRIORITY WEIGHTS IN PROPOSED AND EXISTING SCHEME

S. No .	High PriorWeights		Medium Prior Weights		Less Prior Weights	
	Proposed scheme	Existing scheme	Proposed scheme	Existing scheme	Proposed scheme	Existing scheme
1	0.300	0.300	0.300	0.300	0.400	0.400
2	0.300	0.300	0.300	0.300	0.400	0.400
3	0.300	0.300	0.300	0.300	0.400	0.400
4	0.300	0.3190	0.300	0.300	0.4017	0.3810
5	0.3211	0.3483	0.300	0.300	0.3799	0.3517
6	0.3420	0.3767	0.300	0.300	0.3576	0.3233
7	0.3627	0.4159	0.300	0.300	0.3370	0.2841
8	0.3831	0.4422	0.300	0.300	0.3175	0.2578
9	0.4034	0.4675	0.300	0.300	0.2971	0.2325
10	0.4234	0.4918	0.300	0.300	0.2763	0.2082
11	0.4433	0.5151	0.300	0.300	0.2571	0.1849
12	0.4630	0.5375	0.300	0.300	0.2373	0.1625
13	0.4824	0.5590	0.300	0.300	0.2173	0.1410
14	0.5017	0.5797	0.300	0.300	0.1986	0.1203
15	0.5208	0.5995	0.300	0.300	0.1794	0.1005
16	0.5397	0.6186	0.300	0.300	0.1601	0.0814
17	0.5584	0.6369	0.300	0.300	0.1414	0.0631
18	0.5769	0.6545	0.300	0.300	0.1233	0.0455
19	0.5952	0.6714	0.300	0.300	0.1050	0.0286
20	0.6133	0.6876	0.300	0.300	0.0867	0.0124
21	0.6313	0.700	0.300	0.300	0.0687	0.000
22	0.6491	0.700	0.300	0.300	0.0509	0.00
23	0.7000	0.7000	0.300	0.300	0.0000	0.00

A. Design and Analysis of DBWS as first model (Model-1)

QoS-alert packet scheduling scheme is required towards service provisioning. Service provisioning can be provided by assigning priorities to traffic and scheduling them with apt algorithm. In this section, QoS aware packet scheduling scheme is discussed.

A new Dynamic benefit weighted scheme (DBWS) is described and compared with Adaptive weighted scheme.Packets are divided into high, medium and low priority service classes using a traffic classifier. The proposed DBWS scheme uses three different queues of sizes 6, 100 and 100 for high, medium and low priority traffic respectively. It calculates scheduling weights for each classdepending upon average queue length of respective buffers. Amount of packets to be scheduled is selected depending upon scheduling weights and are transferred to WRR Scheduler placed inside a router. Finally, these packets are forwarded to subsequent router where similarprocess is conducted.The DBWS method ensures improved high priority service by dynamic allotment of weights in a controlled way. The important features of DBWS are as follows:

Buffer space allocated for emergency services is little higher (6) to contain transient bursts [16].

- The rise/decline in average queue length of high service is projected and weights for the present time are calculated based on this rise or decline.
- Average queue range of prioritized service is computed depending uponExponential Weighted Moving Average (EWMA) [16].
- For maintaining delay jitter within acceptable limits for prioritized traffic (<10ms) [17], lowest and extreme thresholds are allotted which will be the pointer to allocate the proper weights.
- Medium services are provided with similar weight values and only be minimized for hugeemergency traffic.

Average queue length of EF is calculated as [16]:

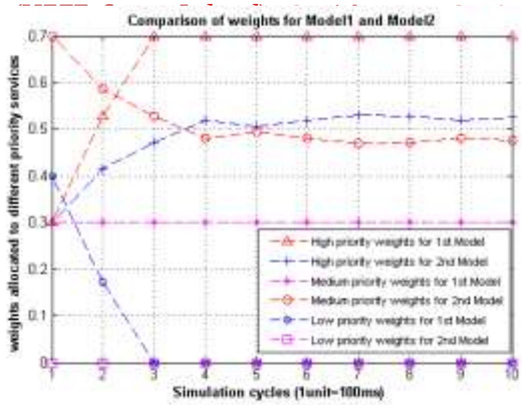

Fig. 4. Weight Standardization and differences in weights for different packet sizes in Exp 1(b)

$$avg_{qsize} = (1 - 0.01) * avg_{qsize} + 0.01 * inst_{qsize} \quad (1)$$

where avg_{qsize} is average queue size, $inst_{qsize}$is instantaneous queue size and 0.01 is scaling value to decrease fluctuations in $inst_{qsize}$.

In DBWS scheme, the uniformity of weights is sustained by approximating the variance between the average queue sizes at successive time slots and by adding benefit in terms of weights calculation depending on this variation. The current weight *Wt* is reliant on the difference of the average queue length at time *t* from preceding average queue length at instant (*t-1*). Lower and upper tolerances defined as lowest and extreme thresholds, are put to examine the performance of high priority data for queuing delay parameter. Minimum threshold signifies the required queuing delay and maximum threshold signifies allowable queuing delay.

If average queue size is under least threshold, the high precedence weight is considered as 0.3.Itoffers small queuing delay. On the other hand, if average queue size is within least and highest threshold, the weights are proportionately speckled with rise or decline in average queue length at successive time slots. If it rises beyond the extreme threshold, adequate but bounded to maximum limit, bandwidth is assigned.

In suggested scheme, weights are computed in the order of priority. As high service is quantified to value , so the scheme starts by computing weight of high priority service by means of the subsequent equation:

$$highpri_{wt} = \begin{cases} high\_wt_{init} highavg_{qsize} \epsilon \{0, 0.5) \\ \frac{(0.3)*\left(high_{New_{Avg}} - high_{Prev_{Avg}}\right)}{(highMax_{Th} - highMin_{Th})} + highPrev\_wt, \quad highavg_{\text{qsize}} \epsilon \{0.5, 2.2) \\ high_{upper\_value}, \quad highavg_{\text{qsize}} \epsilon \{2.2, 6\} \end{cases}$$
(2)

where the $high_{upper\_value}$ is 0.7, the upper limit of high priority service weights. $high_{new_{avg}}$ is the average length of buffer space at present instant *t*. $high_{prev_{avg}}$is the average length of buffer space at preceding instant*(t-1)*. $highMax_{Th}$ is taken as 2.2 so as to maintain delay jitter in bearable limits. $highMin_{Th}$ is 0.5 as average. length of queue is under 0.5, we can attain small delay jitter for current applications. $highPrev_{wt}$ is the weight of high priority class at instant*(t-1)* if $highpri_{wt}$is taken as weight of high priority service at current instant *t*. $high\_wt_{init}$ is the original weight allotted to high priority service, with weight of 0.3 when average queue length is between 0 to 0.5.

The scheme holds the summation of high, medium and low services weights as *1* where *1*indicates total bandwidth of the connectionfrom which packets are forwarded. The entire bandwidth of the connection is proportionately segregated into high, medium and low classesdepending upon weight calculation [8]. If high priority weight is 0.7, the utmost weight assigned to medium classwill be 0.3, in caseearlier weightallocated to low classisshifted to high services. The calculation is as follows:

TABLE III
TRAFFIC ADOPTED FOR CONDUCTING VARIOUS EXPERIMENTS

S.No.	Parameters				
	Service Type	High Priority	Medium Priority (Exp1(a))	Medium Priority Exp1(b)	Low Priority
Exp 1	Traffic Type	Poisson			
	Number of Flows	3 flows at same time			
	Packet size in bytes	50 to 1000 Bytes	5 to 450 bytes		5 to 450 bytes
	Channel Bandwidth	710Kbps			
	Datarate	120Kbps	610 Kbps	1220 Kbps	610Kbps
Exp 2	Packet size in bytes	50 to 1500 bytes	5 to 950 bytes		5 to 950 bytes
	Channel Bandwidth	900Kbps			
	Datarate	400Kbps	640Kbps	1280 Kbps	640Kbps
Exp 3	Packet size in bytes	900 Bytes	100 Bytes		100 bytes
	Number of flows	3 flows at same time			
	Channel Bandwidth	900Kbps			
	Datarate	400Kbps	640Kbps	1280kbps	640Kbps
Exp 4	Packet sizes	136 bytes to 5000 bytes	5 to 950 bytes	5 to 950 bytes	5 to 950 bytes
	Traffic Type	MPEG	Poisson	Poisson	Poisson
	Channel Bandwidth	2.1Mbps			
	Data Rate	Upto 3Mbps	0.5Mbps		0.5Mbps

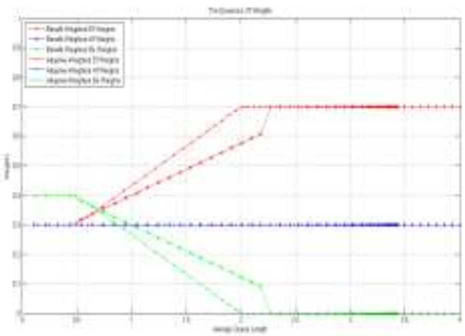

Fig.2. Weight Standardization and differences in weights

- Initialize $highpri_{wt}$ = 0.3, $medpri_{wt}$ = 0.3, $lowpri_{Wt}$ = 0.4
- If $highpri_{wt}$ augments from 0.3, decrease $lowpri_{Wt}$ by amount of addition in $highpri_{wt}$ and if adjusting $lowpri_{Wt}$ is insufficient, then a little portion of $medpri_{wt}$ can be shifted to $highpri_{wt}$.
- The highest weight of medium priority class is restricted to 0.3.
- The weights are computed for low priority class with the subsequent formula.

$$lowpri_{Wt} = 1 - (highpri_{Wt} + medpri_{Wt}) \qquad (3)$$

To estimate the performance of proposed packet scheduling algorithm, we first equated this algorithm with adaptive weighted, priority and static WRR. Since adaptive weighted algorithm has better results in contrast to priority and static, we have evaluated our findings with adaptive-weighted scheduling algorithm paper[16]. Adaptive weighted is one of the most suitable mechanism for multimedia applications. The major benefit of this scheme is that not only it helps in producing better performance to guaranteed applications but also offer fairness to medium and low priority applications. The comparison of high priority weights computed in both DBWS scheme and adaptive-weighted scheme is depicted in Fig.2. It can be seen that the DBWS algorithm issues high priority weights more

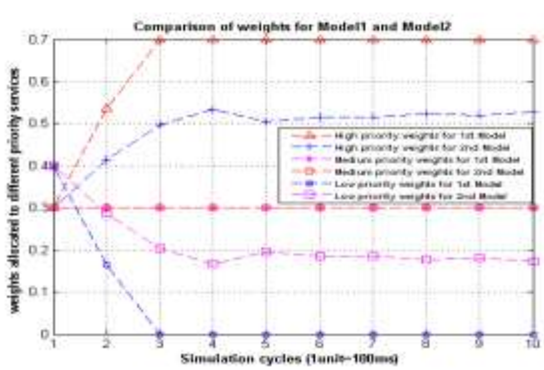

Fig. 3. Weight Standardization and differences in weights for different packet sizes (Exp1(a)) and Exp2(a)

efficiently as related to existing algorithm. The computed weights for high, medium and low are documented in Table II. Discussing this in the table, beginning with initial readings of weights of high, medium and low as 0.3, 0.3 and 0.4; the adaptive weighted algorithm is consuming a little more bandwidth (weights)in comparison with benefit weighted algorithm and henceforth high weight readings are attaining the maximum values slightly before anticipated algorithm. Subsequently channel bandwidth for a shared link is static that is, overall weights for a shared link is constant, the calculated weights which are assigned to high priority class in turn can be applied for medium and low services. Soloss of packets reduces for these services.

B. Design and Analysis of Prediction Based Scheduling Scheme as Second Model (Model-2)

In this segment, we establish dynamic bandwidth allocation prediction algorithm as second model to augment its usage for diverse applications. This algorithm includes following added features and variances from first model:

- Traffic from diverse applications separated depending upon several data features like data bit rate, PLR and acceptable delay into various queues and prioritize them into high, medium and low priority data depending upon these characteristics.
- Buffer space utilized for high priority application is 10 and is maintained slightly higher to adjustbursty data. Also, the buffer space for medium and high priority queue is taken large (100) to reduce the packet loss.
- Highest and lowest thresholds are allocated which would be the pointer to assign the suitable weights.
- Algorithm envisages average rise or fall in average queue length of high & medium priority applications and for high priority class, sum up these values with preceding weight values of high priority data. This aid in calculating weights for current time slot in optimized manner depending upon this rise or fall to assign bandwidth needed for high and medium priority services.
- Medium priority weights are computed to adjust

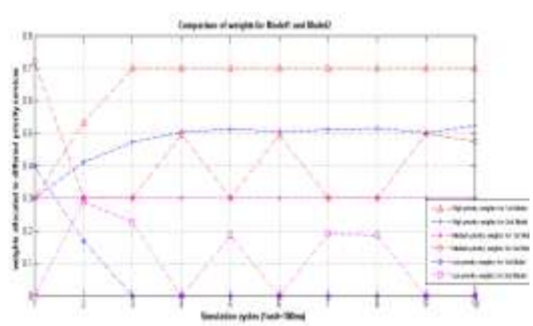

Fig. 5. Weight Standardization and differences in weights for Exp3(a)

soft real time applications if hard real time applications are taken as high priority in certain applications.

- Average queue size of high priority queue is found using Eq. (1) and average buffer space of medium priority queue is computed with the help of general averaging technique.

Threshold levels are considered for buffer managing. If average buffer length is below lowest threshold, high priority weight is approximated to 0.3, which will produce less delay, if average buffer length is between least and highest threshold, weights are proportionately assigned with rise or fall in the average queue length at successive time slots. If it goes beyond highest threshold, adequate but bounded to high limit bandwidth is assigned. In this algorithm, the weights are computed in a circular manner, beginning from highest to lowest priority queue. Bandwidth for highest precedence queue is calculated as:

$$high_{pri_{wt}} = \begin{cases} high_{wt_{init}} highavg_{qsize} \epsilon \{0,0.833\} \\ \frac{(0.3)*\left(high_{New_{Avg}} - high_{Prev_{Avg}}\right)}{(highMax_{Th} - highMin_{Th})} + highPrev_{wt} highavg_{qsize} \epsilon \{0.833,3.667\} \\ high_{upper\_value} highavg_{qsize} \epsilon \{3.667,10\} \end{cases} \quad (4)$$

Bandwidth for medium precedence queue is calculated as:

$$med_{pri_{wt}} = \begin{cases} med_{wt_{init}} medavg_{q_{size}} \epsilon \{0,60\} \\ \left(1 - high_{pri_{wt}}\right) medavg_{q_{size}} \epsilon \{61,100\} \end{cases} \quad (5)$$

med_wt_{init} is alloted to 0.3 if the average buffer length of medium precedence queue is from 1 to 60, for more than 60, $med_{pri_{wt}}$is allocated as variation in total bandwidth and the bandwidth assigned to high precedence application. Lastly, if the weight is not utilized completely from higher and medium queues, it gets allocated to low precedence data. The weight of low priority queue can be obtained as (3).

To evaluate the performance of this second packet scheduling scheme (model-2), we compared the results with first model. We conducted several experiments with message divided into fixed size packets and variable packet sizes. The comparison of high priority weights, medium priority weights and low priority weights calculated in both first and second scheme for different experiments are shown Fig.3, and 4 and 5. Fig.3 show the weights allocated in models-1 and 2, for simulation scenario discussed in Table III (Exp1(a) and 2(a)). We observed that in model-2, high priority weights are even more smoothly distributed as compared to model-1 in a situation where average but continuous high priority data packets keep arriving to the centralized server in a network. In this model, we tried providing better performance to medium priority packets as compared to first model (to satisfy soft deadline). When medium priority data is less, weights allocated to medium priority are almost similar in both models. If we increase medium data (Fig.4) to double (Exp1(b)), more bandwidth (weights) is allocated to medium data and hence model-2 will provide better services to medium priority data including high priority data but at the cost of low priority packets. We conducted another experiment (Exp 3(a) and 2(a)) to realize how packet sizes manipulate the allocation of weights. We noticed that if packet sizes are same and bigger than 100 bytes, it consumes more bandwidth as compared to variable packet sizes of smaller length under similar simulation environment with this scheme. In Fig.5 and Fig.3, we can observe the difference. For the same data rate if packet size is variable (Fig.3) the weights are allocated more evenly as compared to same packet sizes in Fig.5.

To validate our results, we compared our models among each other and with the standard adaptive weighted scheme [16] in terms of weight allocation. The packet lost is mainly dependent on buffer size. If the buffer assigned for a particular service is full, packets are lost. To analyze how buffer management is important for different priority mechanism we developed these models with slight modification in buffer sizes.

IV. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

In order to verify the performance of the proposed QoS-models, a sequence of experiments were conceded using MatLab R2013. The accomplished results are presented and discussed here. We focused mainly on two essential parameters required for providing QoS to delay and loss sensitive applications in IoT environment namely packet lost ratio and waiting time of packets. In order to record excellent performance, different experiments were conducted by modifying data rates of different services, buffer sizes and categorizing different applications according toits QoS requirements. IoT has a wide research possibility in numerous fields like transportation, healthcare, smart environments like smart home, smart city, and structural health monitoring etc. In all these applications lightweight intelligent objects are active participants which are proficient of sensing discrete events and transfering it to various other devices. This generally consist of a memory device (a few tens of kilobytes), a CPU (8, 16 or 32 bit microcontroller) and a wireless communication device of low power (appr. a few Kbps to few 100Kbps). Many IoT technologies include sensors, RFID and Bluetooth technologies which all have low data rates. The data rates vary depending upon the manufacturers or consumers point ofview. The data transfer rates of RFID vary from 26Kbps to 424Kbps. Bluetooth operates at a rate which is not more than 1Mbps. IEEE 802.15.4 transceivers in sensor networks, operate at a maximum raw data rate of 250Kbps. Security surveillance cameras in an intelligent home requires a data rate of 2 to 3Mbps. At the manufacturer's side, data from million devices are aggregated and the data rate will be significantly different.

To compare the performance of different models and

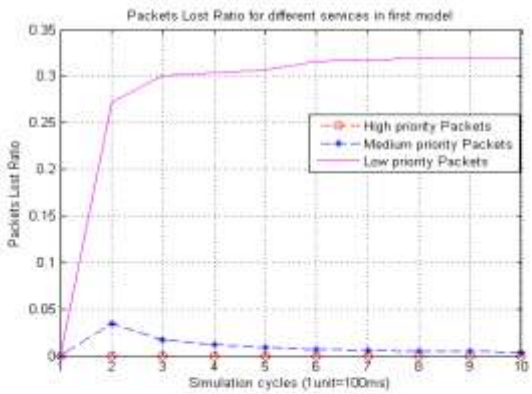

Fig. 6. Packet Lost Ratio for different services in First Model (Exp1(a))

depicting their feasibility through range of IoT applications we first categorize different IoT applications [18][19] according to its QoS requirements. Table IV provides necessary details on different IoT applications along with data rates and priority assignments. Accordingly we have divided the traffic into high, medium and low traffic and performed certain experiments based on the data rates mentioned in Table III and compared different models according to their suitability in different applications.

In Experiment 1, 2, 3 and 4 we reflect on sensors urgent data as higher priority data. Poisson arrival method can be assumed for data model to each individual sensor node. So, for higher, medium and lower precedence services, Poisson traffic with variable sizes packets need to be considered. In Exp-4, high priority service class in which packets are expected to be delivered in time are taken as video packets. MPEG format is chosenso that CCTV captured video images can be usually altered into MPEG format and is generally used. Emergency service class includes all non detrimental applications such as peer to peer applications. Query initiation scheme is an example of peer to peer application and can be modeled as Poisson data. Packets are segregated as higher, medium and low precedence packets, and are accumulated in dissimilar queues of length 10,100,100 in Model-2 and 3 (6,100 and 100 in Model1 and 4) respectively. Lastly they are scheduled using planned packet scheduling scheme. In the simulation environment considered, we divide the data into three applications in which high priority service classes contain packets that are expected to be delivered on time and are considered as video packets. Best efforts service class include every non detrimental services like peer to peer services. As Query instigation

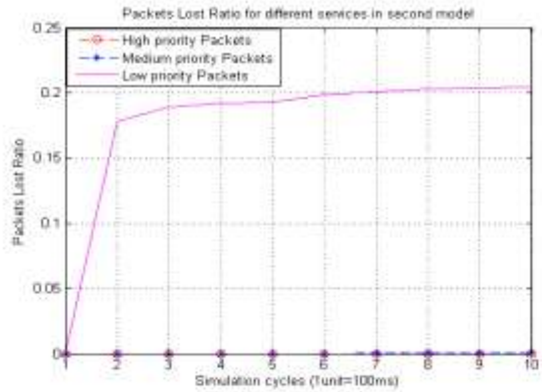

Fig. 7. Packet Lost Ratio for different services in Second Model

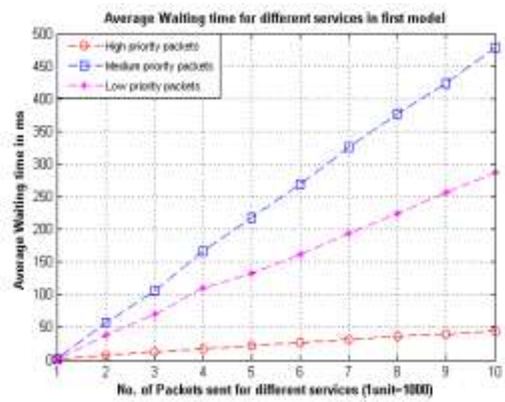

Fig. 8. Average waiting time for different services in first model (Exp2(b))

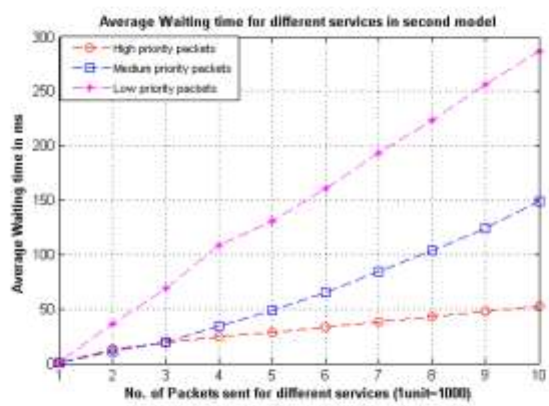

Fig. 9. Average waiting time of high priority packets (Exp2b)

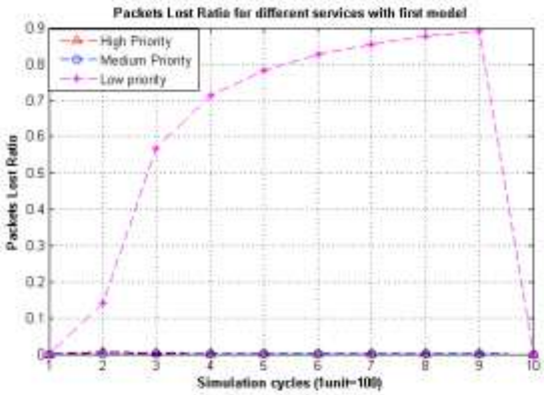

Fig. 10. Packet lost ratio in first model for Exp 4(b)

model is an example of peer to peer service and could be represented like Poisson data. Hence, lower precedence data can also be represented like Poisson traffic by means of uneven packet sizes and for variable system loads. Features of different scenarios are specified under Table III.

A. Comparison on Packet Lost Ratio for Different Services in Different schemes

At certain condition, if buffer queue is full, the packets which are entering that buffer will be dropped and this can be computed by Average packet dropped ratio (APDR). APDR for all queues is calculated as:

$$Average\ Packet\ Dropped\ Ratio = \begin{cases} \frac{\sum_{c=1}^{N}(q_c^1 - B1)}{\sum_{c=1}^{N} q_c^1} & if\ q_c^1 > \ B1\ and\ n = 1 \\ \frac{\sum_{c=1}^{N}(q_c^n - B2)}{\sum_{c=1}^{N} q_c^n} & if\ q_c^n > B2\ and\ n = 2\ or\ 3 \end{cases} \quad (6)$$

Fig.6 and7 shows packet lost ratio in first and second model respectively for Exp. 1(a) where we observe that high priority packets has no loss in both models, approximately 2% loss of medium priority in first and zero loss in second model and 30% loss of low priority packets in first and 20% packet lost of low priority packets in model-2.

B. Comparison of waiting time for different services in different schemes

Delay can be increased if packets are stored in the buffer for more time. So, emergency services should be stored for least time interval to attain less overall delay and to solve the difficulty of delay jitter.

For WRR scheduling scheme, Average waiting time (A_n) of data packets in n^{th} buffer can be calculated as[20]:

$$A_n = \frac{\left[(\mathrm{Total}_{\mathrm{link}_{\mathrm{BW}}} - (BWn_{link})) * (\mathrm{avg}_{\mathrm{pkt}_{\mathrm{sz}}})\right]}{\mathrm{Total}_{\mathrm{link}_{\mathrm{BW}}} * 2(1 - \left(\frac{\rho_n}{\mu_n} * \frac{(BWn_{link})}{\mathrm{Total}_{\mathrm{link}_{\mathrm{BW}}}}\right)) * \mathrm{Total}_{\mathrm{link}_{\mathrm{BW}}}}$$

(7)

where BWn_{link} is the bandwidth consumed through n^{th} queue to send packets which is computed based upon suggested algorithm, $\mathrm{avg}_{\mathrm{pkt}_{\mathrm{sz}}}$is average size of scheduled packet; $Total_{link_{BW}}$ is bandwidth of link used, ρ_n is entrance pace of packets in n^{th} queue, μ_n is service rate of packets leaving the n^{th} queue.

Fig.8 and 9, compares waiting time of all applications packets in both models which provides us a clear detail that waiting/stored time for most prior service is lesser than other services in all models because of allocating priorities to those packets.

For IoT applications where video file is taken as high priority file Exp-4 is suitable. We have plotted several graphs for weights allocated, packet lost ratio and waiting time to verify the models performance when high data rate applications are running. There is no loss of medium priority packets in first model (Fig.10) and low priority lossesare intolerable but reduces as soon as the high and medium packets data rate is reduced. On the other hand, little losses for medium priority packets can be seen in second model but low priority losses are reduced drastically (Fig.11). The finalcomparison of models and various applications is discussed in Section V.

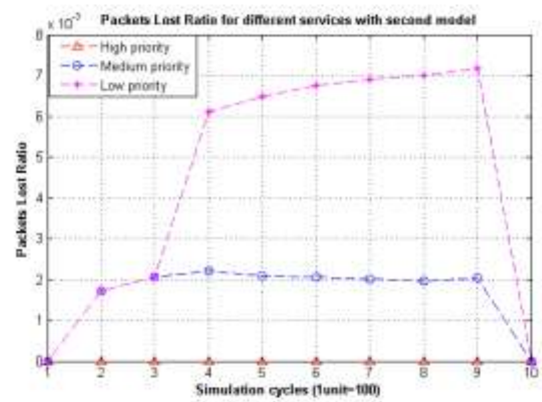

Fig.11. Packet Lost ratio in second model for Exp 4(b)

TABLE IV
QoS Parameters at different IoT layers

Application Domain	Application	QoS Requirements	Type of service class (high, medium and low priority)	Data rates
Transportation and logistics	Assisted Driving	Requires continuous flow of data with guarantees	High	2-3mbps
	Logistics	Either interactive or non-interactive and in many cases require soft real time guarantee.	Medium	250 to 500Kbps
	Value added advertisement	Not critical and doesn't require any real time data.	Low	0.1mbps
	Digital map downloading	Does not need real time information.	Low	0.1 mbps
Smart Environment	Smart homes and offices (emergency theft or fire condition)	Doesn't require real time guarantees in normal conditions but in some situations like alarms for fire or theft or break into a smart home/offices makes it emergency data .	High	2-3Mbps
	Smart museum and gym	Does not need real time guarantees	Low	250Kbps
	Industrial plants	Requires soft real time guarantees	Medium	100-900Kbps
eHealth Services	Tracking	Requires continuous flow of data of medical condition of different patients	High	64Kbits to 1Mbit
	Identification and Authentication	It is interactive data and doesn't require real time guarantees but sometimes soft real time guarantees may be required due to security requirement of patient's data.	Medium	26kbps to 424Kbps
	Sensing	Requires continuous flow of data	High	250kbps
	Data Collection(RFID)	Data collected from health surveys, health plans etc. It is not real time.	Low	26kbps to 424Kbps
Emergency eHealth services (Teleconsultation in accident situations)	Audio	Audio conferencing among patients and doctors	High	4-25Kbps
	Video	Video (H.263 encoded) streaming of victim to specialized doctors	High	32-384Kbps
	ECG	Real time transfer of victims physiological information (ECG data, sugar level, BP)	Medium	1-20Kbps
Personal and social	Social networking	This application is interactive and doesn't need any real time guarantees	Low	2-3Mbps
	Historical queries (RfID)	This application is interactive and doesn't need any real time guarantees	Low	26Kbps to 424Kbps
	Losses and Thefts	Needs real time guarantees	High	250Kbps

V. IOTAPPLICATIONSAND PROPOSED SERVICE MODELS

In next part, we conclude with the best utilization of proposed models in supporting different IoT applications. Different IoT applications with its standard data rates are summarized and listed in Table IV. It also recapitulates, categorization of data as high , medium and low. based on their QoS requirements and data rates. The similar standard data rates are considered in Table III for plotting various graphs. Based on our study and observations from different experiments, for each type of application i.e transportation and logistics, smart environment, e-Health services, emergency e-Health services and personal & social, we suggest suitable model as shown in Table V.

TABLE V
SUITABLE MODEL FOR DIFFERENT IOT APPLICATIONS

Models	Transportation and Logistics	Smart Environment	e-Health Services	Emergency e-Health Services	Personal and Social
Proposed Model 1			Yes	Yes	
Proposed Model 2		Yes	Yes	Yes	Yes

The general overview is that if overall (high+meidum+low) datarate is same and in that high and medium priority data rate is more, second model gives good results. If high priority data is bursty and very high, third model provides best results. For average high data rate all three models can be applied. Based on these results we can conclude the use of different models in different IoT applications as shown in Table V.

VI. CONCLUSION

In this paper, novel QoS-aware schemes have been proposed. The schemes provide service classification and offer QoS to crisis services in IoT. Investigation and model results show that the schemes are effective in attaining dynamic service classification in IoT applications with different QoS parameters. We categorize diverse IoT applications based on their QoS requirements, data rate and considering the performance of all services in total and defined best suited model which is feasible for each application. We firstly investigated different QoS models suitable for delay and loss sensitive applications that are involved in IoT. Then we proposed three different models with three priority levels each of which aims to enhance the QoS in IoT. All three models are effective in providing services to high priority delay and loss sensitive applications but the difference lies in saving packet losses from medium and low priority applications i.e.to provide superior treatment to real time applications without causing much degradation to the performance of other network traffic.

REFERENCES

[1] Vigna, Benedetto, "*Smart Systems for Internet of Things*" IEEE Conference on Design Automation & Test in Europe Conference & Exhibition, pp-1, 2013.

[2] Yuqi Wang, Hui Qi, "*Research of Intelligent Transportation System Based on Internet of Things Frame*" Scientific Research Journal on Wireless Engineering and Technology, Vol. 3, pp 160-166, May 2012.

[3] Anurag, Moosavi S.R, Rahmani, "*Pervasive health monitoring based on Internet of Things: Two case studies*", IEEE EIA 4th International Conference on Wireless Mobile Communication and Healthcare, pp 275-278, Nov. 2014.

[4] Souza, Alberto. M.C, Amazonas, Jose R.A, "*A Novel Smart Home Application using an Internet of Things Midleware*", IEEE Proceedings of European Conference on Smart Objects, Systems and Technologies, pp 1-7, 2013.

[5] Ling Zheng, Shuangbao Chen, Shuyue Xiang, Yanxiang Hu, "*Research of Architecture and Application of Internet of Things for Smart Grid*", IEEE International Conference on Computer Science & Service System, pp 938-941, 2012.

[6] Klepec B, Kos A, "*Performance of VoIP applications in a simple differentiated services network architecture*," IEEE international conference Eurocon 2001 at Bratislava, Slovakia, Vol 1, pp-214-217, 2001.

[7] Jia-Ming Liang, Jen-Jee Chen, Hung-Hsin Cheng, and Yu-Chee Tseng "*An Energy-Efficient Sleep scheduling with QoS consideration in 3GPP LTE advanced networks for Internet Of Things*", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol 3 Issue 1, pp13-22, March 2013.

[8] Irfan Awan and Muhammad Younas, "*Towards QoS in Internet of Things for Delay Sensitive Information*", MobiWIS 2013 workshops, CCIS 183,, Springer International Publishing Switzerland pp-86-94, 2013.

[9] Chien-Liang Fog, Christine Julien, Gruia-Catalin Roman, Chenyang Lu, "*Challenges of Satisfying Multiple Stakeholders: Quality of Service in the Internet of Things*" SESENA ACM, Waikiki, Honolulu, USA, May 2011.

[10] Xianmin Geng, Aiwen Luo, Zhinjun Sun, and Yu Cheng, "*Markov Chains Based Dynamic Bandwidth Allocation in Diffserv Network*" IEEE communication letters, Vol. 16, (10), pp-1711-1714, 2012.

[11] Jiong Jin, Jayavardhana Gubbi, Tie Luo, Marimuthu Palaniswami, " *Network Architecture and QoS Issues in the Internet Of Things for a Smart City*", IEEE International Symposium on Communications and Information Technologies (ISCIT), pp-956-961,2012.

[12] Marie-Aurelie Nef, Leonidas Perlepes, Sophia, Panayotis K. Kikiras, " *Enabling QoS in the Internet of Things*" Proceedings of Fifth International Conference onCommunication Theory, Reliability and Quality of Service (CTRQ), Mont Blanc, France, pp-33-38, 2012.

[13] Saima Abdullah, Kun Yang, "*A QoS aware message scheduling algorithm in Internet of Things environment*" IEEE Online Conference on Green Communications" , pp-175-180, Oct. 2013.

[14] Chien-Liang Fog, Christine Julien, Gruia-Catalin Roman, Chenyang Lu, "*Challenges of Satisfying Multiple Stakeholders: Quality of Service in the Internet of Things*" SESENA ACM, Waikiki, Honolulu, USA, May 2011.

[15] Nourhene Maalel, Enrico Natalizio, Abdelmadjid Bouabdallah, Pierre Roux, Mounir Kellil, "*Reliability for emergency applications in Internet of Things*", IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS), pp 361-366, May 2013.

[16] Haining Wang, Chia Shen, Kang G Shin, "*Adaptive Weighted Packet Scheduling For Premium Service*", IEEE International Conference on Communications, Vol 6, pp. 1846-1850, 2001.

[17] Mirjana Stojanovic, Vladanka Acimovic-Raspopovic, "*A Novel Approach for Providing Quality of Service in Multiservice IP Networks*", Facta Universitatis: Series " Electronics and Energetics, pp. 261-274, 2004.

[18] L.Atzori, A.Iera, G.Morabito, "*The Internet of Things: A Survey*," Computer Networks Journal 2010, Springer.

[19] Marie-Aurelie Nef, Leonidas Perlepes, Sophia, Panayotis K. Kikiras, " *Enabling QoS in the Internet of Things*" Proceedings of Fifth International Conference onCommunication Theory, Reliability and Quality of Service (CTRQ), Mont Blanc, France, pp-33-38, 2012.

[20] Valadimir Hottmar, Bohumil Adamec, "*Analytical Model of a Weighted Round Robin Service System*", Hindawi Publishing Corporation, Journal Of Electrical and Computer Engineering, pp.1-6, Vol. 2012.