بسم الله الرحمن الرحيم

University of Khartoum

Faculty of Engineering and Architecture

Internet Backbone Network Traffic in Sudan

A Thesis Submitted for Partial Fulfillment of the Requirements for the Degree of MS.c in Computer Architecture and Networks

By: ***Hiba Mohammed Osman***

Supervisor: ***Dr. Sami Mohammed Shariff***

Dedication:

To You ……………….

Acknowledgments:

I would like to thank Dr. Sami Sharif, staff of faculty of Engineering, my family and the staff of Data Cloud sector in Sudatel.

Table of Contents

Abstract

The growing demand of Internet service in Sudan makes it important to have accurate traffic statistics in order to help in the planning and provision of Internet links. This thesis presents statistics of the Internet traffic for Users connected through the public frame relay network .It also presents some observations that may help in improvement of the service to be of a better quality and affordable cost.

Measurements and statistics in this research showed that most of the Internet users are concentrated in Khartoum center and Khartoum south areas, and are allocated percentages of the total bandwidth between 43% for users in Khartoum center area and 37% for users in Khartoum south area. Statistics of bandwidth distribution per user category showed that 42.9 % of the bandwidth is allocated to Service providers. Statistics of the daily traffic variations showed that the Overall traffic has its peak hour at 2 pm and drops to its minimum at 6 a.m. it also showed that the mean of traffic intensity is twice its standard deviation and the variance is very high if related to the mean. Per weekday analysis of the traffic showed that the average traffic utilization per weekday for the aggregate links maintains the same level with slight drop on Fridays. From the study of traffic variations per user category it is observed that traffic generated by local service providers has a great impact on the overall traffic pattern. The percentage of traffic uploaded and downloaded by each of the international links implies that the routing policy used to load share the upload traffic between the international links resulted in an equal load sharing. And that the shortest path of 18 % of downloads traffic is through Emirate link. A straight line with a slope of 0.5 can represent the relation between download and upload traffic for aggregate links. The packet arrival model of the aggregate links do not follow Poisson distribution, and is proved to be self similar .The queuing delay increased sharply for utilization values greater than 55%, and maintains the same level for utilization values less than 53%.

خلاصه

الطلب المتزايد علي خدمة الانترنت في السودان يجعل من الضروري عمل دراسات احصائية دقيقه للحركة في شبكة الانترنت حاليا و ذلك لاستخدامها في التخطيط لتوسعة سعات الانترنت و لتقديم خدماتها بالجوده المطلوبه .تقدم هذه الرساله احصائيات للحركة بشبكة الانترنت الرئيسيه في السودان و التي توفر الخدمه للمشتركين عن طريق تقنية ترحيل الاطر . الرساله تقدم ملاحظات قد تساعد في تطوير خدمة الانترنت من حيث الجوده و التكلفه .

توصلت هذه الدراسه الاحصائيه الي ان مستخدمي الخدمه يتمركزون في منطقتي الخرطوم وسط و الخرطوم جنوب و يستغلون نسب من السعات الحاليه تتراوح بين ٤٣% للخرطوم وسط و ٣٧% للخرطوم جنوب . بينت الدراسة

التي اجريت لتوزيع السعات بحسب نوع المشترك ان مقدمي الخدمة المحليين يستغلون ٤٢,٩% من السعات الحاليه .كما ظهر ان استغلال الخطوط الحاليه يصل الي اقصي درجاته عند الثانيه ظهرا و يقل الي اقل درجه له عند الساعه السادسه صباحا . تبلغ قيمة متوسط الحركة ضعف قيمة انحرافها الطبيعي . اظهرت الدراسه ان مستوي الحركة في جميع ايام الاسبوع متطابق ما عدا في يوم الجمعه حيث ينخفض قليلا . يؤثر التباين في حركة الانترنت الصادره من مقدمي الخدمه المحليين في الشكل العام لتباين الحركة الاجماليه . تؤدي السياسه المتبعه في تقسيم الحركة بين خطوط الانترنت الصاعده الي نسبة تقسيم متساويه .كما ظهر ان ١٨% من الحركة الوارده تاتي عن طريق خط الربط مع الامارات . بينت النتائج امكانية تمثيل العلاقه بين الحركة الصاعده و الوارده بخط مستقيم ميله يساوي ٠,٥ . بينت دراسة التوزيع الاحصائي للحركة انها لا تتبع لتوزيع بويزن و وجد انها لها خصائص الحركة ذات التطابق المثلي . وضح من دراسة علاقة نسبة استغلال خطوط الانترنت بنسبة تاخير البيانات ان التاخير يزيد بصوره سريعه عند نسب استغلال تزيد عن ٥٥% وانه ثابت و قليل في نسب استغلال اقل من ٥٣ %.

Chapter 1

Introduction

The Internet service in Sudan is provided to organizations and internal service providers through the public frame relay network. There are four international links to ISPs in Europe, Canada and Emirates; satellites are used to provide international connectivity through Umharaz Earth station. Two types of Internet connectivity services are provided, committed information rate and burstable rate. These services are based on frame relay technology.

Demand to Internet service continues to increase; users began to develop critical applications that will be running over the Internet. Applications like ecommerce, video conferencing, voice over IP and distance learning are expected to come into being.

The planning phase of the introduction of new broadband Internet links has started and an ATM public network will be introduced shortly.

In order to assist on the planning, design and provisioning of the new Internet service with the required quality of service and price, a statistical analysis for the current Internet traffic is required. In order to adapt the traffic to the demand knowledge about the previous links behavior is required.

Having knowledge about the Internet users and capacity distribution will give a good idea about where and how to locate the Internet point of presence (POP) in the country, how to optimize the topology design, and how to design the required routing and redundancy of the network. It will also give a good idea about which categories of community does need support in this field.

Knowledge about Long and short term traffic variations will help in the prediction of busy hour and free hours during the day, which in turn will help in capacity planning. Also it may lead to the introduction of new cheaper services during free hours. Long-term statistics would help in the trend forecasting. Statistics about arrivals model will help in the queuing design, which in turn help in providing the required quality of service in terms of packet loss and queuing delay. The relation between download and upload will

help in capacity planning.

Statistics of the Internet Traffic could be found by gathering data from the Internet core router and switches, the data could be in terms of number of bytes or frames per time unit. Also a measure of delay and packet loss in routers and switches could be done. A lot of tools that use the standard SNMP protocol or any vender proprietary protocol are used to collect such information. A long-term data could be stored in large storage devices and then analyzed using proper statistical analysis tools to get the required analysis.

There are no previous researches in this field done in Sudan; hence there is no data from the Internet backbone of Sudan to compare with.

Many researches have been done for networks outside Sudan; some researches show evidence that the subset of network traffic that is due to World Wide Web transfers can show characteristics that are consistent with self-similarity [1], while other says that, as the traffic load increases --- that is, as the number of simultaneous transport connections increases --- arrivals tend to the Poisson model [2].

This thesis presents the work done to find the required statistics, and presents some recommendations to be taken into account during the design and implementation of the new Internet links.

Chapter two of this thesis contains some background information about the principle of Internet traffic engineering and the concept of grade of service, traffic variations and busy hour. Also it shows the Poisson distribution and self similarity concept and how to test whether certain traffic is self similar or has a Poisson model.

In chapter three the network studied is described and the points used in the measurement process are also mentioned. The chapter also contains information about the tools used in measurements and analysis and the type of information gathered from the network.

In Chapter four the results of the statistical analysis are discussed. The way the Internet users and bandwidth distributed through out the country is shown. The traffic variations are shown in daily, weekly and monthly basis. In this chapter the relation between the upload and download traffic have been represented using a linear equation .a traffic model estimation is stated and it is proved that the traffic is self-similar. Also in this chapter the optimum value of the link utilization that would give the best delay is stated. Chapter five contains comments and conclusions of the work done, with some recommendations for the provision of the Internet service, it also show the limitations of this research and what needs to be considered for further researches in this issue.

The attached CDROM contains the traffic data gathered from the network.

Chapter 2

Theory of Internet Traffic

The Internet traffic is an important parameter for network design and management. This traffic is generally studied by a number of models and theories. In this chapter the theoretical background of the Internet traffic concepts is presented.

2.1 Overview and Principles of Internet Traffic Engineering

The Internet traffic engineering is a very challenging field; that is due to the following reasons:

- The Internet support very critical communications infrastructure and applications like banking, voice and video applications.
- End users are demanding a very high quality service from their service providers

- The optimization of Internet backbones is an important issue

Internet networks need to be engineered in a way that insures the delivery of packets between nodes efficiently and cost effectively. Also resources should be effectively shared between users by configuring and adjusting sharing parameters.

2.1.1 Internet Traffic Engineering:

One of the important functions of the Internet is the routing of traffic between nodes. Internet traffic engineering is used to control the routing of traffic between nodes and to optimize it effectively.

In traffic engineering one should take into account performance as seen and required by end users, also economical considerations should be taken into account.

As stated in [3], the concept of Internet traffic Engineering is as follows:

“Internet traffic engineering is defined as that aspect of Internet network engineering dealing with the issue of performance evaluation and performance optimization of operational IP networks. Traffic Engineering encompasses the application of technology and scientific principles to the measurement, characterization, modeling, and control of Internet traffic. “[3].

The objective of teletraffic theory can be formulated as follows:

- “To make the traffic measurable in well defined units through mathematical models” [4] .
- “To derive the relationship between grade-of-service and system capacity in such a way that the theory becomes a tool by which investments can be planned. “ [4]

The following points, define the main tasks that could be achieved by using the teletraffic theory:

- “ to design systems as cost effectively as possible with a pre-defined grade of service when we know the future traffic demand and the capacity of system Elements “[4].
- “To specify methods for controlling that the actual grade of service is fulfilling the requirements, and also to specify emergency actions when systems are overloaded or technical faults occur “[4].

Generally, the above tasks can be achieved by the following methods and techniques:

- Demand forecasting methods (e.g. based on traffic measurements).
- System capacity calculations methods
- Specification of quantitative measures for the grade of service.

2.1.2 Modeling of telecommunication systems

To be able to analyze the Internet system a model must be setup to describe the system. This model consists of the following elements:

- The system structure
- The operational strategy
- The statistical properties of the traffic

The system structure is defined by the topology and concept and is defined in literature as: "The system structure is given by the physical or logical system, which is described in manuals in every detail. " [4]

The operational strategy is a set of methods that are generally used in order to adapt the traffic system to the demand. In computer engineering, this adaptation takes place by means of the operation system and by operator interference. In a telecommunication system, strategies are applied in order to give priority to call attempts and in order to route the traffic to the destination. [4]

The statistical properties of the traffic are used to model the user's demand. The statistical models are obtained by measurements on real systems; the measurement process must be of iterative nature. A mathematical model is build up from a thorough knowledge of the traffic. Properties are then derived from the model and compared to measured data. If they are not in satisfactory accordance with each other, a new iteration of the process must take place [4].

2.1.3Traffic Concepts and Grade of Service

Generally Communication systems are planned in away that adjust the amount of equipment so that the customer demand and grade of service is satisfied while keeping the cost as minimal as possible. Traffic engineering is used to adjust the amount of equipment depending on the amount of traffic. In this section some fundamental concepts of traffic are introduced.

The term traffic is Italian and means business. Traffic intensity describes to some extent the utilization of the resources, and is measured in data networks by bits per second (bps). The formal definition of traffic intensity in literature is as follows:

"The instantaneous traffic intensity in a pool of resources is the number of busy resources at a given instant of time." [4] .

Traffic variations are the changes in traffic patterns in hourly, weekly, monthly or yearly basis. Traffic variations have a relation with the behavior of users and their activities. Generally the traffic is generated by any user independently of each other.

Following is the definition of busy hour and traffic intensity concepts:

- Busy hour:

" A one-hour period within a specified interval of time (Typically 24 hours) in which the traffic loads in a Network or sub-network is greatest."[4]

- Traffic intensity:

"A measure of traffic loading with respect to a resource capacity over a specified period of time. In classical telephony systems, traffic intensity is measured in units of Erlang."[4]

The variations in teletraffic can are observed over different times and periods, the following observation periods are generally used:

- 24 hours variation.
- Weekly variations
- Variation during a year.

The traffic increases year by year due to the development of technology and economics in the society.

2.2 Poisson distribution:

The distribution of arrivals has an effect on the performance of the queuing system. One of the important distributions is Poisson distribution which is given by the following equation:

$$Pr\,[k \text{ items arrive in time interval } T] = \frac{(\lambda T)^k\, e^{-\lambda T}}{K!}$$

Expected number of items to arrive in time interval $T = \lambda T$.

Mean arrival rate, in items per second = λ .[1].

The above distribution is defined based on the following assumptions:

1-the probability of arrivals of a call in a small interval is proportional to the length of the interval and is independent of the amount of elapsed time since the arrival of the last call.

2-no two arrivals can occur at the same time.

2.3 Self-Similar Traffic

Studies [5, 1] showed that Internet arrivals are following self similar rather than Poisson distribution.

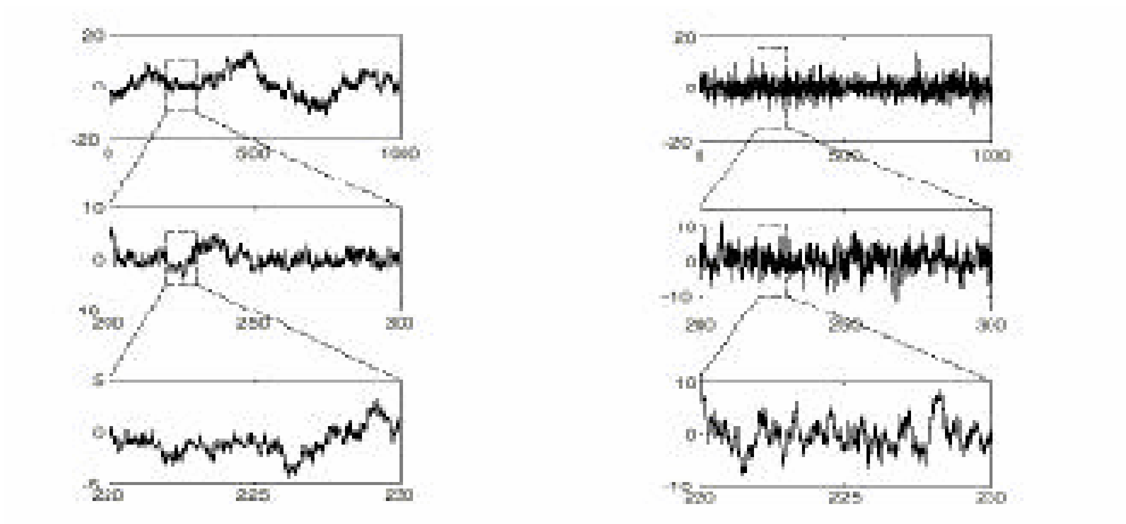
Self similar phenomenon is defined as follows:

"A phenomenon that is self-similar looks the same or behaves the same when viewed at different degrees of "magnification "or different scales on a dimension .this dimension can be space or time" [5]

If one compared between Poisson traffic and self similar traffic it is clear that

(1) in Poisson traffic only modest-sized buffers are needed and queue may build up in the short run but over a longer period the buffers cleared out, while in case of self similar traffic queue sizes are build up more than one expect from Poisson traffic.

(2) Statistics of the self-similar process do not change with a change in time scale as shown in fig2.1 below.



(a) Self-Similar Process (b) Non-Self-Similar Process

fig 2.1 Comparison between self-similar and non-self-similar processes.

Self similarity has a profound impact on performance; these impacts which are defined in [5] can be summarized as follows:

- . The higher the load on the networks, the higher the self-similarity.
- If high levels of utilization are required, larger buffers are needed for self similar traffic than would be predicted based on classical queuing analysis.

A number of approaches have been taken to determine whether a given time series of actual data is self similar and if so, to estimate the self similarity parameter H.
One of the more common approaches is Variance-time plot [5], which gives the following relationship between the variance of the aggregated time series $x^{(m)}$ of a self similar process and the aggregation parameter m. The variance obeys the following for large m,

$$Var(x^{(m)}) \sim Var(x)/m^{\beta}$$

Where the self similarity parameter H = 1-(β/2). This can be rewritten as

$$Log\ [Var(x^{(m)})] \sim log\ [Var(x)] - \beta log(m)$$

Because log [Var(x)] is a constant independent of m , if we plot $Var(x^{(m)})$ versus m on a log log graph , the result should be a straight line with a slope of -β . the plot can easily generated from a data series x(t) by generating the aggregate process at different levels of aggregation m and then computing the variance .
Slope values between -1 and 0 indicate self similarity [5].

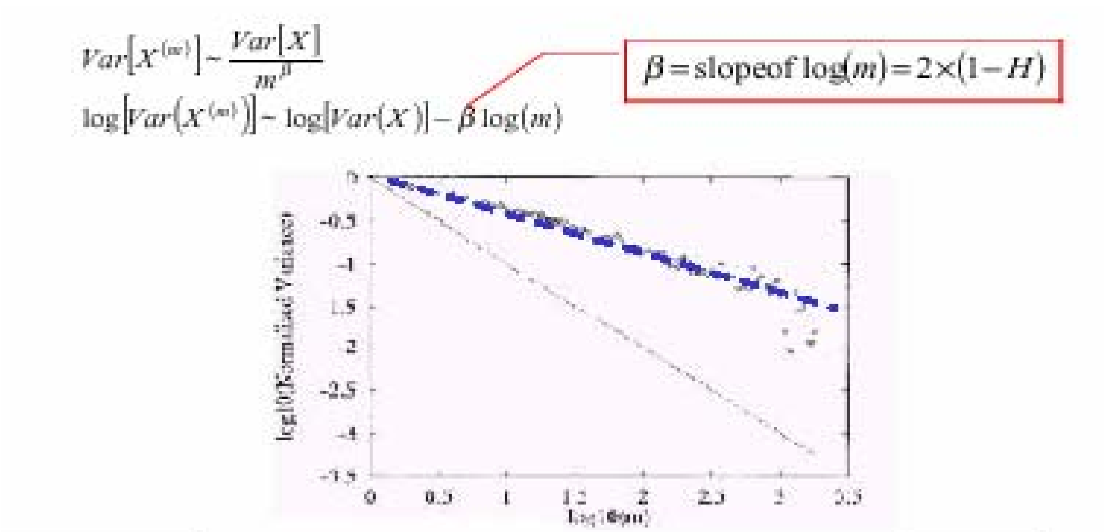


fig 2.2 Variance-time plot

2.4 Measurement Recommendations for Internet Traffic Engineering

There are some recommendations for the measurements and collections of the traffic statistics from the Internet network, in order to do the task of traffic engineering.

Following are the IETF recommendations for the measurements and statistics mechanisms collections:

“. The actions of these mechanisms should not adversely affect the accuracy and integrity of the statistics collected. The mechanisms for statistical data acquisition should also be able to scale as the network evolves.

Traffic statistics may be classified according to long-term or short-term time scales. Long-term time scale traffic statistics are very useful for traffic engineering. Long-term time scale traffic statistics may capture or reflect periodicity in network workload (such as hourly, daily, and weekly variations in traffic profiles) as well as traffic trends. Aspects of the monitored traffic statistics may also depict class of service characteristics for a network supporting multiple classes of service. Analysis of the long-term traffic statistics MAY yield secondary statistics such as busy hour characteristics, traffic growth patterns, persistent congestion problems, hot-spot, and imbalances in link utilization caused by routing anomalies."[3]

"Measured traffic statistics should provide reasonable and reliable indicators of the current state of the network on the short-term scale. Some short term traffic statistics may reflect link utilization and link congestion status."[3]

Chapter 3

Measurements:

The Internet network has been monitored for the period 17 April 2002 to 30-September 2002, some network monitoring tools were used to capture the required data and then a statistical analysis tool is used to analyze the data. In the coming section a description of the study network, the point used to measure the traffic and the tools used to monitor and analyze the traffic as well as the type of information gathered.

3.1 Description of the Measurement Points

The study network consists of four international links as shown in fig 3.1, which are named Emix, BT1,BT2 and Teleglobe , the first three are connected through Umharaz earth station and the last is connected using DVB/IP technology via a small dish at Khartoum Center . All of the links are terminated in the Internet core router (see table 3.1 for the capacity of each link). Emix, BT1and BT2 are connected using serial ports while Teleglobe is via Ethernet. These four connections in the router form part of our measurement points.

The other measurement points named as in fig 3.1, s1/1,s1/3,s2/0 these are serial ports on the Internet core router that connect the router to the public frame relay network and used to aggregate the users links in terms of frame relay PVC's (permanent virtual circuits). Beside s2/2, which is a point-to-point link. Table 3.2 describes the customer links specifications.

Table 3.1 Internet Links Specifications

Link Name	**capacity**	**Type**	**International ISP**	**Protocol**
EMIX	2Mbps	Full Duplex	Etisalat El imarets	**HDLC**
BT1	2Mbps	Full Duplex	British Telecom	**PPP**
BT2	2Mbps	Full Duplex	British Telecom	**PPP**
Teleglobe1	6Mbps CBR/ 12Mbps BBR	Simplex Downlink	Teleglobe	**Ethernet**

Fig 3.1 Top level Internet Connectivity Diagarm(sept 2002)

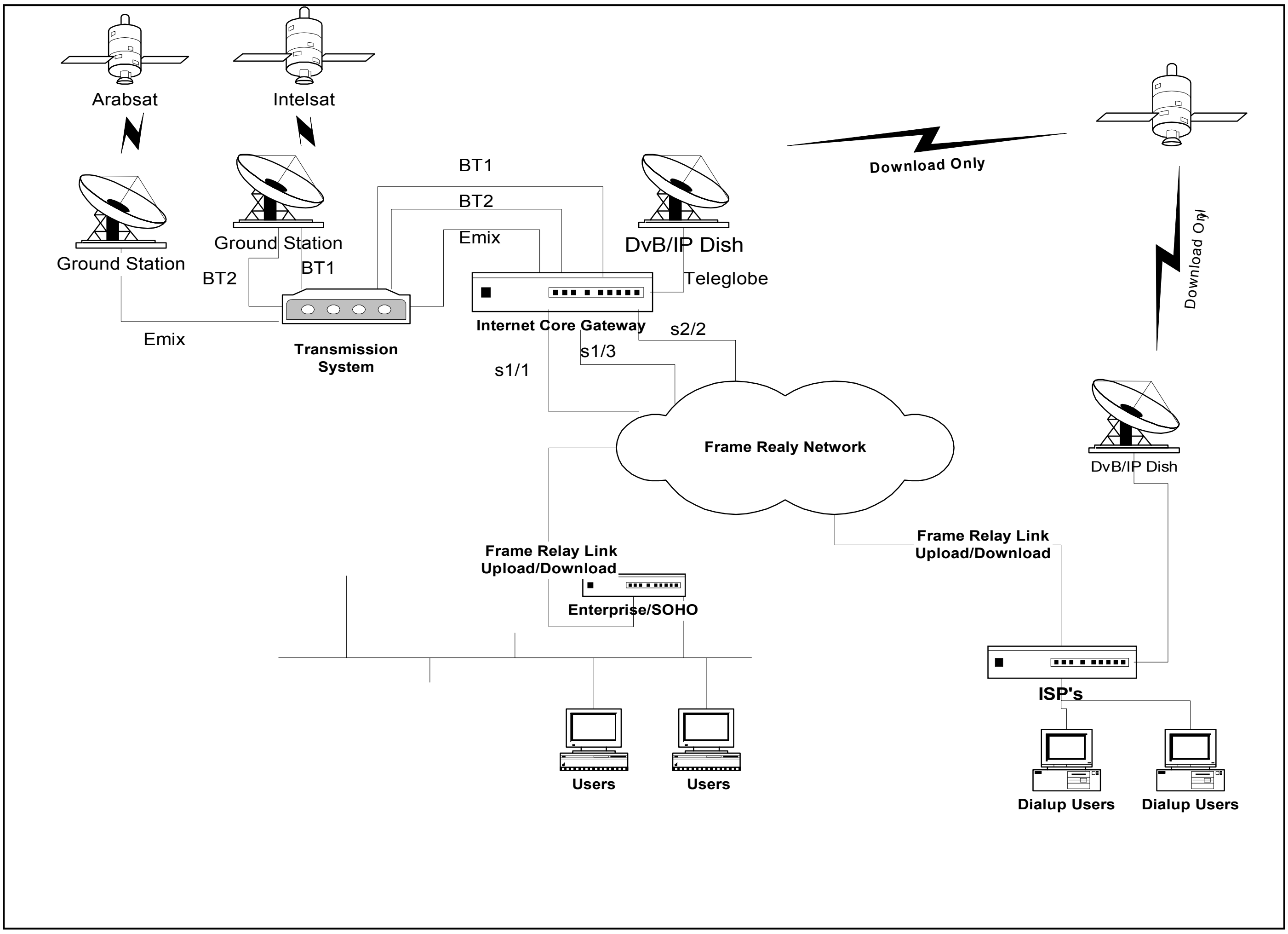


Table 3.2 frame relay aggregate links specifications

Link Name	capacity	Type	Protocol
S1/1	2Mbps	Full Duplex	**Frame Relay**
S1/3	2Mbps	Full Duplex	**Frame Relay**
S2/0	2Mbps	Full Duplex	**Frame Relay**
S2/3	2Mbps	Full Duplex	**Frame Relay**
S2/2	2Mbps	Full Duplex	**ppp**

3.2 Tools Used:

MRTG

MRTG (Multi Router Traffic Grapher), a software used for the collection of the data; it uses SNMP protocol to collect the traffic from the Internet gateway router for each frame relay PVC and international ports, the traffic is collected in a text file format.[7]

SPSS

For the statistical analysis of the Traffic SPSS software is used, it is loaded with information about each customer ports , aggregate ports and international links ports as well as the traffic intensity values gathered from the MRTG . It is then used to find all he statistical properties of the traffic and traffic distributions.[8]

46020 Alcatel frame relay management system:

Is management software used to manage the frame relay switches, and used to collect some information about the queuing delay in the frame relay switches.[9]

3.3 Network Information Gathered:

The type of information gathered from the network is as follows:

For the Internet users :

- Name
- Customer category (educational, ISP…)
- Bandwidth
- Location (Khartoum center, medani, Omdurman…)

- Upload and download bytes/second averaged over 5 minutes ,30 minutes ,2 hours and 1 day for the period from 17-4-2002 to 30-9-2002

For International Links:

- Link type (upload/download)
- Link name
- Protocol used
- International service provider
- Satellite used
- Bandwidth
- Upload and download bytes/second averaged over 5 minutes ,30 minutes ,2 hours and 1 day for the period from 17-4-2002 to 30-9-2002
- Queuing delay and utilization in frame relay switch for faculty of engineering for one day

For the frame relay aggregate ports:

- Link name
- Protocol used
- Bandwidth
- Upload and download bytes/second averaged over 5 minutes ,30 minutes ,2 hours and 1 day for the period from 17-4-2002 to 30-9-2002
- Number of upload and download packets/second for two of the links for 1 day period
- Queuing delay and utilization in frame relay switch for one link for one day period.

Chapter 4

Statistics of the Traffic

SPSS and Excel software programs are used to find the statistics of the traffic information gathered, following is a description of the statistics resulted.

4.1 Statistics of the Local Internet Users

4.1.1 User categories

The number of users is 46; they are divided into 11 classes as shown in the following table.

Table 4.1 frequency of Internet users per category

User Classes	Frequency	Percent of Total users
Governmental Organization	9	19.6
NGO	7	15.2
Telecom	7	15.2
Educational	5	10.9
Oil International Company	5	10.9
Local ISP	4	8.7
Oil Local Company	3	6.5
Airways	2	4.3
VoIP	2	4.3
Banks	1	2.2
Press	1	2.2
Total	46	100.0

4.1.2 Distribution of Internet users per aggregate port

The users are distributed between the aggregate frame relay links using frame relay PVC's as shown in table 4.2 below, each group of users share the bandwidth of the corresponding frame relay port.

Table 4.2 Frequency of Internet users per aggregate port

Serial Port Name	**Customer Class**	**Frequency**	**% of Total Sum**
	Governmental Organization	5	13.8%
	Telecom	3	7.3%
S1/1	Educational	2	1.8%
	Oil Local Company	1	3.7%
	Oil International Company	1	.9%
	Local ISP	1	.9%
	Airways	1	1.8%
	Press	1	.9%
S1/3	Educational	2	6.4%
	Oil International Company	2	2.8%
	NGO	2	4.6%
	Local ISP	2	5.5%
	Governmental Organization	1	.9%
	Telecom	1	1.8%
	NGO	5	18.3%
	Governmental Organization	3	8.3%
	Oil Local Company	2	1.8%
	Oil International Company	2	2.8%
	Telecom	2	2.8%
	VoIP	2	3.7%
S2/0	Banks	1	.9%
	Educational	1	4.6%
	Airways	1	.9%
S2/3	Telecom	1	.9%
S2/2	Local ISP	1	1.8%

4.1.3 Distribution of Internet Users per Port Location

Most of the Internet users are located in Khartoum area; table 4.3 below shows the number of Internet users per location

.

Table 4.3 frequencies of Internet users per location

Port Location	**Number of users**	**% of Total Sum**
Khartoum Centre	22	43.7%
Khartoum South	11	30.0%
Arkawit	5	13.3%
Khartoum North	3	4.3%
Omduramn	2	4.3%
Aljaili	1	1.0%
Medani	1	.7%
El sahafa	1	2.7%
Total	46	100.0%

4.2 Internet bandwidth Distribution between Internal Users

The total Internet Bandwidth is distributed between the users as per their required bandwidth, below is a mentioning of the bandwidth allocated to each user category and each location.

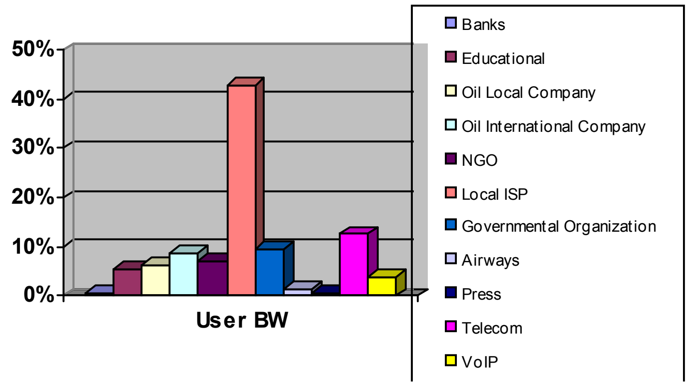
4.2.1 Distribution per user category

The total Internet capacity is distributed between the Internet users and locations, table 4.4 and fig 4.1 below show the Internet capacity distribution per user category as percentage of the total bandwidth.

Table 4.4 Bandwidth Distribution per user category

User Class	User BW
Local ISP	42.9%
Telecom	12.7%
Governmental Organization	9.5%
Oil International Company	8.7%
NGO	7.1%
Oil Local Company	6.3%
Educational	5.6%
VoIP	4.0%
Airways	1.6%
Banks	.8%
Press	.8%
Total	100.0%

Fig 4.1 Bandwidth Distribution per user category

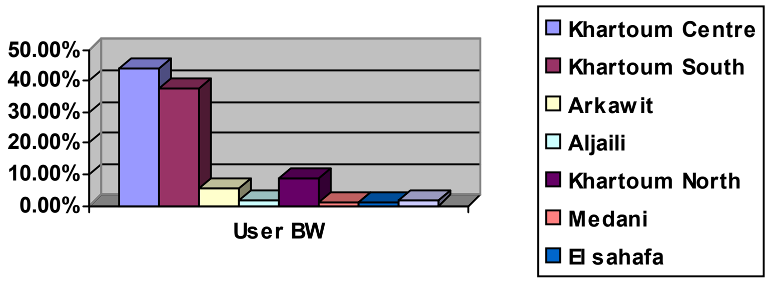


4.2.2 Distribution per location

The Internet service is connected to the customers through the public frame relay network, which expands most of the main cities in Sudan. Table 4.5 and fig 4.2 below show the Internet capacity distribution per customer location as percentage of total bandwidth.

Table 4.5 Bandwidth Distribution per location

Fig 4.2 Bandwidth Distribution per



Location	User BW
Khartoum Centre	43.7%
Khartoum South	37.3%
Khartoum North	8.7%
Arkawit	5.6%
Aljaili	1.6%
Omduramn	1.6%
Medani	.8%
El sahafa	.8%
Total	100.0%

4.3 Traffic Variations for International Internet Links

Following is a description of the traffic variations per day, week and month for the international Internet links.

4.3.1 Daily Traffic Variations

i- Input Traffic variation

As shown on figure 4.3 the input traffic intensity starts to grow from 8 am and reaches it peak between 12 and 2 pm then the curve slope down dramatically from 4 pm to 7 pm and have other peak between 10 pm to 1 am, the other peak at night might be generated by the traffic from internal ISPs because most of the traffic is assigned to the local ISP's with percentage of 42.9 % and the ISPs serve home users who always access the Internet at night. The cause of the increased access at night is the cheapest night telephone and Internet service. Business users generate the evening peak, while the drop between 4 to 7 pm is the period of the end of business day.

ii- Output Traffic variation

The output traffic follows the same behavior as the input.

iii- Peak hour:

Shown in table 4.6 below is the peak hour for each link, also the table shows values of intensity and utilization of the peak hour.

Table 4.6 Peak hour Intensity and Utilization for International Links

ISP	Peak Hour	Intensity (byte)	Utilization
EMIX	3pm	73193	28.8%
BT1	3 pm	125331	51%
BT2	1 pm	122692	49.9%
Teleglobe	2 pm	97507	18.6%
Total	2 pm		36.53%

iv- Comparison between traffic patterns of the International ISP links:

The four links maintain the same daily traffic variations for input and out put traffic, with slight difference in the values of incoming traffic.

From the statistics gathered we observed that the incoming and outgoing traffic are shared between the four links as shown in table 4.7 below.

Table 4.7 Load Distributions between International Links

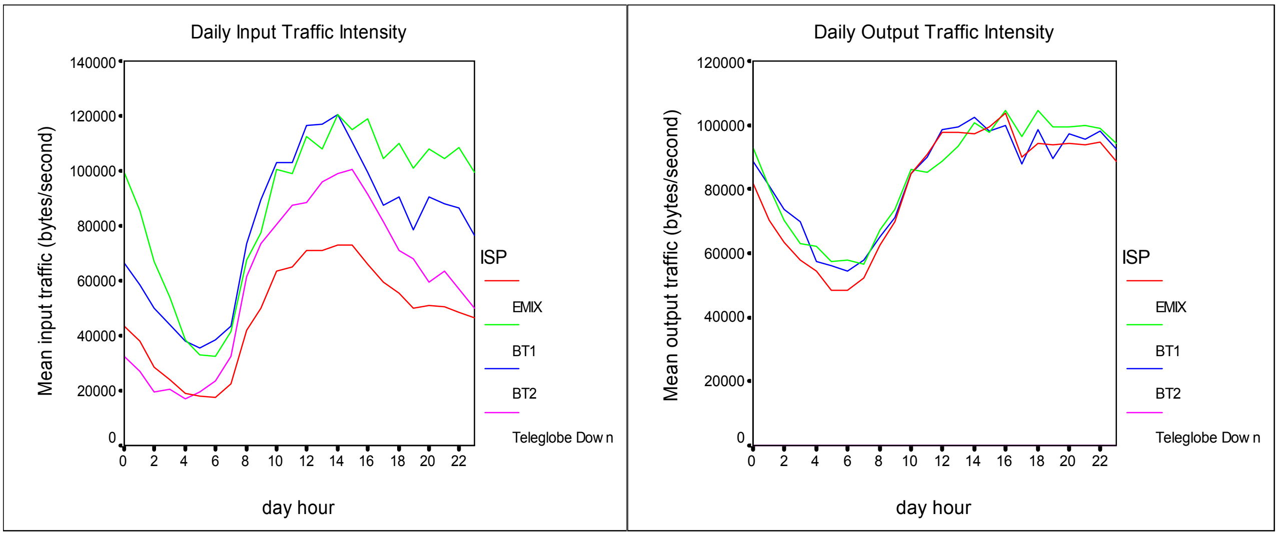
ISP	Input Traffic Load share	Output Traffic Load share
EMIX	18%	33%
BT1	32%	33%
BT2	30%	34%
Teleglobe	20%	0 %
Total	100%	100%

From table 4.7 it is clear that the load balancing technique used to load balance the out put traffic is efficient with exact load sharing, for the incoming traffic the BT links carry the same load which indicate that the metrics of BGP protocol used to control the traffic gave

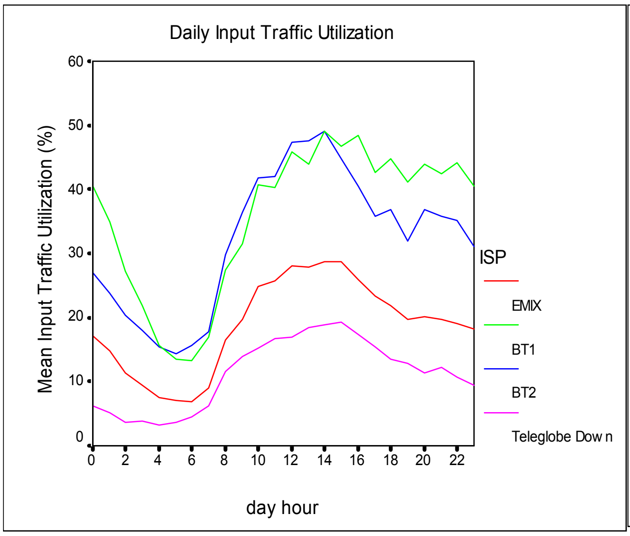
good results. for the case of emix the load is less than BT because no routing policy is used to load share the traffic of this link it just depend on
as- path length metric .

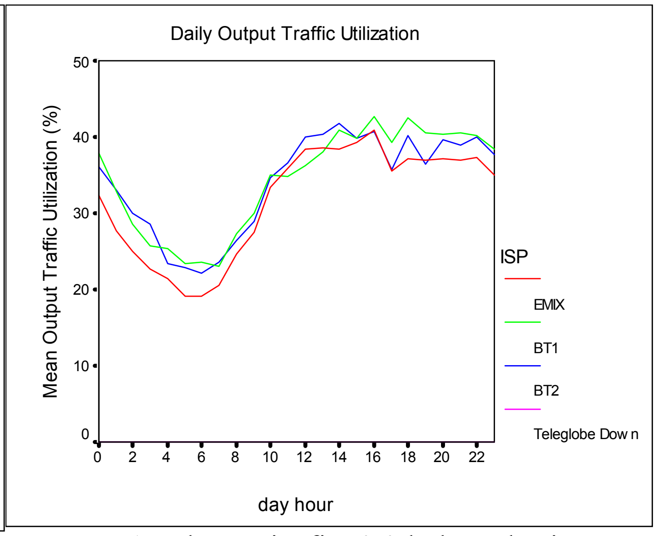
Teleglobe link is static, no dynamic routing protocol is used and the traffic of the customers connected to teleglobe is not shared with the other links, the 20% indicate that they are utilizing less traffic than the other customers.

Fig 4.3 Daily traffic variations for international Internet links, for the period APRIL 2002 to September 2002



4.3.2 Weekly Traffic Variations

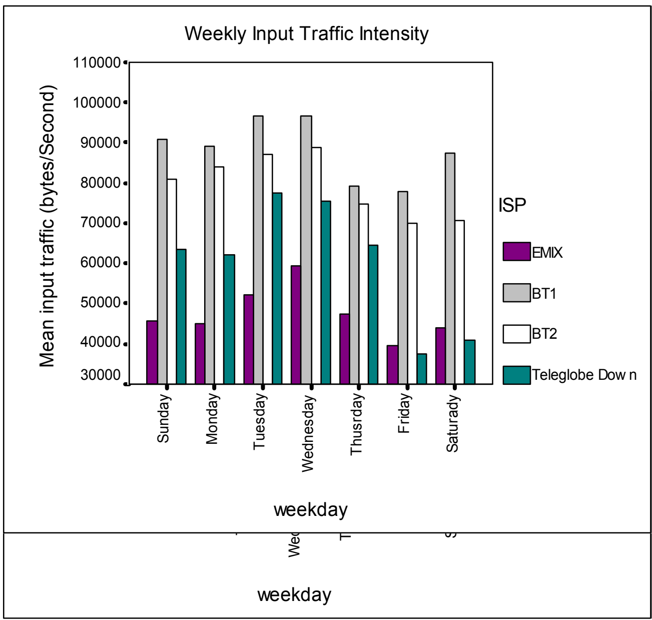


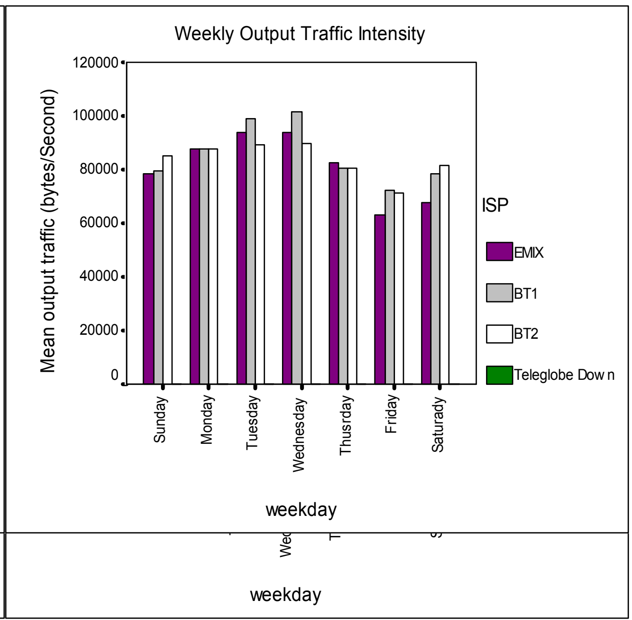


As shown in fig 4.4 below the input and output traffic has its lowest values on Friday and peak values on Wednesday.

The difference between Friday and Wednesday traffic is always less than 10%, because of which we can say that weekly variations on the traffic are of no significance now, because most of the users are home users and their access on Friday is not less than the other days. In the future if business users increase then we expect to have great variation on the traffic.

Fig 4.4 Weekly Traffic Variations for international links (statistics taken for the period 17-4-2002 to 30-9-2002)





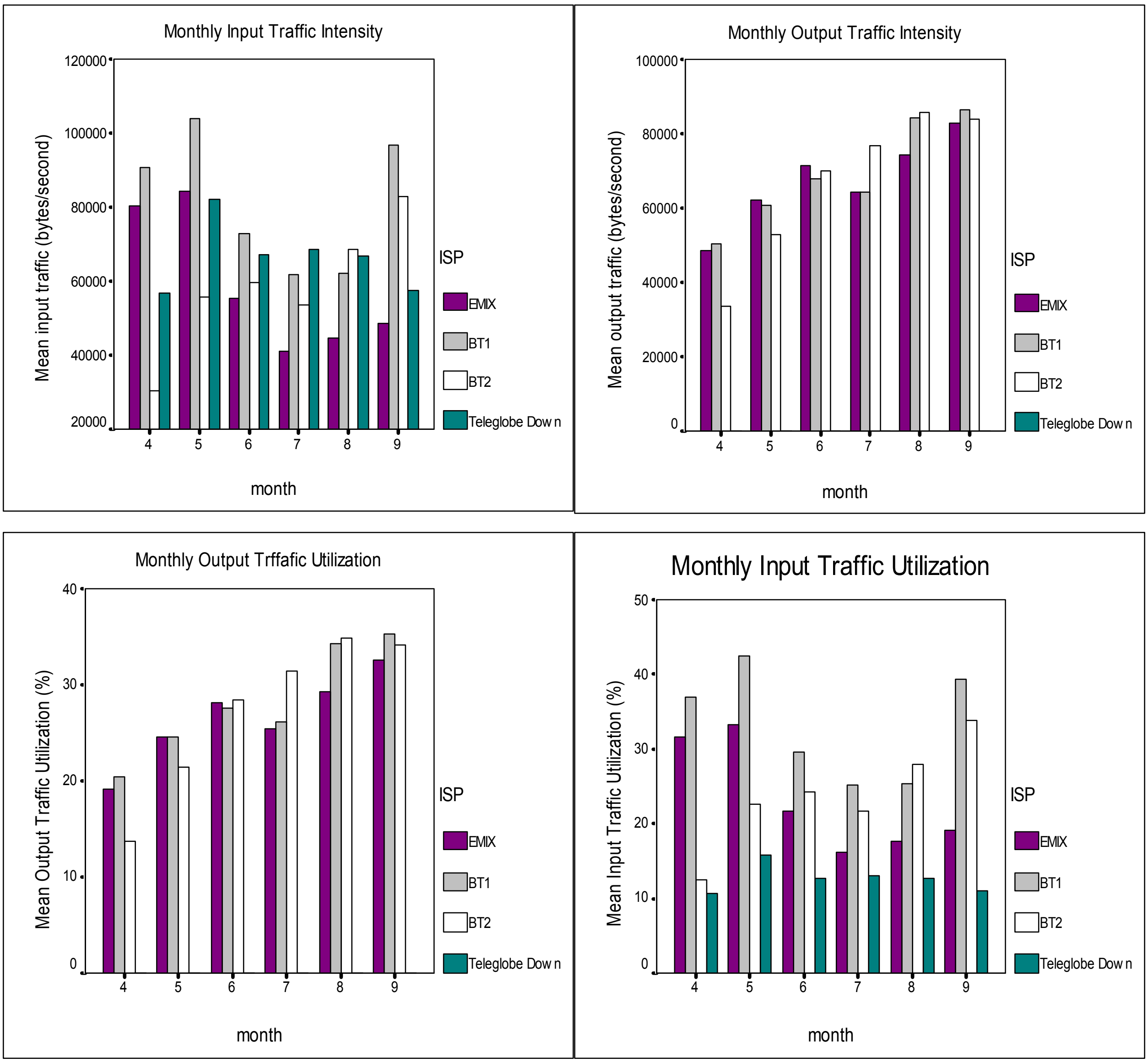
4.3.3 Monthly Traffic Variations

The monthly growth of the Internet traffic is shown on table 4.8 and fig 4.5 . The data on the table shows the average traffic utilization per month for the period of April to September 2002.There was no observable change on the overall incoming traffic while the outgoing traffic is doubled. That is because during this period most of the ISPs began to have their own download links and relay on sudatel backbone for the upload only.

Table 4.8 Monthly traffic Variation for International links

ISP	month	Mean Input Utilization	Mean Output Utilization
EMIX	4	32.24	18.60
	5	34.03	22.12
	6	22.01	27.75
	7	16.41	26.01
	8	17.64	29.24
	9	19.01	32.41
BT1	4	37.27	19.89
	5	41.82	23.93
	6	29.76	27.56
	7	25.28	26.33
	8	25.27	34.33
	9	39.32	35.19
BT2	4	11.12	9.34
	5	22.51	20.15
	6	23.42	28.04
	7	22.81	27.49
	8	27.98	34.83
	9	33.51	34.22
Teleglobe Down	4	10.61	
	5	15.60	
	6	12.41	
	7	14.51	
	8	12.77	
	9	10.93	

Fig 4.5 Monthly Traffic Variations for international Internet links (statistics taken on 17-4-2002 to 30-9-2002)

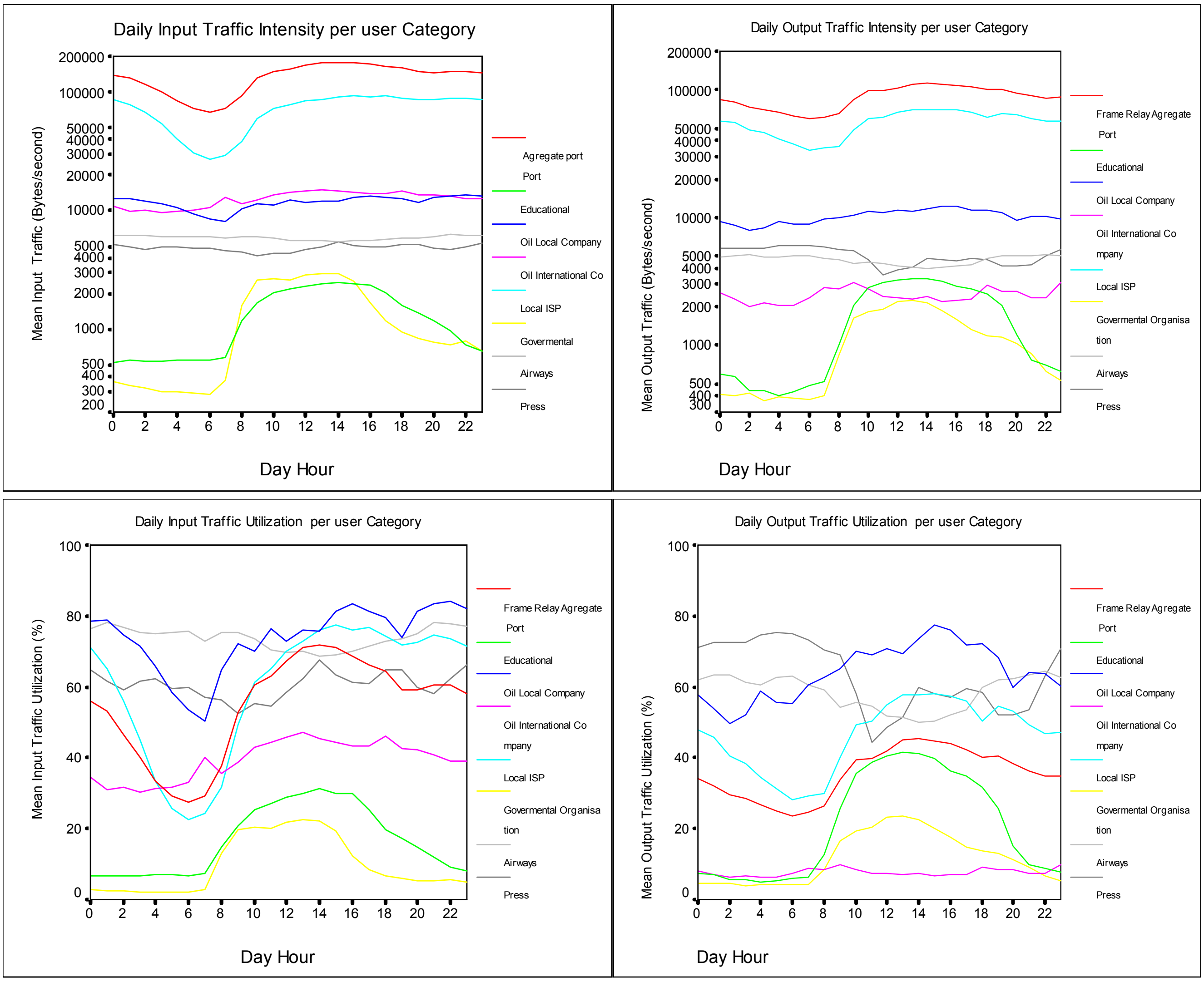


4.4 Traffic Variations for Local Internet Customers Links

4.4.1 Daily Traffic Variations

Shown below is the daily incoming and outgoing Internet traffic intensity and utilization for a sample of each user category. The scale of the y axis is logarithmic

Fig 4.6 Daily traffic variation per user category



The traffic of the users show in fig 4.6 is aggregated by the frame relay aggregate port, from the curve it is clear that ISP’s have a great impact on the aggregate port traffic pattern .all of the users have their peak hour between 1 to 4 pm except for the press and airways categories, have their peak at night (10 pm).

The consumption of educational and governmental organizations is very low if compared to the other categories.

Table 4.9 below shows the peak hour utilization per customer category.

Table 4.9 peak hour utilization per user category

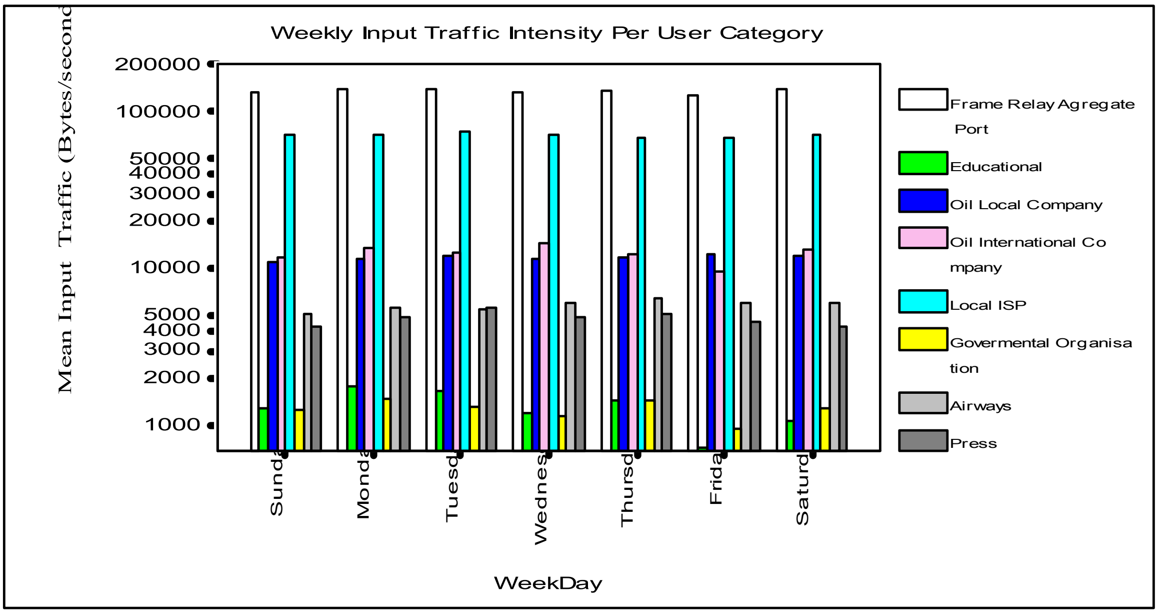
Cust_Class	Peak Hour	Output Traffic Utilization (%)	Input Traffic Utilization (%)
Frame Relay Aggregate Port	14	45.53976692	71.85122935
Educational	15	49.44848602	36.87798913

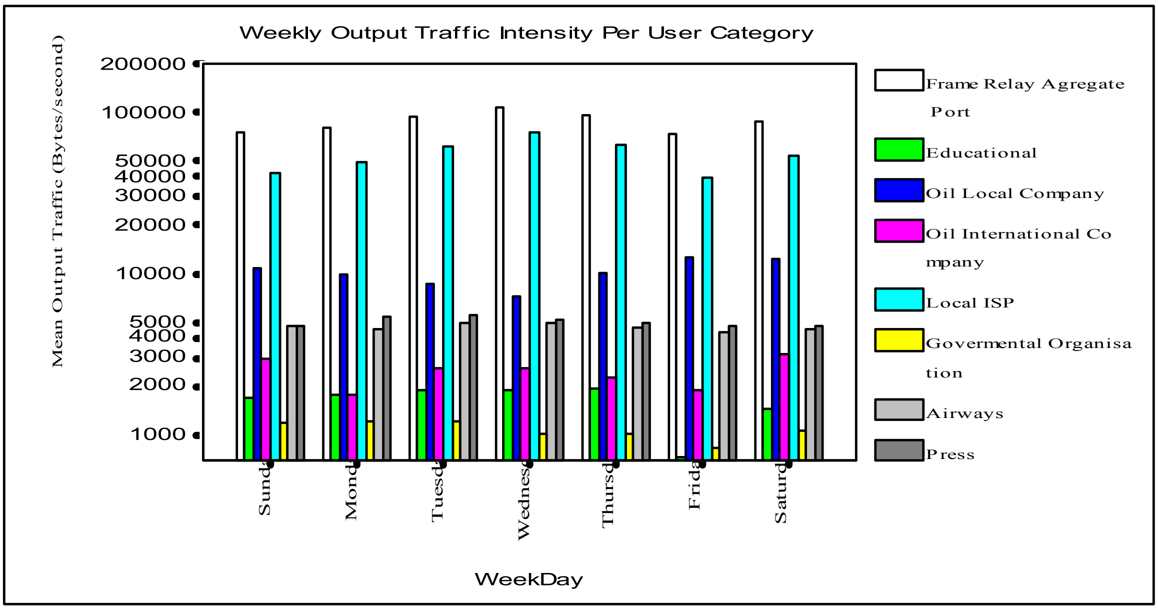
Oil Local Company	16	76.22827744	83.5933689
Oil International Company	13	8.743031832	47.90803571
Local ISP	16	57.55376866	76.11620647
Governmental Organization	13	20.59411627	24.34467737
Airways	22	64.58519737	77.76332237
Press	22	63.20074405	62.17127976

4.4.2 Weekly Traffic Variations

Shown in fig below is weekly traffic variations for some customers categories , it is clear that the traffic pattern remains the same for all of week day except a slight reductions on Fridays . the Y axis in fig 4.7 is in logarithmic scale .

fig 4.7 weekly traffic variations per user category

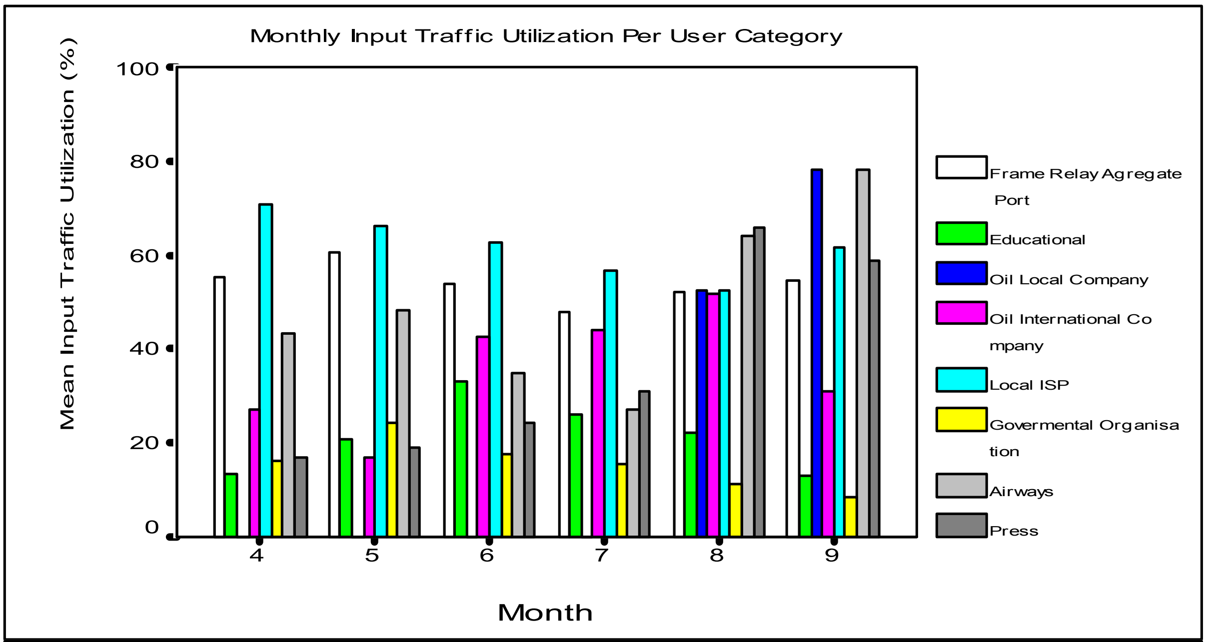


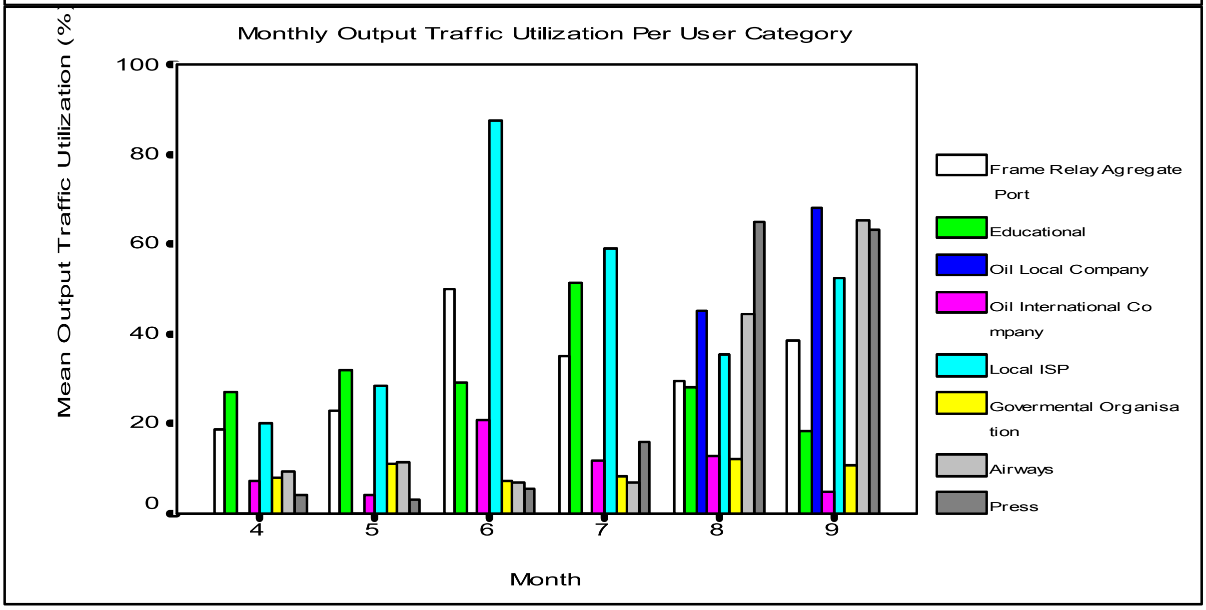


4.4.3 Monthly Traffic Variations

Shown on fig below is the monthly traffic variation for some customer's categories. It is clear that there is no constant pattern for the monthly growth. The y axis of fig 4.8 is in logarithmic scale.

fig 4.8 Monthly traffic variation per user category





4.5 Estimation of the relation between downloads and uploads traffic

The relation between the outgoing and incoming traffic have been calculated for some of the links during the period of August 2002, the resultant curve have been estimated using a curve estimation tool in the SPSS software. Table 4.10 shows the result of each link when compared with linear, quadratic and cubic equations.

The coefficients of the equations are b0, b1, b2, and b3 as on table 4.10.

. Fig 4.9 shows the diagram of the relation between output and input traffic for the links mentioned on table 4.10, the readings for these links taken during august 2002, for each input traffic utilization we found the mean value of the output.

In the case of Estimation using linear equation we observed that the relation between the download and upload for the links which carry an aggregated traffic (EMIX,BT1,BT2,S11) could be represented by a straight line with a slope of 0.5 and have a shift on the y axes , and the output value will have 12% standard deviation .

We can conclude that for the case of aggregate links the upload traffic at any instant of time is half the download plus some constant value and could be represented by the linear equation :

Average Upload = constant +0.5*Download

Maximum Upload = constant +0.5*Download +12%

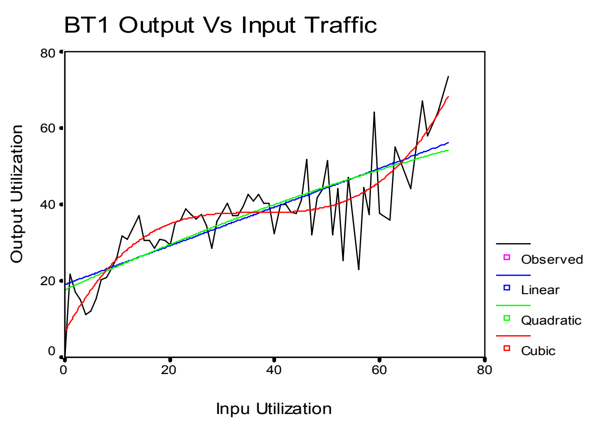
Minimum Upload = constant +0.5*Download - 12%

Table 4.10 relations between upload and download traffic for international links

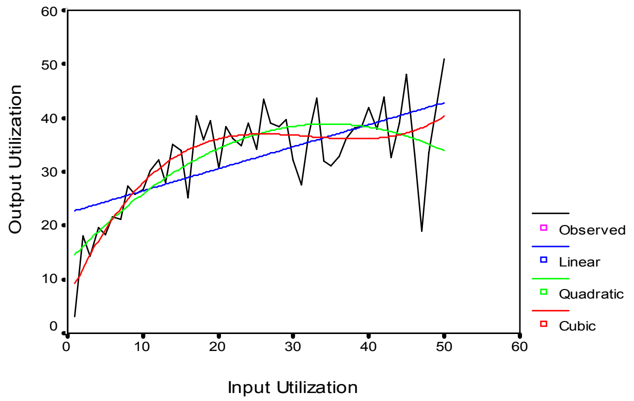
Link Name	Mathematical Function	b0	b1	b2	b3
EMIX	Linear	22.3642	0.4101		
BT1	Linear	19.037	0.5072		
BT2	Linear	13.9939	0.5635		
S11	Linear	5.6212	0.489		
ISP	Linear	10.9086	0.4965		
UofK	Linear	21.1135	0.0905		
Educational	Linear	69.3133	0.3587		
Oil	Linear	6.4944	0.1227		
Airways	Linear	10.0219	0.5259		
Press	Linear	46.1027	0.3479		
EMIX	Quadratic	13.2006	1.4839	-0.0214	
BT1	Quadratic	17.6378	0.6288	-0.0017	

BT2	Quadratic	13.7646	0.5794	-0.0002	
S11	Quadratic	5.695	0.4849	4.20E-05	
ISP	Quadratic	7.685	0.6716	-0.0017	
UofK	Quadratic	27.3901	-0.2957	0.0039	
Educational	Quadratic	64.4044	0.7462	-0.005	
Oil	Quadratic	4.0422	0.2728	-0.0015	
Airways	Quadratic	-0.8022	1.088	-0.0053	
Press	Quadratic	35.9682	0.9465	-0.0059	
EMIX	Cubic	6.2878	3.051	-0.0984	0.001
BT1	Cubic	6.6185	2.5457	-0.0693	0.0006
BT2	Cubic	5.3552	1.6332	-0.029	0.0002
S11	Cubic	-4.2483	1.4487	-0.0233	0.0002
ISP	Cubic	-4.2036	1.8804	-0.0295	0.0002
UofK	Cubic	17.7017	0.9164	-0.0268	0.0002
Educational	Cubic	56.0798	1.9808	-0.0424	0.0003
Oil	Cubic	2.3378	0.4848	-0.0069	3.60E-05
Airways	Cubic	-0.5854	1.0635	-0.0047	-4.00E-06
Press	Cubic	22.259	2.5628	-0.046	0.0003

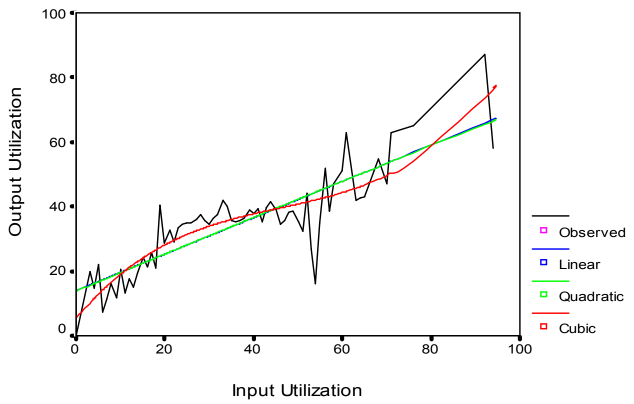
<u>fig 4.9 Upload Vs Download Traffic</u>

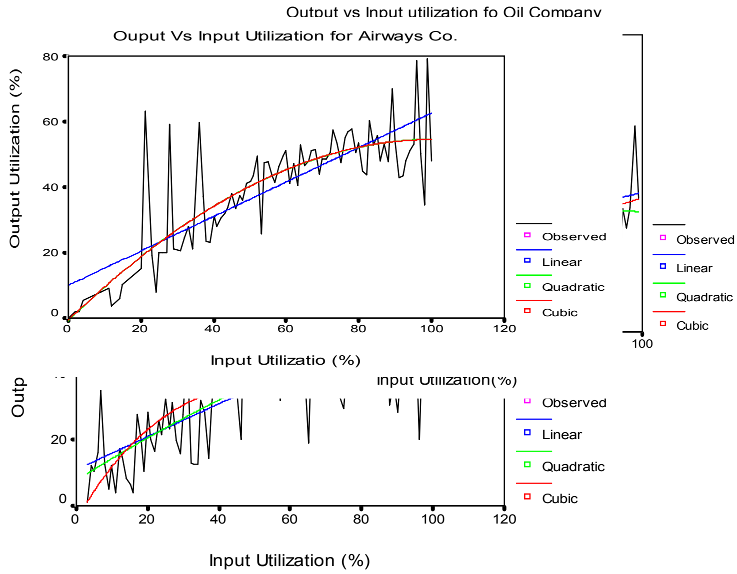


EMIX Output Vs Input Traffic

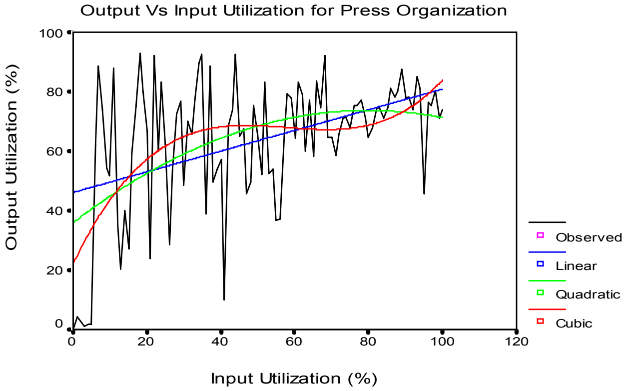


BT2 Output Vs Input Traffic





Output Vs Input Utilization for Press Organization



S11 Output Vs Input Utilization

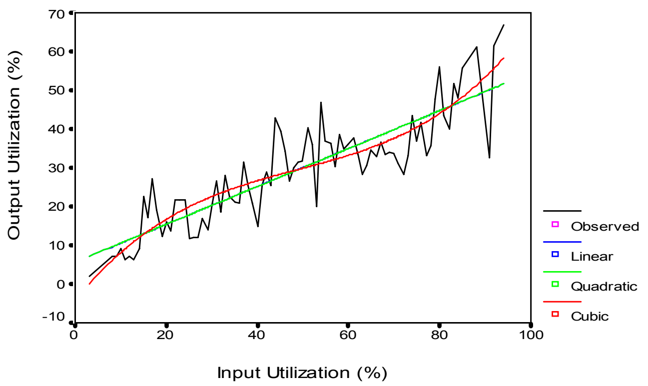
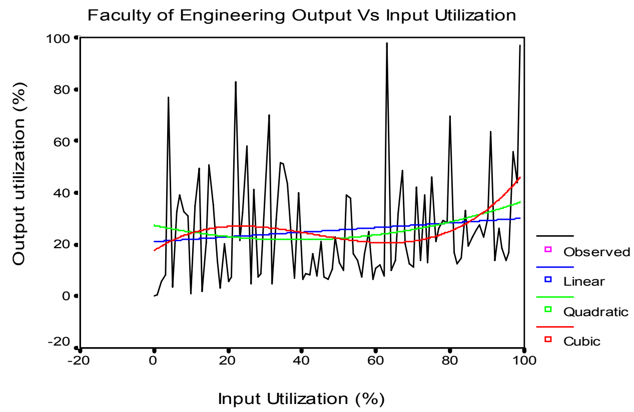
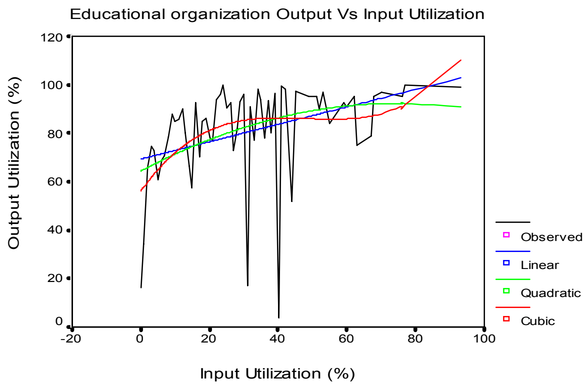


Fig 4.9 Upload Vs Download Traffic





4.6 Arrivals model Estimation

4.6.1 Mean and Standard Deviation of the arrival rate

The mean and standard deviation of the arrivals rate are calculated for EMIX, BT1 and BT2 links , for the period between 15-8-2002 to 30 -9 2002 . Table 4.11 below, shows the mean and standard deviation of the input and output Utilization. N represents the number of readings taken.

Table 4.11 mean and standard deviation fo the Upload and Download Utilization

Link		Input Utilization	Output Utilization
EMIX	Mean	18.64	32.55
	Std. Deviation	11.88	16.64
	N	6987	6987
BT1	Mean	37.88	36.24
	Std. Deviation	18.54	17.59
	N	6371	6371
BT2	Mean	32.55	34.78
	Std. Deviation	16.37	16.80
	N	7004	7004

All of the links have the same mean and standard deviation in the upload traffic.

For the download BT1 and BT2 are having nearly the same mean and standard deviation values, but emix traffic is less than one half BT traffic, this means that the number of web sites that have the shortest path through emix are much less than those through BT.

The ratio of the web sites that have the shortest path through BT to those have the shortest path through emix is 4:1.

The technique used to distribute upload traffic between the three links work fine, and gave 1:1 load balancing.

4.6.2 Comparison with Poisson distribution

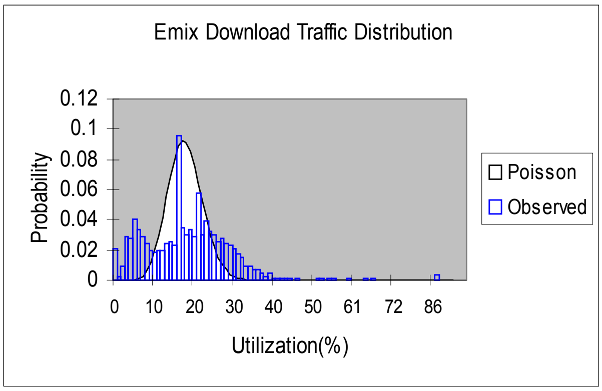
To find the distribution of arrivals, the probability of each traffic utilization value is calculated, the probability is taken from the 5 minute averaged utilization readings for the period of 15-8-2002 to 30-9-2002. Mean, standard deviation and number of readings used are shown in table 4.11 above for each link.

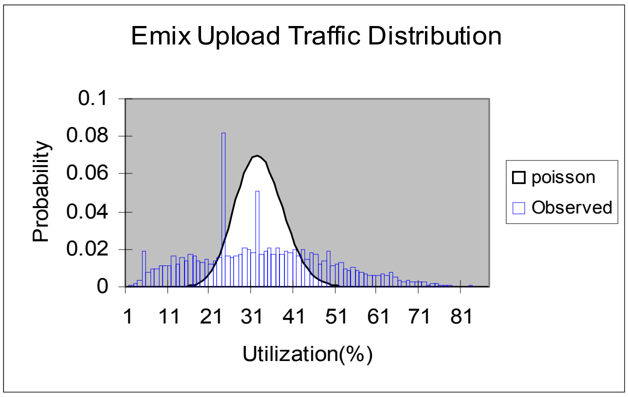
Fig 4.10 below shows a plot of the observed distribution for each link compared with the corresponding Poisson distribution.

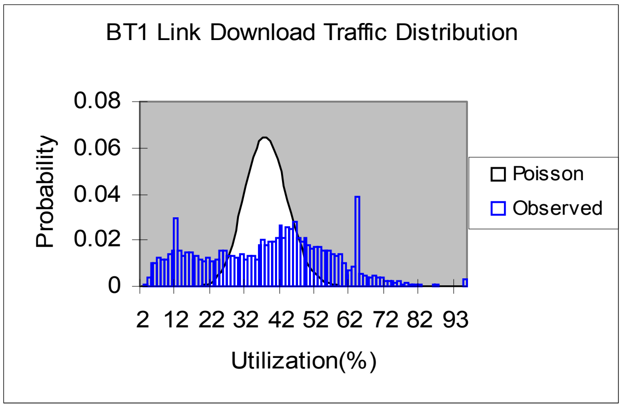
Poisson distribution for each graph is calculated from the utilization values and the mean Utilization values mentioned in table 4.11.

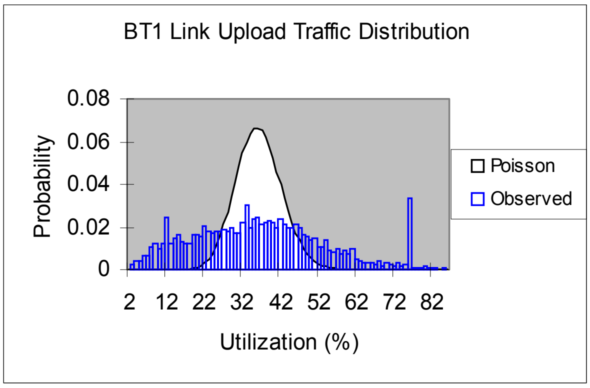
As shown in fig 4.10 the arrival model for all of the links does not follow Poisson arrival model, hence the queuing strategies that apply to traffic of Poisson arrival model may not be suitable with this kind of traffic model.

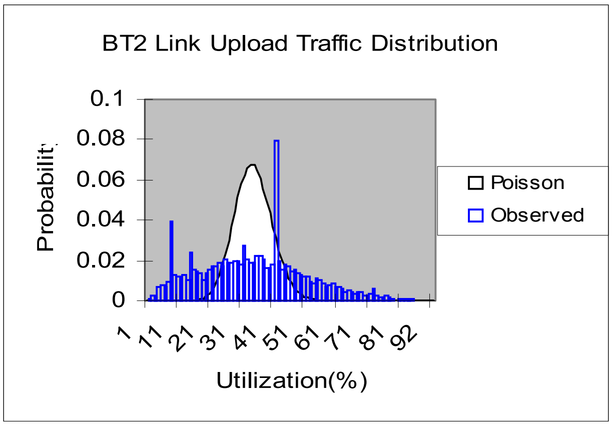
Fig 4.10 Traffic Distribution compared with Poisson distribution

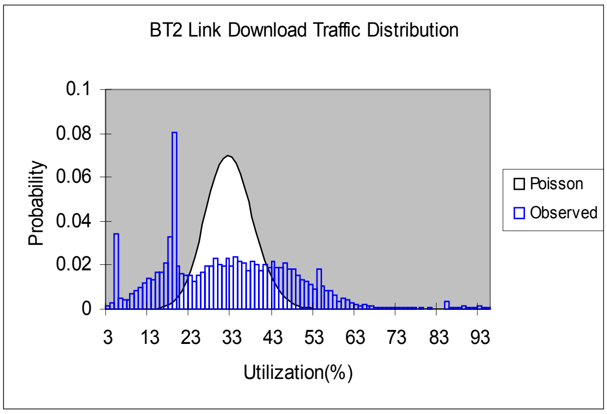












4.7 Self Similarity Check

The data used for self similarity check was calculated from one of the aggregate frame relay links that aggregates user's Internet links. The data was taken for an 8 hours period, with a 12 second time unit .

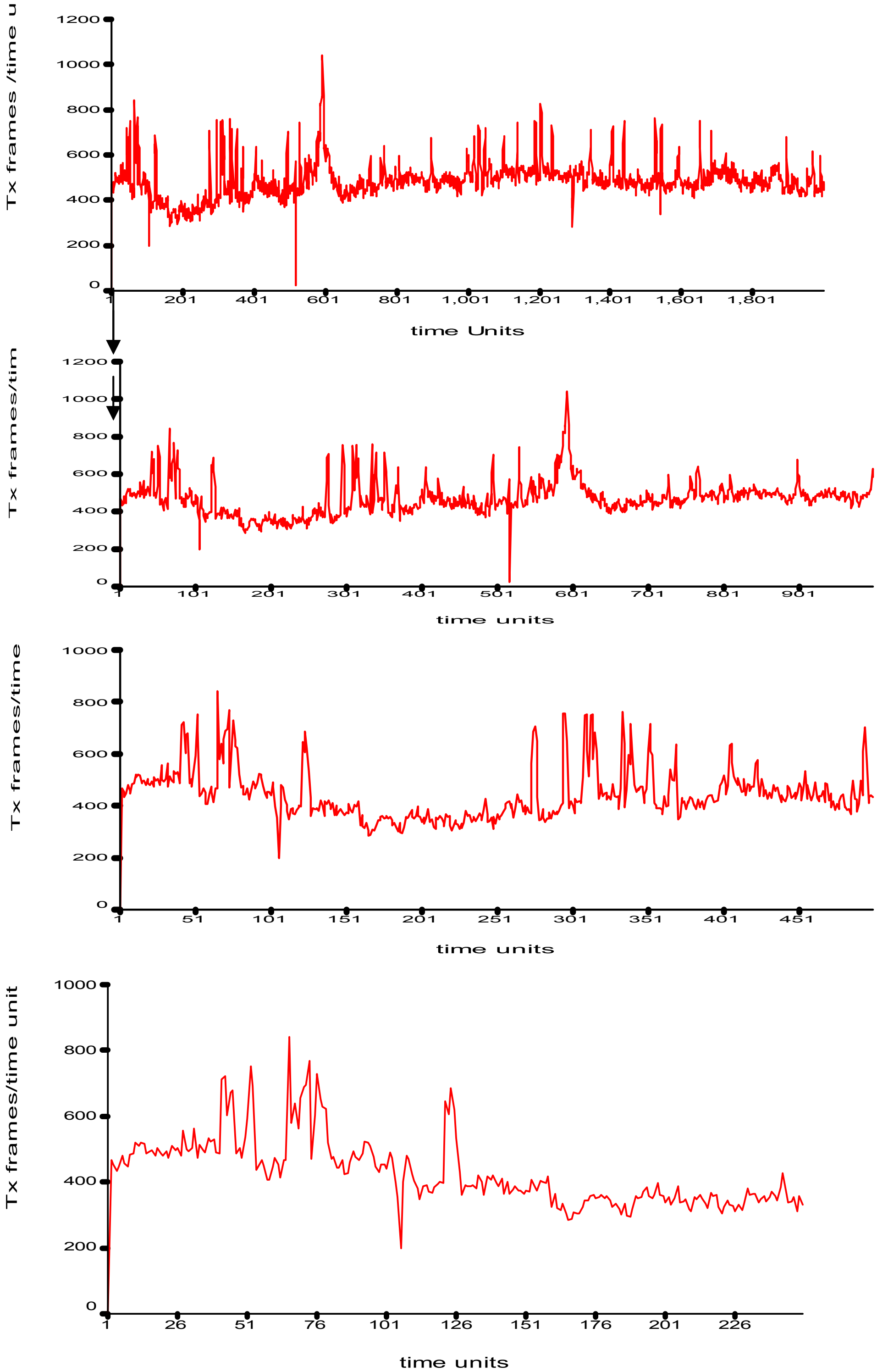
Traffic Behavior on different Time scales:

Fig 4.11 below shows the traffic variations on different time scales. On the X axis each time unit = 12 seconds, the traffic on fig 4.11 maintains the same behavior when viewed at different scales of aggregation.

Fig 4.12 shows plots of the number of packets per unit time for a measurement set from 24-12-2002 , which consists of over 7 hours of continuous monitoring of the Internet traffic , on the plot each subsequent plot is obtained from the previous one by increasing the time resolution by a factor of 2 and displaying a chosen sub interval .the first line covers a period of 7 hours , the second is 3.5 hours , and so on .

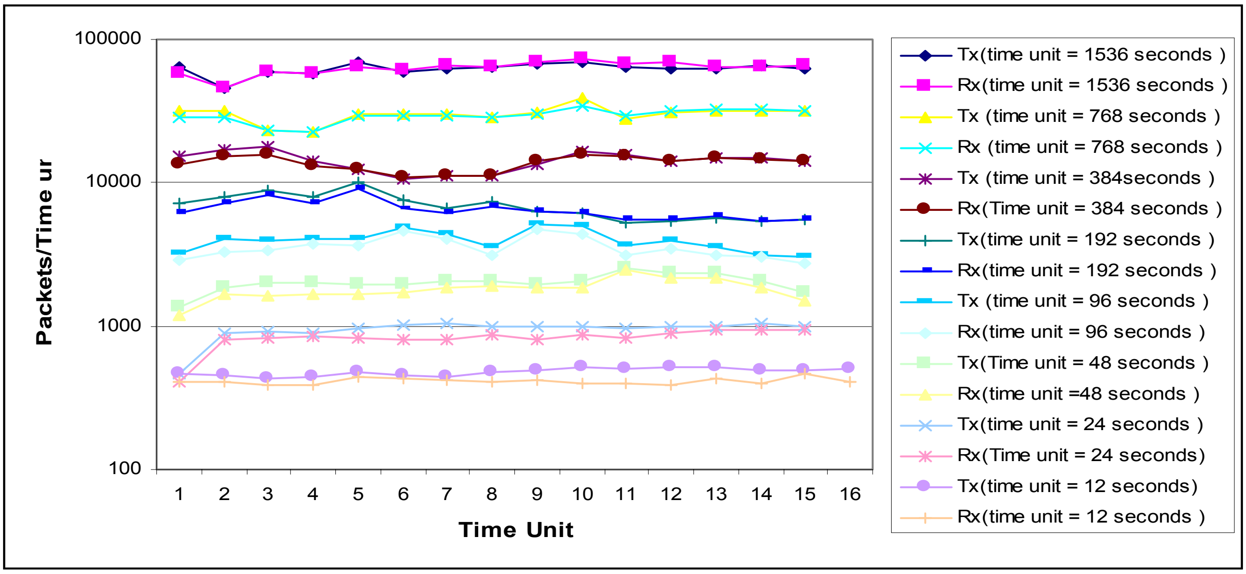
We observed that all of the plots look similar to one another, all of the plots involve a fair amount of burst ness thus Internet traffic tend to look the same at large scales (hours) and at small scales (minutes, seconds) .

fig 4.11traffic in different time scales



Tx frames /time u
time Units
Tx frames/tim
time units
Tx frames/time
time units
Tx frames/time unit
time units

Fig 4.12 shows plots of the number of packets per unit time for a measurement set from 24-12-2002



The variance time plot have been calculated for the same data using the same aggregation levels as on fig 4.12 above , and the resultant variance-time plot is shown in fig 4.13 below .

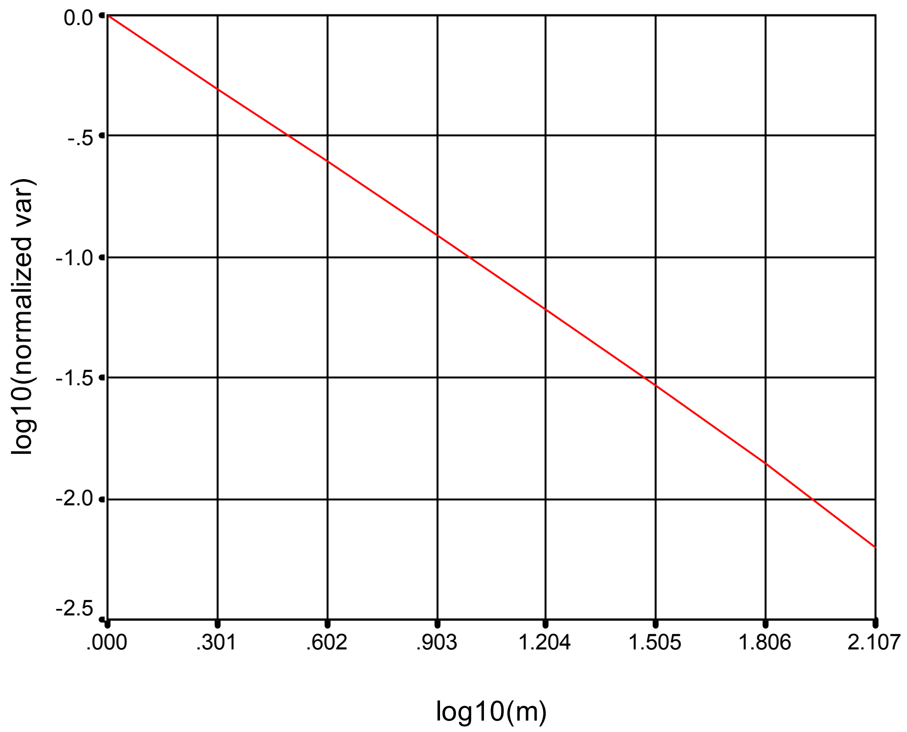
The values of log variance and log m are used to estimate the straight line equation in order to find the slope; we came with the following equation

Log 10(normalized var) = 0.0159 – 1.0379 log10 (m)

From the above equation , the slope of the line that represent log var and log m is -1 , this gives that $\beta=-1$ and H=1 .

This indicates that the traffic is self similar with H=1 .

Fig 4.13 variance-time plot



Log m	Log normalized var
.000	.000
.301	-.303
.602	-.606
.903	-.910
1.204	-1.218
1.505	-1.529
1.806	-1.857
2.107	-2.199

4.8 Relation between Utilization and queuing delay

FIG 4.14 shows delays versus utilization plot for a measurement set from 17-10-2002 which consists of over 8 hours of continuous monitoring of university of Khartoum Internet frame relay link.

The y axis shows the normalized queuing delay values averaged over 14 second polling intervals and divided by the maximum allowable delay value of the switch. , category x axis show percentage utilization of the link. The dark line shows general trend of the obtained plot. We can observe that the delay increase sharply for utilization values above 55%.

Fig 4.14 Delay-Utilization plot

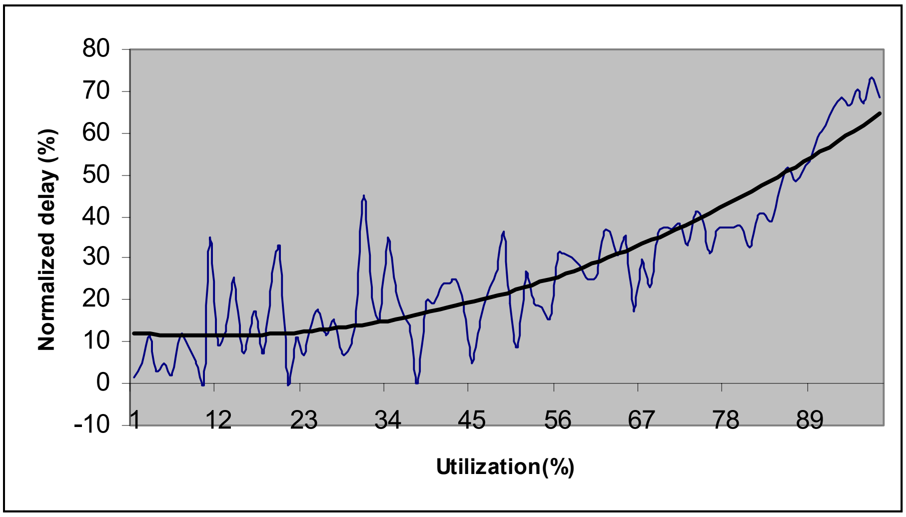
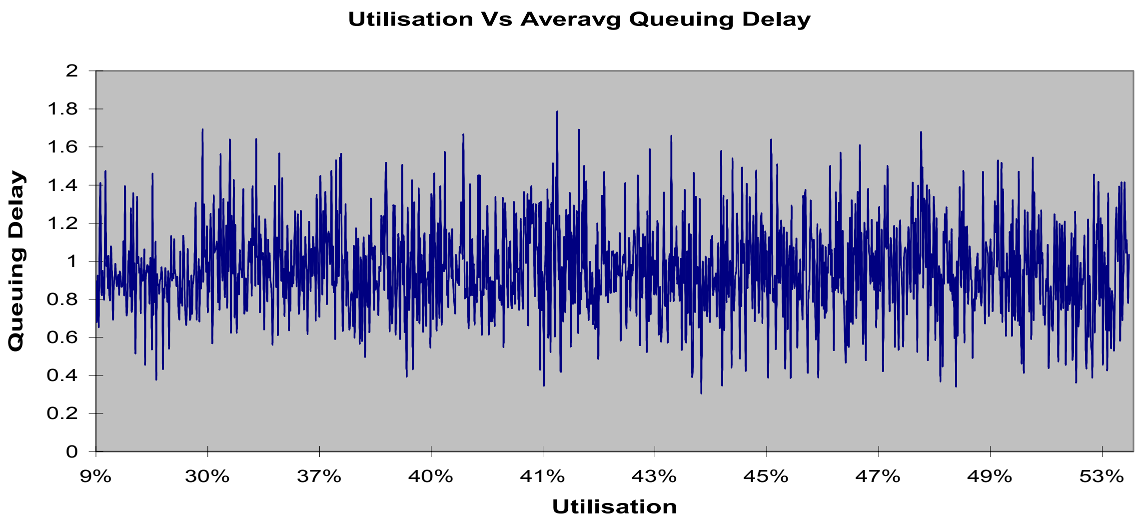


fig 4.15 shows delay versus utilization plot for a low loaded Internet aggregate link and it is clear that the delay maintains the same level for all utilization values of 53% or less

Fig 4.15 delay utilization plot



Chapter 5

Conclusions and comments:

The statistics done in this research reflects the nature of the Internet community in Sudan , in spite of the existence of frame relay data network in most of the cities we found that most of the users are concentrated in Khartoum area . Due to lack of knowledge about the technologies used to provide capacity and to manage their links most of the users buy capacities which they don't need, for example most of them have uplink and downlink with the same capacity while they don't utilize the uplink.

Educational organizations need some financial support to enter the Internet field. Since there is only one bank now uses a dedicated Internet link, and knowing that banks began to put their informatics infrastructure into being, we expect in the near future an increase on the demand from this category. Hence issues like security and VPN application should be taken into account to support the need of this category.

Since there is 43.7 % of the Internet users and 43.7 % of bandwidth in Khartoum center area and 30% of users, 37.3 % of bandwidth located in Khartoum south area, two Internet point of presence (POP's) need to be located in Khartoum south and center .the international links should be distributed between theses areas to reduce the number of transited switches . The remaining users and bandwidth could be shared between the two POP's. Broad band links with redundancy should be installed between the links to carry the traffic between the two POP's and to route the traffic in case of failures.

The four local Internet service providers are carrying 42.9% of the traffic, so we have to think for the introduction of broadband solutions to carry their traffic. Since they serve southlands of home users and small organizations their links should be well protected by the provision of backup links to increase the availability.

The peak hour of utilization is on 2 pm, a new cheapest service during the day need to be introduced to utilize the unused periods, this service could be very cheep and almost free to users from rural areas.

Maintenance and upgrade of the Internet network could be scheduled on the low utilization periods.

The load sharing policy, which is used currently to load share the upload traffic between the international links, is giving very good results and could be implemented in future.

In the case of the download traffic 18% of the traffic is coming through EMIX link, this indicate that the access to Arabic web sites and sites that are from Asia is 18%, so we have to take into account the introduction of 18% international links from ISPs like EMIX or any ISP the connect our network to Arabic world.

Per weekday analysis of the traffic does not give any variation of the traffic, except that on Friday there is slight drop, further long term statistics are needed to prove this issue.

In a monthly basis the upload traffic was doubled in four months, and the download does not increase.

The daily, weekly and monthly traffic patterns showed that the traffic pattern variation depends on the user type and the overall traffic pattern is shaped by the local ISP patterns.

The relation between download and upload traffic utilization could be represented by a straight line with a slope of 0.5 this could be used to predict the amount of upload if we know the download. These results might not be accurate because a long term statistics are needed to prove the relation.

Since the packet arrival model does not follow Poisson distribution and is found to be self-similar, an adjustment on the queue length in the data switches and routers is needed in order to reduce the packet loss and delay. Further studies are needed on this issue because this test has not been done for long period and because the self-similarity of Internet traffic is still a research issue.

The packet delay in frame relay switch queues that carry Internet traffic found to

increase sharply when the utilization exceeds 55% and maintains the same level for 53% or less utilization, this should be taken into account for the links that carry a delay sensitive applications.

This is the first time to do such a research in the Internet backbone of Sudan , hence there is no previous data to compare with . Because the data used in this research are gathered in few months, the results obtained may need to be recalculated using a long-term statistics. a lack of gathering tools and statistical analysis applications limited the work obtained on this research. Despite of all theses problems we could take most of the results into account when provisioning the new Internet services and for the optimization of the current network.

For further researches, huge storage devices, statistical analysis tool, packet sniffers and tool to gather the traffic information from the routers and switches should be provided.

Refferences and Web Pages

References

[1] - Crovella, Bestavros , Self-Similarity in World Wide Web Traffic: Evidence and Possible causes , 1996

[2]Jin Cao, William S. Cleveland, Dong Lin, and Don X., Internet Traffic Tends To Poisson and Independent as the Load Increases, 2001

[3] Overview and Principles of Internet Traffic Engineering, IETF draft, www.ietf.com, March 2002

[4] TELETRAFFIC ENGINEERING HANDBOOK, www.tele.dtu.dk/teletraffic , ITU {D SG 2/16 & ITC}, 2002-10-31

[5] William Stallings, "High-Speed Networks and Internets", First Edition, 1998

[6] A Framework for Internet Traffic Engineering Measurement, IETF draft, www.ietf.com , March 2002

Web Pages:

[7] MRTG (multi router traffic grapher), http://people.ee.ethz.ch/~oetiker/webtools/mrtg/

[8] SPSS, http://www.spss.com

[9] Alcatel 5620 Network Manager, www.alcatel.com