Analysis of IPv4 vs IPv6 Traffic in US

Mahmood-ul-Hassan [1,4], Muhammad Amir Khan [2,3,4], Khalid Mahmood[2, 4], Ansar Munir Shah[2, 4]
[1]Al-Jouf University, College of Science and Arts, Tabarjal Saudi Arabia
[2] College of Science and Arts, Department of Information Systems, Mahayel Aseer
King Khalid University, Abha Saudi Arabia
[3]Department of Electrical Engineering,
COMSATS Institute of Information Technology, University Road, Tobe Camp, 22060, Abbottabad, Pakistan
[4]Department of Computer Science
IIC University of Technology Cambodia

***Abstract*—It is still an accepted assumption that internet traffic is dominated by IPv4. However, due to introduction of modern technologies and concepts like Internet of Things (IoT) IPv6 has become the essential element. So keeping in mind the advancements in new technologies and introduced concepts to update the adoption of IPv6 on the internet. We want to find out what percentage of the IPv6 traffic is present in the Internet from last 6 years (2008-14) and to obtain the adoption curve of IPv6 native traffic by years to analyze if it is slow or fast. Also what are factors, constraints and limitations involve in the adoption of IPv6. Therefore, we have taken two data sets from the Caida website. The dataset belongs to OC-48 and OC-192 links from two data center of Equinix located at Chicago and Sanjose in the US. Finally compare the final curve with infograph of World IPv6 Launch to know how realistic it is and applied Linear Prediction techniques to see the future trend of the dataset obtained from the US population.**

Keywords—IPv4; IPv6; Mobile node; IP Traffic; IID Testing

I. INTRODUCTION

Everything on Internet is identified with the IP and IoT has made the new paradigm of the Internet by recognising the devices with IP even if it is your camera or TV. And old IPv4 is not at all capable to provide IP addresses to such increasing number of connected devices which is more in number than the users themselves. According to cisco there will be 25 billion devices by 2015 and 50 billion by 2020 which is more than the population. Also from World IPv6 Launch there will be 20 billion devices online by the end of 2016. So we want to know how far the dreams have become the realities.

Internet was designed as an experiment in 1973 and launched in 1983. However, since 1981 IPv4 has been the de facto standard in the world of Internet routing. In the early 1980s, the benefits of IPv4 were unambiguous and that is why its adoption rates increased enormously. Later on in early 1990s, due to accrued demand for Internet end users, the adoption was grown exponentially, which has lifted concerns within the Internet Engineering Task Force (IETF) and other ISPs. At this point, we realized that a point will come when all of the IPv4 address space will be out of capacity. The questionable, 2^{32} i.e. 4.3 billion largest address space, now seemed to be finite. This caused the necessity for superior technology that would be able to cater the need for a larger address space and also allow for improved services. [1]

Prediction done by Cisco in September 2005, that the unadvertised address space of IPv4 will be exhausted in more or less 10 years [2]. US IPv4 policies revision began in 2008 to see what arrangements can be done to utilize the used pool of IPv4 [3]. But, still the last block was sold in 2011, there is some address space with major Internet service providers that have not yet been used, for instance according to an article publicized by MIT; US still has 1.5 billion IPv4 addresses out of 3.7 billion IPv4 addresses. This is the main reason, the US has a lot of these unused IPv4 address space and this caused a slow transformation to IPv6. [4]

Researchers are of the view, according to statistical analysis that by the end of 2012, there will be a need for IP addresses to connect approximately 3.6 billion devices throughout the world [5]. Several of these devices due to technology advancements would need multiple IP addresses to perform functions accordingly. Hence, 4.3 billion IPv4 address space is even less than sufficient for the whole networking world to work efficiently. Moreover, scalibilty of Network Address Translation (NAT), to translate multiple private addresses is limited. It provides with a disadvantage of limiting a peer to peer communication session with VOIP (Voice Over Internet Protocol) and IPsec (IP security) [6]. Also, some researchers have found that NAT transversal cost for the vendor of the application is so high to reach 500 million dollars per year [7,10]. However, using the IPv6 which allows peer to peer communication without the need of the intermediate server involving no cost.

Route aggregation is another benefit provided by IPv6 in an organised manner to analyze the network with geographically significance features. Summurization of routes can be done in a sequential manner. Such ease, reduces the total size of the routing information or table and hence casue s the reduction in the load throughout the network. Memory usage is minimized to reduced rates if such an efficient IP proocol is to be utilized. In this way, less network resources are consumed in an efficient way with minimized associated cost.

Day by day, the availableness of the IPv4 address space is detractive so internet is in need for an enhanced technology for its survival. IPv6 has thousand trillions of a address space that would be sufficient enough to cater the exponentially increasing advancements of multiple IP usage technology

[8,9]. However, it is not the case that IPv6 only provides theadvantage of enormous address space. Mobility is one of the important feature of the IPv6 that helps in providing more than one IP addresses to multiple network interface[11]. By netwrok interface we mean to say if let say we have a smart phone, then it has to network interface one is the WiFi and the other is the GSM internet interface. Seamless roaming can be achieved through the use of IPv6 as comapred to IPv4. Moreover, both vendor and operator of telecom companies has the big benefit of utilizing the IPv6. Wireless Sensors devices used in BCI is significantly a decent example of IPv6 mobility.

Our contribution in this paper starts from the description of the some important features of IPv6 protocol. Then we analyse the dataset taken from the Caida website to know the percentage of the IPv6 and IPv4 traffic in the data networks of Chicago and Sanjose. In the final section of the article we have analysed and applied the Linear Prediction tools on the same dataset for further analysis.

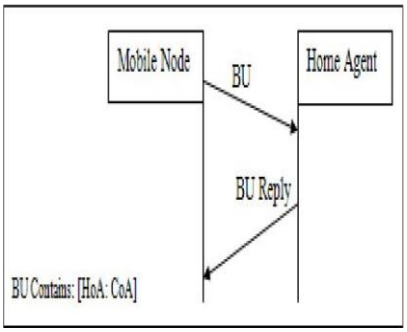


Fig. 1. Mobility Handshake between MN and HA using BU Messages

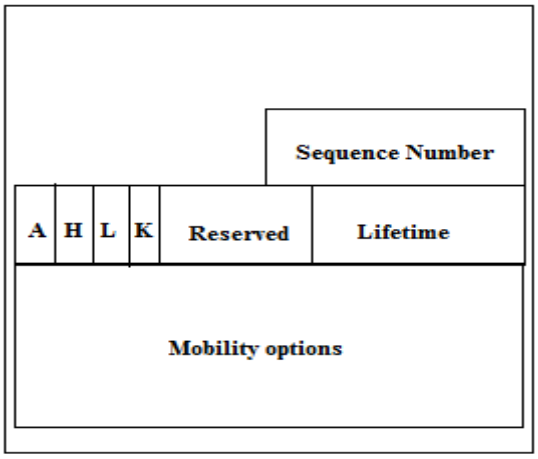


Fig. 2. BU Message send by a MN to HA or Correspondence Node

II. Mobility in IPv6

In this section, we have assumed that a mobile device is IPv6 enabled and has moved to some other network. Now, we will describe that what steps are involved in its mobility procedure.

Registration in Home Agent (HA) Signaling messages: Information included and why it is needed.

Home Agent Registration: Once a Mobile Node (MN) has completed agent discovery, it knows whether it is on its home network or a foreign network. If on its home network it communicates as a regular IP device, but if on a foreign network it must activate Mobile IP. This requires that it communicate with its home agent so information and instructions can be exchanged between the two. This process is called Home Agent Registration, or more simply, just registration. Describe through Figure 1.

Included Information: The BU (Binding Update) message is used and send by MN to notify the HA or the correspondent node of the binding information of CoA and HoA of the MN. A MN sends the BU message with its CoA and its HoA whenever it changes the point of attachment to internet and changes its CoA. The receiving node after receiving the BU message will create an entry to keep the binding information that also includes CoA and HoA. More information included in the single BU message is described below as seen in the Figure 2.

Sequence Number: It contains a sequence number (16 bit variable) for BU message to avoid replay attack. Also if present sequence number in the database is smaller then it will be removed with new greater sequence number having the recent update information.

Lifetime: This field specifies the proposed lifetime of the binding information included. When BU message is used for HA registration, the value must not be greater than the remaining lifetime of either HoA or the CoA of the MN. The value is in units of 4 s.

Mobility Options: BU message may have following mobility options:

- The Binding Authorization Data Option
- The Nonce Indices Option
- The Alternate CoA Option

Why this Information is Included:

This information is needed in order to tell the HA about the current location or more correctly current point of attachment of the MN with the respective network. This current point of MN refers to the temporary network in which MN is assigned a CoA to act as Locator. Whereas the permanent address of MN is HoA that will continue to act as Identifier. In this way, HA knows that MN is in other network and HA has to send the packets to MN using the CoA.

Security of the procedure: Security in this procedure is to have a unique key which is only known to MN and HA. This key is to be encrypted by using some cryptographic algorithms. The BU message must be secured by IPSec. IPSec stands for Internet Protocol security.

This protocol (IPSec) makes the virtual tunnel of each packet by encryption.

It ensures the confidentiality and integrity of each packet by ensuring a Virtual Private Network between nodes.

In the network given above in Figure 3, if Node1 wants to send data to MN which has been moved to a new point of attachment with the Router Rb, then it has two routes to use. Keeping in mind the cost of each link is 1.

Route1: Node1-Ra-HA-Rb-MN with Cost:4 and

Route2: Node1-Ra-Rb-MN with Cost:3.

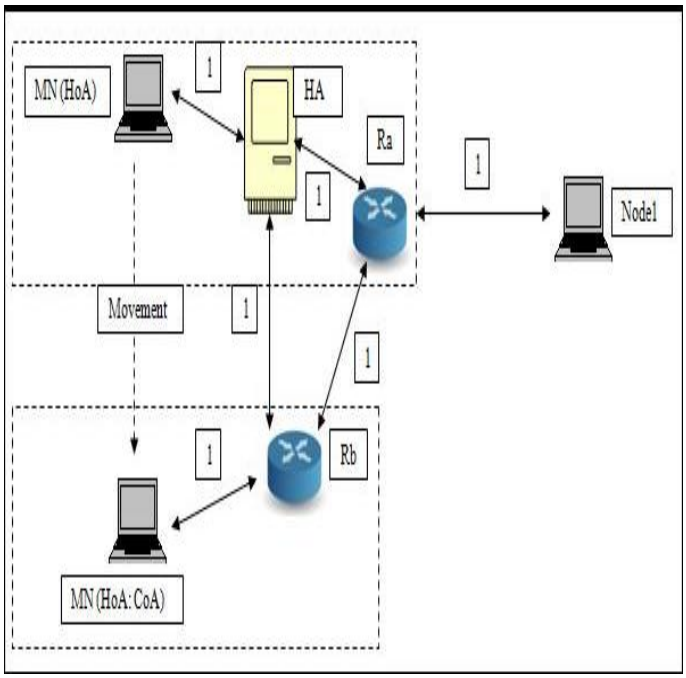


Fig. 3. Mobility in IPv6

Due to the mobility procedure adopted in IPv6, Route2 can not be utilized because it is not suitable due to security concerns and other limitations. However, Route2 has minimum cost then Route1. Therefore, traffic should only go through the HA to the MN which is actually the traffic exchange withoutroute optimization. (A tunnel is established between the Home Agent and a reachable point for the Mobile Node in the foreign network.)

How the HA gets the packets sent to the HoA:

Consider the Figure 3 again, If Node1 wants to send data to MN which is in the temporary Network and having HoA and CoA respectively. Then it will send the data through HA using the IPv6 addresses.

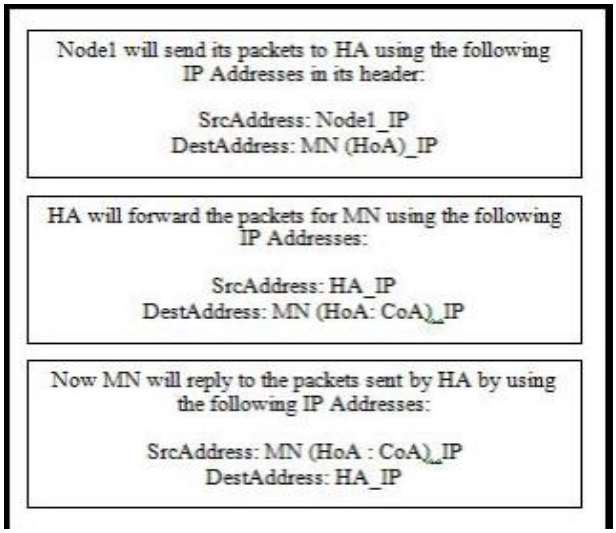


Fig. 4. Packets sent to HoA

III. Analysis of IP Traffic

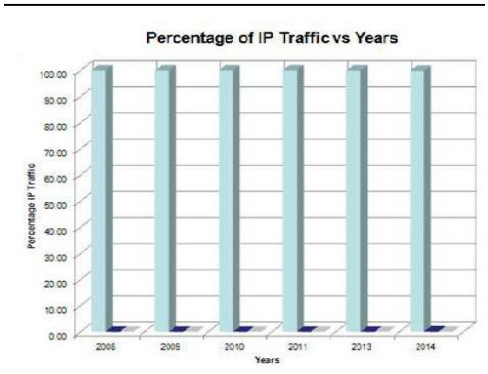
Before we start the analysis of the dataset taken from the Caida website, it is important to know and understand the terms we are going to use in our analysis which are as follows:

Allocation: The Entire IPv4 address space is maintained by IANA (Internet Assigned Number Authority)

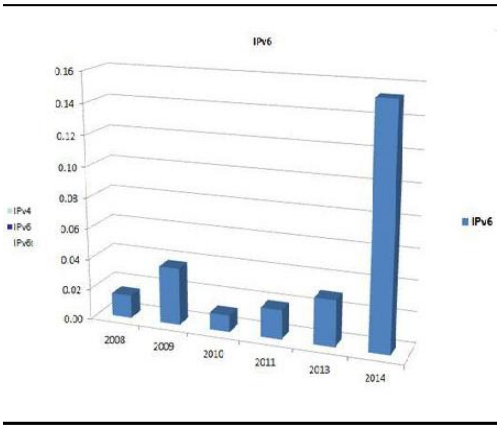
IANA allocates blocks of addresses to 5 Regional Internets Registries (RIRs); US-Europe-Asia-LatinAmerica-Africa.

Assignment: The RIRs assigns IPv4 addresses to several ISPs or Network operators.

Advertisement: It denotes the part of assigned IPv4 addresses that are in real use in traffic.



(a)



(b)

Fig. 5. (a). Percentage of IP Traffic vs Years in Chicago monitor (b).IPv4 vs IPv6

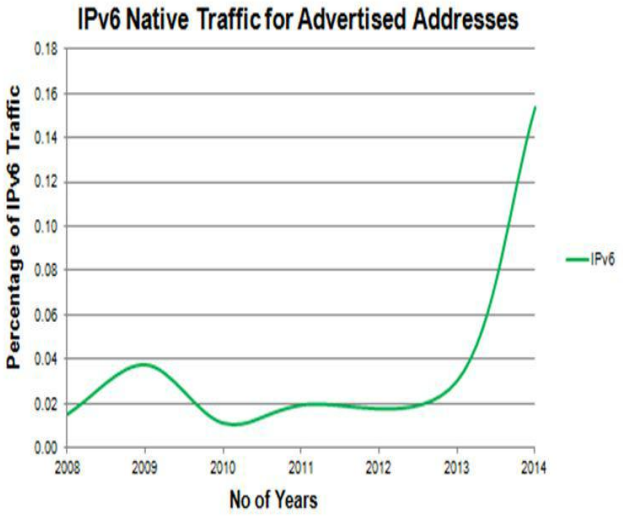
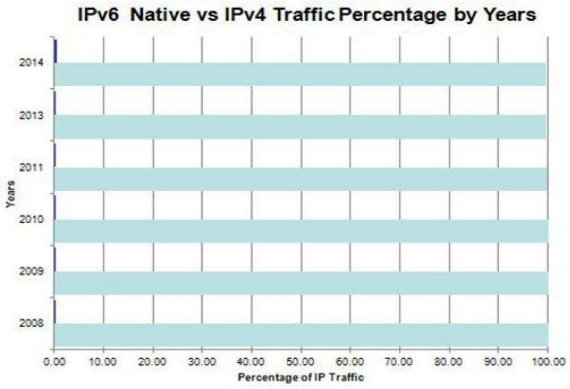


Fig. 6. Percentage Growth of IPv6 Native Traffic by Years in Chicago

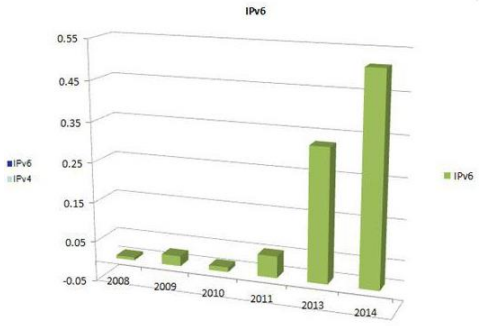
In Figure 5, the data set belongs to the Chicago monitor and as I have mentioned in the first presentation that Chicago is the third largest city in the US after NY and LA with 9.5 million people. So analysis of this data is very important as it belongs to the huge population. IPv4, IPv6 and IPv6t (tunnel) traffic shown in the bar graph. As you have seen that over the last 6 years IPv4 is maintain the IP traffic percentage of more than 98-99 percent on average. However, if you analyse the Figure 6 as shown below you will come to know that overall progress of IPv6 traffic over six years is in increasing manner. Although increasing manner is shown but still the percentage is not more than 0.16 at the end of the Month of April 2014. IPv6 traffic tends to increase since the July 2012 as shown in the graph. And the affective reason for this is the event WorldIPv6Launch that has happened in the same time span. And in this launch was conducted by Internet Society in which major ISPs, AT and T, Google, Akamai and many more united to redefine the global internet by enabling all of their devices and equipment to use IPv6 permanently. More and more big companies are still joining this launch up till now and that is the reason that IPv6 traffic is continuously increasing.

TABLE I. PERCENTAGE IP TRAFFIC BY YEARS FROM CAIDA DATASET [1]

Chicago	Sanjose
IPv4:99.84 IPv6:0.15	IPv4:99.49 IPv6:0.51
IPv4:99.94 IPv6:0.03	IPv4:99.69 IPv6:0.31
IPv4:99.97 IPv6:0.02	IPv4:99.94 IPv6:0.64
IPv4:99.98 IPv6:0.01	IPv4:99.99 IPv6:0.01
IPv4:99.96 IPv6:0.04	IPv4:99.98 IPv6:0.02
IPv4:99.98 IPv6:0.01	IPv4:99.99 IPv6:0.01



(a)



(b)

Fig. 7. (a) Percentage of IP Traffic vs Years in Sanjose monitor (b) Percentage of IPv4 and IPv6 Traffic

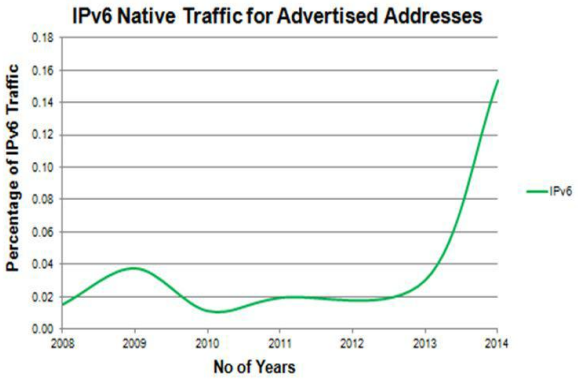


Fig. 8. Percentage Growth of IPv6 Native Traffic by Years in Sanjose

IV. RESULTS AND DISCUSSION

Reasons for Minimal IPv6 Traffic in the Backbone

As our data set belongs to the data centre in US so I am

only going to explain the specific reasons belongs to the region and some over all general issues in the IPv6 slow growth.

A. IPv4 Unadvertised Block

Referring again to my second presentation in which I have shown some statistics from the ARIN website also shown in Figure 7. If you see the bar from the ARIN which is responsible for allocation of IP blocks in US, there are more than 30 percent of the IPv4 addresses that are still unadvertised at the end of 2012 onwards. According to Colorado State University this percentage of unadvertised block was 42 percent i.e. 1.5 billion of addresses. So the first reason of very little IPv6 traffic is that in US IPv4 addresses are still unadvertised.

B. Transition Cost

It is one of the important issue raised and discussed by many analysts while talking about the migration to IPv6. There is a report issued by Arbor Networks in 2012 in which US Department of Commerce says that to implement an end to end IPv6 network the estimated cost is 25billion dollars for all the ISPs in the US.

C. Governmental Initiative

There are no official initiatives reported until 2012 from US government for the adoption of IPv6. Might be because of availability of IPv4 blocks or may be some other reasons as well. US government is not very active as of European, Japan and China. Japan has taken his lead in the adoption since 1999 and very next year IPv6 RFC 2460 been published. There government has taken a solid initiative by the name e-Japan initiative for the adoption of IPv6 since 2001 and declared as national mission. In the same manner China government has taken the initiative in the same year with the name China Next Generation Internet. So the point to make here is that US government should take some initiatives to announce some funds and technical support for ISPs to provide end to end IPv6 networks if they want to be in race of IPv6 adoption.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

D. Linear Prediction Analysis

The first step that was taken in applying linear analysis and prediction to our dataset was to remove any trends, seasonal components, etc. associated with the dataset. In particular, the data used for this study does, in fact, have a trend associated with it. To estimate the trend, a polyfit function of order 1 was used. The trend was then removed from the data. After doing so, it needed to be determined whether or not the data was independently and identically distributed (IID). There are five hypotheses that can be tested to determine whether a process is IID. If the majority of the hypotheses are rejected, the data cannot be considered to be IID. Section 4.a, below, describes the process in which the dataset is tested to determine whether or not it is IID.

E. IID Testing

The first of hypothesis is the Sample ACF. An autocorrelation function is performed on the data. Then, if three or more stems in the autocorrelation have a value greater than 1.96/(n) (approximately 0.27), the process is said to not be IID. The second hypothesis that is tested is the Portmanteau test which takes a single statistic Q which is defined in the equation below.

$$Q = n\sum_1 \epsilon H\, p^2 \ldots\ldots\ldots (1)$$

If the value Q is greater than the inverse of the chi-square distribution with h degrees of freedom, the hypothesis is rejected. This is because a higher Q signifies a higher correlation. The third hypothesis that is employed is the Turning Point Test. Where C is the number of turning points in the data, the following two equations show the average value for C and its variance for any IID sequence with length n.

$$\mu c = \frac{2}{3}(n-2) \ldots\ldots\ldots\ldots(2)$$

$$\theta_C^2 = \frac{16n-29}{90} \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

The hypothesis is rejected if $\frac{|C-\mu c|}{c} > 1.96$. The fourth hypothesis, the Difference-Sign Test, is similar to the third hypothesis except that instead of counting a value C, a new value S is counted. S corresponds to the number of instances where there is an increase in the value of the data at time index i, compared to the value of the data at time index i-1. (i.e. $y_i > y_{(i-1)}$) For a sequence to be considered IID, the following two equations must hold. $\mu c = \frac{1}{2}(n-1)$ and $\theta_s^2 = \frac{n+1}{12}$If $\frac{|S-\mu s|}{\theta s} > 1.96$, the hypothesis is rejected. The fifth and final hypothesis to be tested is called the Rank Test. This hypothesis is very similar to the Difference-Sign Test. The difference is that instead of finding a value S (the number of times when there is an increase from one time instance to the next time, i-1, instance, i), the test finds a value P, where P is the number of times that there is an increase from the value at one time index to another value at some future time index. (i.e. $y_i > y_{(i-1)}$) For an IID process, the following two equations give the statistical mean and variance for P. $\mu_{p=\frac{1}{4n(n-1)}}$

$\theta_p^2 = \frac{n(n-1)(2n+5)}{8}$the hypothesis is rejected. The following is a section from the output of the MATLAB program, displaying the results of each IID hypothesis test.

IID hypothesis not rejected: Sample ACF IID hypothesis not rejected: Portmanteau IID hypothesis rejected: Turning Point

IID hypothesis rejected: Difference-Sign IID hypothesis not rejected: Rank

F. Autocorrelation and Partial Autocorrelation Func-tions

With the trend removed from the dataset, the autocorrelation and partial autocorrelation functions are used to determine which type of process the data can be modelled after. The options available are as fol-lows: autoregressive

(AR), moving-average (MA), and autoregressive-moving-average (ARMA). Although, it is possible to use more than one of these models for making predictions, it is better to search for and identify the one that is suited best to the dataset of interest. Once, we have settled on a choice, we can then employ that model to make a prediction about the adoption of IPv6 based on the traffic of IPv4. Also of note, it can be observed that the readings do appear to represent an IID process, despite the fact that one would expect a dependency on a number of factors such as time, transition cost, equipment cost, government policies, etc. The figure 9 is the result of the autocorrelation being performed on the data. As seen in the figure, the autocorrelation slowly decays to zero. After performing the partial autocorrelation function to the dataset, the following figure 10 was produced.

As demonstrated by Figure 10, it is suggested there exists a high correlation in the dataset between distant entries. However, this is not the case since the dataset is relatively small (53 entries) and, therefore, the most reliable statistics come from the first 12 lags. For this reason, the partial autocorrelation statistics after lag 20 have been removed, resulting in Figure 11 below. From both Figures 10 and 11, it can be determined that the appropriate model to use for making predictions based on this dataset is the AR (autoregressive) model of order AR Processes: We are given two procedures for predicting AR processes: the Yule-Walker method and the Burg method. For the analysis performed, both methods were employed and compared. Discussion of the comparison can be found in the following section.

Error Measurement and Comparison: To compare the Yule-Walker and the Burg method, the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) of each method were taken. To accomplish this, the predicted values of each technique were compared against the test values (the last three values) of the

- MSE in Yule-Walker = 0.0022998
- MSE in Burg = 0.0022978
- MAPE in Yule-Walker = 0.024427 percent

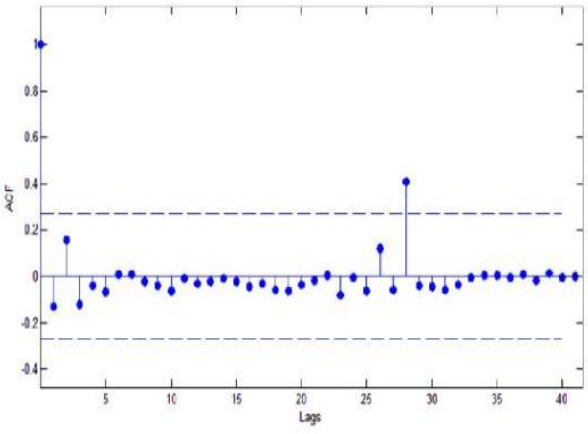


Fig. 9. Autocorrelation function

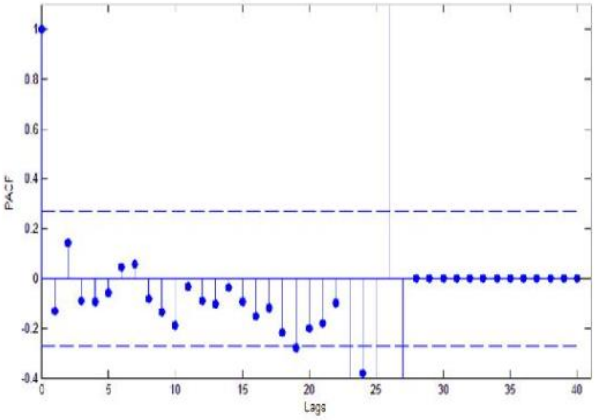


Fig. 10. Partial Autocorrelation function

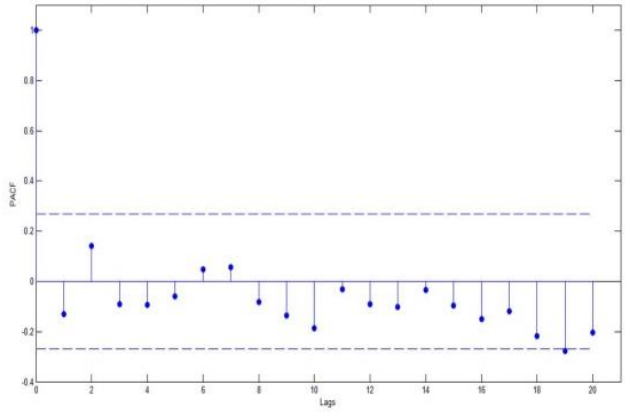


Fig. 11. Partial Autocorrelation function

V. CONCLUSION

As seen from these results, the Burg method yields a slightly lower error for both statistics. However, the difference is almost negligible, indicating that either method would be suitable from linear prediction. Additionally, both statistics (MAPE and MSE) are equally useful since the data is being compared does not come from different datasets and, therefore, the scale is not an issue in this case. Also from the analysis of the IP traffic, it is very clear that IPv6 traffic growth rate is increasing year by year. There is a constant increasing manner shown in the traffic, however it is the fact that IPv4 still the dominant protocol of the internet traffic today.

REFERENCES

[1] M. Gomez et al., IPv6: A New Business Opportunity, 2008

[2] T. Hain, A Pragmatic Report on IPv4 Address Space Consumption, The Internet Protocol Journal, vol. 8, pp.2-19, Sep.2005

[3] http://www.icann.org/en/news/announcements/announcement-2-05feb09-en.htm

[4] M. Orcutt, End of the Internet as we know it, June 24, 2011. http://www.technologyreview.com/printer_friendly_article.aspx?id=37859

[5] C. Marsan, IPv6 vs. Carrier-grade NAT, June 7, 2010. http://www.networkworld.com/news/2010/060710-tech-argument-ipv6-nat.html

[6] S. Shiuh-Pyng, H. Fu-Shen, H. Yu-Lun, L. Jia-Ning , ”Network address translators: effects on security protocols and applications in the TCP/IP stack,” Internet Computing, IEEE , vol.4, no.6, pp.42-49, Nov/Dec 2000; doi: 10.1109/4236.895015

[7] M. P. Gallaher and B. Rowe, IPv6 Economic Impact Assessment, NIST, October, 2005.

[8] N. Jauhari, Advantages of IPv6 - The next generation Internet, http://linuxpoison.blogspot.com/2009/01/advantages-of-ipv6-next-generation.html

[9] M. Nikkhah, R. Guerin, Y. Lee, and R. Woundy, Assessing IPv6 Through Web Access: A Measurement Study and Its Findings, in Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies, 2011.

[10] N. Sarrar, G. Maier, B. Ager, R. Sommer, and S. Uhlig, Investigating IPv6 Traffic - What Happened at the World IPv6 Day? in Proceedings of the 13th International Conference on Passive and Active Network Measurement, 2012.

[11] E. Karpilovsky, A. Gerber, D. Pei, J. Rexford, and A. Shaikh, Quantifying the Extent of IPv6 Deployment, in Proceedings of the 10th International Conference on Passive and Active Network Measurement,2009.