DiPLO
www.diplomacy.edu

Bandwidth management: the public policy approach in a university campus network

Maduka Kingsley Attamah, Nigeria

Abstract

This research presents an example of how evidence-based policy-making can lead to bandwidth optimisation in a university network, giving rise to improved network performance and cost savings. The research data were based on network event logs from a campus network. The logs were analysed to draw useful insights about the needs and behavioural patterns of users on the network. The results show how this approach can lead to effective policies for bandwidth management in a university network.

Keywords: log mining; network optimisation; bandwidth management; university campus network; data.edu; ICT policy; Nigeria

Introduction

Bandwidth management aims at forestalling network congestion and improving service quality. Network congestion could arise due to the limitations of the network infrastructure and as a result of the activity of users on the network. In some cases, time-sensitive traffic such as those generated by video conferencing and Internet telephony tend to congest bandwidth (Welzl, 2005; Janevski, 2003). In some other cases the traffic is undesirable. These include: traffic generated by malware, as well as heavy peer-to-peer (p2p) traffic at peak times (Kondakci, 2003; Gummadi *et al.* 2003; Sandvine, 2002). In order to optimise available bandwidth and to improve network performance, network administrators often resort to prioritisation, throttling, and sometimes blocking of select network traffic (Mitra and Ramakrishnan, 1999).

Bandwidth availability can be managed by the use of artificial intelligence techniques, or by making public policies that seek to influence the use of the network.

Some techniques based on artificial intelligence include: traffic shaping (Georgiadis *et al.,* 1996), scheduling algorithms (Brzezinski, 2007), congestion avoidance schemes (Jacobson, 1988; Bauer, 2008; Allman *et al.* 2009) and bandwidth reservation algorithms and protocols (Kim and Varshney, 2005). Some public policy techniques include: bandwidth prioritisation schemes which are often trailed by heavy debates relating to net neutrality and privacy issues (Kosiur, 2001; Raul, 2002; Lenard and May, 2006; Nunziato, 2009; Cohen, 2010), promotion of local content, peering and interconnection caching (AAISOA and CIPACO, 2005), use of economic measures (e.g. traffic is metered and users pay for what they use) (Dermler, 2000; Bauer, 2008).

But whichever approach is adopted towards providing effective results, there is a need to clearly understand the characteristics of user activity on the network in question. Indeed, with a proper analysis of user activity, bandwidth management need not always be overly technical in order to achieve improved network performance.

This research paper presents an example of how the analysis of network activities in a university environment can lead to policies that improve network performance and administration.

This document has been produced by DiploFoundation with the financial assistance of the European Union. The contents of this document are the sole responsibility of the author and can under no circumstances be regarded as reflecting the position of the European Union or DiploFoundation.

Method of data collection

Network monitoring was done by logging a selection of events on the network over a period of five weeks. These logs were collected using Syslog (2011) across select nodes of the University of Ibadan (UI) campus network. Research nodes were selected based on the stability of electric power supply to the nodes and the capability of the node equipment. The collected logs were analysed at the end of the monitored period to reveal hard facts that become useful in evidence based policy-making.

A Syslog-based log repository is capable of accepting log messages from remote devices on a network and sorting them based on message type and source. The routers (Node A, B, … N) have in-built utility for remote Syslog-based logging. This facility enables the forwarding of log messages from each node to a remote Linux server. The forwarded messages include: firewall logs generated when p2p traffic is detected; user session statistics (hotspot messages); and web-proxy logs (visited URLs). In all, over 13 Gigabytes of logs were collected.

The hotspot session statistics includes such details as session start and end time, total downloaded and uploaded bytes, MAC-address of device used, hotspot login parameters, and the node and client IP addresses. At the start of each session, which is triggered by a hotspot login, a custom router-embedded script collects initial parameters of the client. It then forwards them to the server for use in creating a new session entry in the database. The initial parameters includes: the client's IP and MAC-addresses, time of login and the network node.

Post-data-collection activities

The collected logs were first converted into database entries for further analysis. The following analyses were carried out: p2p analysis, to investigate the contribution of p2p activities to the overall network traffic; web–proxy analysis, to investigate the web resources that are of interest to users; and hotspot analysis, to discover patterns in network use within the University. In all, SQL queries were used to synthesize relationships and trends among the data collected. In the following sections we focus on the particulars of each of these post-data-collection activities, and the results that ensued.

Data pre-processing

Entries in the log files were grouped according to their source. The groups were: web-proxy, hotspot and firewall. Messages related to each of these groups were processed accord-

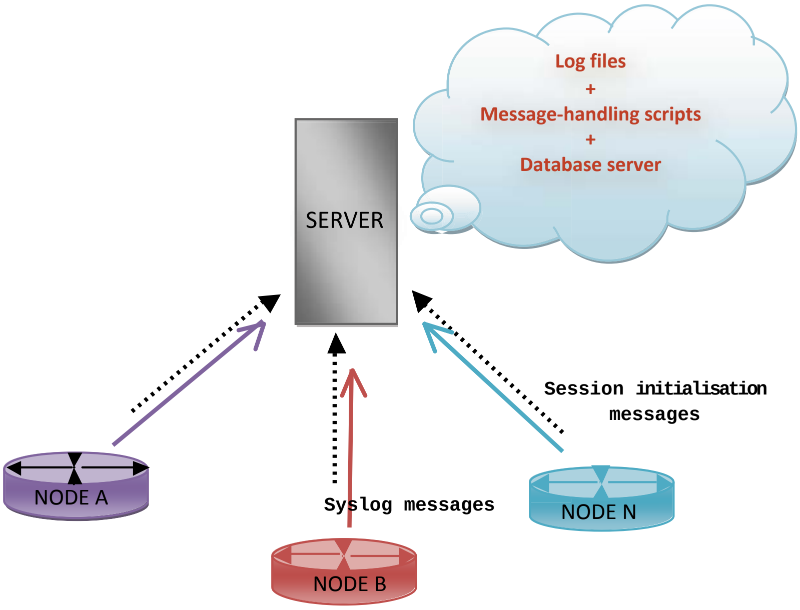

Figure 1. Schematic of network monitoring setup on UI network.

ingly as follows: Web proxy messages were broken down into source IP address and visited URL fields. Hotspot messages were broken down into start and end-of-session times, total downloaded and uploaded bytes, network node, and IP and MAC address fields. These fields were then stored in a database.

The processing and merging of connected details in the log files with those in the database required that the log entries be properly tagged ahead of time. Such tags indicated node names and services which generated the messages, as well as the IP address of the users. Since the usernames and IP addresses were also collected at session initialisation, these same parameters were matched with time frames – within the node concerned – to generate database fields that complemented session initialisation entries.

Since the network nodes utilised Dynamic Host Configuration Protocol (DHCP) (Droms, 1997), it was necessary to increase the lease time to at least one day in order to ensure the uniqueness of the user IP address while processing the log for a given day and node. For this to be feasible, the DHCP IP address pool was widened to make provisions for a large number of users within a node in a given day. Typical subnet mask of /22 or /23 was sufficient for this purpose.

P2P analysis

A Firewall rule was set up to flag p2p traffic across each monitored node. These p2p events were then analysed in order to derive the intensity of p2p activity in each node at any given time of day.

The results identified nodes with prominent p2p activities and the corresponding time of day in which they were more pronounced. Figure 2a identifies the three main network nodes from which these activities originate. Figures 2b shows that p2p activities peak

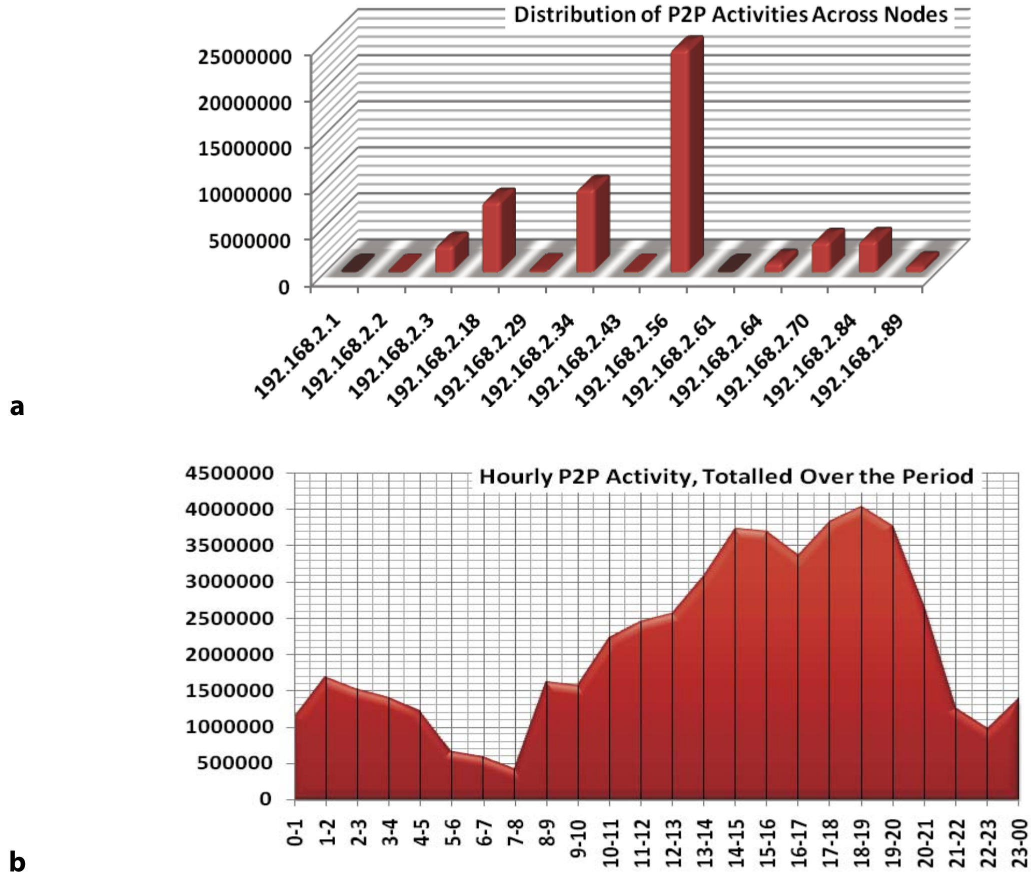

Figure 2. (a) P2P flags (vertical axis) compiled at various network nodes (horizontal axis), over the monitored period. (b) Total p2p flags (vertical axis) at each hourly interval of day (horizontal axis), over the monitored period.

at about 6-7pm; but generally, they are more pronounced from 1-2pm to around 8pm.

Traffic analysis

Traffic usage statistics of all user sessions were logged throughout the monitored period. This enabled the calculation of total bandwidth (upload plus download) consumed by each user in each period, as well as total traffic usage in each of the monitored nodes. Overall traffic distribution across the entire network was also determined at various times of the day. In Figure 3a, total traffic is plotted against time of day, clearly indicating the peak traffic period. Figure 3b indicates the nodes that made the most contribution to the overall traffic.

User session analysis

Over the monitored period, a total of 13 556 distinct user sessions were observed. A session starts when a network user signs into the campus hotspot, and ends when that user signs out. The user signs out explicitly by following a sign out link, or by a hard disconnection from the network, which triggers a session timeout event.

At session initialisation, some details about the user are collected and uploaded to a central database server. These details include: hotspot username, MAC address, user IP address, the network node at which the user is stationed, and the timestamp (at session initialization). At the end of a session, a session termination message is sent to a log file. This message contains the duration

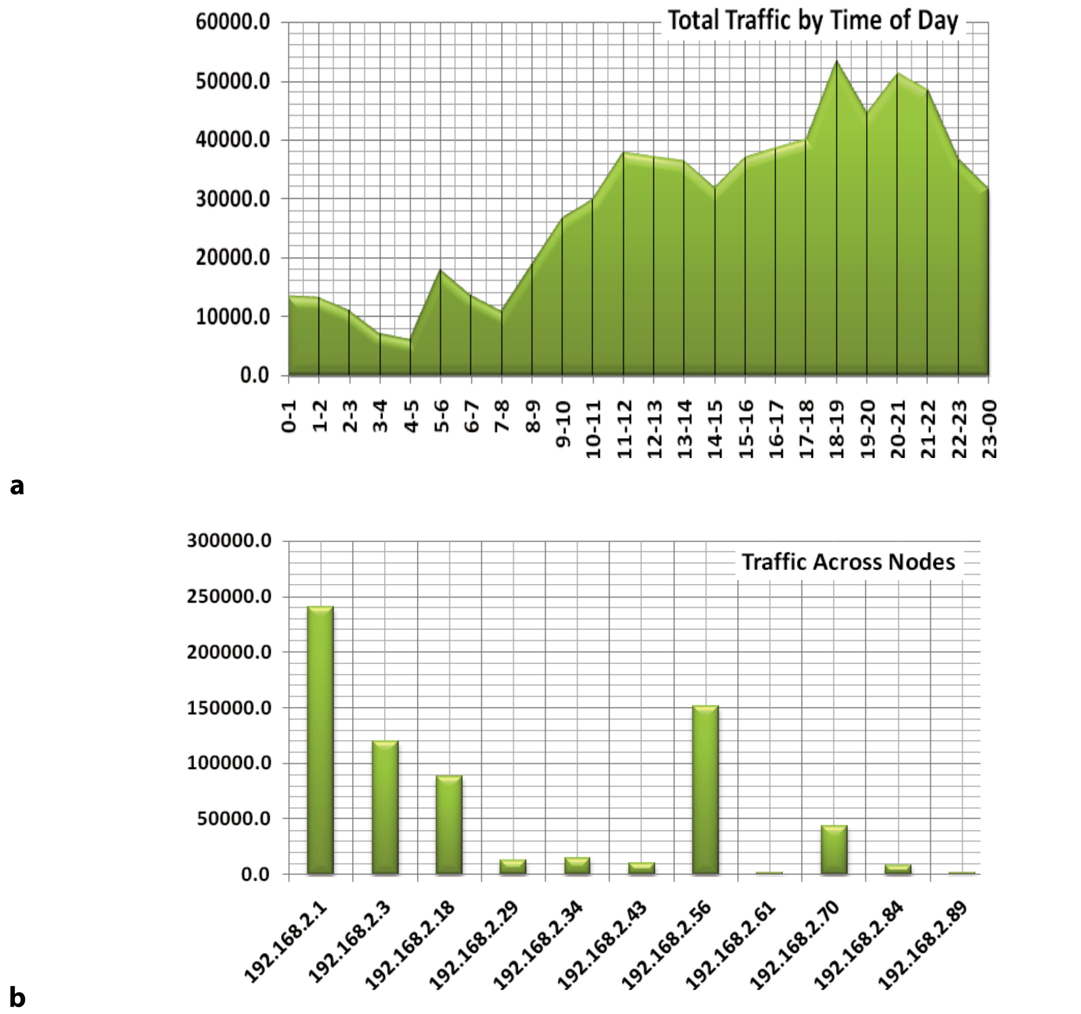

Figure 3. Traffic statistics during the monitored period. (a) Total traffic in megabytes (vertical axis) at each hourly interval of day (horizontal axis), over the monitored period. (b) Total traffic in megabytes (vertical axis) compiled at various network nodes (horizontal axis), over the monitored period.

of the session, the timestamp (at session termination), and the traffic statistics of that user session.

A count of active user sessions within each hour of the day was done across all the nodes (Table 1). Naturally, some sessions spilt into adjacent hours. These sessions were counted in whichever hour they overlapped, thus giving rise to a total of 17 765 total hourly session count. Figure 4a shows this hourly dis-

Table 1. *User session count by hour of day and by network node.*

NODE| TIME	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-00	TOTALS
192.168.2.1	66	52	20	3	3	7	15	29	139	297	477	452	446	550	522	613	680	580	495	397	413	333	195	195	6979
192.168.2.3	63	44	22	13	9	11	26	40	93	80	150	193	151	140	180	207	236	213	258	258	275	239	216	208	3325
192.168.2.18	6	8	6	1	2	9	1	2	13	9	32	55	74	90	92	93	87	87	74	56	48	28	23	20	916
192.168.2.29	-	-	-	-	-	-	-	-	11	51	73	100	73	108	109	123	93	73	68	37	7	2	6	-	934
192.168.2.34	-	-	-	-	-	-	-	-	2	1	2	1	3	5	2	2	2	4	5	2	1	1	-	-	33
192.168.2.43	-	-	-	-	-	-	-	-	1	11	24	36	59	52	26	53	54	24	3	2	1	-	-	-	346
192.168.2.56	94	41	15	11	13	18	29	69	98	100	169	137	105	136	163	137	139	124	128	92	158	183	184	165	2508
192.168.2.61	2	2	2	-	-	1	-	2	8	5	5	1	1	-	3	2	2	9	16	5	5	12	18	13	114
192.168.2.70	-	-	-	2	1	1	4	7	44	95	228	219	231	273	269	218	202	123	73	67	52	45	14	7	2175
192.168.2.84		2	1	-	-	-	2	2	1	7	24	26	30	25	39	28	40	37	25	18	19	9	11	6	352
192.168.2.87	-	-	-	-	-	-	-	-	-	5	4	10	12	5	10	17	10	6	2	1	1	-	-	-	83
TOTALS	231	149	66	30	28	47	77	151	410	661	1188	1230	1185	1384	1415	1493	1545	1280	1147	935	980	852	667	614	17765

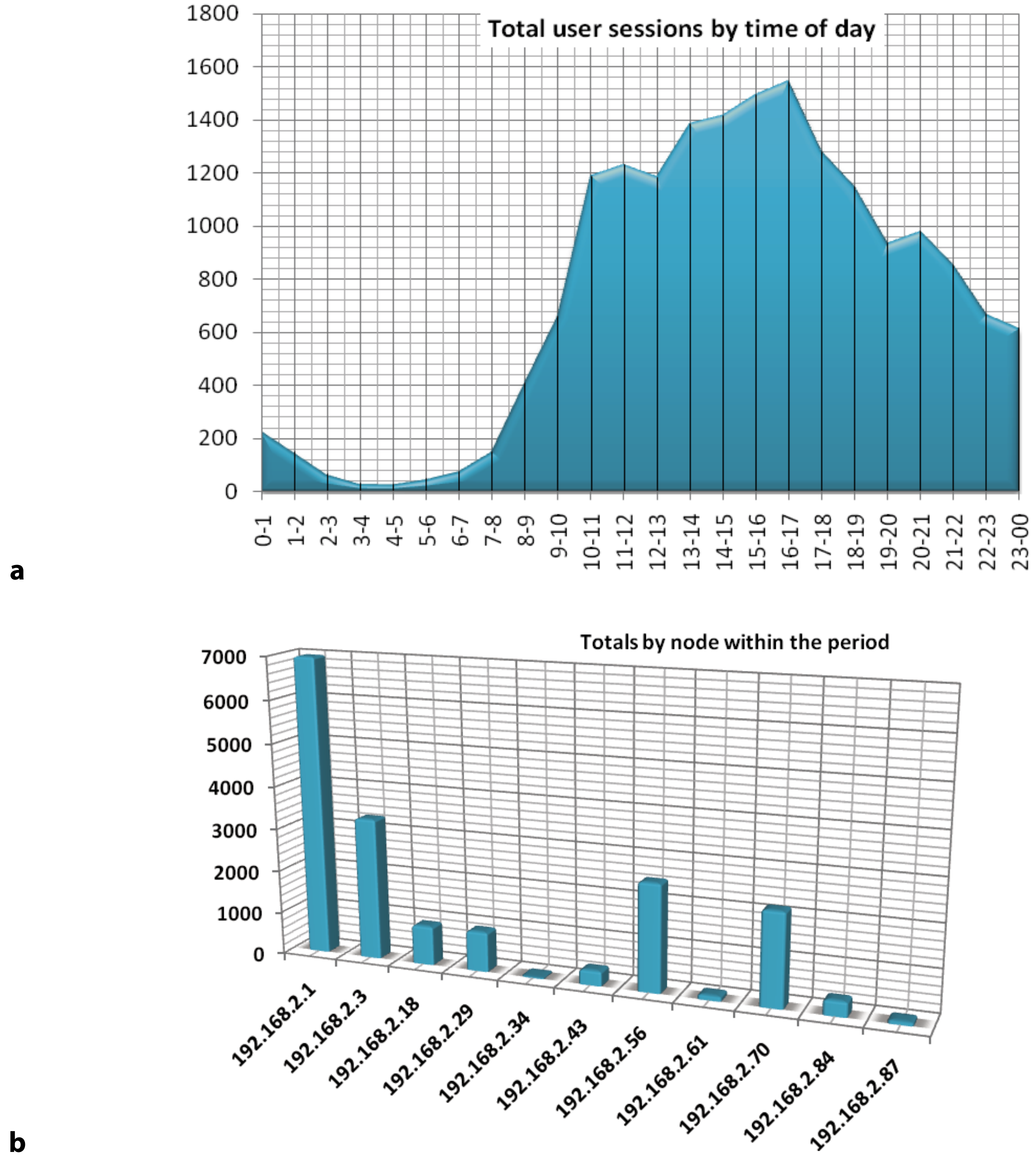

Figure 4. Traffic statistics during the monitored period. (a) Total count of active user sessions (vertical axis) at each hourly interval of day (horizontal axis), over the monitored period. (b)Total count of active user sessions (vertical axis) compiled at various network nodes (horizontal axis), over the monitored period.

tribution of users across the entire network in the monitored period. Figure 4b indicates which nodes have the largest density of users.

Web proxy analysis

A web proxy server was setup in each of the monitored nodes to track URLs requested by users on the network. Analysis of these URLs revealed the popular domains and online utilities on campus, ascertained by the analysis of URLs captured in the web proxy logs over the monitored period. Table 2 shows a list of some of the most popular utility brands among the network users.

Table 2. *Most popular utility brands observed during the monitored period.*

Most popular mail server	Yahoo! Mail
Most popular search engine	Google
Most popular antivirus software	Avast
Most popular news site	Google news
Most popular computer brand	Hp

Discussion

Correlation of factors: p2p intensity, number of users, and overall traffic

The results of p2p analysis present us with three prominent nodes in terms of intensity of p2p activity: *192.168.2.56*, *192.168.2.18*

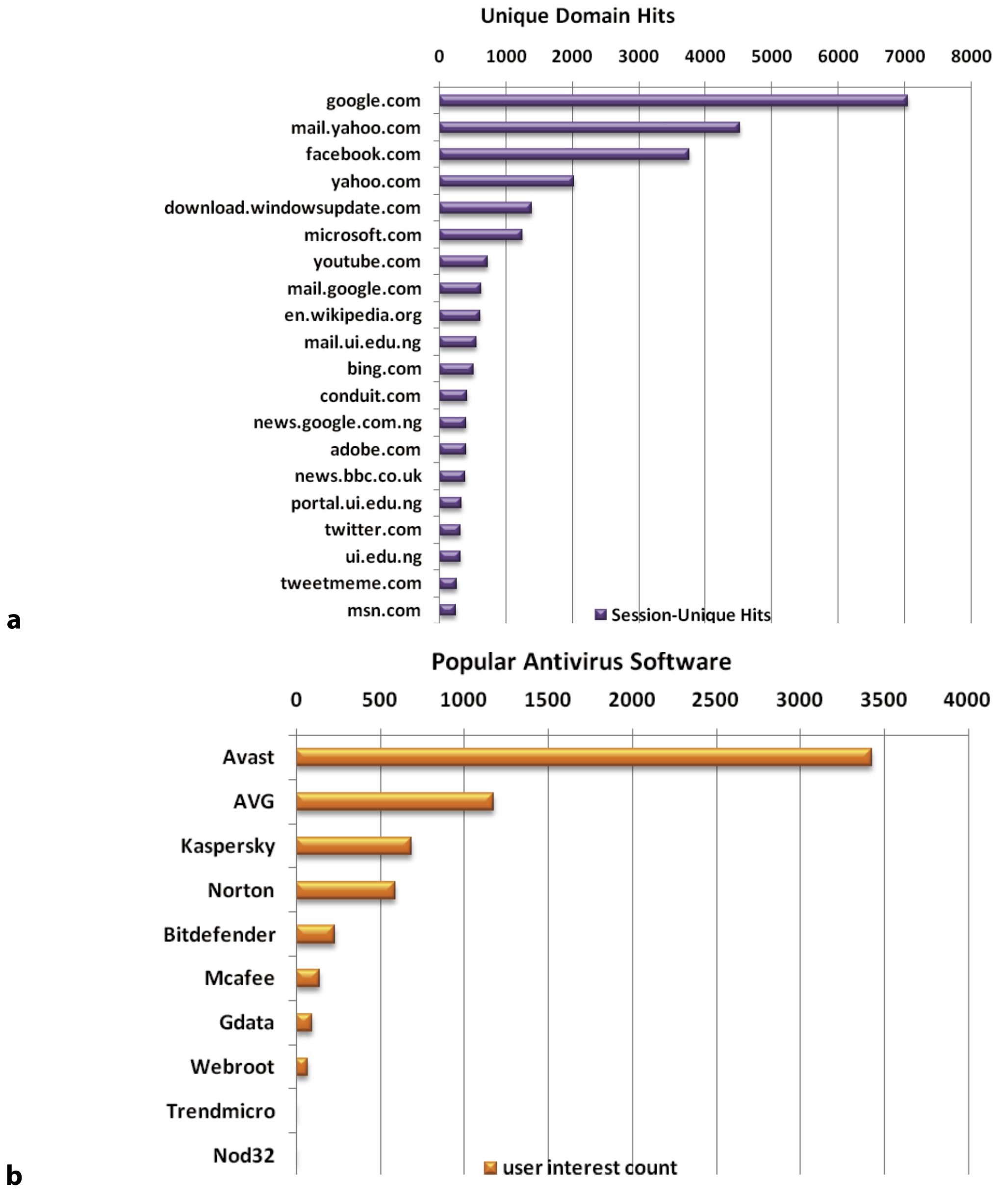

Figure 5. Popular (a) domains and (b) antivirus software observed during the monitored period.

and *192.168.2.34*. The nodes with the heaviest traffic were: *192.168.2.1*, *192.168.2.56*, *192.168.2.3* and *192.168.2.18*. Finally, the nodes with the greatest users were: *192.168.2.1*, *192.168.2.3*, *192.168.2.56*, *192.168.2.70*.

The correlation between the number of users and p2p activity clearly indicates that the stronger factor contributing to traffic statistics is number of users. This is also evidenced in the disproportion between the intensity of p2p on *192.168.2.56* and *192.168.2.1* (in Figure 2), against the final traffic outcomes (in Figure 3). It seems that the number-of-user component outweighed the p2p-activity component in those nodes. Furthermore, looking at Figures 3 and 4, the nodes with very few users are also the nodes with least traffic.

However, we cannot overlook the obvious proportion in the intensity of p2p activity on the nodes and the eventual traffic outcome (Figures 2 and 3). The exception is node *192.168.2.1,* which turns out to be the network of a staff-only Internet café.

There are two main factors that drive traffic on a node and therefore affect user bandwidth availability, namely: number of active users and intensity of p2p activity on that node.

Intense p2p activity starts around mid-day, peaks between 6pm and 7pm, and continues high until around 9pm. On the other hand, we see a sharp surge in number of users at the start of office hours (8am), hitting the peak between 4pm and 5pm, and then declining fairly steadily until midnight.

User sessions versus p2p activity

There is a good amount of overlap between p2p time on the entire network and overall user session time. This indicates that users might generally be interested in p2p as part of a normal network access session.

However, a look at the logs for *192.168.2.56* – the node with the highest p2p activity – revealed that a few users were persistently generating p2p traffic. The source of these events was traced to browser toolbar p2p applications from conduit.com. Such apps include: p2p games, file sharing, TV, radio, etc., all mounted on the browser and part of the often large p2p swarms arising from the user base of conduit.com network. This linkage with browser add-on applications was detected by tracing the history of URLs up to the p2p entries in the log files.

The observations in *192.168.2.56* cannot be generalised yet as further analysis is necessary on all nodes. However, there is at least an indication that browser-toolbar apps are playing a strong role in overall p2p traffic.

Some security observations

It is interesting to observe in Table 1 that users were found on the network at times when indoor facilities were already closed, indicating considerable open field and outdoor browsing activity. Also, a closer look at hotspot accounts usage revealed a high incidence of account sharing among staff and students. For example, results showed a particular staff account that was used by over 130 different users during the monitored period. This indicates that students are seeking more flexibility in network access as they currently have to access the network with auto-generated hotspot access codes, which can only be purchased and used at designated Internet cafes within the campus. Staff hotspot accounts, on the other hand, can be used anywhere on campus.

Policy recommendations

From the foregoing discussions, the following recommendations should be considered in order to improve the overall user experience of available bandwidth:

1. Incentives that foster even distribution of users across all nodes: primarily, a stable and backed-up power supply at all network nodes especially in offices and in residential areas. Network users also have to be kept informed of the efficiency they gain in accessing the network from low density nodes and at off-peak periods. If users could access the network at their homes, offices and classrooms (within the campus) without having

to migrate to specific facilities or locations, the number of users per node would even out naturally following the network plan. Alternatively, more bandwidth could be allocated to those nodes with high user density. Nonetheless, there is a clear need to invest in power backups across all nodes to increase overall network availability.

2. A follow-up to (1) could be the shaping of p2p traffic at peak user times. In addition, shaping or blocking of p2p traffic at nodes in which they occur, rather than at the network border gateway, will have the advantage of secluding the troubled nodes (in the case of malicious activity) and freeing overall bandwidth at the gateway.
3. In order to avoid security issues that could arise from sharing hotspot accounts, all network users should be provided with their own regular accounts. These accounts, while providing flexibility in network access, could still be metered and charged by usage.

Insights from user activity and traffic statistics

A log-based evaluation of the upload to download ratio of traffic revealed an estimate of the bandwidth proportions required by the overall user community. This data is useful in planning for Internet bandwidth expansion. See Table 3 for details.

An average user session duration of 56.4 minutes gives an insight into the nature of activities of many users on the network (Table 3). Coupled with the rather low average download/upload rate, users are limited to browsing web pages, checking mails and downloading small files.

However, from the top 20 visited domains presented in Figure 5a, we observe a keen interest in YouTube.com, which is a rich source of educational videos. Clearly, users could not have had a very smooth viewing experience judging from the average rate of download (Table 3), and looking at user distribution over nodes and over time.

Furthermore, from results in Figure 5a, the domains of the University (ui.edu.ng, mail.ui.edu.ng, portal.ui.edu.ng, etc.) lie further down the list of most popular domains. In addition, the local online newspapers lie outside the top 20 domains, whereas Google News features prominently. Both Google Mail and Yahoo! Mail are shown to be preferred to the institutional mail server. Social networking on Facebook is also a very prominent activity on the campus network. Unexpectedly, however, visits to research oriented domains are unremarkable, together with general dearth of interest in local content. The predominant interest in international content tends to reinforce the need for increased international bandwidth, with consequent increase in Internet transit expenditure.

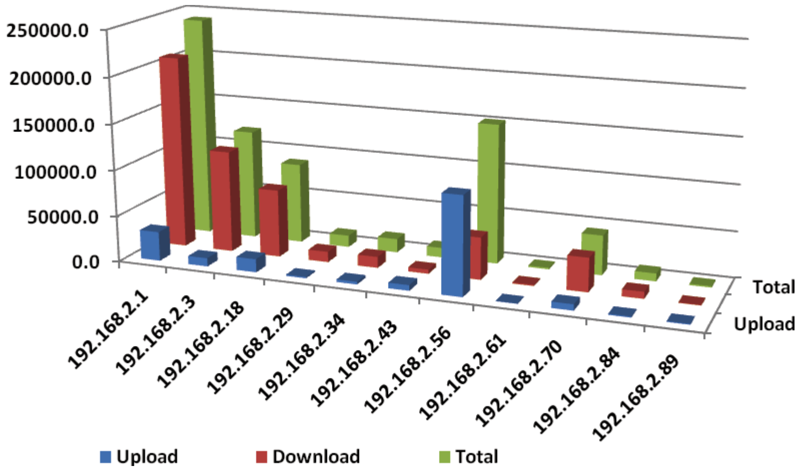

Figure 6. Upload/download traffic distribution (in megabytes) across nodes during the monitored period.

Table 3. Summary of traffic statistics.

Sum of Session Durations (Hours)	No. of Distinct Sessions	Total Download Traffic (KB)	Total Upload Traffic (KB)	Average Rate of Download (KB/s)	Average Rate of Upload (KB/s)	Proportion of Download/ Upload	Average Session Duration (Minutes)
12 724.47	13 536	512 842 630.928	179 878 921.713	11.2	3.9	3:1	56.4

Policy recommendations

Reduced expenditure on unit bandwidth is an essential goal of bandwidth management. To achieve this goal, the following areas of policy intervention are particularly important:

1. There is need for increased Internet bandwidth, in the University. Foregoing traffic statistics and user activity provide relevant guidelines.
2. There is need to create incentives for greater access to local institutional domains. Such incentives include enriching the local domains and services with utilities relevant to the user's academic and social pursuits.
3. Creation of local cache of popular foreign content. For example, antivirus updates for popular antivirus software could be cached on a local update server for faster access and reduced international traffic. Other educational resources such as research papers and videos could also be made available on the campus network with agreement with content owners. . The increasing online repository of open educational courseware is particularly relevant here.

Conclusion

Optimum bandwidth management can be achieved by means of evidence based policy research and implementation. Network logs proved to be a rich source of such evidence as demonstrated in this paper. The Network data used for this research was obtained from the campus network of University of Ibadan, Nigeria. Analysis of the gathered data yields results that prove to be strongly relevant to the making of policies geared towards network bandwidth optimisation. Policy recommendations were made for the campus network, following the observations from the research.

References

1. AAISOA and CIPACO (2005) Institut Panos Afrique de l'Ouest, Association of African Internet Service Provider Associations and CIPACO *Development of local Internet traffic in West and Central Africa and beyond: synthesis of an e-discussion*. 2005, Dakar: Panos Institute West Africa. 53 p.
2. Allman M, Paxson V and Blanton E (2009) *TCP Congestion Control*. RFC 5681, IETF, September 2009.
3. Bauer SJ (2008) *Congestion on the Internet: operator responses, economic analysis, and improving the network architecture*. Boston, MIT, Dept of Electrical Engineering and Computer Science, 156 p.
4. Brzezinski A (2007) *Scheduling algorithms for throughput maximisation in data networks*. Boston: MIT Department of Electrical Engineering and Computer Science. 226 p.
5. Cohen ED (2010) *Mass surveillance and state control: the total information awareness project*. 1st ed., New York: Palgrave Macmillan. vi, 252 p.
6. Dermler G (2000) *Towards a scalable system for per-flow charging in the Internet*. Research report RZ. International Business Machines Corporation. Research Division. Yorktown Heights, NY: IBM T.J. Watson Research Center. 8 p.
7. Droms R (1997) *Dynamic Host Configuration Protocol*. RFC 2131, IETF, March 1997.
8. Georgiadis L *et al.* (1996) Efficient network QoS provisioning based on per node traffic shaping: In *Proceedings of the Fifteenth annual joint conference of the IEEE computer and communications societies conference on The conference on computer communications - Volume 1* (INFOCOM'96), Vol. 1. IEEE Washington, DC, USA: Computer Society, pp 102-110.
9. Gummadi KP, Dunn RJ, Saroiu S, Gribble SD, Levy HM and Zahorjan J (2003) *Measurement, modeling, and analysis of a peer-to-peer file-sharing workload*. In SOSP'03: Nineteenth ACM Symposium on Operating Systems Principles, 2003, pp. 314–329.
10. Jacobson V, *Congestion avoidance and control. ACM SIGCOMM Computer Communication Review* 18 (4) (1988), pp. 314–329.
11. Janevski T (2003) *Traffic analysis and design of wireless IP networks*. 2003, Boston: Artech House. xvi, 368 p.
12. Kim S and Varshney PK (2005) *Adaptive online bandwidth allocation and reservation for QoS*

sensitive multimedia networks. Computer Communications. Volume 28, Issue 17, 17 October 2005.

13. Kondakci S (2009) *A concise cost analysis of Internet malware*. Computers & Security, Volume 28, Issue 7, October 2009, *Pages 648-659.*
14. Kosiur DR (2001) *Understanding policy-based networking*. Wiley Networking Council series. New York: John Wiley & Sons Ltd. xx, 348 p.
15. Lenard TM and May RJ (2006) Progress & Freedom Foundation (USA) *Net neutrality or net neutering: should broadband internet services be regulated*. New York, NY: Springer. xii, 225 p.
16. Mitra D and Ramakrishnan KG, *A case study of multiservice, multipriority traffic engineering design for data networks*. Proceedings of IEEE GLOBECOM 99, Rio de Janeiro December (1999), pp. 1087–1093.
17. Nunziato DC (2009) *Virtual freedom: net neutrality and free speech in the Internet age*. Stanford, Calif.: Stanford Law Books. xv, 194 p.
18. Raul AC (2002) Progress & Freedom Foundation (USA) *Privacy and the digital state: balancing public information and personal privacy*. Boston: Kluwer Academic Publishers. x, 148 p.
19. Sandvine (2002) *Peer-to-Peer File Sharing: The Impact of File Sharing on Service Provider Networks,* in *Industry White Paper*. Sandvine Inc.: Ontario.
20. syslog.org (2011) *Syslog*. Available at http://www.syslog.org/wiki/Main/Syslog [accessed 15 March 2011].
21. Welzl M (2005) *Network congestion control: managing Internet traffic*. Chichester, West Sussex, and Hoboken, NJ: John Wiley & Sons Ltd. xviii, 263 p.