IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. V (Jan. 2014), PP 51-55
www.iosrjournals.org

A Study on Quality of Service for Computer Networks

D. Rama Krishna Reddy[1], D. Hemalatha [2], Azmath Mubeen[3]
[1] *Dept of Mathematics, University College of Science, Osmania University, Hyderabad, India.*
[2] *Dept of Computer Science, Osmania University College for Women, Hyderabad, India.*
[3] *Dept of Computer Science, Osmania University College for Women, Hyderabad, India.*

Abstract : *This paper presents some of the basic concepts of Quality of Service. The major research areas of Quality of Service for Computer Networks are represented. The paper also correlates and compares few of the current and evolving and popular Quality of Service Routing techniques.*
Keywords: *GoS, QoS, QoS Routing.* *Keywords*

I. INTRODUCTION

The network Quality of Service (QoS) is defined as: "The capability to control traffic-handling mechanisms in the network such that the network meets the service needs of certain applications and users subject to network policies and strategies". To provide and afford the abilities, capabilities and potentiality of measure and control required by either definition, QoS networks should have methods and techniques to control the allotment of resources among applications and users.

The notion of QoS has come as a response to the new challenges and requests imposed on the network performance by novel applications, especially and specifically multimedia real-time and online applications. These real time and online applications made it compulsory to set constraints on what can be defined as an admissible time delay when routing or sending information over a network. Those time challenges and demands are classified into three main categories.

-The first is the subjective human needs for interactive computing such as chatting sessions and other interactive web applications.

-The second is the automated assignments under time restrictions such as the automatic once-per-day backups during a limited pre-assigned time span.

-The third category is the necessity of few applications for a transmission rate with limited time jitter along with a temporal arrangement of the transmitted packets. This is the case when transmitting or streaming multimedia data over a network. The transmission rate is required to keep the transmitted material meaningful and apparent while the preserved temporal order is necessary for synchronization.

The temporal requirements presented above are constitutional to QoS that some references define QoS in terms of those requirements. In the Webster's New World Dictionary of Computer Terms the QoS is defined as "the guaranteed or assured data transfer rate". The word "guaranteed or assured" is of special importance since QoS can only be implemented through guarantees on the limitations of few network parameters as will be explained below.

It is significant here to annotate that although QoS became an issue only in the past few years, but the idea of QoS had been anticipated earlier before modern applications mandated the use of QoS. In the early or initial Internet Protocol (IP) specification, a Type of Service (ToS) byte is reserved in the IP header to facilitate QoS. Until the late 1980s, almost all IP implementations ignored the ToS byte since the want for QoS was not yet clear.

Comparison of GoS and QoS

It is a difficult work to find the GoS(Grade-of-Service) standards required to support an assured QoS. This is because the GoS and QoS notions have varied viewpoints. the GoS takes the situation from the network point of view, while on the other hand the QoS views the situation from the customer's point of view.

Citation Diagrams

In order to obtain an overview of the network under consideration, it is often necessary to create a so-called citation diagrams. This consists of one or few simplified drawing(s) of the path or route or a call (or connection) can take in the network including acceptable, suitable and appropriate reference points, where the interfaces between entities are defined.

Consider a telephone network with terminals, subscriber or end user switches and transit or traverse switches. In the example the signaling network is ignored. Suppose the call can be routed in one of three ways:

1. Terminal → subscriber switch→ terminal

This is drawn as citation diagram shown in Fig. 1.1.

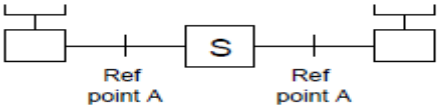

Figure 1.21: *Reference configuration for case 1.*

2. terminal → subscriber switch→ transit switch→ subscriber switch → terminal

This is drawn as a citation diagram shown in Fig. 1.2.

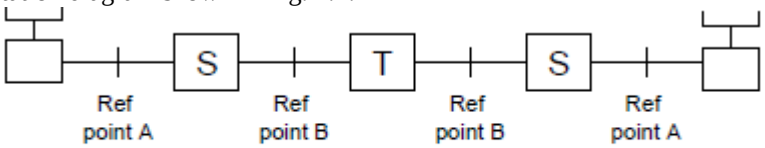

Figure 1.22: *Reference configuration for case 2.*

3.terminal→subscriber or switch transit switch transit switch subscriber switch terminal

This is drawn as a citation diagram shown in Fig. 1.3.

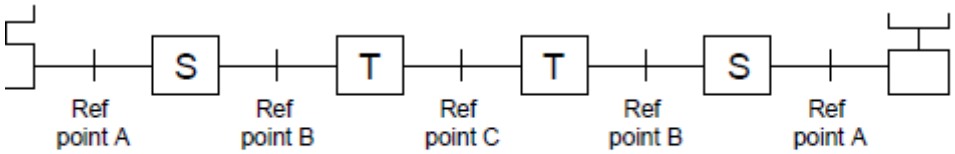

Figure 1.23: *Reference configuration for case 3.*

Based on a given set of QoS requirements, a set of GoS parameters are chosen and defined on an end-to-end basis within the network boundary and limitations, for each major or main service category provided by the network. The preferred and selected GoS parameters are specified or expressed in such a way that the GoS can be derived at well-defined reference points, i.e. significant or denoted points in the network traffic. This is to allow the partitioning or fragmenting of end-to-end GoS objectives to obtain the GoS objectives for each network stage or component, depending on some well-defined reference connections.

II. QoS Performance Measures

In order to provide QoS, some quantitative calculations of that constitute QoS must be described. As mentioned above, QoS is quantitatively defined in terms of guarantees or limits on certain network performance criteria and principles. The most general network performance criteria are: bandwidth, packet delay and timing jitter, and packet loss.

2.1 Bandwidth

The phrase bandwidth defines the transmission capacity of an electronic line. Theoretically, it describes the range of possible transmission rates, or frequencies. In general practice, it describes the size or width of the wire or cable that an application program requires to transmit over the network .The importance of a channel or cable bandwidth is that it arbitrates the channel capacity, which is the maximum data rate that can be transmitted. The relationship between channel capacity and data transmission rate was specified in the Information Theory of Claude Shannon in the 1940s.

According Shannon's channel capacity theory, if data transmission rate is R and channel capacity is C, then, it is always possible to find a method to transmit data with arbitrarily low probability of error provided R ≤C and, conversely, it is not possible to find such a technique if R > C.

2.2 Packet Delay and Timing Jitter

The delay, also known as latency, consists of three different types of delays : serialization delay, propagation delay, and switching delay.

- Serialization delay, also called transmission delay, is the time taken by a device to synchronize a packet on a precise output rate. This transmission delay is a function of the bandwidth and the packet size. For example, a

packet with size of 64 bytes would take 171 μs when transmitted at the rate of 3Mbps. The same packet would take 26 ms when transmitted at the rate of 19.2 kbps.

-Propagation delay is the time taken by a bit to travel from a source device to destination device. According to the physics theory, the propagation delay is dependent on speed of light (and it is 3x10-8ms). So the propagation delay is the speed of bit, making at best a fraction of the speed of light. Hence, propagation delay is a function of the distance that the bit traveled through the communication link or medium.

-Switching delay is the time drag between packets received and retransmits it. The switching delay is a function of the device speed.

In addition to these three types of delay, other delays also contribute and support to the overall performance and accomplishment of the network. Depending on the traffic, network circumstance, network topology and the nature of the data or information being transmitted, various packets will experience various delays. The term packet jitter refers to this variation in packet delay.

When the network is congested or clogged, queues will develop up at the routers and start affecting the end-to-end delays. Queuing delay may be trivial when the network is fast and not experiencing congestion. However, when the network is congested or blocked up, the queuing delay grows and becomes significant. The number of clients in a queue is a random variable and its distribution depends on r, the ratio of arrival rate to service rate. The probability p of having n clients in the queue is calculated as follows:

$$P(n)=(1-r)*rn \quad (1)$$

The queuing delay is a function of the number of packets in the queue and the service time for each queue. When the service rate is μ, the average queuing delay AQD can be calculated as follows:

$$AQD = 1/(1-r)\ \mu \quad (2)$$

2.3 Packet Loss

Packet loss is another essential QoS performance measure. Some applications may not work or perform properly, or sometimes may not work at all, if the packet loss exceeds a specified number, or rate. For example, the video streaming may become waste , after certain streaming video frames are lost. This number may be zero in certain cases. Therefore, certain assurances on the number of rate of lost packets may be required by certain applications for QoS to be considered. Packet loss may occur because packet is deleted or dropped at congestion points when the number of packets arriving significantly exceed the size of the queue. Corrupt packets on the transmission channel or cable can also cause packet loss.

III. QoS Levels

There are still many applications that do not require any QoS, even though after realizing the unavoidable need to QoS networking.. Moreover, those applications that do require QoS vary in the degrees of priorities and assurances that they require to implement QoS. Therefore, on the other extreme we have tasks and applications that do not need any guarantees. On the other extreme, we have tasks and applications that need complete assurances that may not be agreed or compromised.

In between those two extremes there are abundant levels of QoS. However, those levels of QoS have been clustered into three main categories: best effort service, soft QoS, and hard QoS.

3.1 Best effort service

This level of service does not provide any assurances at all. It expresses the first extreme mentioned above. It may not actually be considered a QoS. Many or maximum network applications work very well in a satisfactory manner with best effort service. An example of such similar applications is the File Transfer Protocol (FTP). No assurances or performance measures are expected for FTP. The only criteria is whether the transfer was completed successfully or not.

3.2 Soft QoS

This QoS is also known as distinguished or differentiated service. In this QoS level, complete assurances are not given. Rather, different priorities are alloted to different tasks. Hence, applications are clustered into various classes of priorities. Many application traffics work very well with this policy of complete assurances are not required. For example, network control traffic should always be given higher priority over other data transmission to ensure the availability of the basic connectivity and functionality at all times.

3.3 Hard QoS

Hard QoS is also called assured service. It describes the level of QoS for applications that require complete assurances on the minimum required resources of the network in order to work correctly. Prior network resource reservation over a path is usually performed to enable the network to provide, or deny, the required assurance. Example of applications that require Hard QoS include multimedia applications, where audio and/or video data streaming is done in real-time. It is necessary to note that here internet currently can provide only one level of

QoS that is first level, though the IP protocol supports other levels of QoS but still the Internet does not offer any of the other two levels of QoS. However, the other networks such as ATM networks do support other two levels of QoS .

IV. QoS Routing

The emergence of QoS networking created many challenges to network developers of many fields. It is true that physicists and engineers contributed significantly to the development of faster networks.

On the new prospect now we have optical networks that use Wavelength Division Multiplexing (WDM) technology to increase the capacity of optical transmission systems by transmitting multiple wavelengths over a single fiber, reaching transmission rates are of terabits per second. However, as the physical abity and competence of the networks grow, the demands by new applications to exploit those ability and proficiencies also grow. This necessitates the need for computer scientists to constantly develop algorithms and solutions to provide the required exploitation. Numerous aspects or factors entangle the difficulty or problem of QoS routing. One of those factors or aspects is the diversity of the requirements and guarantees of different distributed computing applications running concurrently. This problem amplifies and increases to include applications with minimum or zero constraints requirements, making it compulsory to construct or develop routing techniques and mechanisms that handle all the three levels of QoS presented above. The other crucial factor is the problem of maintaining accurate network state information in a big dynamically changing network. This later factor will be a very important subject of matter in the current dissertation or thesis work. Each node in the network has to maintain its local state in order to maintain network state information. All local states can be combined to form the global state information. Typically, a node or an intermediate device maintains the network global state information using one of two algorithms: link-state algorithm and distance-vector algorithm. This is done by using the selected algorithm to exchange the local states between all nodes in the network on periodical or regular basis. The resulting global state information may not be accurate due to lot of factors that will be discussed later. Dealing with this uncertain and indeterminate network global state information is one of the major problem this dissertation is trying to solve.

4.1 Routing Classification

QoS routing algorithms can be categorized according to the main destination of the searched path into two main categories: unicast routing algorithms and multicast routing algorithms.

4.1.1 Unicast Routing

In a unicast routing algorithm, the problem is to find the best attainable and feasible route from a source node to a destination node, satisfying a pre-designated or pre-nominated set of coercions or constraints.

4.1.2 Multicast Routing

In multicast routing, the problem is to find the best attainable and feasible tree that covers a source node and a set of destination nodes, satisfying a pre-designated and pre-nominated set of coercion or constraints. The essence of the problem forces or impels the use of algorithms from one category or the other with no trade-off between the two.

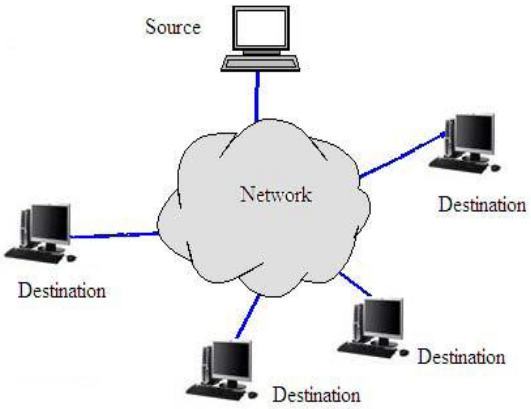

Figure 1.4 an example for multicastouting

The other way of classifying QoS routing algorithms is based on path search and deployment policy. There are main three routing approaches or algorithms: source routing, distributed routing, and hierarchical routing. These three approaches or algorithms can be used in exchangeable fashion or interchangeably in many cases in keeping view of trade-offs between the advantages and disadvantages of each approach.

4.1.3 Source Routing

In source routing, the attainable or feasible route is calculated locally at the source node, which is expected or predicted to have its own, or personal maintenance method or performance of the network global state information. The source routing provides or facilitates localized storage of the network state information and the centralized computation of the path; this is the main advantage of source routing. The local maintenance of the global state enables the source node to calculate the path locally as well. The computational complexity in this case is much smaller or less than that of the distributed computing. This makes source routing algorithms easy, simple and comfortable to design and implement. In addition, it guarantees or ensures loop-free routing. However, source routing suffers with two major problems. The first one is the inaccuracy and unreliable of state information. The degree of precision of the global state at each node is directly proportional to the frequency of updates. Nevertheless, the updating frequency is also directly proportional to the updating overhead, which is inversely proportional to the availability of resources for the actual network activities. This unavoidable imprecision in the global state information may result in the failure of finding an existing feasible or attainable path.

4.1.4 Distributed Routing

The second routing strategy, distributed routing, rely on using distributed computing to calculate the path. The computation is done by exchanging global state information and the control messages stored or reserved locally at each node. Nevertheless, some distributed routing algorithms may not require the maintenance of global state of information at all .The major advantage of distributed routing is the distributed computation of the route between source and destination, which enables minimum response time and better and maximum scalability(ranking and progression). The shorter response time and better and maximum scalability are achieved at the expense of higher network traffic due to more message swapping and exchanging. Furthermore, distributed routing cannot be loop-free, especially when unbounded and global states at various nodes are inconsistent.

4.1.5 Hierarchical Routing

In this, hierarchical routing, nodes are combined into clusters, which are furthermore or still grouped into a higher level clusters. This recursive clustering continues to assemble up or build up to form a multi-level hierarchy. Instead of maintaining and conserving global state information at each node, the aggregated or accumulated state of information is maintained, where an elected or designated node in each cluster or group maintains the global state of the nodes in the cluster in which it is local to in addition to the aggregated or accumulated states of the other clusters. The use of partial or incomplete global states maintained by logical nodes enhances or improves the scalability of the hierarchical routing significantly when compared with other routing schemes. In addition, the overall traffic in the network does not be as excessive as in distributed routing. Thus, hierarchical routing combines advantages of both source and distributed routing. The only noticeable or blatant problem in this hierarchical routing, is this is not a trivial one, means that the aggregation or accumulation of the network states introduces or inserts additional imprecision or indeterminate state.

V. Conclusion

This paper concludes that the performance, functioning and efficiency of routing algorithms which are not designed specifically to take imprecision and indeterminate features into account degrades or deteriorates significantly as the imprecision increases or grows. Most QoS routing algorithms available today do not take this uncertainty and ambiguity into consideration. Instead, they expect and believe that this imprecision does not exist, regardless of the basic and congenital nature of this uncertainty and ambiguity. Even though, research has been done to estimate and evaluate the impact of neglecting this uncertainty on the performance of different routing algorithms. In addition, fewer or some routing algorithms have been proposed with the main objective of handling the intrinsic and basic imprecision and reducing and lessening its effect.

REFERENCES

[1] X. Yuan and W. Zheng, "A Comparative Study of Quality of Service Routing Schemes That Tolerate Imprecise State Information," Florida State University Computer Science Department, Technical Report. [Online].Available:http://websrv.cs.fsu.edu/research/reports/TR-010704.pdf.
[2] "Quality of Service Based Routing: A Performance Perspective", ACM SIGCOMM, 1998.
[3] Chao Peng, Hong Shen, "New Algorithms For Fault-Tolerant QoS Routing",submitted to the International Conference Dependable Systems and Networks(DSN-2006)
[4] S. Chen and K. Nahrstedt, "An Overview of Quality of Service Routing for Next-Generation High-Speed Networks: Problems and Solutions,".
[5] "Predictive routing to enhance QoS for stream-based flows sharing excess bandwidth" Xun Su, Gustavo de Veciana.