❒   1442

Internet data traffic analysis for identifying usage trends on each day of the week in a university

Aderibigbe Israel Adekitan, Claudius O. A. Awosope
Department of Electrical and Information Engineering, Covenant University, Nigeria

Article Info	ABSTRACT
	Internet data traffic monitoring and management are important requirements for ensuring top notch quality of service in a network. Data traffic logs contain useful hidden information that can be harnessed and interpreted as a resource for making informed network management decisions. In this study, logged internet data traffic for both the upload and download traffic in a university for one year was analysed using statistics and partial least squares approach to structural equation modelling (PLS-SEM). Time series plots, statistical properties and trends for each day of the week over a 51-week period were developed. The result shows that the most data was downloaded on Thursdays while the most upload occurred on Mondays. A path model was developed using Smart PLS3, and the performance of the model was evaluated using the construct reliability and validity of the model. The results reveal that the weekly variance is majorly accounted for by usage variations on Tuesdays, Fridays and Saturdays. An overall model R-square value of 0.876 was observed.

Corresponding Author:

Aderibigbe Israel Adekitan,
Department of Electrical and Information Engineering,
Covenant University, Ota, Ogun State Nigeria.
Email: aderibigbe.adekitan@covenantuniversity.edu.ng

1. INTRODUCTION

Data is generated daily by every aspect of human and machine interaction on a global scale. The data generated from a field of study or from a specific domain contains valuable information and statistics. The Internet has gradually transformed the world into a global village providing speedy connectivity, access to volumes of information and bringing people together across different regions of the world. The use of the internet has also been associated with product price reduction in some industries [1]. The ability to measure activity-related data has provided platforms for gaining new and pertinent knowledge which has helped to enhance decision making processes. According to McAfee et al. (2012) [2], about 2.5 exabytes of data are generated daily, and according to the Telegraph's Technology Intelligence article, the world's internet data traffic in 2016 passed one zettabyte and Cisco has since predicted that this figure will double by 2019 [3]. The collection and accurate analysis of data create an opportunity for reducing cost, loss, and waste via improved management's decision making capacity through data-based knowledge acquisition [4].

Decision making, that is data driven, yields better and more productive decisions [2] due to the evidence-based approach which is factual, and also reflects operational realities. Obtaining relevant information from a dataset requires the use of the right tools, technology and skill set. A dataset loaded with hidden useful information, even with the availability of the right tools, will still require human insight for appropriate analysis toward extracting the hidden knowledge.

Internet usage behaviour is an important factor that influences the level of internet data traffic. In academia, students use the internet for both academic and non-academic purposes, and sometimes there

may be tendencies for addictive internet usage as revealed by research studies [5-7]. Increasing internet traffic creates a strain on the internet network facilities, and excessive traffic may impair network connectivity and service delivery [8]. The ability to track and monitor internet traffic by Internet Service Providers (ISP) has helped in generating traffic data that is representative of the network under study and as such, valuable information can therefore be extracted for trending and performance evaluation [8-9].

Internet traffic can be monitored by observing packet or flow traces. For continuous traffic monitoring, flow aggregation or packet sampling can be deployed in order to reduce the volume of data generated thereby yielding coarse traffic statistics [10]. In the early days, Internet traffic classification relied on the use of transport layer port numbers but the accuracy of this method is challenged by the applications with hidden identity, achieved by dynamically assigning ports or by deliberately using anonymous port numbers [11-13]. A better alternative is through the use of tools that monitor packet payloads by identifying the patterns of strings of known applications [14-15]. This method is challenged by high resource requirements, its inefficiency on encrypted traffic and privacy related issues. Host behaviour-based [14, 16], and flow features-based [11, 17] which analyse packet flow [18] and flow duration are alternative approaches that do not require payload inspection [12]. Recently, methods using Transport Layer Statistics features and Machine Learning approaches have been developed [19], and there are continuous studies toward improving their performance.

Extensive studies have been carried out on internet traffic monitoring, and this is an indicator of its importance in network management for identifying heavy bandwidth internet traffic for controlling applications that utilize high bandwidth [11] and ultimately for ensuring high quality of service [8, 20-23]. In this study, the internet traffic data generated in a smart university in Nigeria is statistically analysed to identify usage trends on each of the seven days of the week over a period of one year. The study classifies each day of the week from Monday to Sunday in terms of internet upload and download traffic over a 51-week period in a bid to identify the most traffic intensive and sensitive day of the week in the University.

2.   METHODOLOGY

The internet traffic data of Covenant University in Nigeria for a period of one year, spanning over 51 weeks was logged during the empirical study by Adeyemi et al. (2018) [21], and this provides an opportunity to carry out trend analysis on a real internet traffic data obtained practically by logging the internet traffic data using Mikrotik Hotspot Manager, FreeRADIUS, and web-based Radius Manager application. In this study, the internet upload and download traffic dataset was pre-processed and sorted according to the particular day of the week i.e. Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday thereby creating day-based traffic data across 51 weeks of the study period. Statistical analyses were carried out to identify data pattern and trend across the days of the week for the full year under study. A path analysis model was also developed by applying Partial Least Squares (PLS) approach to Structural Equation Modelling (SEM) in order to explain how internet data traffic variation for each day of the week determines the weekly internet protocol (IP) traffic variance, and also to classify the seven days of the week in terms of the strength of their causal effect on weekly variation trends for the full year under study using the path coefficient.

3.   STATISTICAL CHARACTERISTICS OF THE IP TRAFFIC DATASET

Statistical attributes of the internet download and upload traffic data are presented in tables and plots in this section. The variation of the internet download traffic data for each of the days of the week (i.e. Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday) across 51 weeks is presented in Figure 1, while that of the upload traffic data is shown in Figure 2.

Figure 3 and Figure 4 show the boxplot of the IP download data traffic and IP upload data traffic respectively for each day of the week. The boxplots show the traffic variation across four main quartiles i.e. the 25th percentile, the median, the 75th percentile and the maximum value. The normal probability plots for the download and the upload data traffic are presented in Figure 5 and Figure 6 respectively.

The average download and upload data traffic for each of the 51 weeks are displayed in the bar charts of Figure 7 and Figure 8 respectively. Figures 9(a) and 9(b) present the box plot of the total download and upload data traffic per week.
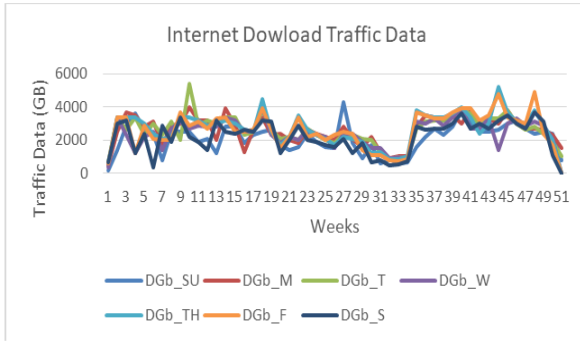
Figure 1. Internet download traffic data trend for each day of the week for 51 weeks
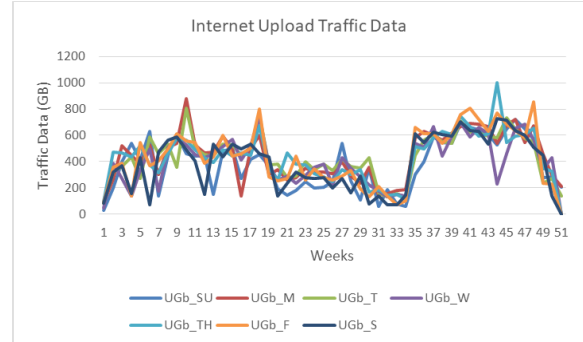

Figure 2. Internet upload traffic data trend for each day of the week for 51 weeks

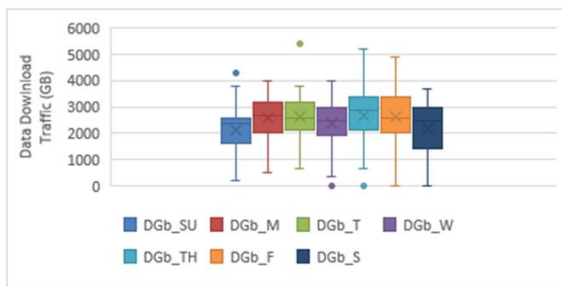Figure 3. A box plot showing the quartile variations for the download traffic data for the seven days
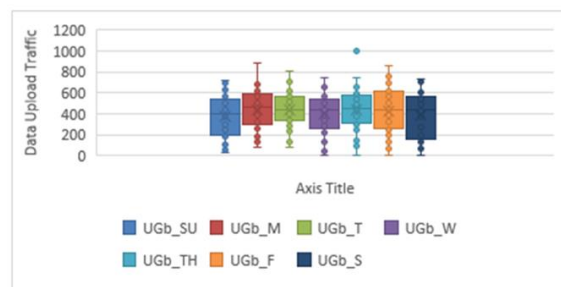

Figure 4. A box plot showing the quartile variations for the upload traffic data for the seven days
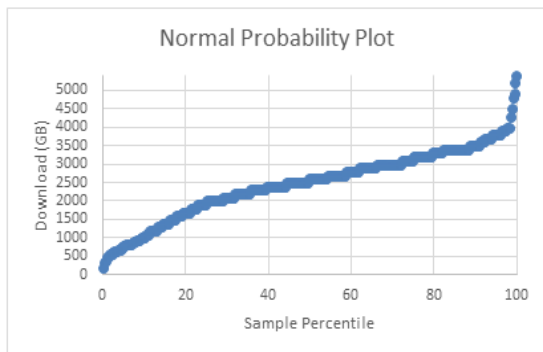

Figure 5. Normal probability plot for the download data traffic
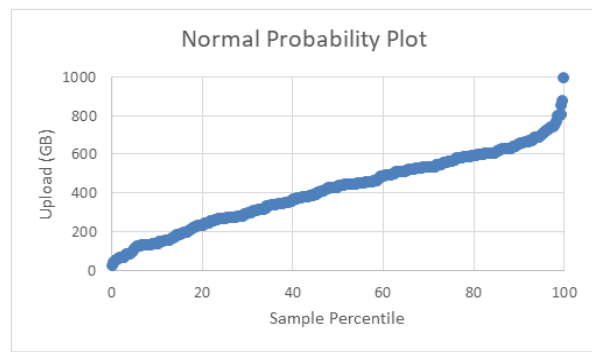
Figure 6. Normal probability plot for the upload data traffic
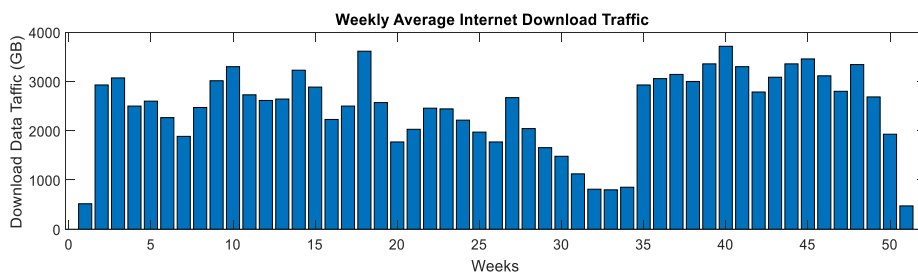
Figure 7. Bar chart of the average IP download data traffic per week
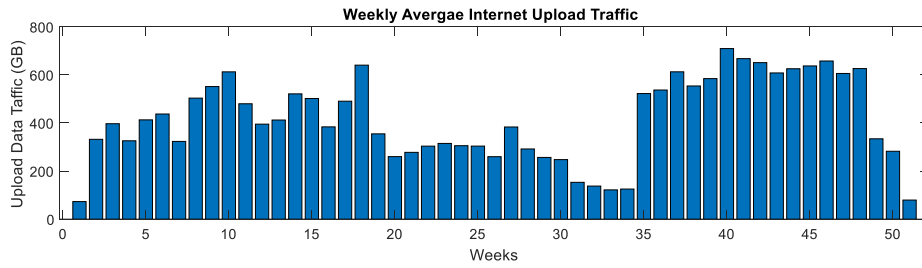
Figure 8. Bar chart of the average IP upload data traffic per week

(a)   (b)
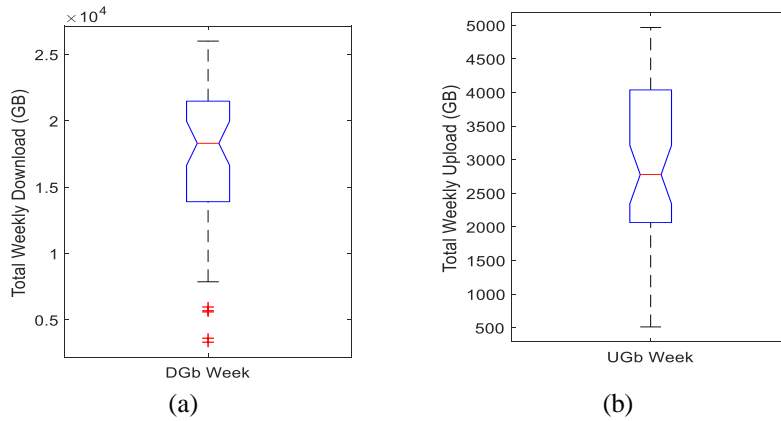
Figure 9. Box plot of the (a) Total weekly download internet traffic (b) Total weekly upload internet traffic

The total download and upload internet traffic data on a weekly (7 days) basis are presented as box plots in Figures 9a and 9b respectively, while in Figure 10, Figure 11, Figure 12, Figure 13, Figure 14, Figure 15 and Figure 16 the internet traffic data for both the download and upload are presented for each day of the week showing trends and traffic variation patterns.
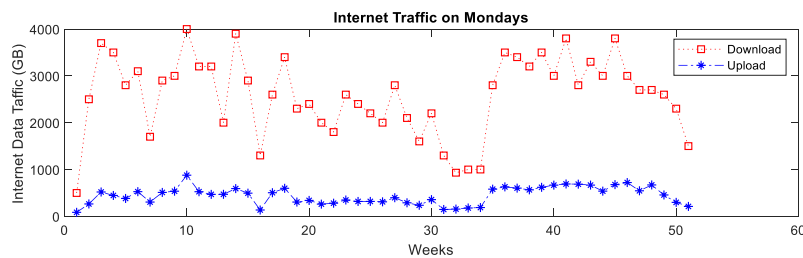
Figure 10. Internet data traffic trend for all the Mondays within the study period
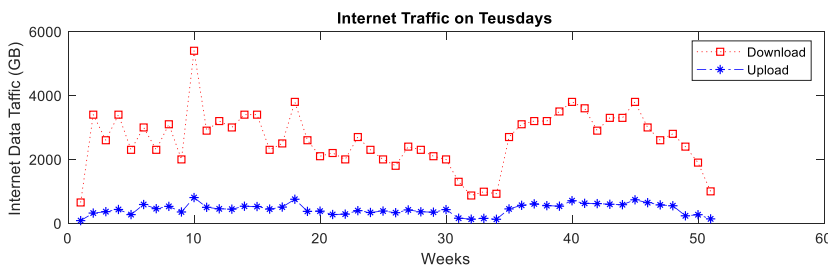
Figure 11. Internet data traffic trend for all the Tuesdays within the study period
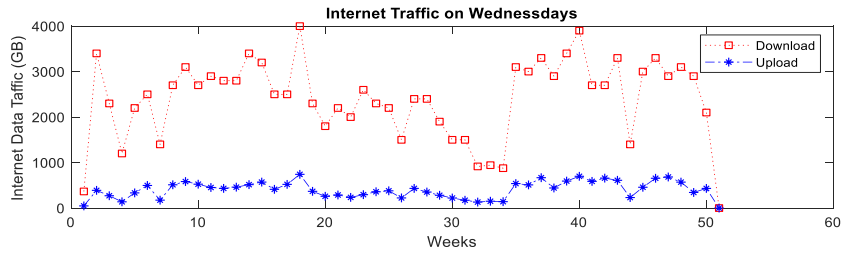
Figure 12. Internet data traffic trend for all the Wednesdays within the study period
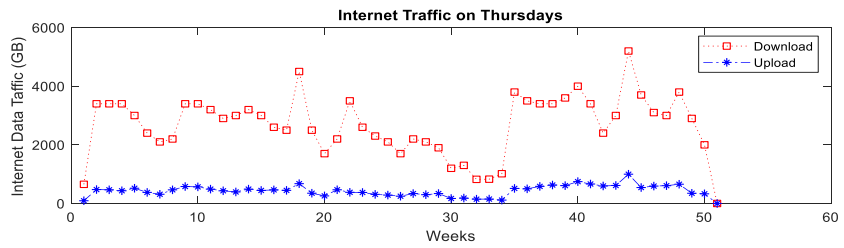
Figure 13. Internet data traffic trend for all the Thursdays within the study period
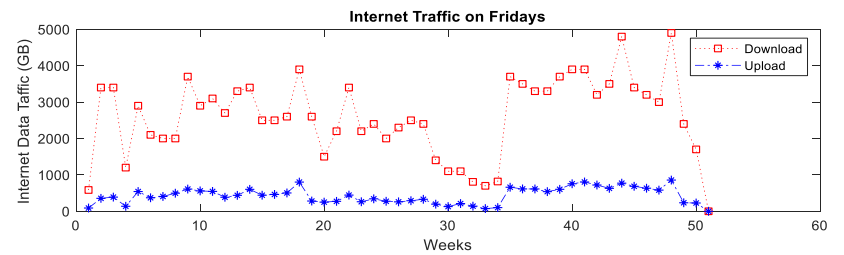
Figure 14. Internet data traffic trend for all the Fridays within the study period
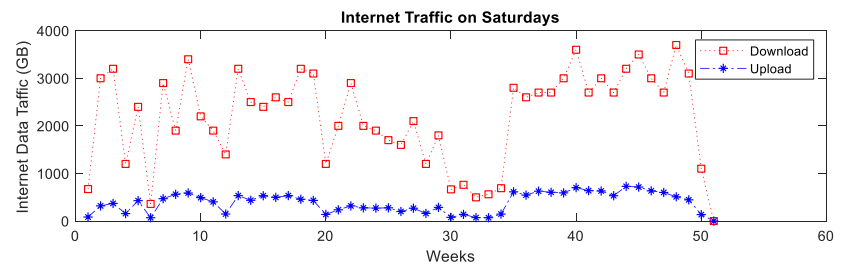
Figure 15. Internet data traffic trend for all the Saturdays within the study period
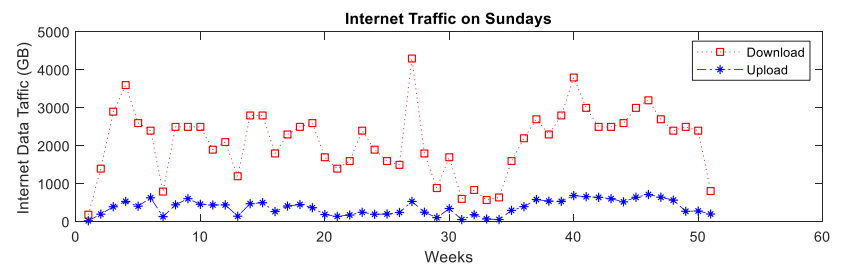
Figure 16. Internet data traffic trend for all the Sundays within the study period

For the seven days of the week; the correlation matrix and the covariance matrix of the internet download data traffic are presented in Table 1 and Table 2 respectively while the correlation matrix and the covariance matrix of the internet upload data traffic are presented in Table 3 and Table 4.

Table 1. Correlation Matrix of the Download Traffic Data

|	DGb_SU	DGb_M	DGb_T	DGb_W	DGb_TH	DGb_F	DGb_S
DGb_SU	1						
DGb_M	0.799576	1					
DGb_T	0.67461	0.859821	1				
DGb_W	0.614065	0.704149	0.740886	1			
DGb_TH	0.641574	0.755975	0.789076	0.739835	1		
DGb_F	0.570947	0.681736	0.688424	0.778132	0.911544	1	
DGb_S	0.520165	0.531075	0.596339	0.723522	0.812338	0.872035	1

Table 2. Covariance Matrix of the Download Traffic Data

	DGb_SU	DGb_M	DGb_T	DGb_W	DGb_TH	DGb_F	DGb_S
DGb_SU	749119.5						
DGb_M	577644.1	696706.6					
DGb_T	517131.6	635632.1	784415.2				
DGb_W	460259.1	508981.3	568246.7	749937.7			
DGb_TH	564589.6	641568.5	710563	651415.4	1033765		
DGb_F	531800	612376	656154.8	725175.4	997392	1158122	
DGb_S	434554	427866.7	509792.4	604771.8	797213.8	905812.4	931654.1

Table 3. Correlation Matrix of the Upload Traffic Data

	UGb_SU	UGb_M	UGb_T	UGb_W	UGb_TH	UGb_F	UGb_S
UGb_SU	1						
UGb_M	0.839863	1					
UGb_T	0.791664	0.877544	1				
UGb_W	0.749896	0.785733	0.790411	1			
UGb_TH	0.746667	0.768865	0.776502	0.718496	1		
UGb_F	0.728648	0.815367	0.808204	0.808819	0.898802	1	
UGb_S	0.644511	0.734918	0.72104	0.724241	0.836383	0.866414	1

Table 4. Covariance Matrix of the Upload Traffic Data

	UGb_SU	UGb_M	UGb_T	UGb_W	UGb_TH	UGb_F	UGb_S
UGb_SU	37641.37						
UGb_M	29826.46	33505.86					
UGb_T	26274.2	27477.98	29262.5				
UGb_W	26348.46	26046.98	24486.72	32797.71			
UGb_TH	26846.8	26082.15	24616.76	24114.54	34345.16		
UGb_F	30724.88	32437.92	30048.04	31835.6	36202.32	47236.69	
UGb_S	25764.73	27717.99	25414.25	27025.08	31937.46	38799.65	42454.6

A one-way ANOVA can be carried out to test the null hypothesis that the means of the internet data traffic for all the days of the week are equal, and if F > F crit, the null hypothesis is rejected. To further perform a comparative analysis of the internet data traffic for each day of the week, a single factor or one-way ANOVA was carried out. The results for the internet download data traffic are presented in Table 5 and Table 6. From Table 6, it is shown that F (2.8938) > F crit (2.12451). Hence, it implies that the means of the internet download data traffic for the seven days of the week are not equal which implies that there is a significant statistical difference between, at least, two of the daily data groups. A similar analysis for the internet upload data traffic as shown in Table 7 and Table 8 reveals that F (0.915171) < F crit (2.1245) with a P-value of 0.4839 (p>0.05) implying that there is no strong statistical daily variation for the upload data traffic. In Table 9, the total traffic data for each day of the week across the 51 weeks is presented for both download and the upload traffic data.

Table 5. Anova: Single Factor Summary for the Daily Download Traffic Data

Groups	Count	Sum	Average	Variance
DGb_SU	51	107839.8	2114.506	764101.90
DGb_M	51	131730.9	2582.959	710640.70
DGb_T	51	133336.7	2614.445	800103.50
DGb_W	51	122300.5	2398.049	764936.40
DGb_TH	51	136419.7	2674.896	1054440.00
DGb_F	51	133010.9	2608.057	1181285.00
DGb_S	51	111707.7	2190.347	950287.20

Table 6. Anova: Download Traffic Data for the Seven Days of the Week

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	15442662	6	2573777	2.8938	0.00912	2.12451
Within Groups	3.11E+08	350	889399.2			
Total	3.27E+08	356				

Table 7. Anova: Single Factor Summary for the Daily Upload Traffic Data

Groups	Count	Sum	Average	Variance
UGb_SU	51	19285.4	378.15	38394.20
UGb_M	51	22508.2	441.34	34175.98
UGb_T	51	22273.8	436.74	29847.75
UGb_W	51	20533.4	402.62	33453.66
UGb_TH	51	22069.9	432.74	35032.06
UGb_F	51	21932.3	430.05	48181.42
UGb_S	51	19706.4	386.40	43303.69

Table 8. Anova: Upload Traffic Data for the Seven Days of the Week

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	205826.3	6	34304.379	0.915171	0.4839	2.1245
Within Groups	13119438	350	37484.108			
Total	13325264	356				

Table 9. Total Traffic Data Ranking for each Day of the Week

	Download (GB)	Ranking	Upload (GB)	Ranking	Total (GB)	Ranking
Mon	131730.9	4	22508.2	1	154239.1	4
Tue	133336.7	2	22273.8	2	155610.5	2
Wed	122300.5	5	20533.4	5	142833.9	5
Thu	136419.7	1	22069.9	3	158489.6	1
Fri	133010.9	3	21932.3	4	154943.2	3
Sat	111707.7	6	19706.4	6	131414.1	6
Sun	107839.8	7	19285.4	7	127125.2	7

4. THE PARTIAL LEAST SQUARES MODEL

The concept of Structural Equation Modelling (SEM) has been around since the early 80's and overtime it has evolved in scope and areas of application for testing theories and concepts in different fields of study [24]. SEM methods are either covariance-based SEM or variance-based SEM [25]. In this study, Partial Least Squares (PLS) approach to SEM which is a variance-based method and can be applied for explanatory or predictive analysis [26], is applied in order to explain the variance in weekly internet data traffic for both the upload and download traffic as the dependent latent constructs across the 51 weeks of study, in terms of the IP traffic variation on each day of the week as the independent latent variables. The analysis therefore enables this study to identify and classify the significance of the contribution of each day of the week to the overall weekly IP traffic variance. To achieve this, a path analysis model is developed using Smart PLS3 application [27] as shown in Figure 17, and the performance of the model is evaluated using the R2, the significance of the path coefficient of the endogenous latent variables. This is obtained by running the PLS algorithm and by performing the bootstrapping procedure using a two-tailed test at 10 percent significance level (p<0.1).
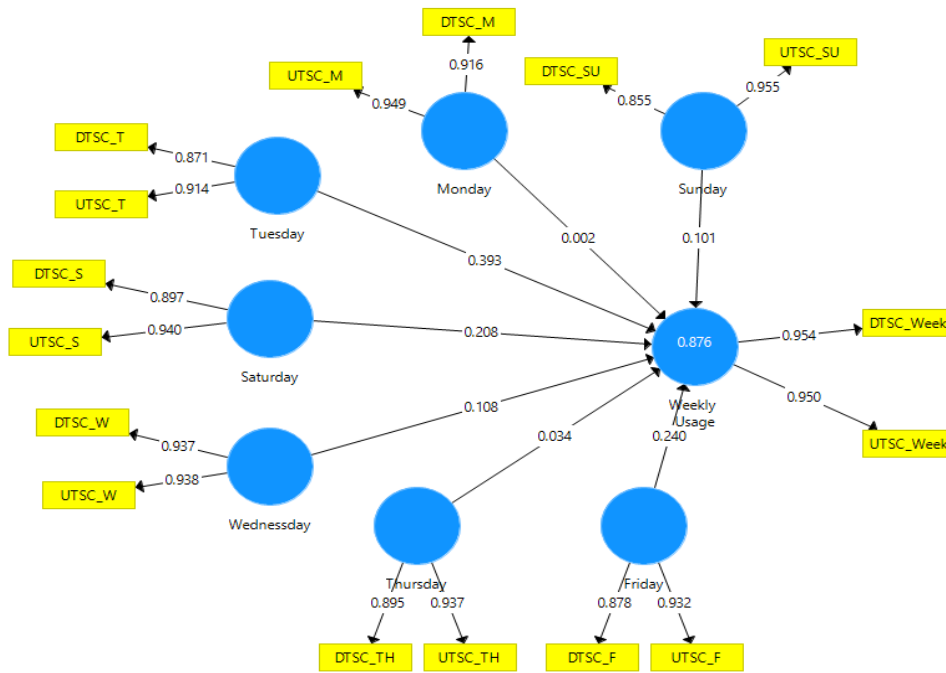
Figure 17. Smart PLS3-based path model

5.   RESULTS AND ANALYSIS

The performance of the PLS model was evaluated using 5000 bootstrap samples which were drawn randomly after replacement from the original study sample. The upload and download internet data traffic for each day were applied as reflective constructs to measure the causal effect of daily variations on weekly trends. In Table 10, the relationship of the outer loadings with the latent variables is presented at 95% confidence level. From Table 10, it is observed that all the p-values are zero (0) and the t-values are far greater than 1.65 which is the critical t-value at 10% significance level for a two- tailed test. Likewise, the outer loadings are all greater than 0.7, and all these factors confirm that all the reflective constructs are significant. In Table 11, the causal effect of each day of the week on the observed weekly variations in internet data traffic is modelled as hypotheses H1, H2, H3, H4, H5, H6 and H7 for each day of the week. Table 11 reveals that H1, H3 and H6 are the only significant hypotheses at 10% significance level with t-values > 1.65. This means that variations in the internet data traffic for both the upload and download traffic on Tuesdays (H6 = 3.609), Fridays (H1 = 1.833) and Saturdays (H3=1.798) in this particular order account for or explain significantly the overall weekly variance of the IP traffic at the university under study.

Table 10. The Relationship of the Outer Loadings with the Latent Variables

Relationship	Std. Beta	Std. Error	[t-value]*	P-values	95%CL LL	95%CL UL
DTSC_F <- Friday	0.872	0.056	15.703*	0	0.769	0.943
DTSC_M <- Monday	0.914	0.029	31.739*	0	0.861	0.952
DTSC_S <- Saturday	0.892	0.04	22.418*	0	0.818	0.944
DTSC_SU <- Sunday	0.848	0.064	13.421*	0	0.727	0.927
DTSC_T <- Tuesday	0.866	0.056	15.591*	0	0.762	0.936
DTSC_TH <- Thursday	0.891	0.041	21.748*	0	0.811	0.941
DTSC_W <- Wednesday	0.936	0.027	35.18*	0	0.887	0.971
DTSC_Week <- Weekly _Usage	0.955	0.015	65.041*	0	0.93	0.978
UTSC_F <- Friday	0.934	0.017	55.812*	0	0.903	0.958
UTSC_M <- Monday	0.95	0.011	87.7*	0	0.931	0.966
UTSC_S <- Saturday	0.942	0.012	81.042*	0	0.922	0.959
UTSC_SU <- Sunday	0.958	0.009	100.85*	0	0.943	0.974
UTSC_T <- Tuesday	0.917	0.017	55.317*	0	0.888	0.942
UTSC_TH <- Thursday	0.939	0.012	77.234*	0	0.918	0.957
UTSC_W <- Wednesday	0.937	0.025	37.493*	0	0.892	0.972
UTSC_Week <- Weekly _Usage	0.948	0.025	38.146*	0	0.9	0.979

* Significant at p<0.1CL LL - Confidence Limit Lower LimitCL UL - Confidence Limit Upper Limit

Table 11. The Relationship of the Latent Variables for Hypothesis Testing

Hypothesis	Relationship	Std. Beta	Std. Error	[t-value]*	P-values	95% CL LL	95% CL UL
H1	Friday -> Weekly _Usage	0.233	0.131	1.833*	0.067	0.028	0.451
H2	Monday -> Weekly _Usage	0.014	0.137	0.015	0.988	-0.213	0.238
H3	Saturday -> Weekly _Usage	0.204	0.115	1.798*	0.072	0.016	0.397
H4	Sunday -> Weekly _Usage	0.096	0.106	0.952	0.341	-0.08	0.268
H5	Thursday -> Weekly _Usage	0.041	0.145	0.232	0.816	-0.182	0.276
H6	Tuesday -> Weekly _Usage	0.376	0.109	3.609*	0	0.198	0.554
H7	Wednesday -> Weekly _Usage	0.117	0.137	0.783	0.434	-0.071	0.373

* Significant at p<0.1

In terms of the path coefficient of the inner model for the original sample, we have the following order for the seven hypotheses respectively: Tuesday (0.393) > Friday (0.24) > Saturday (0.208) > Wednesday (0.108) > Sunday (0.101) > Thursday (0.034) > Monday (0.002) as observed in Figure 17. The model achieved an overall $R2$ value of 0.876 as shown in Table 12, and R^2 values that are greater than 0.75 are deemed substantially significant in explaining the variance of the endogenous variable [24]. The discriminant validity check was carried out using the Fornell-Larcker Criterion and as shown by the matrix in Table 13, the model satisfies the criterion [28-29].

Model performance evaluation was carried out using construct reliability and validity as presented by Cronbach's Alpha, rho_A, Composite Reliability, and the Average Variance Extracted (AVE) results. Table 14 shows that the result of the Cronbach's Alpha is greater than the 0.7 threshold requirement for all the latent variables. In PLS approach to structural equation modelling, the Cronbach's Alpha is often said to give a conservative result. Hence, the Composite Reliability is often recommended [29, 30]. The Composite Reliability (CR) test does not assume that all the indicators are equally reliable unlike the Cronbach's alpha. For exploratory analysis, CR values between 0.6 and 0.7 are deemed satisfactory [24, 31], while CR values greater than 0.7 are preferred. As shown in Table 14, the CR values are all greeter than 0.7 which indicates a very good result. The convergent validity was also examined using the AVE, and the AVE of each of the latent variables is greater than the 0.50 threshold as shown in Table 14 indicating a satisfactory convergent validity. This implies that each of the latent variables adequately explains over 50% of its indicators' variance. Alos, the Rho_A reliability index is plotted and it shows that the index is greater than the 0.7 threshold for all the latent variables [25].

Table 12. Model Evaluation using R Square

	R Square
Weekly _Usage	0.876

Table 13. Fornell-Larcker Criterion

	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday	Weekly _Usage
Friday	0.905							
Monday	0.765	0.933						
Saturday	0.754	0.609	0.919					
Sunday	0.61	0.813	0.581	0.906				
Thursday	0.878	0.762	0.803	0.668	0.916			
Tuesday	0.685	0.762	0.567	0.634	0.682	0.893		
Wednesday	0.728	0.716	0.68	0.67	0.727	0.672	0.937	
Weekly _Usage	0.837	0.797	0.772	0.714	0.827	0.837	0.782	0.952

Table 14. Performance Evaluation using Construct Reliability and Validity

	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
Friday	0.784	0.827	0.901	0.82
Monday	0.853	0.888	0.931	0.87
Saturday	0.818	0.856	0.915	0.844
Sunday	0.797	0.979	0.902	0.821
Thursday	0.812	0.847	0.913	0.84
Tuesday	0.747	0.766	0.887	0.797
Wednesday	0.862	0.862	0.936	0.879
Weekly _Usage	0.897	0.898	0.951	0.907

6.   CONCLUSION

Internet usage behaviour greatly determines the volume of internet data traffic generated within a network. Heavy internet data traffic impacts the quality of service of the internet network due to the increased bandwidth requirement which may be beyond the network capacity, and as such, it may become imperative to identify and terminate undesirable applications triggering the increased traffic within the network. To achieve this, it is vital to have a means of monitoring the network's traffic and adequate tools for analysing the collected data. In this study, the internet traffic data of Covenant University for a year spanning over 51 weeks was logged to enable an extensive analysis of the data generated. The dataset was pre-processed and sorted based on the days of the week for both the internet upload and the internet download data traffic. Various statistical analyses were carried out and time series plots were generated to show trends and traffic variations on each day of the week from Monday to Sunday. For the total period of study, the highest total download occurred on Thursdays with a figure of 136419.7GB while the highest total upload occurred on Mondays with a figure of 22508.2GB. The dataset was further explored by carrying out analysis using PLS approach to SEM which is a variance-based method. The performance of the path model developed was evaluated using R2, the construct reliability and validity of the model. A R2 value of 0.876 was achieved which means that 87.6% of the variance in the weekly data traffic is explained by the daily variations of the upload and download traffic with Tuesdays having the highest significance and Mondays having the least significance in terms of the contribution of each day of the week in explaining the weekly IP traffic data variations. The SmartPLS3-based model satisfies all the vital model evaluation criteria for variance explanatory analysis.

REFERENCES

[1]   Zettelmeyer, F., F.S. Morton, and J. Silva-Risso, "How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data". *Journal of Marketing Research*, 2006. 43(2): p. 168-181.
[2]   McAfee, A., et al., "Big data: the management revolution". *Harvard business review*, 2012. 90(10): p. 60-68.
[3]   Titcomb, J., World's internet traffic to surpass one zettabyte in 2016, in The Telegraph. 2016.
[4]   Ashton, K., "That 'internet of things' thing". *RFID journal*, 2009. 22(7): p. 97-114.
[5]   Anand, N., et al., "Internet use patterns, internet addiction, and psychological distress among engineering university students: A study from India". *Indian journal of psychological medicine*, 2018. 40(5): p. 458.
[6]   Bahrainian, S.A., et al., "Relationship of Internet addiction with self-esteem and depression in university students." *Journal of preventive medicine and hygiene*, 2014. 55(3): p. 86.
[7]   Lin, M.-P., et al., "Prevalence of internet addiction and its risk and protective factors in a representative sample of senior high school students in Taiwan". *Journal of Adolescence*, 2018. 62: p. 38-46. DOI: https://doi.org/10.1016/j.adolescence.2017.11.004.
[8]   Adekitan, A.I., J. Abolade, and O. Shobayo, "Data mining approach for predicting the daily Internet data traffic of a smart university". *Journal of Big Data*, 2019. 6(1): p. 11. DOI: 10.1186/s40537-019-0176-5.
[9]   Schwefel, H.-P., I. Antonios, and L. Lipsky, "Understanding the relationship between network traffic correlation and queueing behavior: A review based on the N-Burst ON/OFF model". *Performance Evaluation*, 2017. 115: p. 68-91. DOI: https://doi.org/10.1016/j.peva.2017.07.002.
[10] Lee, Y., W. Kang, and H. Son. "An internet traffic analysis method with mapreduce". in *Network Operations and Management Symposium Workshops (NOMS Wksps)*, 2010 IEEE/IFIP. 2010. IEEE.
[11] Abdalla, B.M.A., et al., "Impact of packet inter-arrival time features for online peer-to-peer (P2P) classification". *International Journal of Electrical and Computer Engineering (IJECE)*, 2018. 8(4): pp. 2521-2530. DOI: 10.11591/ijece.v8i4.pp2521-2530.
[12] Kim, H., et al. "*Internet traffic classification demystified: myths, caveats, and the best practices*". in Proceedings of the 2008 ACM CoNEXT conference. 2008. ACM.
[13] Moore, A.W. and K. Papagiannaki. "*Toward the accurate identification of network applications*". in International Workshop on Passive and Active Network Measurement. 2005. Springer.
[14] Sen, S., O. Spatscheck, and D. Wang. Accurate, "*scalable in-network identification of p2p traffic using application signatures*. in Proceedings of the 13th international conference on World Wide Web. 2004. ACM.
[15] Karagiannis, T., K. Papagiannaki, and M. Faloutsos. "BLINC: multilevel traffic classification in the dark". in *ACM SIGCOMM computer communication review*. 2005. ACM.
[16] Pang, S., et al., "Malicious Events Grouping via Behavior Based Darknet Traffic Flow Analysis". *Wireless Personal Communications*, 2017. 96(4): p. 5335-5353. DOI: 10.1007/s11277-016-3744-4.
[17] Abdalla, B.M.A., et al. "*Multi-stage Feature Selection for On-Line Flow Peer-to-Peer Traffic Identification*". in *Asian Simulation Conference*. 2017. Springer.

[18] Anderson, D., W.S. Cleveland, and B. Xi, "Multifractal and Gaussian fractional sum–difference models for Internet traffic". *Performance Evaluation*, 2017. 107: p. 1-33. DOI: https://doi.org/10.1016/j.peva.2016.11.001.

[19] Shi, H., et al., "Efficient and robust feature extraction and selection for traffic classification". *Computer Networks*, 2017. 119: p. 1-16. DOI: 10.1016/j.comnet.2017.03.011.

[20] Al-Turjman, F., "Information-centric framework for the Internet of Things (IoT): Traffic modeling & optimization". *Future Generation Computer Systems*, 2018. 80: p. 63-75.

[21] Adeyemi, O.J., et al., "Exploration of daily Internet data traffic generated in a smart university campus". *Data in Brief*, 2018. 20: p. 30-52.

[22] Gadhavi, L.J. and M.D. Bhavsar, "Prediction based efficient resource provisioning and its impact on QOS parameters in the cloud environment". Vol. 8. 2018. 5359-5370.

[23] Audah, L., Z. Sun, and H. Cruickshank, "QoS based admission control using multipath scheduler for IP over satellite networks". *International Journal of Electrical and Computer Engineering (IJECE)*, 2017. 7(6): p. 2958-2969. DOI: 10.11591/ijece.v7i6.pp2958-2969.

[24] Hair, J.F., C.M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet". *Journal of Marketing theory and Practice*, 2011. 19(2): p. 139-152.

[25] Dijkstra, T.K. and J. Henseler, "Consistent and asymptotically normal PLS estimators for linear structural equations". *Computational Statistics & Data Analysis*, 2015. 81: p. 10-23. DOI: https://doi.org/10.1016/j.csda.2014.07.008.

[26] Richter, N.F., et al., "European management research using partial least squares structural equation modeling (PLS-SEM)". *European Management Journal*, 2016. 34(6): p. 589-597. DOI: https://doi.org/10.1016/j.emj.2016.08.001.

[27] Ringle, C.M., S. Wende, and J.-M. Becker, SmartPLS 3. "Boenningstedt: SmartPLS GmbH", http://www.smartpls.com, 2015.

[28] Fornell, C. and D.F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error". *Journal of marketing research*, 1981: p. 39-50.

[29] Wong, K.K.-K., Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. Marketing Bulletin, 2013. 24(1): p. 1-32.

[30] Hair, J.F., et al., The Use of Partial Least Squares Structural Equation Modeling in Strategic Management Research: A Review of Past Practices and Recommendations for Future Applications. Long Range Planning, 2012. 45(5): p. 320-340. DOI: https://doi.org/10.1016/j.lrp.2012.09.008.

[31] Nunnally, J.C. and I. Bernstein, Psychometric Theory. 3rd ed. 1994, New York: McGraw-Hill.