Analysis of HTTP and HTTPS Usage on the University Internet Backbone Links

Shwan Dyllan, Hichem Dahimene, Phillip Wright, and Perry Xiao
London South Bank University/ICT Department, London, UK
Email: {dyllons, dahimenm, wrightps, xiaop}@lsbu.ac.uk

***Abstract*— We present our latest study on monitoring and analysing the behavior of the LSBU (London South Bank University) data network traffic. The analysis of the network activity allows us to understand the network traffic characteristic, on a continuous basis, which identifying the data traffic patterns on the gateways (JANET via Kings College & via Imperial College) and identify and highlight the end-to-end performance of the network path. In this paper, the PRTG network monitoring tool (Paessler AG, Germany) has been chosen as this provides the capabilities of SNMP, port utilisation as well as sFlow. The objectives of the study are to point out the impact of the LSBU network performance and its congestions current state and predict the future data network traffic and congestions; to identify the venerability of the network, due to high usage in particular time of the day; and to check the capacity usage of the current resources to maximise performance and network utilization.**

***Index Terms*—data traffic, sFlow, SNMP, Data network, network monitoring.**

I. Introduction

Network monitoring and analysis are very important, in order to understand the performance of the networks, the reliability of the networks, the security of the networks, and to identify potential problems. LSBU network is based on three tiers network architecture, i.e. CORE layer, Distribution layer and edge layer. The core layer is responsible for the routing protocols, the distribution layer responsible for all the VLAN management as well as spanning tree protocol and loop prevention as well as some level of security i.e. DOS protection. And finally the edge layer responsible for the end user connectivity.

Fig. 1 shows the schematic diagram of the LSBU network. The university computer network is set up by the latest data transmission and control technologies, using fibre-optic communication buses and radio lines. The university's computer network is the part of UK Education Network JANET and connects more than 8000 networked devices consist of 4000 Desktop computers, printers, IP camera, security locks, media control units, Wi-Fi access points, security alarm, fire alarm etc. All data flows of network are controlled by the two main Core Switches in DC1 (Data Centre 1) and DC2 (Data Centre 2), which are connected to the JANET routers via optical channels of 1Gbps and to the Internet. JANET provides computer network and related collaborative services to UK research and education. Through the main university Core Switch (DC1) there are about 70 Tbytes data per month that go across LMN interface. Out of this, about 50 Tbytes data was downloaded from the Internet, while 20 Tbytes data were uploaded. LSBU based on 124 Class B IP address areas, i.e. a total of 31744 public IP addresses and 110 class C IP addresses for switch management and miscellanies, Wi-Fi SSID VLANs. The university network is divided into sub networks creating 124 virtual LANs.

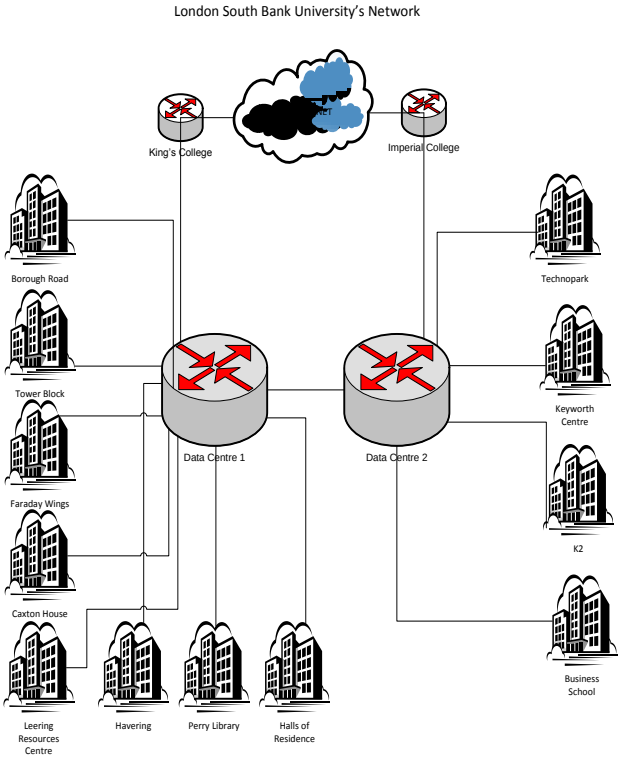

Figure 1. The schematic diagram of London South Bank University

II. Network Monitoring Tools

Data network traffic analysis become more demandable for managing the network infrastructure. Operational efficiency is vital to keep the cost minimum and service level high. As the result of technology improvement and the speed of current network, the service providers are designing their application or service with high network usage. Network monitoring

Manuscript received July 20, 2013; revised September 21, 2013

doi: 10.12720/jiii.2.1.67-70

tool such as SNMP, sFlow and port utilisation [1-4] are widely used currently to monitor and understand the network traffic activities. In this paper, the PRTG network monitoring tool from Paessler AG, Germany, has been selected as the monitoring and reporting tools, as it equipped with SNMP, sFlow and port utilisation, to provide the necessary data that then used to analysis the network.

sFlow protocol is used by many vendors to collect IP traffic information. Today, sFlow has become an industry standard for traffic monitoring. This protocol exports network flow data from the routing devices. Monitoring and analysis system based on this technology consists of:

- sFlow sources or exporters which may be one device, which handles all the organization's flows where all the information about network traffic is exported.
- sFlow collector is a computer and software that collect and store data from the agent device.
- sFlow agent is one of the core switch which pushes data to the sFlow collector.
- PRTG is the software that analysis the collected data.

On contrast Network flow is defined as a sequence of packets that are transferred between given two endpoints within a certain time interval in packet switching networks. The endpoints and the packet are identified by IP addresses as well as transport layer port numbers.
sFlow represent a data source at a granularity level that is scalable for network and security analysis, thus sFlow will be used as the main took tool to analyse the network. A sFlow record, obtained from Extreme BD10K switch, can contain a wide variety of information about the traffic in a given network flow

The sFlow analysis provides the result of the network traffic from the two main links for LSBU for traffic modelling, traffic engineering, capacity planning and forecasting, and anomaly and attack detection. This work allowing us to find approximate average traffic on the both Core Switches and their main link to the outside world, which is based on the sFlow analysis.

In order to systematically manage data network at any institution it is apparent that procedures are set in place to monitor asses and analyse the flow of the data, identifying the peak and off peak time period and any anomalies. This will result in proactively manage a robust network. Internet measurement workgroup's [3] investigation has profound effect using sFlow and SNMP monitoring and analysis of the data network, particularly for determining changes and anomalies.

ACM Sigmetrics have applied wavelet methods to filter traffic flow and carried out various investigations, such as network flow analysis and sFlow to collate and prediction bandwidth, bottleneck and capacity for future expansion [5]. The work has also investigated firewall loges and intrusion detection and port scanning anomalies. Other studies have been carried out addressing and investigating SNMP statistics and methodology for network bandwidth expansion [6]. The methodology investigates trends and changes that have an effect in providing good IP backbone prediction and upgrade to the backbone infrastructure.

To monitor our network traffic thoroughly we have to take the following steps under consideration

- Checking the traffic characteristics for one month
- Resource utilisation
- Evidence of congestion
- Traffic overhead

Fig. 2 and 3 show typical example of 48 hour network traffic at DC1 and DC2 , highlighting the HTTP and HTTPs traffic take up most of the overall traffic. It is highly beneficial to identify the services i.e. VOIP, Video Conferencing, Media streaming and sharing, which make up most of the bursting data traffic on the network and manage these activity or services in a control environment. Therefore SNMP for network performance, port utilization as well as security measurement may be beneficial since this network traffic is apparent to the overloaded networks.

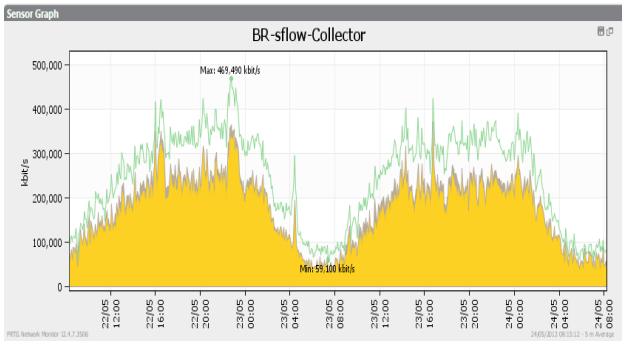

Figure 2. Example sFlow of all traffic and highlighting HTTP & HTTPS traffic over 48 hours periods in DC1

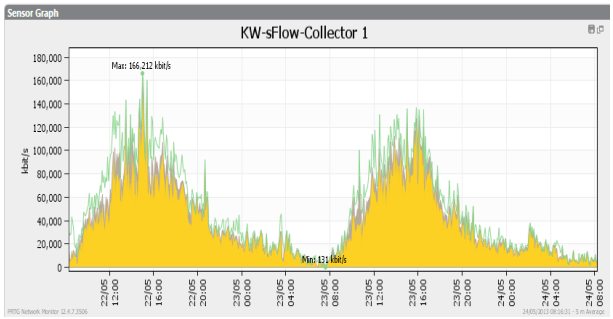

Figure 3. Example sFlow of all traffic and highlighting HTTP & HTTPS traffic over 48 hours periods in DC2

In this paper, we study the traffics of LSBU Network to see the efficiency of the entire infrastructure to identify and manage the traffic both inbound and outbound at the gateway.

III. Results and Discussions

Fig. 4 and 5 show the sFlow report for one month traffic, gathered for the two gateways. The Kings College link connected via DC1 shows a strong day/night effect. This is because it connects (see figure 1) the Student Halls Network (RESNET) and the LRC (Learning Resources Centre). The RESNET apparently generated the most traffic, while LRC opens till 24:00

daily and both operates during weekends. Further granular investigation has to be done to separate the two traffic flows. Imperial College link connected via DC2 shows both the day/night effects, and weekday/weekend effect. This pattern ties in with business operating hours 08:00 till 18:00.

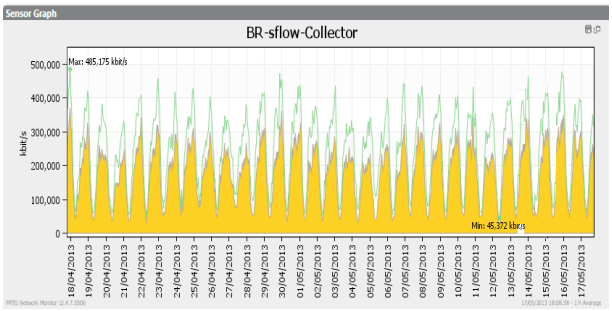

Figure 4. One month of HTTP, HTTPS and Total traffic DC1 Link

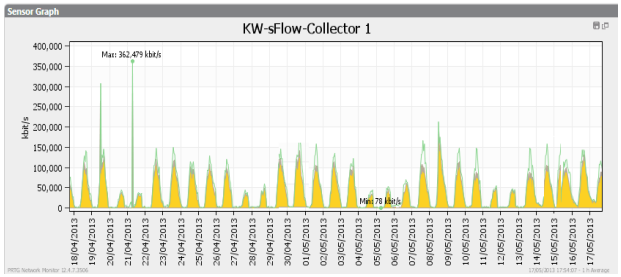

Figure 5. One month of HTTP, HTTPS and Total traffic DC2 Link

The usage of HTTP port (80) is more than port HTTPS (443). Per port analysis helped us interface the nature of the HTTP & HTTPS flow [7]-[8]. Around 40% to 60% of the flow runs on ports below 1024 including port NTP (123), DNS (53), IMAP (143), etc. According to port-based classifier, the top-used ports in terms of sFlow are normally used by P2P applications, such as 6881 (BitTorrent), 6699 and 6257. Depending on the computer network the listening ports are usually in range of 3000 to 5000. Therefore our analysis reveals that beside HTTP and HTTPS, there are many malicious ports are available that can throttle our traffic.

Figure 6. Sample raw data of the traffic

Fig. 6 shows example traffics among the traditional Internet traffic, we have other application using vast list of port available to communicate various other type of data over the same backbone, which leads to unplanned congestion on the network. If we analyse the figure 6 below we can identify the ratio of the traditional and unpredicted additional traffic [9].

Fig. 7 represent a month worth of data passed through the system organised in such way we can see what type of communication that had taken place. This data only show the TCP communication in that we can identify that the minority of the communication is the World Wide Web communication and 29% of it for other communication [10]. This traffic has to be looked at and analysed further to optimise the utilisation of the gateway.

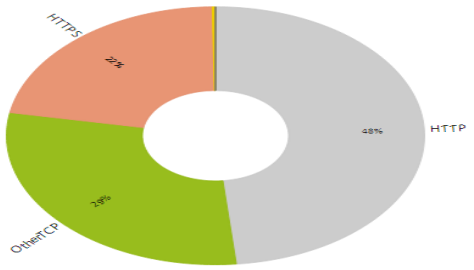

Figure 7. Pie chart representing the ratio of the HTTP, HTTPS and other traffic through the backbone network

We also captured traffic between HTTP & HTTPS simultaneously to test our method of simulation by running the following code as illustrated in Figure 8 & 9

while ((c < 100)); do ab -n 1000 https://www.lsbu.ac.uk/index.shtml && echo "HTTPS complete" & ab -n 1000 http://www.lsbu.ac.uk/index.shtml && echo "HTTP complete" & ((c = c + 1));sleep 60; done

Figure 8. Simulation test on internal traffic

Figure 9. Internal traffic simulation test Result

IV. CONCLUSIONS

In our analysis we focused on the usage of http and https and we discovered that the https is highly used in LSBU network. The study shows that the maximum bandwidth in DC1 road link has picked times of maximum 500 Mbps as well as the link from DC2 picked times of maximum 200 Mbps. This analysis has made it apparent that the bandwidth that both uplinks can run at the single interface if a failure does occur on one of the uplinks it might cause a bottleneck.

Most of our communication is now turning towards cloud based services. I.e. E-Mails, SharePoint, CRMs, Social Media are on the raise and this will have a significant effect on the raise in the Secure Communication specially the HTTPs communication. As we have seen from the http, https and other TCP communication ratio on figure 7 PIE chart representing the ratio of the HTTPS, HTTP and other traffic through the backbone network. This will mount to significant increase

There are project underway, which is to migrate all mail services for LSBU users on the cloud (Office 365), another major project is the expansion of wireless zone to cover the entire university connecting up to 30K devices. There are also plans to make Lync (video/Audio conferencing service) made available to LSBU users. These are some of the services that are cloud based and will increase the load on our network.

Current 1GB uplink on each of our gateways will cope with the traffic with load balancing put in place, but if there wouldn't be resilient if there were a failure occurs on one of the gateways

ACKNOWLEDGEMENTS

We thank London South Bank University for the financial support of this study.

REFERENCES

[1] S. A. K. Tanoli, I. Khan, N. Rajatheva, T. Issariyakul, and T. Erke, "Trace-based analysis for campus-wide wireless LAN over advanced training system," in *Proc. First International Conference on Future Information Networks*, 2009.

[2] A. Balachandran, P. Bahl, and G. Voelker, "Hot-spot cognetion relief and user service guarantees in public-area wireless networks," in *Proc. WMCSA*, June 2002.

[3] User manual AirMagnet, Web site. (2008). [Online]. Available: http://www.airmagnet.com/

[4] E. Garšva, N. Paulauskas, G. Gražulevičius, and L. Gulbinovič, "Academic computer network traffic statistical analysis," in *Proc. 2nd Baltic Congress on Future Internet Communications*, 2012.

[5] Network Animator (ns2) Web Site. (2005). [Online]. Available: http://www.isi.edu/nsnam/nam/

[6] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *Proc. of ACM MobiCOMOO*, August 2000, pp. 1-10.

[7] W. Li, A. W. Moore, and M. Canini,"Classifying HTTP traffic in the new age," in *Proc. ACM SIGCOMM*'08, *Eattle*, USA, Aug17-22, 2008.

[8] M-J. Choi, C-G. Jin, and M. Kim, "Classification of client-side application-level HTTP traffic," in *Proc. Korea Information and Communication Society*, vol. 36 no. 11, Nov 2011, pp. 1277-1284

[9] G. Bloch, S. Greiner, H. de Meer, and K. S. Trivedi, "Queueing networks and markov chains" *Modeling and Performance Evaluation with Computer Science Applications*, John Wiley & Sons, Inc., 1998.

[10] D.-Y. Chen, Y. Hong, and K. S. Trivedi, "Classification of faults, errors and failures in communication systems," in *Technical report, Center for Advanced Computing and Communications*, ECE Department, Duke University, 2002.

Shwan Dyllon: A Network Engineer at London South Bank University, UK. Accomplished BSc in Computer Study (1998) and MSc in Telecommunication and Network Engineering (2005) from London South Bank University, UK. At present undertaking PhD in Data Traffic analysis and Data Mining

Mohamed Hichem Dahimene: A Network and Desktop Support Engineer supporting end user devices and network equipment.at London South Bank University, UK. Did his studies in Science and Technology in Algeria, BEng in Computer System and Networks from London South Bank University, MSc in Telecommunication and Computer Engineer at London South Bank University, currently finishing off his thesis in Data Traffic Analysis Project.

Phillip Wright: A Network Team Leader at London South Bank University, UK.

Dr Perry Xiao CPhys MInstP (Institute of Physics) CEng MIET (The Institution of Engineering and Technology)
Born in Chang Chun, China, May 1968, BEng degree in Opto-electronics in 1990 and MSc degree in Solid State Physics in 1993, from Jilin University, China. Achieved his PhD in Photophysics in 1998 from London South Bank University, London, UK.
He is a Reader at London South Bank University, UK. He has more than 80 publications in journals, conference proceedings, and book sections. His main research interests are novel infrared and electronic sensing technologies, mathematical modeling and data analysis.

