Better resources management flashcard

\n[toc title="Table of Contents"]\n

\n \t

\n[/toc]\n \n

Contents

Chapter Three

3. 0 QoS Frameworks

Given the fact that the modern internet connects to several administrative domains, end-to-end QoS can be ensured by the concatenation of the domain-to-domain data forwarding, which turn can use three separate frameworks. These include (i) over provisioning (ii) better resources management that includes traffic control and generic switch architectures as well as (iii) routing and traffic engineering that also comprises two basic technologies i. e. MPLS and BGP, Collins (2001).

3. 1 Over Provisioning

The most basic technique, which is popular with ISPs is through the over-provisioning of the bandwidth, buffers, servicing traffic from competing users etc, in order to ensure that congestion is avoided in the first place. This is a resource intensive approach, which is only possible with the high capacity fiber links that can support 1. 6 Tbps. Even so, bottlenecks occur because of limited capacity electronic switching, Zamora, Jacobs, Eleftheriadis, Chang, & Anastassiou (2000). Resource allocation QoS differentiates and prioritizes traffic through the control of congestion or admission. Over-provisioning ensures that the capacity of the resources available is greater or similar to the peak traffic loads estimations, and despite the expense of accomplishing this, it solves the problem efficiently in networks with peak loads that can be predicted. Such is reasonable in the majority of applications, and may include demanding applications that make up for bandwidth delays and variations in special-need traffic such as video streaming, Collins (2001).

It however may have very limited utility especially given the transport protocols that exponentially expand the amount of data that is carried over the network until the available bandwidth is exhausted and additional packets are dropped, making sure that all users lose a few packets. The over provisioning amount in interior networks that is necessary to substitute for QoS is dependent on the traffic demands and the number of users, which also serves to minimize over provisioning capacity, Bhaniramka, Sun, & Jain (2009). This is not least, because bandwidth intensive applications as well as the growth in the number of users eliminates the excess capacity availed by over-provisioning. Once this happens, it becomes necessary to physically update the network links, which is expensive and inconvenient.

According to Ergin, Gruteser, Luo, Raychaudhuri, & Liu (2008), while over provisioning is simple, its ability to handly the growing demand for network services as well as multimedia traffic is heavily limited. In addition, the inneficieny of constantly maintaining additional capacity in order to accommodate the increases in traffic is enormous, if viewed on a larger scale. However, it is remains practical for small networking needs, and also possible for local area networks that use fiber glass infrastructure, which can accommodate massive amounts of traffic without the slightest consgestions. On a larger scale though, or in the case of multiple interconnections, over provisioning can still be used, but only alongside other technologies, Wang (2001). In fact, if used appropriately with traffic enginnering as well as resource management QoS technologies, over provisioning offer resource buffer that would offer a QoS guarantee for many applications.

3. 2 Resource Management

It has two major frameworks i. e. intServ per-flow QoS framework is coupled by dynamic allocation of resources, whose base philosophy that requires routers to reserve resources so that they can offer quantifiable QoS for certain traffic flows.

3. 2. 1 The Resource Reservation Signaling Mechanism

The Resource Reservation Protocol (RSVP) acts as a protocol for signaling applications so that they can reserve resources, and it allows reservations to be initiated by the receivers designed for environments that can easily accommodate heterogeneous receiver service requirements. Basically, the flow dispatches a PATH message (which contains information on the nature of traffic) to a flow receiver that it intends on using, Wang (2001). Once the PATH message is propagated, the network routers along the path record multiple characteristics of the resources need, while the intended receiver sends back an RESV message to request for requisite resources along the same route to the router that initiated the PATH message. If resources are available on the routers along the path, they will allocate buffer space and bandwidth to the flow, while at once installing necessary flow-specific information.

Per Flow services ensures that IntServ can offer guaranteed QoS, despite the fact that flow-specific states within routers is a major alteration of the present day internet architecture, which also means that it is largely unavailable, Chiu, Huang, Lo, Hwang, & Shieh (2003). Millions of flows may be in a router, which makes it difficult for the routers to effectively separately queue flows, despite the fact that it can be utilized for aggregation of flows. It is commonly used intra-domain and its incremental deployment is only practicable in controlled load services. IntServ has the benefit of having an explicit end-to-end resource admission control, a per-request admission policy control as well as signals for dynamic port numbers, Davidson, Fox, & et al (2002). It however suffers from the existence of continuous signaling due to the stateful architecture, coupled by the fact that flow-based approaches are not flexible to increasing demand that is necessary for large implementations.

The Differentiated Services (DiffServ) resource management technique is similar to IntServ, except it has greater scalability and it is a per-aggregate-class service discrimination technique employs packet tagging. Tagging of packets uses bits in the packets header to mark priority packets using type-of-service (TOS). TOS byte comprises of three-bit precedence field, maximum throughput and reliability, 4-bit filed that indicates non-urgent requests, an unused and minimum cost, Jaffar, Hashim, & Hamzah (2009). While the bits were not utilized, DiffServ redefined the byte as a DS field, six of which comprised the Differentiated Service CodePoint (DSCP) field, leaving the other two bits unused. The DSCP is used in the selection of the per-hop behavior or a PHB that packets experience at nodes. DiffServ comprises of two basic principles of design i. e. (i) separation of supporting mechanisms and policy and (ii) pushing intricacy to the network boundaries. Network boundaries include application hosts, edge routers, routers and leaf. They have a relatively limited number of flows, which makes it possible carry out operations with greater granularity, Bhaniramka, Sun, & Jain (2009). Network core routers on the other hand, can have many different flows that should make it more suitable in the performance of simple and fast operations.

The ability to separate between the supporting mechanisms and the control policy ensures that they both evolve differently. DiffServ does define multiple per-hop packet forwarding behavior, which subsequently form the foundation of QoS provisioning that makes control policy a matter for further efforts. Control policies may be altered in any way suitable to the network administrator and ultimate client, but the supporting PHBs must remain relatively stable. DiffServ is however, limited by the fact that the details of how the individual routers act on DS filed is specified by the configuration and it is therefore difficult to anticipate end-to-end behavior, Linawati (2005). It is made more intricate if packets cross more than one DiffServ domains prior to its destination. This is a major disability commercially, because it implies that it is technically impossible to provide different classes of end-to-end connectivity. Internet operators can resolve this through the enforcement of standardized policies for varied networks, despite the fact that not many of them are keen on additional levels of intricacy to the already intricate peering agreements, Ferguson & Huston (1998).

The Best Effort technique on the other hand seeks to provide the best possible QoS without guaranteeing it, making it less efficient, robust as well as a less efficient virtual model for communication, limited by the utility of a single path for every destination. Best Effort QoS provides the best possible paths to the destination, requiring an accurate definition of the Paths set as well as a forwarding pane plane that can efficiently support assignment/forwarding of traffic over several paths for each destination, Xiao, Chen, & Li (2010). The path weights comprise of multi-component measurement metrics, which capture crucial performance measures and define the best paths set using a specialized algorithm.

3. 2. 2 Traffic Policing

This refers to the mechanisms that monitor admitted sessions traffic in order to ensure that sessions remain within the provisions of the QoS contracts. Policing mechanisms ensures that traffic goes through according to the accepted/agreed traffic parameters, and when violations occur, the mechanism must reshape the data packets, Martinez, Apostolopoulos, Alfaro, Sanchez, & Duato (2010). Owing to the fact that policing shapes the packets according to the accepted quantitative parameters, multimedia applications will always be compatible to the traffic regulations, not least, because the sound, video etc signals are normally generated using standard coding that offers standard data coding. Policing may be applied to separate multimedia flows. Non-real time traffic hardly offers quantifiable traffic parameters and requires as much bandwidth as can be available, which makes it necessary for traffic policing to restrict such traffic from using up too much traffic at the expense of real time traffic, according to the set network policy. Policing may be implemented on intermediate or end hosts, and the most common forms of policing include (i) Token Bucket and (ii) Leaky Bucket, Collins (2001).

3. 2. 2. 1 Leaky Bucket

The provision of QoS is determined by the ability to define the flow properties, their aggregates and service needs, which are in turn partly dependent on the interactive voice communications delay bounds as well as the individual and business needs. It may be qualitatively and quantitatively be described as relative or absolute respectively. The TSpec token buckets is the most popular flow specification, which combines the peak rate with a token bucket, the minimum policed unit as well as the highest datagram size, Cho & Okamura (2010). The specifications are employed in the filtering of packets, and once a packet is services, it is subsequently eliminated from the bucket. Buckets facilitate efficiency by ensuring that empty packets queue up to a certain pre-specified bucket volume before they are processed. In addition, during the implementation, token buckets are often executed alongside leaky buckets as shown in the figure below, Ahmad (2001).

This technique is used to smooth out bursty traffic through the setting of the highest burst size and the peak rate. It works in the same way as a leaky bucket, whereby traffic parameters including the bucket(burst size) and hole sizes (maximum rate) are set and subsequently used to shape traffic into acceptable sizes and rates. The size of the bucket effectively determines the size of traffic burst, after which it will drop some of the packets. The buckets that arrive into a bucket enter through the top of the bucket (whose size is set (b)), which has a hole that allows traffic to go out a set maximum rate (r) per second.

When the rate of incoming traffic remains below the rate of the bucket leaking (r), then the outgoing traffic rate would be the same as the incoming traffic. Given an empty bucket, then the incoming rate is below the leakage rate i. e. R

3. 2. 3. 1 Token Bucket

This mechanism differs slightly from the leaky bucket mechanism because it maintains the bursty traffic. The size of the bucket (b) takes up traffic at r bytes per second. As soon as a packet arrives, it subsequently retrieves a token from the bucket (subject to availability), followed by the packet being forwarded to the outgoing stream. While there are tokens within the bucket, the outgoing traffic stream rate will be identical to the incoming rate, but if the token bucket runs empty, then incoming traffic streams must wait until the bucket fills up with more packets that subsequently become tokens.

This mechanism maintains the bursty traffic up until a given specific level, during which the outbound traffic maintains a maximum rate of that is equal or lower than the token rate, which ensures that the token bucket controls the rate of traffic, Rosenberg, et al. (2002).

3. 2. 3 Admission Control

It implements decision algorithms that a host or a router uses in the determination of whether new traffic streams may be admitted without hurting the QoS assurances that have already been granted. Each traffic stream requires a particular amount of network resources (router buffer space and link bandwidth) in order to transfer data from a source to the receives, and admission control is applied to manage the allocation of network resources, Martinez, Apostolopoulos, Alfaro, Sanchez, & Duato (2010). It seeks to compute the admission region correctly, in order to ensure that all the existent capacity is utilized in the proper manner, in the event there are algorithms that encourage under-utilization of the existent capacity. Admission control employed three basic techniques in accomplishing its task, and the include statistical, deterministic as well as measurement-based. Statistical and deterministic approaches make use of priori estimations, while the measurement-based techniques use present measurements of key parameters to faciltate the decision-making process by the specialized algorithms, Linawati (2005). Deterministic methods employ the worst case situations in that disallows QoS violations, and they are acceptable for smoothing traffic flows, despite the fact thaty perform poorly in the case of very unpredictable and widely varied data.

Ergin, Gruteser, Luo, Raychaudhuri, & Liu (2008) studied admission control as well as routing mechanisms in multi-rate wireless mesh networks. The technique is dependent on the precise approximations of the existent bandwidth at the nodes involved and the nodes involved and the required bandwidth consumptions by new flows. Estimation of the parameters in such networks is difficult because of the open and shared nature of wireless channels. The existent available bandwidths approximation techniques to not estimate in an approximate way the extent of interference from the neighboring nodes as well as flow bandwidth requirement approximation, which restricts potential for parallel transmissions, Ergin, Gruteser, Luo, Raychaudhuri, & Liu (2008). The study determined the positive influence on the QoS due to the implementation of admission controls. The per flow throughput and the average delays are drawn on a graph before and after the application of the admission controls.

The research evidence clearly demonstrated that the arrival of a third flow resulted in considerable congestion, and massive delays across the three flows together with large variations. The high delays across all the initial two as well as the third flow, was coupled with an unstable throughput that has adverse consequences for multimedia traffic, which requires the least possible amount of delays to make the output meaningful, Ergin, Gruteser, Luo, Raychaudhuri, & Liu (2008). When the admission controls are enabled, the third flow is locked off the network, because of the limited channel capacity that in turn results in a stable throughput, with the delays reduced to an average of 1/100s. In addition, the short delaying of packets and the consistent throughputs indicate that admission controls are effective and positively influential to real time, multimedia QoS over wireless networks, Ferguson & Huston (1998).

3. 2. 4 Policy Control

Policy specifications that regulate the access to the network resources as well as services founded on the administrative criteria chosen that control everythiong from what users, applications and hosts that are allowed access to the network as well as the priorities to be accorded the differentiated, (Zeng (2010). Instead of configurations of separate network devices, corporate administrators and ISPs can regulate the network through their policy infrastructures that offers support for facilitating administrative intentions that is subsequently translated into differential treatment of packet traffic flows. This is depicted in figure 8, which shows a typical policy architecture. Every domain can comprise more than one policy servers that serve to make both policy and configuration decisions for varied network elements. The policy servers have access to the policy database, the accounting and authorization databases. Every policy entry provides rules in every event, with the input of human operators, Dar & Latif (2010). Policy servers comprise of central policy controllers as well as sets of policy decision points that are in turn responsible for determining the actions that are applicable to the individual packets. They ensure the universality in the decisions that are made regarding access and utilization of network resources.

3. 2. 5 Bandwidth Brokerage

Bandwidth brokers represent logical resources management entities that allocate intra-domain resources, while at once arranging inter-domain agreements. The bandwidth brokers for every domain may be configured with the policies of every organization, while at once controlling the edge routers operations. In the policy frameworks view, bandwidth brokers include PDP functions, policy databases, and edge routers act as PEPs, Martinez, Apostolopoulos, Alfaro, Sanchez, & Duato (2010). Bandwidth brokers play a central inter-domain role by negotiating with the neighboring domains, setting up bilateral agreements with everyone of them, while at once sending proper configuration parameters to the edge router domains. The bilateral agreements imply that bandwidth broke to require coordinating with neighboring domains and ensuring end-to-end QoS through the concatenation of the bilateral agreements across varied domains, along with the proper intra-domain allocation of resources. Within domains, the bandwidth brokers perform resource allocations by the use of admission controls, with the choice of intra-domain algorithms being dependent on the intra-domain negotiations, Gheorghe (2006). The bandwidth broker architecture is similar to the present internet routing, where BGP4 acts as a standard inter-domain router protocol and multiple choices are possible.