2014 *№ 1 (79)*

UDC 621.391

OPTIMAL ROUTING: QUALITY OF SERVICE

N.I. LISTOPAD, A.A. HAYDER

Belarusian State University of Informatics and Radioelectronics
P. Brovki 6, Minsk, Belarus

Поступила в редакцию 17 сентября 2013

The article analyses a new approach of optimal routing of information flows in telecommunication networks using of QoS parameters. The state of art of QoS is presented. The general principals of optimal routing are formulated. The average packet delay in the network is analyzed more detail. In conclusion general problem for investigation has been formulated as problem based on Service-Oriented Architecture.

Key words: information flows, optimal routing, Service-Oriented Architecture.

Introduction

The optimal routing problem in a computer network consists of the determination of the optimal routing policy, i.e., the set of routes on which packets have to be transmitted in order to optimize a well-defined objective function (e.g., delay, cost, throughput etc.). Under appropriate assumptions, the optimal routing problem can be formulated as a nonlinear multi commodity flow problem [1].

General techniques for solving multi commodity problems can be found in the mathematical programming methods; however, the straightforward application of these techniques to the routing problem in computer networks proves to be computationally cumbersome. In fact, the algorithms for the determination of optimal topology and channel capacities in a computer network require hundreds of optimal routing computations; therefore, an extremely fast routing technique has to be used. For that reason, considerable effort has been spent in developing heuristic techniques [1], [2]. Quite satisfactory results have been obtained and computational efficiency has been greatly improved; however, all of these techniques are affected by various limitations.

The Optimality Principle and shortest Path Routing

The optimality principle states that if router *J* is on the optimal path from router *I* to router *K*, then the optimal path from *J* to *K* also falls along the same route. As a consequence of that principle, we can see that the set of optimal routes from all sources to a given destination form a tree rooted at the destination. Such tree is called a sink tree [3].

The following technique is widely used in many forms, because it is simple and easy to understand. The idea is to build a graph of the subnet, with each node of the graph representing a router and each arc representing a communication line (link).To choose a route between a given pair of routers, the algorithm just finds the shortest path between them on the graph. The shortest path concept includes definition of the way of measuring path length. Deferent metrics like number of hops, geographical distance, the mean queuing and transmission delay of router can be used. In the most general case, the labels on the arcs could be computed as a function of the distance, bandwidth, average traffic, communication cost, mean queue length, measured delay, and other factors [3].

Quality of Service

Simply we can say that the Quality of Service (QoS) related to different aspects of telephony and computer networks that allow it to work under specific requirements, and working under the existing resources efficiently, so we can have a stable work and result.

In other words we can define QoS as:

– 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;

– QoS criteria are numerous and are highly depend of the application. There are throughput, delay, jitter, loss rate or of the end-user: image resolution, sound quality, appropriate language [4].

In a lot of causes QoS-routing should be provided by including constrain for the average packet delay [5]. The queuing plus transmission delay have frequently been approximated using M/M/1 model. As the results by the Kleinrock-formula [6]. For average packet delay in the network the following constraints should be noted*:*

$$\frac{1}{\gamma}\sum_{e\in E} f_e(s,t,e)\left[\frac{1}{y_e(s,t,e)-f_e(s,t,e)}+\mu\left(P_e+K_e\right)\right]\le T_{\max}(s,t) \text{ for all } (s,t)\in D, \quad (1)$$

where $T_{\max}$ (s,t) – maximum possible delay; $1/\mu$ – the average packet length (bits/packet); λ_e – the average packet arrival rate to link e (packets/second); P_e – propagation delay on link e; K_e – node processing delay entering link e; γ – total traffic in the network (packets/second).

Delay is the amount of time between when a packet enters the network and leaves the network is most common factor considered in QoS metrics. Packet delays can be split into four components [7].

Processing delay is the time needed by network elements such as routers or end systems to process a packet. It depends on the processing speed of the network element hardware and the complexity of the functions to perform. These range from simple packet classification for forwarding and fire-walling, to complex payload modifications for encryption and content adaptation.

Network components normally have input and/or output queues. The time a packet resides in these queues is called queuing delay. Queues become larger when the network becomes congested, which results in a longer queuing delay.

Transmission delay is the time needed to transmit a packet at a specific bit rate. It can be calculated as

transmission delay = (number of bits to transmit(/(transmission ratio). (2)

The propagation delay describes the time needed by the signals to travel (propagate) through the medium. It can be calculated as

propagation delay = (physical distance)/ (propagation velocity). (3)

Propagation and queuing delay are the key contributors to delay as long as no heavy processing like encryption or packetisation by applications is needed. In real-world networks, packets experience a delay on their path from the sender to the receiver, which is not constant but rather varying over time, because conditions on a route and the involved systems change. This is a result of the fluctuation of Internet traffic and resulting queue sizes. The delay is bounded by a minimum and maximum delay. The difference between these bounds is called delay variation (jitter). The delay variation can be compensated by buffering packets, either within the network elements (routers) or the receiving end systems. Since end-system memory is much cheaper than router memory, buffering in the end system is usually preferred.

There are some ways to determine maximum possible delay and bandwidth. First of all, you should allocate $T_{\max}$(s,t) и $y_{\max}$ empirically, for example, from performance required by any application.

Actually $y_{\max}$ is due to technical possibilities of the telecommunication links.

In [5] it proposed the algorithms to allocate maximum delay for each route any links and the network at the whole. An objective function could be more complicated and include other requirements of QoS not only delay, but, for example, cost of delay for each link e [6]:

$$T(s,t,e) = \beta \frac{f_e(s,t,e)}{y_e(s,t,e) - f_e(s,t,e)}, \quad (4)$$

where $T(s,t,e)$ is cost delay for link of the demand (s,t); β is cost factor.

QoS Concepts. Fundamentally, QoS enables you to provide better service to certain flows. This is done by either raising the priority of a flow or limiting the priority of another flow. When using congestion-management tools, you try to raise the priority of a flow by queuing and servicing queues in different ways. The queue management tool used for congestion avoidance raises priority by dropping lower-priority flows before higher-priority flows. Policing and shaping provide priority to a flow by limiting the throughput of other flows. Link efficiency tools limit large flows to show a preference for small flows.

Cisco IOS QoS is a tool box, and many tools can accomplish the same result. A simple analogy comes from the need to tighten a bolt: You can tighten a bolt with pliers or with a wrench. Both are equally effective, but these are different tools. This is the same with QoS tools. You will find that results can be accomplished using different QoS tools. Which one to use depends on the traffic.

QoS tools can help alleviate most congestion problems. However, many times there is just too much traffic for the bandwidth supplied [8].

Basic QoS Architecture. The basic architecture introduces the three fundamental pieces for QoS implementation:

– QoS identification and marking techniques for coordinating QoS from end to end between network elements;

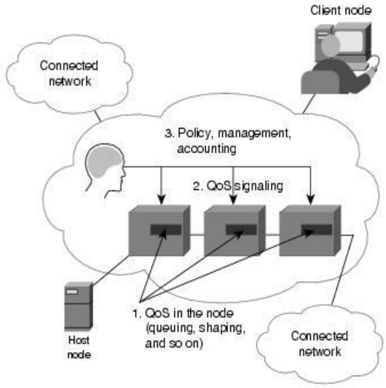
– QoS within a single network element (for example, queuing, scheduling, and traffic-shaping tools);

– QoS policy, management, and accounting functions to control and administer end-to-end traffic across a network [8].

QoS Policy Control. The QoS policy control architecture is being developed as a key initial piece of the CiscoAssure policy networking initiative. This initiative leverages standards-based QoS policy control protocols and mechanisms to implement QoS policy from a single console interface.

At the infrastructure level, packet classification is a key capability for each policy technique that allows the appropriate packets traversing a network element or particular interface to be selected for QoS. These packets can then be marked for the appropriate IP precedence in some cases, or can be identified as a Resource Reservation Protocol (RSVP) [9]. It is a Transport Layer protocol designed to reserve resources across a network for an integrated services Internet. RSVP operates over an IPv4 or IPv6 Internet Layer and provides receiver-initiated setup of resource reservations for Multicast or Unicast data flows with scaling and robustness. It does not transport application data but is similar to a control protocol, like Internet Control Message Protocol (ICMP) or Internet Group Management Protocol (IGMP). RSVP is described in RFC 2205.

RSVP can be used by either hosts or routers to request or deliver specific levels of quality of service (QoS) for application data streams or flows. RSVP defines how applications place reservations and how they can relinquish the reserved resources once the need for them has ended. RSVP operation will generally result in resources being reserved in each node along a path. RSVP is not a routing protocol and was designed to inter-operate with current and future routing protocols [8].



Basic QoS implementation has three main components

Uses and Benefits of QoS

Network administrators can use QoS to guarantee throughput for mission-critical applications so that their transactions can be processed in an acceptable amount of time. Network administrators can also use QoS to manage User Data Protocol (UDP) traffic. Unlike Transmission Control Protocol (TCP), UDP is an inherently unreliable protocol that does not receive feedback from the network and, therefore, cannot detect network congestion. Network administrators can use QoS to manage the priority of applications that rely on UDP, such as multimedia applications, so that they have the required bandwidth even in times of network congestion, but do not overwhelm the network.

QoS provides the following benefits:

– gives administrators control over network resources and allows them to manage the network from a business, rather than a technical, perspective;

– ensures that time-sensitive and mission-critical applications have the resources they require, while allowing other applications access to the network;

– improves user experience;

– reduces costs by using existing resources efficiently, thereby delaying or reducing the need for expansion or upgrades [9].

Conclusions and further extensions

The main features of the presented models of the optimal routing are taking in account the QoS requirements, in particular, delay and bandwidth metrics as very important requirements for a lot of applications. Especially the results of this work should be relevant for network clouds as a main part of the new network architecture for Future Internet.

In majority of the works as mentioned in [2,3] the traffic on the link is presented be M/M/1 model. But more realistic model, that could be described the traffic at the networks is model, based on BMAP–flows. This is allowed for network design procedures to utilize more realistic models and characterizations of traffic behavior both in the calculation of network delay and in the sizing of network links.

Network design problems with QoS consideration are typically difficult solve combinatorial problems. Success in these directions of research will enable network designers to any practical problems for optimal routing information flows with QoS requirements.

General Problem for investigation should be formulated as problem based on Service-Oriented Architecture approach [10]. Using the approach presented in [10]. We define the objective function as the follows:

$$F(x) = w_i \sum_{i=1}^{m_1} \frac{L_{i\max} - L(x)}{L_{i\max} - L_{i\min}} + w_j \sum_{j}^{m_2} \frac{L(x) - L_{j\min}}{L_{j\max} + L_{j\min}}, \quad (5)$$

where $m = m_1 + m_2$.

It’s necessary to solve the following optimization problems LP:

$$\max F(x) \quad (6)$$

subject to the constrains.

In other words the General problems (5)-(6) for QoS requirements of telecommunication networks could be formulated as the following: it’s necessary to provide maximum of bandwidth and minimums of delay, jitter and packet loss ratio; for this requirements m_1=1 and m_2=3.

In a continued evolution toward end-to-end services, Cisco is expanding QoS interworking to operate more seamlessly across heterogeneous link layer technologies, and working closely with host platform partners to ensure interoperation between networks and end systems. QoS is on the forefront of networking technology. The future brings us the notion of user-based QoS in which QoS policies are based on a user as well as application.

References

1. *Frank H., Chou W.* // Networks. 1977. Vol. 1. P. 99–122.
2. *Fultz G.L.* Adaptive routing techniques for message switching computer communication networks. California, 1972.
3. cs.bgu: Computer Communication and Distributed Algorithms Laboratory. [Электронный ресурс]. – Режим доступа: http://www.cs.bgu.ac.il/research/atm_lab/ROUTING_.doc. – Дата доступа:16.09.2013.
4. web.univ-pau: QoSCloud. [Электронный ресурс]. – Режим доступа: http://web.univ-pau.fr/~cpham/PROJETS/PIREGRID/QoSCloud.pdf. – Дата доступа:16.09.2013.
5. *Girlich E., Kovalev M.M., Listopad N.I.* // Otto-von-Guericke-Universität Magdeburg. 209. № 21.
6. *Resende M., Pardalos P.* Handbook of Optimization in Telecommunications. Germany, 2006.
7. *Braun T., Statub T.* End-to-end Quality of Service Over Heterogeneous Networks. Germany, 2008.
8. Cisco DocWiki: Quality of Service Networking. [Электронный ресурс]. – Режим доступа: http://docwiki.cisco.com/wiki/Quality_of_Service_Networking#QoS_Concepts. – Дата доступа:16.09.2013.
9. Technet: What Is QoS?. [Электронный ресурс]. – Режим доступа: http://technet.microsoft.com/en-us/library/cc757120(v=ws.10).aspx. – Дата доступа:16.09.2013.
10. *Cardellini V., Casalicchio E., Grassi V. et. al.* / Proceedings of the 6th International ICST conference on heterogeneous networking for quality, reliability, security and robustness and 3rd International Workshop on advanced architectures and algorithms for internet delivery and applications, Las Palmas, Gran Canaria, November 23–25, 2009. P. 431–447.