An Internet Web Management Policy for Government Organisation

Pattarasinee BHATTARAKOSOL
Department of Mathematics
Faculty of Science
Chulalongkorn University
Tel. 662 2185143, Thailand
E-mail: bpattara@sc.chula.ac.th

Vimolrat NGAMARAMVARANGGUL
Department of Mathematics
Faculty of Science
Chulalongkorn University
Tel. 662 2185141, Thailand
E-mail: vimolrat.N@chula.ac.th

ABSTRACT
The Internet Web Traffic Monitor has been developed to show the flow of data, in order to improve on performance of network technologies as well as new innovation. Now, the efficiency of data transfer within networks is interesting for comparing with other networks. By doing so, useful network information from networks within Chulalongkorn University can be accumulated, and the accuracy of such information will allow the Ministry of Information and Communication Technology to better manage and adjust their networks for proper organization. This paper presents factors that effect the Internet traffic based on web applications. The data that are related to the browsed information indicates the need of bandwidth required by each application area. This accumulated data can then be used when setting the network management policy for granting the bandwidth, so that suitable proportions can be allocated for each type of application, thus reducing the delay time of users.

Keyword: Internet Web Traffic Monitor, bandwidth, network

1. INTRODUCTION
Since the Thai government has a policy to transform all government processes to be on the Internet, the web application usage is an unavoidable need from every government organization. If we consider the variation of government area, it would be similar to the variation of research area in the Faculty of Science, Chulalongkorn University. Additionally, the government had implemented an organization to manage all network system of all government sections. This situation is similar to the network management policy of Science Faculty.

The Sciences Faculty consists of 14 individual departments and separated to be three main areas of study and research: Biological Science, Physical Sciences, and Technological Science. Each area performs web browsing to support their study and research. Therefore, the transactions from 14 departments trying to browse the Internet cause the current collision situation of the faculty core network with a long delay for users.

In order to solve the long delay and heavy traffics of a network, a suitable management policy should be settled. As mentioned by Hemmen [2], each organization must have their own network management policy to maintain the optimal availability of the network system to serve users. Many researches have introduced technical and non-technical policies to manage network's traffics and guarantee QoS of network situations and usages.

Section 2 will present some related work in network management policy. A detailed study of the system environment in this research will be described in Section 3, and the data measurement method will be mentioned in Section 4. Section 5 will have details of data analysis followed by discussion in Section 6. The proposed solutions will be in Section 7, and the last section is concludes the research.

2. RELATED WORKS
Before talking in depth about the Internet Web Traffic Monitor, this paper will begin with a literature review of recent developments in network management. Saad and Bakry [1] identified a variety of network profiles for better managed networks. They outlined the reasoning behind structuring information networks, and why the changes affecting these networks demand more efficient and sophisticated planning. An enhanced method of managing network development was discussed by means of a "comprehensive information profile" [1].

Hemmen [2] listed five models used for network organisation and management. He pointed out the lack of research on the organisational aspects of network control and that many authors tended to look at this problem from a purely technical point of view. An interesting reflection of the five organisational methods is made, by comparing important key factors for successful network management.

A policy based bandwidth management system in Java was introduced by Jha and Hassan [3]. Their implementation interfaced the Linux Diffserv format with the standard policy protocols. Java provides an effective object orientated environment through which they performed two practical experiments. These demonstrated the effectiveness of their method in supporting Quality of Service (QoS) within an enterprise network, by new architectures of policy based dynamic resource allocation.

D.W-K & C.S. Hong [4] developed a quality of service management framework which they applied to an ATM network, to test it's efficiency at co-ordinating a distributed multimedia system. The high performance system was designed from a computational viewpoint using the Open Distributed Processing (ODP) concept. Their method met the QoS necessary to assure complete flow of media from the remote server across the network to the points of delivery (the customers).

By the transition from routers to switches, modern networks can gain from tightly segmented workgroups and boosted network performance. This however comes at the cost of increased segment management for network managers, and runs the risk of overloading the communication bandwidth in the system. To properly solve this quandary while improving management and control for enterprise networks, a policy based management is required (as suggested by Wright [5]). This technique also allows the capability for grouping of users into specific networks, while also applying directory based QoS parameters.

Apart from the logistics of network management, the economic cost plays a very important role. This is discussed by Gupta et al [6] who compare the future growth of the internet with the cost of infrastructure investment. This economic viewpoint, while initially unpopular is become an increased focus of interest with researchers. In order control the various QoS required by different users (and the applications they use), accurate pricing of network traffic is essential. Their article introduced the characteristics, performance, and requirements of a price setting mechanism which is organised in a decentralised manner. By pricing in this decentralised manner the risk of network congestion and failure is minimised.

Burrel et al [7] proposed a CDMA transmission policy, combined with a high-level traffic monitoring protocol. The transmission / traffic management technique used a moving boundaries concept and was applied to the management of a multimedia wireless network, where it proved to be a powerful resource for the accommodation of traffic in the wireless networks. Effective tracking of the wireless networks dynamics was attained, with minimal delays, capacity waste rates, and rejections.

3. SYSTEM ENVIRONMENT

Chulalongkorn Universities Faculty of Science has installed a 2-Gigabit Ethernet system and has connected to the university core network that is based on the ATM topology. The network controlling system is located at the Faculty Computer Center (FCC) only. Additionally all LANs of each department are hooked together to the connection systems at FCC before transactions are transferred out from the faculty.

According to the Faculty's organization, there are 14 departments in Faculty of Science, Chulalongkorn Universy. We can classified these departments into 3 groups: physical sciences, biological sciences, and technological sciences. The physical sciences group consists of Chemistry, Mathematics, Physics, and Geology departments. The biological sciences group consists of Biology, Biochemistry, Botany, Marine Sciences, Microbiology, General Sciences (Environmental Sciences). The technological sciences group consists of Food Technology, Chemical Technology, Material Sciences, and Photographic Science and Printing Technology.

Since the Faculty of Science consists of 14 individual departments where more than 500 academic staffs and 3000 students are working, information searching from web pages is unavoidable. Therefore, in order to reduce the Internet traffics of faculty before sending packets through the university core network, a faculty proxy system was setup at FCC. Thus every web browsing must be filtered by this proxy system. The network architecture of Faculty of Science is presented in Figure 1.

4. DATA MEASUREMENT METHOD

The proxy server was installed at only one place: the faculty computer center. Therefore all web browsing is forced to pass the filtering system of proxy. Thus it is easy to monitor the web usages from each department. This research monitors and captures the network traffic of using web applications from all departments under Faculty of Science for 10 days, 200 hours continuously. All information about browsing web pages were stored in the individual files, using Squid software. Squid stores transactions occurred in one hour into a file, so there are 200 transactions files that contains 4273 transactions using in this research.

Each record in a file contains information of the browsing date, IP address of browsing machine, browsed web site, situation of browse (HIT/MISS), the size of the browsed file, and the elapse time for each browsing. Since each department has a different IP address, the department name can be identified.

If the required web information was stored in the cache of proxy then the situation of browse is marked as "HIT", otherwise it will be "MISS". In situation of HIT, the required information came from cache of proxy. On the other hand, the required information will be retrieved from Internet.

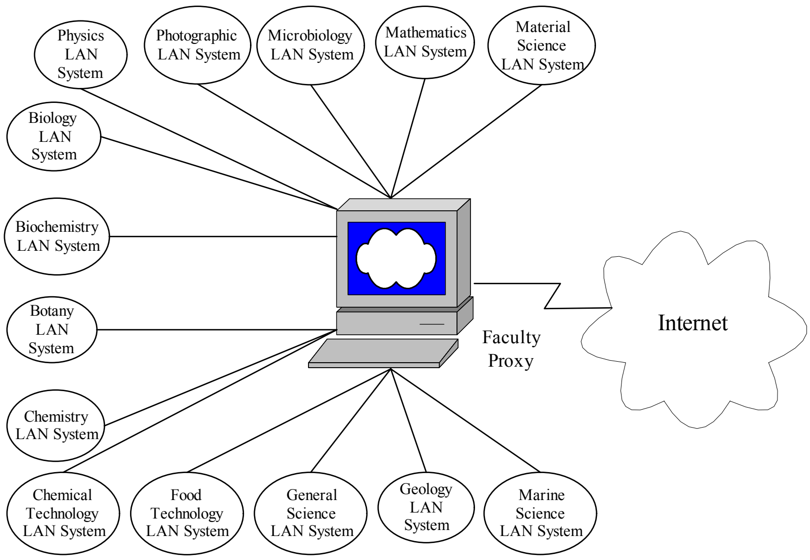

Figure 1. Network Architecture of Faculty of Science

5. DATA ANALYSIS

The aim of installing the proxy server is for most of the web demands to be controlled and limited only in the cache of the installed proxy. Considering 4273 transactions, the rates of HIT and MISS are 2014 and 2259 respectively. From Table 1, we can see that staffs of Department of Biology and Department of Photographic Science and Printing Technology usually retrieve the same web site for their required information (nearly 50% of HITs) while staffs from the Department of General Science prefer retrieving new web sites from Internet (38% of HITs). Additionally the statistical data shows that the retrieved information is mostly obtained directly from Internet (more than 50% of MISS situations).

Considering Figure 2 which shows the sizes of departmental loading data, we can see that there are 3 departments that request data size larger than 40 MB. These departments are Chemistry, Mathematics, and Physics. The three departments are in physical sciences as grouped in the previous section. However when we consider the sizes of web browsing based on the defined groups in the previous section, the average loaded web sizes are concluded in Table 2.

From Table 2 and Figure 2, the highest average loaded data sizes from Internet belongs to the Physical Sciences area, and in contradiction, the Technological Sciences always load small data sizes. The data sizes required by groups of Biological and Technological Sciences are nearly 6 times smaller than the data size of Physical Sciences area.

Table 1. Statistical Data for HIT and MISS rate, classified by departments and sorted by % of HIT.

Department Name	HIT	MISS	SUM	% of HIT
General Science	82	134	216	37.96
Chemistry	136	179	315	43.17
Chemical Technology	124	163	287	43.21
Food Technology	90	107	197	45.69
Microbiology	171	188	359	47.63
Botany	175	190	365	47.95
Mathematics	86	92	178	48.31
Biochemistry	173	185	358	48.32
Physics	188	201	389	48.33
Marine Science	180	192	372	48.39
Material Science	107	112	219	48.86
Geology	193	201	394	48.98
Biology	190	195	385	49.35
Photographic Science and Printing Technology	119	120	239	49.79

Table 2. The average web sizes loaded in 10 days; classified by the group's applications.

Group's applications	average sizes (bytes)
Technological Sciences	10262609.4
Biological Sciences	12838853.2
Physical Sciences	62332827.7

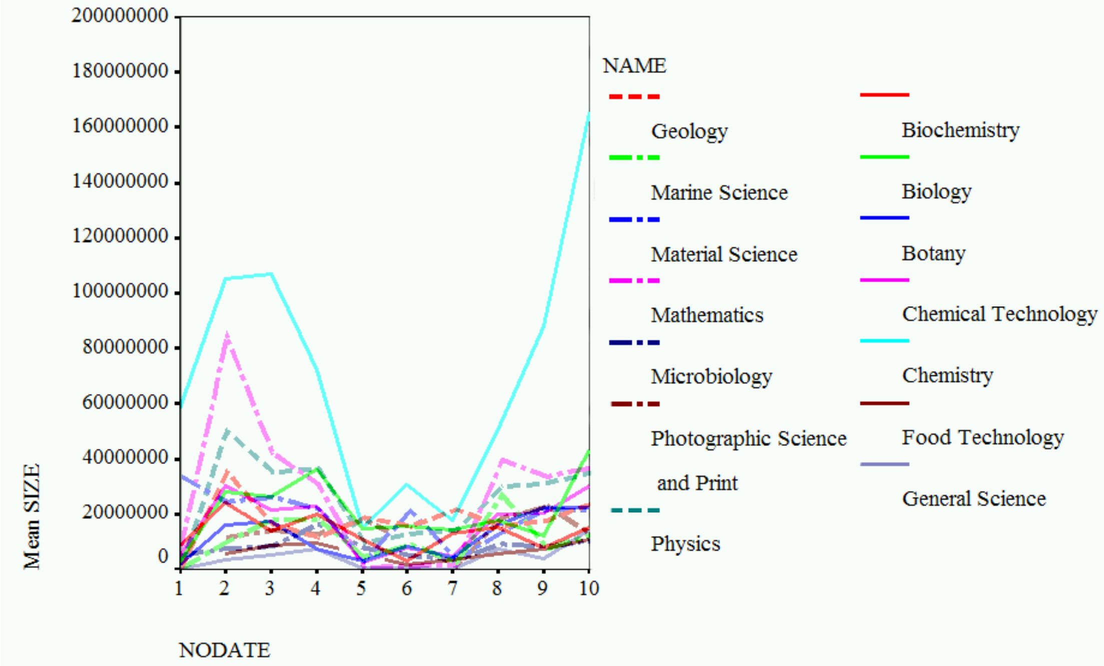

Figure 2. Comparisons of retrieving web sizes between different departments

However, since we consider the HIT rates in Table 1, we can see that these three departments, especially Department of Chemistry, usually retrieve new web sites to load data (HIT rates lower than 50%). Moreover, the lowest HIT rate belongs to Department of General Science (37.96%) and there are 11 departments, which retrieve data from the previous web pages with the HIT rates greater than 45%. These departments are Biochemistry, Biology, Botany, Food Technology, Geology, Marine Science, Material Science, Mathematics, Microbiology, Photographic Science and Printing Technology, Physics.

The largest data streams were seen to occurs from the Department of Chemistry, Mathematics, and Physics. However, data in Table 3 shows that browsing transactions mostly come from the Department of Geology (9.22%), Department of Physics (9.10%), and Department of Biology (9.01%). Nevertheless, Table 4 shows that the number of calls from Biological Sciences is the highest number (nearly 48% of all transactions) while the departments in the Technological Sciences do not often retrieve data (22.05% of all transactions).

6. DISCUSSIONS

Considering data in Table 1, we can see that every department retrieves data from new web locations more than previous locations that are kept in the cache of faculty proxy system. However there are 11 departments that have HIT rates higher than 45%. Thus the installed proxy machine may be too small for handling these types of requirements.

Table 3. Frequencies of calls; classified by department name.

Department Name	% of calls
Mathematics	4.17
Food Technology	4.61
General Science	5.05
Material Science	5.13
Photographic Science and Printing Technology	5.59
Chemical Technology	6.72
Chemistry	7.37
Biochemistry	8.38
Microbiology	8.41
Botany	8.54
Marine Science	8.71
Biology	9.01
Physics	9.10
Geology	9.22

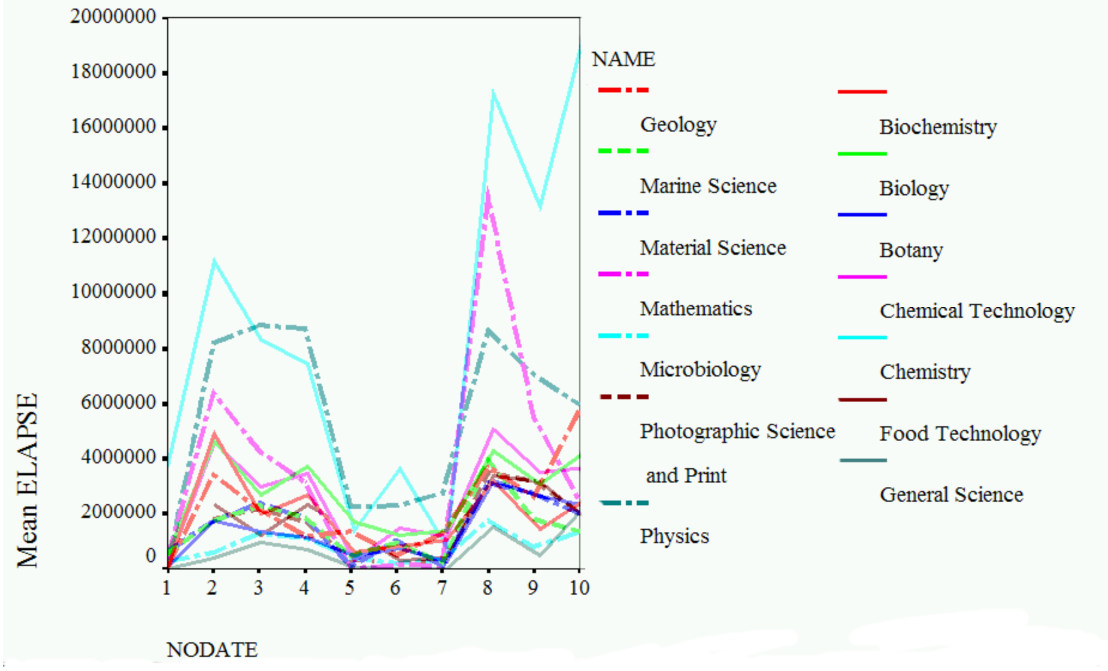

Figure 3. Comparisons of retrieving times between different departments

Table 4. Frequencies of calls; classified by group's applications

Group's Application	% of calls
Technological sciences	22.05
Physical sciences	29.86
Biological sciences	48.10

As we can see from Figure 2, the Department of Chemistry and Department of Mathematics retrieved a large size of data from web system. But when considering the retrieving frequency in Table 2 we can see that these two departments, especially Department of Mathematics, do not frequently retrieve data from the web applications. Therefore, we can conclude that although these two departments do not retrieve data from web applications often but each time of their loading the required data is usually larger size than others. On the other hand, the Physics Department always retrieves data from the web system and the loading data is larger than the applied sciences departments. Additionally, Table 4 illustrates that departments in Biological Sciences always retrieve data from Internet but these data are in small sizes.

Although the Department of General Sciences is classified as Biological area because the main research area of this department is the Environmental Science, there are other studying areas in this department such as applied physics, science communications, applied chemistries, etc. Therefore, there are various types of requests from this department, even though the retrieval rate is only 5.05% of total calls. According to these various requirements, it is difficult to maintain required data in the cache system using proxy. Thus, the loading data of this department usually obtained from Internet.

7. PROPOSED SOLUTIONS

From the presented tables and graphs, each department and each classified group require different bandwidth data transfer but there is no current policy from FCC to manage the different bandwidth requirements. Thus, the elapse times of each department are high.

Considering 11 departments that have their average HIT rates higher than 45%, if these departments install their own proxy then the traffics running in the core network of the faculty will be reduced. Consequently, the elapse times of each retrieval will be reduced because loading data mainly comes from their local proxy machines.

However, it is the fact that not every department can manage their local proxy. So, FCC might set a new policy to split their main proxy system to be 2 types of proxy services. One small to medium proxy size serve departments in Technological Sciences and Biological Sciences. The other proxy machine in a larger size should manage only transactions from departments in Physical Sciences. Figure 4 demonstrates the proposed network architecture that can serve different needs from different groups.

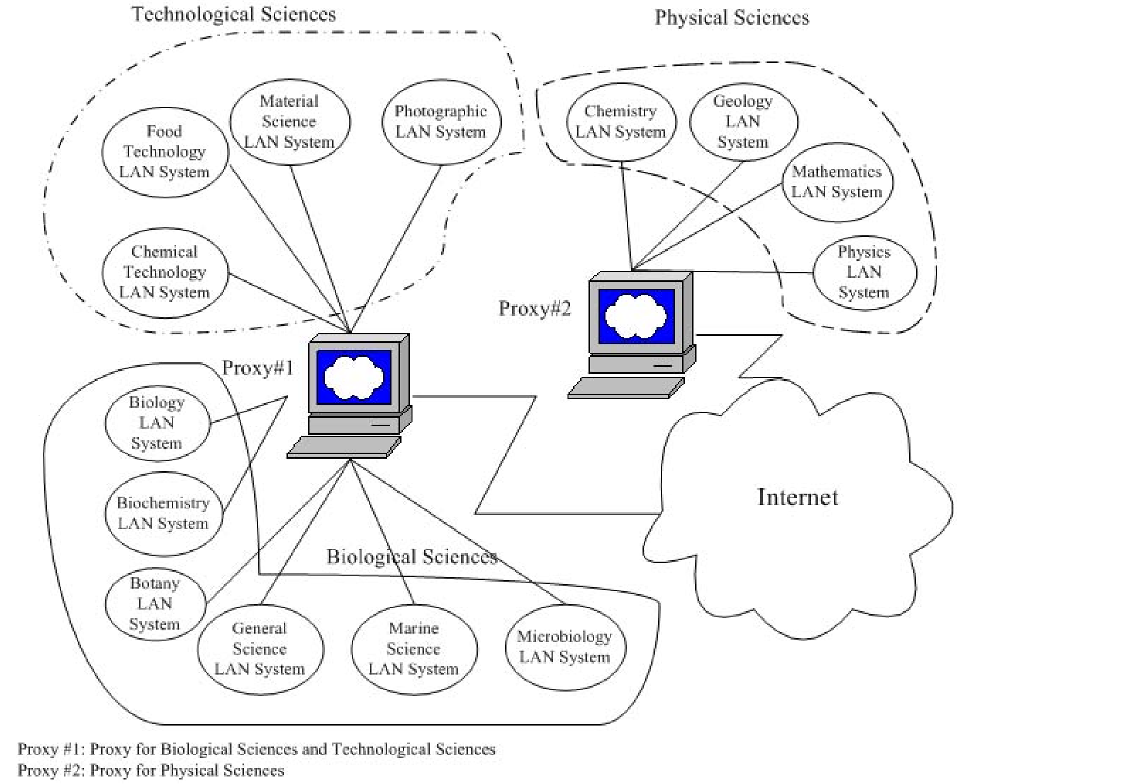

Figure 4. The proposed network architecture

According to the fact that loading data size from Physical Sciences departments is usually larger than others, the high bandwidth is requested. However, this type of requirements does not frequently occurs so it is not suitable to permanently grant the transmission of bandwidth for this requirement. Therefore, a suitable policy is to lock a certain set of transmission ports to serve departments in Physical Science and this port allows high bandwidth to be transferred.

Let T be the total bandwidth of Sciences Faculty provided by university. Let x be the bandwidth requested from the Physical Science departments. Let y and z be the bandwidth requested from the Technological Sciences and Biological Sciences respectively. Thus, $T = x + y + z$. According to statistical data in Table 2, the sizes of loading data from the Technological Sciences and Biological Sciences are small and close to each other. Then, $y = z$. Additionally, from Figure 2, there are some times that the average loading sizes of Physical Science departments is also close to the other two groups. Thus, in normal situations the bandwidth required for each group is $x = y = z$.

Once a high data load arises, according to a request from the Physical Sciences departments to a port of SP, the bandwidth for data transmission of this port must be expanded. The expanding size should be 6 times higher than the other two bandwidths. Thus, if T is the total bandwidth, $T = x + y + z$, where $x = 6y = 6z$. Therefore the total bandwidth can be written as follow.

$$T = x + y + z$$

$$\text{where } \frac{T}{3} \le x \le \frac{6T}{8}$$

$$\text{and } \frac{T}{8} \le y \le \frac{T}{3}, \frac{T}{8} \le z \le \frac{T}{3}$$

8. CONCLUSIONS

This paper has shown the data usages through web applications of departments in Faculty of Science, Chulalongkorn University. Since the faculty separates the academic area to three different areas: Physical Science, Biological Sciences, and Technological Sciences, we can see that the data usages of these three are different. The data analysis shows that most traffics belong to Physical Science departments, while the other areas frequently retrieve data but in small sizes. There are two proposed solutions.

The first solution is to install separated proxy machine for these three groups. One small proxy for Technological Sciences, and Biological Sciences, and the other large proxy for Physical Sciences only. Implementing this solution, the waiting time of users can be reduced, and the

amount of transactions run around the faculty core network system to Internet can be reduced.
The second solution is to implement an algorithm to manage bandwidth as a variable bandwidth rate system. This algorithm will manage bandwidth requirement based on the required data sizes from time to time and since the frequency of retrieving large data sizes is not high, the modification bandwidth policy will only slightly effect transactions from other two areas.

According to this research, it has indicated that each area does not require the same amount of bandwidth usages for web applications. Therefore, instead of granting equal bandwidth to every government's organization, the government network organization management should consider to implement different bandwidth network for organizations in a proper way so that the investment cost for implementing such network will be minimised, and users will find it more worthwhile to use such a network.

9. REFERENCES

[1] Saad and Fahed Haj Bakry paper: "Identifying information network profiles for planning management"
[2] L. J. G. T. van Hemmen paper: "Models supporting the network management Organization"
[3] S. Jha*† and M. Hassan paper: Java implementation of policy-based bandwidth management
[4] Daniel Won-Kyu Hong and Choong Seon Hong paper: "A QoS management framework for distributed multimedia systems"
[5] Michele Wright paper: "Using Policies for Effective Network Management"
[6] Alok Gupta, Dale O. Stahl, and Andrew B. Whinston paper: "The Economics of network management"
[7] Anthony Burrell, Harold P. Stern and P. Papantoni-Kazakos paper: "Transmission policies and traffic management in multimedia wireless networks".