Internet Measurements and Data Study over the Regional Network Ciez@net

Maria-Dolores Cano, Josemaria Malgosa-Sanahuja, Fernando Cerdan, Joan Garcia-Haro
Polytechnic University of Cartagena
Department of Information Technologies and Communications
Campus Muralla del Mar s/n (Ed. Hospital de Marina)
30202 Cartagena, Spain
Ph. +34 968 325953
Fax. +34 968 325338
{mdolores.cano, josem.malgosa, fernando.cerdan, joang.haro}@upct.es

Abstract

In this paper we present the most significant results studying the Internet network traffic measurements obtained in Ciez@net. Ciez@net is a citizen subnet located in the village of Cieza that belongs to the regional network of the Autonomous Community of Murcia in Spain. This subnet is one of the firsts pilot experiences of a Digital City in Europe and the first in the Region of Murcia. The goal is the seamless introduction of the Information Society in a medium-size population. Access to advanced electronic information services is stimulated or subsidized for an effective penetration. These measurements will allow a qualitative and quantitative knowledge of the network traffic in order to achieve a most effective network resource provisioning and Internet traffic forecasting in a real scenario. A suitable dimensioning of the network as well as an adequate provisioning of Quality of Service to users may depend partially on these results. Measurements were taken from a Frame Relay link connecting Ciez@net's users to Internet through the main node located in Murcia city. We used a promiscuous network analyzer that avoids interfering in the network traffic. We report results of traffic load, network performance, percentage composition of traffic by protocol and type of application, and IP packet size distribution in both, up and down communications streams.

1. Introduction

It is well known that Internet traffic is hard to model since the growing number of users, volume, new applications, and topology is causing a tremendous change in its traffic nature [1].

In this context, the necessity of traffic studies about Internet flows and their trends over real scenarios are becoming an important issue. The relevance of finding a real characterization of Internet traffic has produced many research papers in recent years [2][3]. Furthermore, software and hardware tools for analysis and monitoring of networks have flooded the market [4] [5].

In this context, we performed an analysis and monitoring over a real subnet called Ciez@net, located in the village of Cieza, Spain. The project Ciez@net [6] is the first pilot experience of a Digital City in the Region of Murcia. The idea behind the project is to ensure the fast introduction of the Information Society in a population of medium size where several types of inter-relationships are given amongst citizens, administrations and business world. This approach may allow to easily extend the benefits of this trial experience to the whole Region.

The project assumes the fact that the Information Society will only be fully reached when each citizen, institution or company can have access to advanced electronic information services, training to use them and offering prices according to the users willingness to pay. The last aspect may incentive the stimulation of these services, however a starting point is to reach as many people as possible.

Ciez@net provides basic Internet access to users using N-ISDN technology up to the service provider (ISP) and finally through a Frame Relay link as a bearer service. Frame Relay is a packet switching technology that guarantees some resource requirements.

The availability of specific results in real environments as Ciez@net can be used for the better understanding of the real Internet traffic nature, and it is an important help in its characterization since an appropriate Quality of Service and resource provisioning may depend on it in the near future.

We used a promiscuous network analyzer to capture the data traffic. As an alternative, a software tool based on SNMP (Simple Network Management Protocol) could be used. However, it was not employed to avoid interfering the normal working of the net, and to gain more flexibility and independence to capture all major protocols.

In this work, we present the results of Ciez@net Internet traffic such as traffic load, more visited web sites, number of connected users, IP packet

0-7803-7080-5/01/$10.00 ©2001 IEEE

Authorized licensed use limited to: Universidad de Cartagena. Downloaded on October 20, 2008 at 07:17 from IEEE Xplore. Restrictions apply.

length distribution, percentage composition of traffic by protocol and application, and packet length distribution by services. These measurements together with the elapsed time distribution between consecutive packets can be used to consider a more adequate dimensioning and planning of the network resources than the current overdimensioning procedure. We report the results over two time ranges (working and non-working days), in both, up and down streams. The remainder of this paper is organized as follows. Section 2 briefly describes Ciez@net infrastructure and environment. Section 3 and 4 presents the obtained results. Finally, the last section is dedicated to some concluding remarks.

2. Ciez@net infrastructure

In figure 1 we can see the main communications equipments that provide Internet access to both Ciez@net users and other users. Eight ISDN lines allow the first access from home to the ISP (Internet Service Provider) located at Murcia node. Traffic from Ciez@net and other networks join in Murcia node where a Cisco router 2610 provides the Internet routing for both traffics by different paths. Ciez@net traffic get Internet access through a 512 Kbps Frame Relay Link, where measures were performed, while traffic from other networks gain Internet access by a 2 Mbps ATM link.

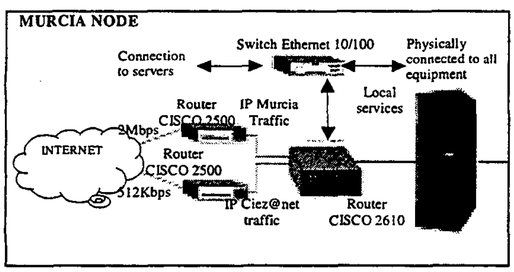
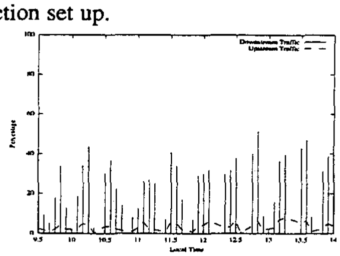


Fig.1. Communications Infrastructure.

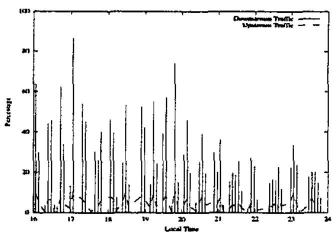
The network analyzer was connected with a sniffer that allows capturing traffic transparently between the router and the Internet access. That is, without overloading the network or the equipment. This methodology provided us the Ciez@net knowledge in terms of network functionalities, quality of service and statistical analysis that will be studied in the next sections.

3. Traffic load and network performance measurements

As it is shown in figure 2, the upstream link (traffic from Ciez@net to Internet) utilization is notably below the maximum capacity of 512 Kbps. Traffic load in a non-working day is displayed in figure 3. The number of connected users has been measured at time intervals of 5 minutes. Figure 4 plots an example for a non-working day. It should be noted that an adequate selection of the measurement interval is critical for this performance figure. On the other hand, table 1 and table 2 show the top 10 visited addresses where *www.terra.es* is the most visited, likely because it is the default web address at the user connection set up.



a) Morning



a) Afternoon

Fig.2. Traffic load in a working day.

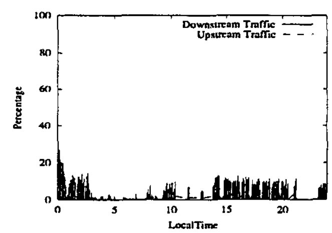


Fig.3. Traffic load in a non-working day.

Destination Address	Captured packets (%)
www.terra.es	34,9
irc.terra.es	14,01
uk1.doubleclick.net	10,63
195.57.98.2	8,51
perso.wanadoo.es	6,37
invertia.com	5,98
216.167.67.172	5,12
206.132.163.167	5,09
charts-d.quote.com	4,82
deriva.f-integra.org	4,48

Table 1. Top 10 web addresses in working days.

Destination Address	Captured packets (%)
www.terra.es	43,00
irc.terra.es	9,41
inforchat.com	9,40
Gd9.doubleclick.net	7,55
62.81.31.8	6,72
Md1.doubleclick.net	5,59
Hg1.hitbox.com	4,97
www.telepolis.com	4,79
A46.g.a.yimg.com	4,51
Artemis.ttd.net	4,00

Table 2. Top 10 web addresses in non-working days.

Authorized licensed use limited to: Universidad de Cartagena. Downloaded on October 20, 2008 at 07:17 from IEEE Xplore. Restrictions apply.

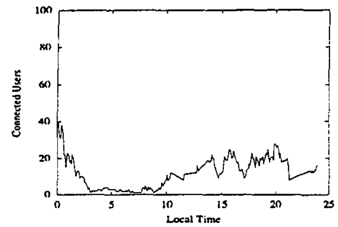


Fig.4. Number of connected users in a non-working day.

4. Analysis of IP packet size distribution

Basically, there are three predominant groups of packet sizes, short packets (from 40 to 150 bytes), medium packets (from 500 to 600 bytes), and relatively large packets (from 1400 to 1550 bytes). Incoming traffic consists of large packets due to the massive transfer of data in the downstream. On the other hand, the upstream traffic consists mainly of short packets (user requests, mouse clicks, etc.). See figures 5 to 8.

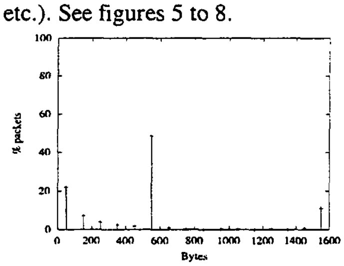


Fig. 5. IP packet size distribution downstream in working days

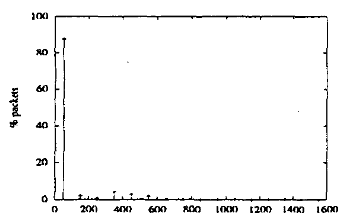


Fig. 6. IP packet size distribution upstream in working days

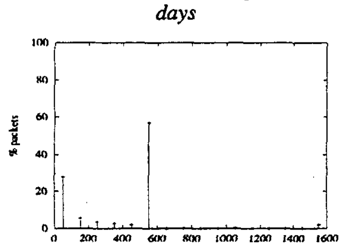


Fig. 7. IP packet size distribution downstream in non-working days

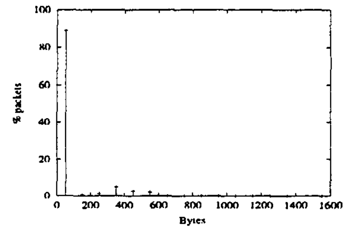


Fig. 8. IP packet size distribution upstream in non-working days

Regarding the most demanding services in Internet, in table 3 we can apreciate that the World Wide Web is currently the dominant application in the downstream with nearly a 65% of packets. Also notice the high percentage of TCP control mechanism traffic that comprises up to 85% of the entire upstream traffic. Other significant services are IRC, FTP, POP3, or RTP. The *Unknown* service represents up to 16% of the captured packets in the downstream. This traffic consists of services whose TCP port identifiers have not been standardized, including applications such as games on line, RealAudio, etc. Furthermore in figures 9 to 12 we display the packet size distribution for the most popular Internet applications where we see again the asymmetric distribution of packets between the up and down streams.

We will like to remark that transferring file service such as ftp, nntp and http have similar characteristics since they are used to transfer large amounts of data traffic employing the same transport protocol. However, medium size packets utilized differ among them. The reason could be found in the particular configurations used by the network servers.

Incoming traffic		Outgoing traffic	
63.5126%	HTTP	80.5280%	Control_TCP
16.5402%	Unknown	10.0966%	HTTP
14.9389%	Control_TCP	07.6905%	Unknown
02.5620%	IRC	00.6673%	IRC
00.9995%	FTP	00.2569%	RTP
00.2532%	RTP	00.2509%	Control_FR
00.2220%	Control_FR	00.1773%	DNS
00.2069%	POP3	00.1339%	NetBIOs-NS
00.1775%	HOST2-NS	00.0681%	POP3
00.1350%	DNS	00.0318%	HOST2-NS
00.1349%	RSRB	00.0269%	AFS3-FILESERVER
00.0896%	RPC	00.0173%	FTP
00.0888%	AFS3_FILESERVER	00.0128%	RPC
00.0450%	GOPHER	00.0118%	CLEARCASE
00.0419%	NetBIOS-NS	00.0112%	RTCP
00.0164%	H.225.0	00.0036%	TPKT
00.0094%	CLEARCASE	00.0036%	LDAP
00.0087%	RTCP	00.0030%	AUTH
00.0049%	NNTP	00.0021%	H.245
00.0034%	AUTH	00.0017%	NNTP
00.0027%	LDAP	00.0016%	GOPHER
00.0024%	TPKT	00.0010%	Telnet
00.0021%	SMTP	00.0010%	SMTP
00.0012%	H.245	00.0005%	H.225.0
00.0005%	Telnet	00.0004%	RSRB
00.0004%	COAUTHOR		

Table 3. % Packet distribution by service in a working day.

Authorized licensed use limited to: Universidad de Cartagena. Downloaded on October 20, 2008 at 07:17 from IEEE Xplore. Restrictions apply.

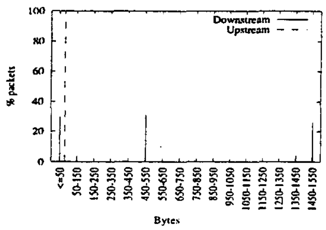


Fig.9: POP3 Service packet size distribution

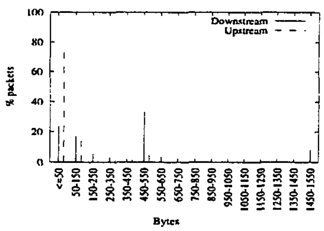


Fig. 10: Unknown Service packet size distribution.

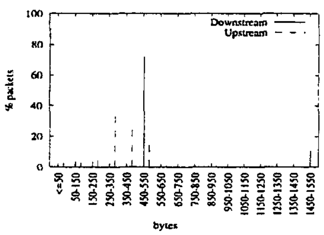


Fig. 11: HTTP.

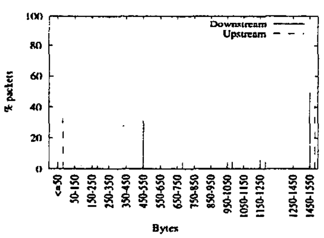


Fig. 12: FTP.

5. Conclusions

In this paper, we measure and initially study the traffic generated in the Ciez@net subnet, which may contribute to a better understanding of Internet traffic in real scenarios, specially those composed by residential citizens.

The maximum load in the downstream arises up to 90% while the upstream one hardly exceeds 10%. The upstream link is clearly underutilized leading to a low bandwidth utilization.

The high percentage of TCP control packets is mainly due to the www applications since them require multiple TCP connections per visited page. The use of an alternative underlaying technology such as ATM does not solve the problem [5]. Re-designing the http protocol or replacing it for a more efficient one might be a possible solution, but with a compatibility drawback with the current applications.

The traffic measurements show the typical scattering in the Internet characterization in terms of time pattern, usage of applications, and IP packet size. The heterogeneity of Ciez@net's users could be the main reason for this variability. This variability lead us to the fact that these measurements can not be considered as conclusive but as only a picture of the period of time when traffic data was captured.

From a theoretical point of view, the analysis of the packet size distribution and the elapsed time between consecutive packets can be used for the specification of traffic models, as well as to aid in the design of control mechanisms for a suitable dimensioning and planning of network resources.

Finally, the availability of these measurements helped us to gain a qualitative and quantitative insight about Ciez@net in order to perform an adequate resource provisioning aimed to provide the expected Quality of Service to the final user.

Acknowledgements

This work was partly supported by the Spanish Research Council under grant TIC2000-1734-C03-03 and by the Fundación Integra under project 0124.

References

[1] S. McCreary. "Trends in Wide Area IP Traffic Patterns". University of California, San Diego. CAIDA. USA. Monterrey, May 2000.

[2] Kc Claffy. "Internet Measurements and data analysis: topology, workload, performance and routing statistics". University of California, San Diego. Cooperative Association for Internet Data Analysis (CAIDA). USA. NAE '99 workshop.

[3] kc Claffy, G. Miller, K. Thompson. "The nature of the beast: recent traffic measurements from an Internet backbone". Proceedings of INET '98.

[4] Brownlee, N., Mills, C. y G. Ruth, "Traffic Flow Measurement: Architecture". RFC 2722. October 1999.

[5] "Ethereal Project". February 13th, 2001. <www.ethereal.com>.

[6] *The Ciez@net project*. February 20th, 2001. <www.f-integra.org/projects.htm#ciezanet>.

[7] M. Alvarez-Campana, A.Azcorra, J. Berrocal, J.R. Pérez, E. Vázquez. "Internet traffic measurements over the Spanish R&D IP/ATM Network Backbone". Proceedings of IFIP ATM '99.

Authorized licensed use limited to: Universidad de Cartagena. Downloaded on October 20, 2008 at 07:17 from IEEE Xplore. Restrictions apply.