QoS Analysis in Data Network: Stability, Reliability, QoS Invoke Rate Perspectives

P K Mishra, Rameshwar Nath Tripathi and Y N Singh
Indian Institute of Technology, Kanpur
pkmishra@iitk.ac.in , rameshn@iitk.ac.in, ynsingh@iitk.ac.in

Abstract---In a data network, a user demands a service and network offers the service. Every user requires varying degrees of service quality. To fulfil the user demand, network must have service quality support mechanism. Internet offers best effort service, but many users demand some kind of guarantee on theservice. Therefore, there is a need of analysis of QoS in the network. This motivates our work to investigate and analyse QoS from different perspectives. QoS can also be interpreted as measure of service quality that the network offers to the user or application. This analysis can be performed in qualitativeor quantitative fashion. QoS can be analysed from different perspectives. In this paper, we present a qualitative aspect of QoS analysis and we do not assume anything about QoS demand of the user. So, we investigate the situations which demand QoS provisions within the network. We also determine the boundary condition on the packet injection rate which is called QoS Invoke Rate (QIR). Below this QIR, it is safe to operate the network without worrying about QoS provisioning in the network. Above the QIR it is desirable to invoke QoS provisions, because limited resources start playing their role. Our main contribution in this paper includes packet injection rate condition for which a network will be stable and reliable while satisfying the user's QoS demand.We devise the boundary condition on packet injection rate (QIR) forQoS support in the network.

Index Terms---QoS, Stability,Reliability, QoS Invoke Rate

I. Introduction

A packet routing network facilitates end to end delivery of information by breaking it into small pieces called packets. These packets are delivered from source to destination by moving them over the network using store and forward methodology. A user takes benefit of the communication services, information resources and entertainment over internet through various applications. These applications may demand single or combination of various contents like text, voice, video and images. Applications requiring combination of multiple contents are termed multimedia applications. The multimedia applications like HDTV, video conferencing and merchandise have a stringent QoS requirement. QoS is built from essential parameters of Band width, Delay, Jitter (delay variation) and Packet loss that an application desires for its content packets.

Besides, for a packet routing network stability and reliability are functional requirements. A network system is stable when number of packets always remains bounded in the network as system runs for arbitrary long period of timeand it is reliable if every packet is delivered across the network in bounded time. Stability and reliability contribute towards QoS in a overlapping manner. Stability covers bandwidth and delay and reliability deals with delay and packet loss parameters. We can well appreciate that overload conditions cause instability in networks. Because, The overload condition is characterized by higher input rate in a network. A network can also become unstable in under load conditions due to a queuing policy. For last two decades under load instabilities due to queuing policies havebeen shown to exist [1, 2, 3]. Reliability in packet delivery can be lost due to many reasons such as connection break, noise and overload. Unreliability can too occur purely due to starvation of packets in queues. Hence, stability and reliability investigation of a network is essential to determine feasibilityof QoS provisioning.

In a network QoS can be handled at different granularity of routing, admission control, resource reservation and scheduling. One mechanism to assure QoS is packet classification and prioritization for purpose of class based scheduling in time. Packets desiring same QoS make a class. These classes of packets are assigned some priority ordering in a queue dedicated to a forwarding entity called station. A station handling multiple classes of packets is termed multiclass station. In a general network, as packets wait in the queue, a scheduling policy like Nearest To Go (NTG) or First in First Out (FIFO) is chosen to pick and pass packets through the station one by one. For QoS guarantee a QoS scheduling policy like Fair Queuing, Weighted Average Queuing or SP/FIFO policy is used. The QoS Scheduling policy on a priority queue allows control over bandwidth and delay to a class of packets by giving quick service to highest priority class at any time. Eventually, by packet classification and prioritization a multiclass network system is implemented for provisioning of QoS.

A .**Movtivation**

The previous work on multiclass queuing networks by Kelly [5], Kumar and Seidman[6], Lu and Kumar [7], RebkoandStolyar [8] and Bramson [9, 10] are based on stochastic inputtraffic and service times and have used fluid model of networkas a tool of study. Thus the analysis suffers from limitationsof assumptions of traffic modeling and becomes less generalas compared to AQM [1, 2, 3, 4] AQM makes as many fewassumptions about the input traffic and service times in thenetwork as possible which makes it more general and elegant.

Marsanet. al. [4] discussed AQM for multiclass queuingnetworks with SP/FIFO policy but used tools of stochastic andfluid models for their results. Our set up differs in approachfor the analysis being based purely on AQM and uses a simplemulticlass station as network. SP/FIFO as QoS queuing policysignificantly resembles the approaches being considered forQoS provisioning in Differentiated Services [11, 12, 13, 14]for internet. Differentiated services or DiffServ is a computernetworking architecture that specifies a simple mechanism forclassifying and managing network traffic and providing QoSon modern IP networks. DiffServ can, for example, be usedto provide low-latency to critical network traffic such as voiceor streaming media while providing simple best-effort serviceto non-critical services such as web traffic or file transfers.

The SP/FIFO policy on a priority queue allows control overdelay to a class of packets by giving quick service to highestpriority class at any time. This ensures improved band widthand latency to a priority class in accordance with its QoSdemand.We note that little or no effort has been made towards QoSanalysis of a multiclass network under AQM. This motivatesus to analyze stability and reliability at a multiclass stationusing AQM. We investigate network stability and reliabilityunder conditions where network is not overloaded and packetloss is only due to starvation of packets in queues. Under AQMa network system is defined by a triple (G,A, Q). G is theunderlying directed graph of the network, A is a hypotheticaladversary injecting set of packets in network G at some rate rand Q is a queuing policy [1]. We use multiclass single stationas trivial multiclass network and SP/FIFO scheduling policyas it resembles realistic approach in diffserve. The adversaryis characterized by its rate. So, one might be interested inknowing if QoS mechanisms of classification and prioritizationneeds to be invoked for all adversarial injection rates or it canbe relaxed up to some critical rate up to which

QoS guarantee is not affected.

B. Contribution

In this paper we have constructed two class single station and a general multiclass single station as most elementarymulticlass networks. For multiclass packet classification wehave taken case of simplest multiclass i.e two class classificationof packets, for obtaining results. Then we extended theresults over general case of multiclass. Results are presentedin form of theorems and corollaries. We have shown that twoclass Single station network is stable and reliable against alladversaries of rate strictly below one. And an adversary ofrate one can make two class single station networks unreliable,while keeping the network stable. These results have been generalizedfor multiple classes. Then, we provide two resolving issue of unreliability in a stable multiclass station.

Finally we give QoS Invoke Rate for an acyclic network.The paper is organized as follows. In section II we discusssystem model to explain AQM and SP/FIFO and definitions.Section III presents stability and reliability analysis of twoclass and multiclass single station networks. In section IV weprovide two protocols for avoidance of unreliability of a stablemulticlass station. Section V gives the QoS invoke rate. Finally, we have concluded the paper with observations and future research directions for further work.

2.Definitions and Preliminaries

In this section we give out important definitions and conventions used for the study and representation. We formally define Stability and Reliability.

Definition 1.Stable System. A system is said to be stable when total number of packets remains bounded in the system as system runs for arbitrary long period of time.

Definition 2.Reliable System. A system is said to be reliable if every packet in the system experiences bounded delay. The reliable nature of network is called reliability.

Definition 3:Station.A network is a directed graph G(V,E). V is set of vertices or nodes of the network and E is set of directed edges connecting two nodes in the network. Each edge at its tail has a queue-server pair called station

Definition 4.Single class station.A station which is not capable of classification of packets is called single class station.

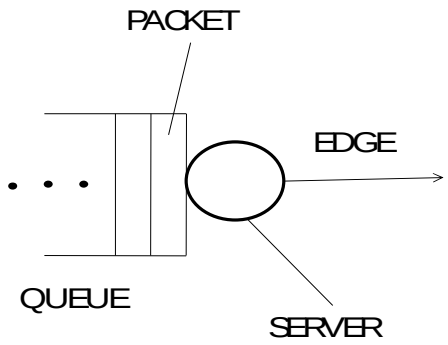

Figure 1. A Single class Station

Definition 5.*Multiclass station*.A station which is capable of classification of packet is called multiclass station.

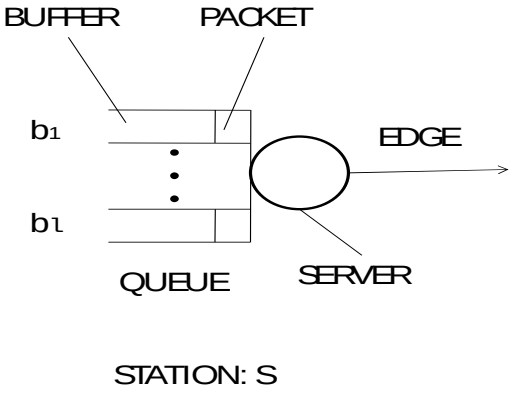

Figure 2. A Multiclass Station

3 .QOS INVOKE RATE

It has been discussed that to meet the stringent QoS demand of various services over packet routing network, the packetsentering the network from a source are classified into different classes at the entry node. This classification is done based onQoS demand. The packets of a class are treated in preferential manner by allocation of priority over the classes. Each priorityclass is assigned a buffer at a queue and a single server servers the queue. This is the basic description of a multiclass scheduling. Conclusively, a multiclass scheduling policy supports QoS by Classification and prioritization mechanism.

Our network system is (G, A, Q) that supports QoS . Where adversary A injects set of packets during any time step with injection rate r _ [0; 1) into the network G. Various packet classes demand different QoS. So, a multiclass queuing policy support the demanded QoS by classification and prioritization of packets. We use SP/FIFO multiclass queuing policy.The QoS mechanisms of classification and prioritization are resource consuming in terms of processing, time and energy.

At this juncture an intuitive question arises that does the system (G, A, Q) have to provide QoSsupport at all rates of adversarial injections or the QoS support mechanisms are to be invoked only above a certain adversarial injection rate?

If the critical rate of adversarial injection up to which QoS support mechanisms are not required to be activated is foundthen one can save on the committed resources. This critical rate is termed QoS Invoke Rate (QIR).

A**. Main Idea**

Parameters like bandwidth, delay, jitter, and packet loss constitute a QoS. Besides bandwidth, delay is another desirableparameter of QoS that comes before jitter and packet loss. As many packets contest for one edge, packets have to wait and delay becomes inadvertent. Thus delay is natural indication of congestion build up or backlog. Congestion in turn demands classification and prioritization of packets for priority treatment. Hence, for the purpose of determining QIR in a stable and reliable network system we choose singleparameter of delay. The delays experienced by a packet in a network are of two types. One is propagation delay thatis equal to time taken to traverse the path. The second is waiting delay that is the total time spent by a packet waiting in queues for servicing. Propagation delay is unavoidable unlike waiting delay. Waiting delay is function of backlog and prioritytreatment in a queue. Delay of a packet is the sum total of thepropagation and waiting delays. QoS is provided to packets byclassification and then offering different delays as per functionof backlog and priority. Hence, total delay of a packet too isa function of backlog and priority.

To evaluate QIR there is a need to develop two basic conceptsof Bottle neck topology and edge stress. These two help in determining maximum packet arrival at an edge in a timestep. Firstly we introduce bottleneck topology by reviewingconcepts of tree, directed tree, rooted tree and defining reverserooted tree. Secondly, we conceptualize edge stress of an edgein a network. Finally we determine QIR of an acyclic packetrouting network by determining maximum congestion at anyedge in the network and give out observations based on QIR.

B. Bottleneck Topology

Network topology is arrangement of network elements like vertices and edges. It essentially gives how data flows in a network. Two or more networks with different physical layouts can be topologically identical. A topology that limits the network performance by bottleneck formation is calledbottleneck topology.

Lemma 1 . There is a unique path from a node to the root in a reverse rooted tree.

Proof: A tree has a unique path from a node toanother node. The underlying graph of a reverse rooted tree is a tree. All edges are implicitly

directed towards root node.Therefore, there exists a unique path from a node to the root node.

Bottleneck Network: A bottleneck network is a reverse tree where the root node is connected by only one edge. The single edge incident to the root is called bottleneck edge. Possible bottleneck topologies are given in figure 3.

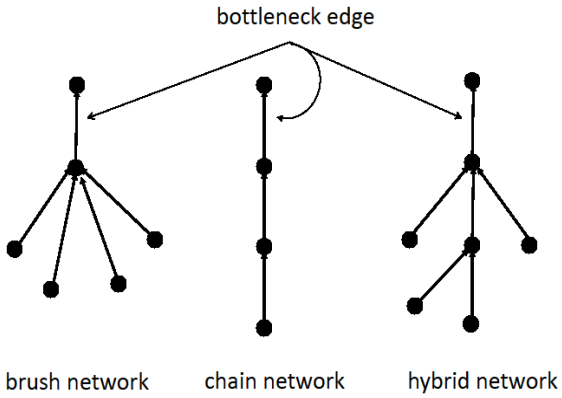

Figure 3: Three possible Bottleneck topologies

Lemma 2: Bottleneck edge is shared by exactly (n-1) paths in a bottleneck network of n nodes.

Proof: In a n node bottleneck network there are exactly (n-1) nodes reaching the root node. Any bottleneck network is a reverse tree, so each node reaches through a unique path to the root as per lemma 11. Besides, there is only one edge called bottleneck edge incident to the root, therefore exactly (n - 1) paths share this edge.

C. Edge stress

In a network packets arrive is at the tail of an edge. These packets are forwarded over the edge one by one per time slot. The packets have predetermined path. Path is sequence of edges from source to destination. So, many packets following different paths can arrive at the tail of an edge in a time slot and contest for the edge. These packets are forwarded over the edge one by one per time slot. One packet out of contestingpackets is selected by the queuing policy for forwarding over the edge and rests of the packets have to wait in queue at the tail of the edge. More the number of paths sharing an edge more is the congestion at the tail of the edge. A notion of stress at an edge at any given time can be developed based on paths sharing the edge. We describe edge stress of an edge for an acyclic network now.

Edge Stress: Edge stress of an edge in any acyclic packet routing network is the maximum number of paths that share the edge at any point of time.

Theorem 1. In a n node acyclic network maximum edge stress of an edge is (n-1).

Proof: To obtain maximum edge stress on an edge in a node acyclic network we need to maximize number of paths through an edge. Each source reaches a destination by single path in the network. Therefore, we need to maximize number of sources to maximize number of paths that claim an edge. Besides, at least one destination is mandatory. Therefore in nnode acyclic directed network there can be at most (n - 1) distinct paths originating from (n-1) distinct sources to one destination. Now, for maximizing edge stress these paths have to do maximum sharing of edges. Hence, these paths reach the common destination by converging on a single edge incident to the destination node. This edge becomes the bottleneck edge in the bottleneck topology created by (n -1) sources and one destination. The destination node becomes the root node.The bottle neck edge experiences the maximum edge stress of (n -1). This is verified by lemma 1 that edge stress of bottleneck edge is exactly (n - 1).

Lemma 3. The maximum number of packets that can arrive at bottleneck edge in n node bottleneck network over a roundof k time steps when adversary injects at rate r is knr .

Proof: When adversary injects at rate r in a network, eachpath receives packets at most rate r. Bottleneck edge in a n node bottleneck network is shared by (n-1) paths. Therefore, over a round of k time steps bottleneck edge receives at most k(n - 1)r packets from the sharing paths. Besides, uptokr packets can be directly injected by the adversary at the tailof bottleneck edge in the same round. Hence, bottleneck edge receives total of knr packets over round of k rounds.under AQM to model QoS environment. Next we discuss SP/FIFO.

D. QoS Invoke Rate

Many packets arriving in a time slot at an edge create congestion. Bottleneck topology creates the worst case of congestion. This worst case of congestion occurs precisely at the bottleneck edge. The condition satisfying avoidance of congestion at bottleneck edge will also hold for any other edge in the network and any other topology. One way to avoid congestion is to restrict the adversary. Determination of peak adversarial rate up to which no congestion occurs at the bottleneck edge will give the QIR. We now formally define QIR.

Definition 6.QoS Invoke Rate. It is the threshold adversarial injection rate above which QoS policies need to be invoked in a network to guarantee QoS requirements of a traffic class.

To obtain QIR we state and prove condition of equivalence between single class and multiclass scheduling.

Theorem 2. Single class and multiclass packet scheduling are indistinguishable in terms of QoS up to QoS Invoke Rate.

Proof: We prove the theorem by considering single class traffic and multiclass traffic one by one and then establishing equivalence in delay suffered by a packet in both cases of scheduling.

Single Class Packet Routing Network: Consider a n node single class bottleneck network with all queues empty initially. If adversary injects at rate r in the network then during any time slot maximum nr packets can simultaneously arrive atthe tail of bottleneck edge. However, only one packet can be forwarded in a time slot. For a round of k time steps at most knr packets will arrive at the tail of bottleneck edge. Hence, for all packets to get transmitted to the root in the same round leaving no packets to wait beyond the round duration, the condition is knr ≤ k, hence r ≤1=n. No packet of around waits in any subsequent round.

Multiclass Bottleneck Network: Consider a*l* class *n* node bottleneck network with adversarial injection rate *r*. The initial condition of network is zero. Any edge has buffers b_1 to $b_l$in its queue. For all packets arriving in a round of k timesteps to get transmitted living no packets to wait at the tailof bottle neck edge is폴 팔1폴폴=knr≤r i.e. r≤ 1/n .

For r > 1=n the packets have to experience delay more thanthe backlog delay created in the same round and congestionoccurs at bottleneck edge and then QoS traffic needs priorityin treatment and we need to invoke QoS policies. We also realize from theorem 1 that bottleneck topology is worst casefor congestion to occur and hencer = 1/n is the critical rateabove which congestion starts to build up. So, we define $rater_q$ = 1/n of adversary to be QoS Invoke Rate.

E. Interpretation of QIR

QIR depends on network size and edge capacity in thenetwork. We assume all edges in network to be of samecapacity. As number of nodes increase, QIR decreases. Whenedge capacity increases, QIR also increases. For a packetrouting network of n nodes and Z bps,rq = Z=n bps. Asexample,200 nodes and 40 Mbps link capacity network hasQIR of 200 Kbps. QIR is more significant for high bandwidthsmall networks and starts to loose significance for low bandwidth large networks.

F. Observation

We note some important some important observations from QIR. Firstly, QoS network resources are non-effectively usedbelow QIR. Secondly, as a outcome of first observation single class and multiclass traffic have equivalence in treatment by the network below QIR. Thirdly, QIR gives the minimum threshold rate above which only effectiveness of QoS measures in a network can be verified by process like simulation. Energy can be saved in a network that runs below QIR by idling QoS routing resources.

This section has formally introduced the sense of QIR. Irrespective of any bound on QIR, occurrence of such a parameter is established by this section. QIR depends up on the network size inversely edge band width directly. We have also noted its use.

7. Conclusion

The paper explores stability and reliability in realistic QoS environment under adversarial queuing model. Our analysisprovides meaningful insight of the multiclass traffic behavior based on priority treatment in a network. In stable networks nonempty initial conditions and bursty injections of packets are root cause of unreliability. We also obtain two protocols to offset unreliability in stable but unreliable multiclass single station network. We also analyzed QIR for an acyclic network with delay as QoS parameter.

An important open issue for study is stability-reliability analysis for more complex multiclass networks like and trees, meshes and cycles. Stability-reliability study of other multiclass queuing policies like Fair queuing and Earliest-Deadline-First is also an interesting open area

References

[1] A. Borodin, J. Kleinberg, P. Raghavan, A. Sudan, and D. Williamson, Adversarial queueing theory, In Symposium on Computer Science, 1996.

[2] M. Andrews, B. Awerbuch, A. Fernandez, J. Kleinberg, T. Leighton, and Z. Liu, Universal stability results for greedy contention-resolution protocols, 37th IEEE Symposium on Foundations of Computer Science, 1996, pp. 380–389.

[3] A. Goyel. Stability of Networks and Protocols in the Adversarial Queuing Model for Packet Routing Networks, 37(2001) 219-224.

[4] M.AjmoneMarsan, M.Franceschinis,E.Leonardi, F.Neri, A.Tarello,InstabilityPhenomena in Underloaded Packet Networkswith QoS Schedulers,TechnicalReport,http://www.tlcnetworks.polito.it/emilio/netrep/instability-tec-rep.pdf, 2003.

[5.]F. P. Kelly. Networks of queues with customers of different types.J.Applied Probability.12:542–554, 1975

[6] P. R. Kumar and T. I. Seidman. Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems.

[7] S. H. Lu and P. R. Kumar. Distributed scheduling based on duedates and buffer priorities. IEEE Transactions on Automatic Control,36:1406–1416, 1991

[8] A. N. Rybko and A. L. Stolyar.Ergodicity of stochastic processes describing the operation of open queueing networks. Problems of Information Transmission, 28:199–220, 1992.

[9] M. Bramson. Instability of FIFO queueing networks. Annals of AppliedProbability, 4:414–693-718, 1994.

[10] M. Bramson. Instability of FIFO queueing networks. Annals of AppliedProbability, 4:414–431, 1994.

[11] S. Blake et.al..al. An Architecture of of Differentiated Services .RFC2475,1998.

[12] K. Nicholaset.al. Definition of Differentiated Services Field (DS Field) in the IPv4 and IPv6 Header. RFC2474,1998.

[13]V.Jacobson, K.Nichols, K.Poduri, An Expedited Forwarding PHB, RFC 2598, June 1999.

[14] J.Heinanen, F.Baker, W.Weiss, J.Wroclawski, Assured Forwarding PHB Group, RFC 2597, June 1999.

[15] A. Ros´en, A note on models for non-probabilistic analysis of packet switching networks, Information Processing Letters, 84 (2002) 237 -240.