1. Quality of service

From Wikipedia, the free encyclopedia

This article **needs additional citations for verification**. Please help improve this article by adding reliable references. Unsourced material may be challenged and removed. *(August 2007)*

This article **may have too long an introduction for its overall length**. Please help by moving some material from it into the body of the article. For more information please read the layout guide and Wikipedia's lead section guidelines. *(September 2010)*

In the field of computer networking and other packet-switched telecommunication networks, the traffic engineering term **quality of service** (*QoS*) refers to resource reservation control mechanisms rather than the achieved service quality. Quality of service is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow. For example, a required bit rate, delay, jitter, packet dropping probability and/or bit error rate may be guaranteed. Quality of service guarantees are important if the network capacity is insufficient, especially for real-time streaming multimedia applications such as voice over IP, online games and IP-TV, since these often require fixed bit rate and are delay sensitive, and in networks where the capacity is a limited resource, for example in cellular data communication.

A network or protocol that supports QoS may agree on a traffic contract with the application software and reserve capacity in the network nodes, for example during a session establishment phase. During the session it may monitor the achieved level of performance, for example the data rate and delay, and dynamically control scheduling priorities in the network nodes. It may release the reserved capacity during a tear down phase.

A best-effort network or service does not support quality of service. An alternative to complex QoS control mechanisms is to provide high quality communication over a best-effort network by over-provisioning the capacity so that it is sufficient for the expected peak traffic load. The resulting absence of network congestion eliminates the need for QoS mechanisms.

In the field of telephony, quality of service was defined in the ITU standard X.902 as “A set of quality requirements on the collective behavior of one or more objects”. Quality of service comprises requirements on all the aspects of a connection, such as service response time, loss, signal-to-noise ratio, cross-talk, echo, interrupts, frequency response, loudness levels, and so on. A subset of telephony QoS is grade of service (GoS) requirements, which comprises aspects of a connection relating to capacity and coverage of a network, for example guaranteed maximum blocking probability and outage probability.[1]

QoS is sometimes used as a quality measure, with many alternative definitions, rather than referring to the ability to reserve resources. Quality of service sometimes refers to the level of quality of service, i.e. the guaranteed service quality. High QoS is often confused with a high level of performance or achieved service quality, for example high bit rate, low latency and low bit error probability.

An alternative and disputable definition of QoS, used especially in application layer services such as telephony and streaming video, is requirements on a metric that reflects or predicts the subjectively experienced quality. In this context, QoS is the acceptable cumulative effect on subscriber satisfaction of all imperfections affecting the service. Other terms with similar meaning are the quality of experience (QoE) subjective business concept, the required “user perceived performance”,[2] the required “degree of satisfaction of the user” or the targeted “number of happy customers”. Examples of measures and measurement methods are Mean Opinion Score (MOS), Perceptual Speech Quality Measure (PSQM) and Perceptual Evaluation of Video Quality (PEVQ). See also subjective video quality.

2. Contents

[hide]

3. [edit] History

Conventional Internet routers and LAN switches operate on a best effort basis. This equipment is less expensive, less complex and faster and thus more popular than competing more complex technologies that provided QoS mechanisms. There were four "Type of service" bits and three "Precedence" bits provided in each IP packet header, but they were not generally respected. These bits were later re-defined as DiffServ Code Points (DSCP) and are sometimes honored in peered links on the modern Internet.

With the advent of IPTV and IP telephony, QoS mechanisms are increasingly available to the end user.

A number of attempts for layer 2 technologies that add QoS tags to the data have gained popularity during the years, but then lost attention. Examples are Frame relay and ATM. Recently, MPLS (a technique between layer 2 and 3) have gained some attention. However, today Ethernet may offer QoS through its 802.1p. Ethernet is, by far, the most popular layer 2 technology.

In Ethernet, Virtual LANs (VLAN) may be used to separate different QoS levels. For example in fibre-to-the-home switches typically offer several Ethernet ports connected to different VLAN:s. One VLAN may be used for Internet access (low priority), one for IPTV (higher priority) and one for IP telephony (highest priority). Different Internet providers may use the different VLANs.

4. [edit] Key qualities of traffic

When looking at packet-switched networks, quality of service is affected by various factors, which can be divided into "human" and "technical" factors. Human factors include: stability of service, availability of service, delays, user information. Technical factors include: reliability, scalability, effectiveness, maintainability, Grade of Service, etc.[3]

Many things can happen to packets as they travel from origin to destination, resulting in the following problems as seen from the point of view of the sender and receiver:

Low throughput
: Due to varying load from other users sharing the same network resources, the bit rate (the maximum throughput) that can be provided to a certain data stream may be too low for realtime multimedia services if all data streams get the same scheduling priority.

Dropped packets
: The routers might fail to deliver (*drop*) some packets if their data is corrupted or they arrive when their buffers are already full. The receiving application may ask for this information to be retransmitted, possibly causing severe delays in the overall transmission.

Errors
: Sometimes packets are corrupted due to bit errors caused by noise and interference, especially in wireless communications and long copper wires. The receiver has to detect this and, just as if the packet was dropped, may ask for this information to be retransmitted.

Latency
: It might take a long time for each packet to reach its destination, because it gets held up in long queues, or takes a less direct route to avoid congestion. This is different from throughput, as the delay can build up over time, even if the throughput is almost normal. In some cases, excessive latency can render an application such as VoIP or online gaming unusable.

Jitter
: Packets from the source will reach the destination with different delays. A packet's delay varies with its position in the queues of the routers along the path between source and destination and this position can vary unpredictably. This variation in delay is known as jitter and can seriously affect the quality of streaming audio and/or video.

Out-of-order delivery
: When a collection of related packets is routed through a network, different packets may take different routes, each resulting in a different delay. The result is that the packets arrive in a different order than they were sent. This problem requires special additional protocols responsible for rearranging out-of-order packets to an isochronous state once they reach their destination. This is especially important for video and VoIP streams where quality is dramatically affected by both latency and lack of sequence.

5. [edit] Applications

A defined quality of service may be desired or required for certain types of network traffic, for example:

- Streaming media specifically
  - Internet protocol television (IPTV)
  - Audio over Ethernet
  - Audio over IP
- IP telephony also known as Voice over IP (VoIP)
- Videoconferencing
- Telepresence
- Circuit Emulation Service
- Safety-critical applications such as remote surgery where availability issues can be hazardous
- Network operations support systems either for the network itself, or for customers' business critical needs
- Online games where real-time lag can be a factor
- Industrial control systems protocols such as Ethernet/IP which are used for real-time control of machinery

These types of service are called *inelastic*, meaning that they require a certain minimum level of bandwidth and a certain maximum latency to function. By contrast, *elastic* applications can take advantage of however much or little bandwidth is available. Bulk file transfer applications that rely on TCP are generally elastic.

6. [edit] Obtaining QoS

- In advance: When the expense of mechanisms to provide QoS is justified, network customers and providers typically enter into a contractual agreement termed a service level agreement (SLA) which specifies guarantees for the ability of a network/protocol to give guaranteed performance/throughput/latency bounds based on mutually agreed measures, usually by prioritizing traffic.
- Reserving resources: Resources are reserved at each step on the network for the call as it is set up. An example is RSVP, Resource Reservation Protocol.

7. [edit] Over-provisioning

An alternative to complex QoS control mechanisms is to provide high quality communication by generously over-provisioning a network so that capacity is based on peak traffic load estimates. This approach is simple and economical for networks with predictable and light traffic loads. The performance is reasonable for many applications. This might include demanding applications that can compensate for variations in bandwidth and delay with large receive buffers, which is often possible for example in video streaming. Over-provisioning can be of limited use, however, in the face of transport protocols (such as TCP) that over time exponentially increase the amount of data placed on the network until all available bandwidth is consumed and packets are dropped. Such greedy protocols tend to increase latency and packet loss for all users.

Commercial VoIP services are often competitive with traditional telephone service in terms of call quality even though QoS mechanisms are usually not in use on the user's connection to his ISP and the VoIP provider's connection to a different ISP. Under high load conditions, however, VoIP may degrade to cell-phone quality or worse. The mathematics of packet traffic indicate that network requires just 60% more raw capacity under conservative assumptions.[4]

The amount of over-provisioning in interior links required to replace QoS depends on the number of users and their traffic demands. This is an important factor that limits usability of over-provisioning. Newer more bandwidth intensive applications and the addition of more users results in the loss of over-provisioned networks. This then requires a physical update of the relevant network links which is an expensive process. Thus over-provisioning cannot be blindly assumed on the Internet.

8. [edit] QoS mechanisms

Early work used the “IntServ” philosophy of reserving network resources. In this model, applications used the Resource reservation protocol (RSVP) to request and reserve resources through a network. While IntServ mechanisms do work, it was realized that in a broadband network typical of a larger service provider, Core routers would be required to accept, maintain, and tear down thousands or possibly tens of thousands of reservations. It was believed that this approach would not scale with the growth of the Internet, and in any event was antithetical to the notion of designing networks so that Core routers do little more than simply switch packets at the highest possible rates.

The second and currently accepted approach is “DiffServ” or differentiated services. In the DiffServ model, packets are marked according to the type of service they need. In response to these markings, routers and switches use various queuing strategies to tailor performance to requirements. — At the IP layer, differentiated services code point (DSCP) markings use the 6 bits in the IP packet header. At the MAC layer, VLAN IEEE 802.1Q and IEEE 802.1p can be used to carry essentially the same information.

Routers supporting DiffServ use multiple queues for packets awaiting transmission from bandwidth constrained (e.g., wide area) interfaces. Router vendors provide different capabilities for configuring this behavior, to include the number of queues supported, the relative priorities of queues, and bandwidth reserved for each queue.

In practice, when a packet must be forwarded from an interface with queuing, packets requiring low jitter (e.g., VoIP or VTC) are given priority over packets in other queues. Typically, some bandwidth is allocated by default to network control packets (e.g., ICMP and routing protocols), while best effort traffic might simply be given whatever bandwidth is left over.

Additional bandwidth management mechanisms may be used to further engineer performance, to include:

- Traffic shaping (rate limiting):
  - Token bucket
  - Leaky bucket
  - TCP rate control—artificially adjusting TCP window size as well as controlling the rate of ACKs being returned to the sender[*citation needed*]
- Scheduling algorithms:
  - Weighted fair queuing (WFQ)
  - Class based weighted fair queuing
  - Weighted round robin (WRR)
  - Deficit weighted round robin (DWRR)
  - Hierarchical Fair Service Curve (HFSC)
- Congestion avoidance:
  - RED, WRED — Lessens the possibility of port queue buffer tail-drops and this lowers the likelihood of TCP global synchronization
  - Policing (marking/dropping the packet in excess of the committed traffic rate and burst size)
  - Explicit congestion notification
  - Buffer tuning

As mentioned, while DiffServ is used in many sophisticated enterprise networks, it has not been widely deployed in the Internet. Internet peering arrangements are already complex, and there appears to be no enthusiasm among providers for supporting QoS across peering connections, or agreement about what policies should be supported in order to do so.

One compelling example of the need for QoS on the Internet relates to this issue of congestion collapse. The Internet relies on congestion avoidance protocols, as built into TCP, to reduce traffic load under conditions that would otherwise lead to Internet Meltdown. QoS applications such as VoIP and IPTV, because they require largely constant bitrates and low latency cannot use TCP, and cannot otherwise reduce their traffic rate to help prevent meltdown either. QoS contracts limit traffic that can be offered to the Internet and thereby enforce traffic shaping that can prevent it from becoming overloaded, hence they're an indispensable part of the Internet's ability to handle a mix of real-time and non-real-time traffic without meltdown.

Asynchronous Transfer Mode (ATM) network protocol has an elaborate framework to plug in QoS mechanisms of choice. Shorter data units and built-in QoS were some of the unique selling points of ATM in the telecommunications applications such as video on demand, voice over IP.

9. [edit] Protocols that provide quality of service

- The Type of Service (ToS) field in the IP(v4) header (now superseded by DiffServ)
- IP Differentiated services (DiffServ)
- IP Integrated services (IntServ)
- Resource reSerVation Protocol (RSVP)

- Multiprotocol Label Switching (MPLS) provides eight QoS classes
- RSVP-TE
- Frame relay
- X.25
- Some ADSL modems
- Asynchronous Transfer Mode (ATM)
- IEEE 802.1p
- IEEE 802.1Q
- IEEE 802.11e
- HomePNA Home networking over coax and phone wires
- The ITU-T G.hn standard provides QoS by means of “Contention-Free Transmission Opportunities” (CFTXOPs) which are allocated to flows which require QoS and which have negotiated a “contract” with the network controller. G.hn also supports non-QoS operation by means of “Contention-based Time Slots”.
- Audio Video Bridging

10. [edit] QoS solutions

The research project “Multi Service Access Everywhere” (MUSE)[5] defined a QoS concept in Phase I which was further worked out in another research project PLANETS. The new idea of this solution is to agree on a discrete jitter value per QoS class which is imposed on network nodes. Including best effort, four QoS classes were defined, two elastic and two inelastic. The solution has several benefits:

- End-to-end delay and packet loss rate can be predicted
- It is easy to implement with simple scheduler and queue length given in PLANETS
- Nodes can be easily verified for compliance
- End users do notice the difference in quality

The MUSE project finally elaborated its own QoS solution which is primarily based in:

- The usage of traffic classes
- Selective CAC concept
- Appropriate network dimensioning

11. [edit] Quality of service procedures

Unlike the Internet2 Abilene Network, the Internet is actually a series of exchange points interconnecting private networks and not a network in its own right.[6] Hence the Internet's core is owned and managed by a number of different network service providers, not a single entity. Its behavior is much more stochastic or unpredictable. Therefore, research continues on QoS procedures that are deployable in large, diverse networks.

There are two principal approaches to QoS in modern packet-switched IP networks, a parameterized system based on an exchange of application requirements with the network, and a prioritized system where each packet identifies a desired service level to the network.

- Integrated services (“IntServ”) implements the parameterized approach. In this model, applications use the Resource Reservation Protocol (RSVP) to request and reserve resources through a network.
- Differentiated services (“DiffServ”) implements the prioritized model. DiffServ marks packets according to the type of service they desire. In response to these markings, routers and switches use various queueing strategies to tailor performance to expectations. DiffServ Code Point (DSCP) markings use the first 6 bits in the ToS field of the IP(v4) packet header.

At the MAC layer, VLAN IEEE 802.1Q and IEEE 802.1p can be used to carry essentially the same information as used by DiffServ.

Cisco IOS NetFlow and the Cisco Class Based QoS (CBQoS) Management Information Base (MIB) can both be leveraged within a Cisco network device to obtain visibility into QoS policies and their effectiveness on network traffic. [7]

Non-IP protocols, especially those intended for voice transmission, such as ATM or GSM, have already implemented QoS in the core protocol and don't need additional procedures to achieve it.

12. [edit] End-to-end quality of service

End-to-end quality of service usually requires a method of coordinating resource allocation between one autonomous system and another. Research consortia such as EuQoS[8] and fora such as IPsphere[9] have developed mechanisms for handshaking QoS invocation from one domain to the next. IPsphere defined the Service Structuring Stratum (SSS) signaling bus in order to establish, invoke and (attempt to) assure network services. EuQoS conducted experiments to integrate Session Initiation Protocol, Next Steps in Signaling and IPsphere's SSS.

The Internet Engineering Task Force (IETF) defined the Resource Reservation Protocol (RSVP) for bandwidth reservation. RSVP is an end-to-end bandwidth reservation protocol that is also useful to end-to-end QoS. The traffic engineering version, RSVP-TE, is used in many networks today to establish traffic-engineered MPLS label-switched paths.

The IETF also defined NSIS[10] with QoS signalling as a target. NSIS is a development and simplification of RSVP.

13. [edit] Quality of service circumvention

Strong cryptography network protocols such as Secure Sockets Layer, I2P, and virtual private networks obscure the data transferred using them. As all electronic commerce on the Internet requires the use of such strong cryptography protocols, unilaterally downgrading the performance of encrypted traffic creates an unacceptable hazard for customers. Yet, encrypted traffic is otherwise unable to undergo deep packet inspection for QoS.

14. [edit] Doubts about quality of service over IP

Gary Bachula, Vice President for External Affairs for Internet2, asserts that specific QoS protocols are unnecessary in the core network as long as the core network links are “over-provisioned” to the point that network traffic never encounters delay. In “quality of service” engineering, this formulation is guaranteed by the *admission control* feature. It is important to note that this only refers to core networks and not end-to-end connections. Recent studies point to a relatively low end-to-end bandwidth availability even on Internet2.

The Internet2 QoS Working Group concluded that increasing bandwidth is probably more practical than implementing QoS.[11][12]

The Internet2 project found, in 2001, that the QoS protocols were probably not deployable inside its Abilene network with equipment available at that time. While newer routers are capable of following QoS protocols with no loss of performance, equipment available at the time relied on software to implement QoS. The Internet2 Abilene network group also predicted that “logistical, financial, and organizational barriers will block the way toward any bandwidth guarantees” by protocol modifications aimed at QoS.[11][12] In essence, they believe that the economics would be likely to make the network providers deliberately erode the quality of best effort traffic as a way to push customers to higher priced QoS services.

The Abilene network study was the basis for the testimony of Gary Bachula to the Senate Commerce Committee's Hearing on Network Neutrality in early 2006. He expressed the opinion that adding more bandwidth was more effective than any of the various schemes for accomplishing QoS they examined.[13]

Bachula's testimony has been cited by proponents of a law banning quality of service as proof that no legitimate purpose is served by such an offering. This argument is dependent on the assumption that over-provisioning isn't a form of QoS and that it is always possible. Cost and other factors affect the ability of carriers to build and maintain permanently over-provisioned networks.

15. [edit] Mobile cellular QoS

Main article: Mobile QoS

Mobile cellular service providers may offer **mobile QoS** to customers just as the fixed line PSTN services providers and Internet Service Provides (ISP) may offer QoS. QoS mechanisms are always provided for circuit switched services, and are essential for non-elastic services, for example streaming multimedia. It is also essential in networks dominated by such services, which is the case in today's mobile communication networks, but not necessarily tomorrow.

Mobility adds complication to the QoS mechanisms, for several reasons:

- A phone call or other session may be interrupted after a handover, if the new base station is overloaded. Unpredictable handovers make it impossible to give an absolute QoS guarantee during a session initiation phase.
- The pricing structure is often based on per-minute or per-megabyte fee rather than flat rate, and may be different for different content services.
- A crucial part of QoS in mobile communications is Grade of Service, involving outage probability (the probability that the mobile station is outside the service coverage area, or affected by co-channel interference, i.e. crosstalk) blocking probability (the probability that the required level of QoS can not be offered) and scheduling starvation. These performance measures are affected by mechanisms such as mobility management, radio resource management, admission control, fair scheduling, channel-dependent scheduling etc.

16. [edit] Standards activity

- **Quality of service**, or **QoS**, in the field of telephony, was defined in 1994 in the ITU-T Recommendation E.800[14]. This definition is very broad, listing 6 primary components: Support, Operability, Accessibility, Retainability, Integrity and Security.

- In 1998 the ITU published a document discussing QoS in the field of data networking, ITU-T Recommendation X.641[15]. X.641 offers a means of developing or enhancing standards related to QoS and provide concepts and terminology that will assist in maintaining the consistency of related standards.

- The main QoS-related IETF RFCs are Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers (RFC 2474), and Resource ReSerVation Protocol (RSVP) (RFC 2205); both these are discussed above. The IETF has also published two RFCs giving background on QoS: RFC 2990: Next Steps for the IP QoS Architecture, and RFC 3714: IAB Concerns Regarding Congestion Control for Voice Traffic in the Internet.

17. [edit] Open source QoS software

- Linux Advanced Routing & Traffic Control
- Bandwidth Arbitrator
- Zero Shell
- Mod qos adding QoS to web applications

18. [edit] See also

- Application service architecture
- Best-effort
- BSSGP
- Bufferbloat
- Class of Service
- Deep packet inspection (DPI)
- Grade of service (GoS)
- Mean Opinion Score (MOS)
- Mobile QoS
- Network neutrality
- QPPB
- Quality of Experience (QoE)
- Series of tubes
- Streaming media
- Subjective video quality
- Tiered Internet
- Traffic shaping

19. [edit] Notes

1. ^ Teletraffic Engineering Handbook ITU-T Study Group 2 (350 pages, 4·48MiB)(It uses abbreviation GoS instead of QoS)
2. ^ Leonard Franken. Quality of Service Management: A Model-Based Approach. PhD thesis, Centre for Telematics and Information Technology, 1996.
3. ^ Peuhkuri M., IP Quality of Service, Helsinki University of Technology, Laboratory of Telecommunications Technology, 1999.
4. ^ Yuksel, M.; Ramakrishnan, K. K.; Kalyanaraman, S.; Houle, J. D.; Sadhvani, R. (2007). "IEEE International Workshop on Quality of Service (IWQoS'07)" (PDF). Evanston, IL, USA. pp. 109–112. doi:10.1109/IWQOS.2007.376555.[dead link]
5. ^ Multi Service Access Everywhere
6. ^ An Evening With Robert Kahn, from Computer History Museum, 9 Jan 2007
7. ^ Using CBQoS & NetFlow to Manage QoS Policies in Your Environment
8. ^ EuQoS
9. ^ IPSphere: Enabling Advanced Service Delivery
10. ^ “Next Steps in Signaling” Charter
11. ^ [a] [b] Oram, Andy (2002-06-11). "A Nice Way to Get Network Quality of Service?". O'Reilly. Archived from the original on 2010-09-12. Retrieved 2006-07-07.
12. ^ [a] [b] Benjamin Teitelbaum, Stanislav Shalunov (3 May 2002). "Why Premium IP Service Has Not Deployed (and Probably Never Will)". Internet2 QoS Working Group. Archived from the original on 2010-09-12.
13. ^ Bachula, Gary (2006-02-07). "Testimony of Gary R. Bachula, Vice President, Internet2" (PDF). pp. 5. Retrieved 2006-07-07.
14. ^ ITU-T Recommendation E.800
15. ^ ITU-T Recommendation X.641

20. [edit] References

- *Technical, Commercial and Regulatory Challenges of QoS: An Internet Service Model Perspective* by Xipeng Xiao (Morgan Kaufmann, 2008, ISBN 0-12-373693-5)
- *Deploying IP and MPLS QoS for Multiservice Networks: Theory and Practice* by John Evans, Clarence Filsfils (Morgan Kaufmann, 2007, ISBN 0-12-370549-5)
- Lelli, F. Maron, G. Orlando, S. Client Side Estimation of a Remote Service Execution. 15th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2007. MASCOTS '07.
- *QoS Over Heterogeneous Networks* by Mario Marchese (Wiley, 2007, ISBN 978-0-470-01752-4)
- RFC 1633: Integrated Services in the Internet Architecture: an Overview
- RFC 2475: An Architecture for Differentiated Services
- RFC 3209: RSVP-TE: Extensions to RSVP for LSP Tunnels

21. [edit] External links

Listen to this article (info/dl)

This audio file was created from a revision of Quality of service dated 2008-07-18, and does not reflect subsequent edits to the article. (Audio help)

More spoken articles

- QoS Tutorial
- Cisco's Internetworking Technology Handbook
- Network QoS
- Quality of Service on Microsoft TechNet
- Technical, commercial, and regulatory challenges of QoS by XiPeng Xiao (Google Books)
- Web based traffic shaping bridge/router

22.

23. Class of service

From Wikipedia, the free encyclopedia

This article **has no lead section**. Please help by adding an introductory section to this article. For more information, see the layout guide, and Wikipedia's lead section guidelines, and join in a discussion of the issue on the talk page. *(June 2011)*

24. Contents

[hide]

25. [edit] Class of Service (CoS)

Main article: IEEE 802.1p

As related to network technology, CoS is a 3 bit field within an Ethernet frame header when using 802.1Q tagging. The field specifies a priority value of between 0 and 7 inclusive that can be used by Quality of Service (QoS) disciplines to differentiate traffic.

While CoS operates only on Ethernet at the data link layer, other QoS mechanisms (such as DiffServ) operate at the network layer and higher. Others operate at the physical layer.

Although 802.1Q tagging must be enabled to communicate priority information from end station to switch or switch to switch, some switches use CoS to internally classify traffic for QoS purposes.

26. [edit] Classification of service

The term can be used generically to refer to the classification of network traffic within network equipment based on packet inspection. Cisco implements such classification through Network-Based Application Recognition (NBAR). NBAR works with the existing QoS system.[1]

27. [edit] Class of Service (COS)

As related to legacy telephone systems, COS is often used to define the permissions an extension will have on a PBX or Centrex. The Class of Service acronym is normally written as COS vs. CoS as is often used in data networking parlance. Certain groups of users may have a need for extended voicemail message retention while another group may need the ability to forward calls to a cell phone, and still others have no need to make calls outside the office. Permissions for a group of extensions can be changed by modifying a COS variable applied to the entire group.

COS is also used on trunks to define if they are full-duplex, incoming only, or outgoing only.

28. [edit] References

- "Deploying IP and MPLS QoS for Multiservice Networks: Theory and Practice" by John Evans, Clarence Filsfils (Morgan Kaufmann, 2007, ISBN 0-12-370549-5)
- Supporting differentiated classes of service in Ethernet passive optical networks, Glen Kramer, Biswanath Mukherjee, Sudhir Dixit, Yinghua Ye and Ryan Hirth

1. ^ [1] Cisco NBAR

29. [edit] See also

- Type of Service
- DiffServ
- Quality of service

30. GRADE OF SERVICE

From Wikipedia, the free encyclopedia

In telecommunication engineering, and in particular teletraffic engineering, the quality of voice service is specified by two measures: the **grade of service** (**GoS**) and the **quality of service** (**QoS**).

Grade of service is the probability of a call in a circuit *group* being blocked or delayed for more than a specified interval, expressed as a vulgar fraction or decimal fraction. This is always with reference to the busy hour when the traffic intensity is the greatest. Grade of service may be viewed independently from the perspective of incoming versus outgoing calls, and is not necessarily equal in each direction or between different source-destination pairs.

On the other hand, the **quality of service** which a *single* circuit is designed or conditioned to provide, e.g. voice grade or program grade is called the quality of service. Quality criteria for such circuits may include equalization for amplitude over a specified band of frequencies, or in the case of digital data transported via analogue circuits, may include equalization for phase. Criteria for mobile quality of service in cellular telephone circuits include the probability of abnormal termination of the call.

30.1. Contents

[hide]

30.2. [edit] What is Grade of Service and how is it measured?

When a user attempts to make a telephone call, the routing equipment handling the call has to determine whether to accept the call, reroute the call to alternative equipment, or reject the call entirely. Rejected calls occur as a result of heavy traffic loads (congestion) on the system and can result in the call either being delayed or lost. If a call is delayed, the user simply has to wait for the traffic to decrease, however if a call is lost then it is removed from the system.[1]

The Grade of Service is one aspect of the quality a customer can expect to experience when making a telephone call.[2] In a Loss System, the Grade of Service is described as that proportion of calls that are lost due to congestion in the busy hour.[3] For a Lost Call system, the Grade of Service can be measured using *Equation 1*.[4]

$$\text{Grade of Service} = \frac{\text{number of lost calls}}{\text{number of offered calls}} \qquad (1)$$

For a delayed call system, the Grade of Service is measured using three separate terms:[1]

- The mean delay t_d – Describes the average time a user spends waiting for a connection if their call is delayed.
- The mean delay t_o – Describes the average time a user spends waiting for a connection whether or not their call is delayed.
- The probability that a user may be delayed longer than time t while waiting for a connection. Time t is chosen by the telecommunications service provider so that they can measure whether their services conform to a set Grade of Service.

30.3. [edit] Where and when is Grade of Service measured?

The Grade of Service can be measured using different sections of a network. When a call is routed from one end to another, it will pass through several exchanges. If the Grade of Service is calculated based on the number of calls rejected by the final circuit group, then the Grade of Service is determined by the final circuit group blocking criteria. If the Grade of Service is calculated based on the number of rejected calls between exchanges, then the Grade of Service is determined by the exchange-to-exchange blocking criteria.[1]

The Grade of Service should be calculated using both the access networks and the core networks as it is these networks that allow a user to complete an end-to-end connection.[4] Furthermore, the Grade of Service should be calculated from the average of the busy hour traffic intensities of the 30 busiest traffic days of the year. This will cater for most scenarios as the traffic intensity will seldom exceed the reference level.

The grade of service is a measure of the ability of a user to access a trunk system during the busiest hour. The busy is based upon customer demand at the busiest hour during a week month or year.

30.4. [edit] Class of Service

Different telecommunications applications require different Qualities of Service. For example, if a telecommunications service provider decides to offer different qualities of voice connection, then a premium voice connection will require a better connection quality compared to an ordinary voice connection. Thus different Qualities of Service are appropriate, depending on the intended use. To help telecommunications service providers to market their different services, each service is placed into a specific class. Each Class of Service determines the level of service required.[4]

To identify the Class of Service for a specific service, the network's switches and routers examine the call based on several factors. Such factors can include:[2]

- The type of service and priority due to precedence
- The identity of the initiating party
- The identity of the recipient party

30.5. [edit] Quality of Service in broadband networks

In broadband networks, the Quality of Service is measured using two criteria. The first criterion is the probability of packet losses or delays in already accepted calls. The second criterion refers to the probability that a new incoming call will be rejected or blocked. To avoid the former, broadband networks limit the number of active calls so that packets from established calls will not be lost due to new calls arriving. As in circuit-switched networks, the Grade of Service can be calculated for individual switches or for the whole network.[5]

30.6. [edit] Maintaining a Grade of Service

The telecommunications provider is usually aware of the required Grade of Service for a particular product. To achieve and maintain a given Grade of Service, the operator must ensure that sufficient telecommunications circuits or routes are available to meet a specific level of demand. It should also be kept in mind that too many circuits will create a situation where the operator is providing excess capacity which may never be used, or at the very least may be severely underutilized. This adds costs which must be borne by other parts of the network. To determine the correct number of circuits that are required, telecommunications service providers make use of Traffic Tables.[4] An example of a Traffic Table can be viewed in *Figure 1*.[4] It follows that in order for a telecommunications network to continue to offer a given Grade of Service, the number of circuits provided in a circuit group must increase (non-linearly) if the traffic intensity increases.[4]

30.7. [edit] Erlang's lost call assumptions

To calculate the Grade of Service of a specified group of circuits or routes, A.K. Erlang used a set of assumptions that relied on the network losing calls when all circuits in a group were busy. These assumptions are:[4]

- All traffic through the network is pure-chance traffic, i.e. all call arrivals and terminations are independent random events
- There is statistical equilibrium, i.e., the average number of calls does not change
- Full availability of the network, i.e., every outlet from a switch is accessible from every inlet
- Any call that encounters congestion is immediately lost.

From these assumptions Erlang developed the Erlang-B formula which describes the probability of congestion in a circuit group. The probability of congestion gives the Grade of Service experienced.[4]

30.8. [edit] Calculating the Grade of Service

To determine the Grade of Service of a network when the traffic load and number of circuits are known, telecommunications network operators make use of *Equation 2*, which is the Erlang-B equation.[4]

$$\text{Grade of Service} = \frac{\left(\frac{A^N}{N!}\right)}{\left(\sum_{k=0}^{N} \frac{A^k}{k!}\right)} \qquad (2)$$

A = Expected traffic intensity in Erlangs, *N* = Number of circuits in group.

This equation allows operators to determine whether each of their circuit groups meet the required Grade of Service, simply by monitoring the reference traffic intensity.

(For delay networks, the Erlang-C formula allows network operators to determine the probability of delay depending on peak traffic and the number of circuits.[4])

30.9. [edit] References

1. ^ *a b c* Kennedy I., Lost Call Theory, Lecture Notes, ELEN5007 – Teletraffic Engineering, School of Electrical and Information Engineering, University of the Witwatersrand, 2005
2. ^ *a b* Peuhkuri M., IP Quality of Service, Helsinki University of Technology, Laboratory of Telecommunications Technology, 1999.
3. ^ Farr R.E., Telecommunications Traffic, Tariffs and Costs – An Introduction For Managers, Peter Peregrinus, 1988.
4. ^ *a b c d e f g h i j* Flood, J.E., Telecommunications Switching, Traffic and Networks, Chapter 4: Telecommunications Traffic, New York: Prentice-Hall, 1998.
5. ^ Ritter, M., Phuoc, P., Multi-Rate Models for Dimensioning and Performance Evaluation of ATM Networks, COST 242, Institute of Computer Science, University of Würzburg, June 1994