Network Traffic Observations in Data Centers and Forecasting Techniques for Resource Utilization

Samar Raza Talpur
UCD
Email: samar.talpur@ucdconnect.ie

***Abstract*—Data Center have decisive role in online corporate world. At present these data centers are not only house of servers, switches and routers, but to provide speedy services to vendors and uninterrupted network connectivity to client's websites. The importance of managing data center traffic, forecasting techniques for resource utilization are more challenging in effective data centers. This paper focuses to observe and analyses the live traffic in real-world data center networks, apply forecasting techniques for traffic optimization and proper resource utilization in data centers. We propose forecasting model for data centers to predict and estimate proper bandwidth utilization in real-world situations. Our model can be useful and identify the upcoming network trend, bandwidth demand and the essential growth to predict the futuristic assessment. The paper is based on day to day network traffic engineering and observation through exponential smoothing method of time series, the approach optimizes the upcoming network traffic for data centers.**

Keywords*—Network engineering; traffic optimization; exponential smoothing method; forecasting techniques

I. Introduction

World population is growing day by day, as a result of that, additional internet users are also increasing at an incredible pace. Approximately, 4.5 children are born in a second across the world. On the other hand, nearly 8 new internet users in a second are added on the planet. Furthermore, the data centers are also growing in a large number. In contrast of that, a huge volume of network traffic is increasing due to the growth of various services [1], [2], [3].

The network performance totally rely on accuracy of traffic engineering and predictions. Almost 40% population of the globe has access to the internet and as a result, flow of traffic between data centers and end-users is increasing day-to-day. Once we look into past, it was just 1% in 1995, while the growth of internet users increased by ten times from 1999 to 2013. Subsequently, one billion of users touched in 2005, two billion in 2010 and three billion reached in 2014 [4]. Annual rapid jump of IP traffic is expected from terabytes to zettabytes. According to Cisco statistics, the yearly worldwide IP traffic for data centers will exceed from 3.4 zettabytes to 10.4 zettabytes at the end of 2019 [5].

The contemporary data centers hold tens of thousands of servers to perform the processing for multiple internet business applications. These data centers are increasingly using it for the fixation of traffic demand in order to streamline management and make better use of resources.

In spite of advancements made in the area of network usage monitoring and management, there still persist a number of key limitations that need to be overcome. Based on the research, some key challenges faced by operators and data center service providers are described as under:

1) Primary goal of data center designing is to sustain failures of network equipment and also deliver satisfactory bandwidth. However, network maintenance and electric power often make tens to thousands of servers' offline when network crashes and congestion in network can also increase network latency [6].
2) Installation of reliable network traffic predication models can spot problems before they occur. In order to deliver better quality of services to the clients, service providers can take advantage of it and handle unexpected bursty traffic for effective performance. It is mandatory to propose a well-organized management for proactive network management systems as there is substantial demand of distributed computing systems. The administrative tasks may facilitate the process of mechanism to monitor and control the activities of existing network and also synchronize the activities already executed in a network [7].
3) Service providers are struggling to meet demand and supply of bandwidth to clients' requirement in order to overcome end user network traffic requirements from existing and emerging data centers. Even though service providers have tremendous bandwidth to offer but the intention for demand and supply is still unsatisfied, typically the requirement is as much as high though bottleneck traffic happens [8].
4) Most of the time, the organizations have certain bureaucratic layers to immediately purchase exclusive and limited edition of hardware for the purpose of procurement and inventory of IT assets. However, the specialized hardware needs to preorder and reserve for purchasing and upgrading before time and cost can be reduced accordingly.
5) Network Management also works to gather supply and demand information from business entities for the purpose of purchasing IT equipment. Therefore, it is hard to collect the information for demand and supply from concerned entities. The collected information can be analyzed to overcome the upcoming purchase of business entities and provided shared data can conduct reverse cataloging for assets management, network management and procure of equipment [9].
6) However, an increase of unexpected traffic consumption causes interruption and delay in sensitive applications, and also affects the performance of networks to run smoothly. Meanwhile, planning and upgrading

to buy more bandwidth at the eleventh hour is more likely challenging and also problematic for both client and customers. Similarly if the bandwidth consumption is lower than actual load of contract, then the reduction of cost can be bargain accordingly.

7) Employing additional IT professionals (if needed) can be planed before the time, usually hiring new staff takes time in routine; this process required enormous quantity of work to be done by HR, often multiple rounds of interviews for suitable candidate needs to be done. Moreover rostering and scheduling for existing employees can also be done effortlessly.
8) Sometimes the unusual network activities are hectic and creates problem for management, with forecasting techniques, these types of infrequent activities can be addressed ahead of its time and without any stress.

All the above-mentioned points are considered as the key challenges in the growth of conventional data centers faced by operators, decision makers and administrators. However at the heart of evaluations the upcoming data centers are looking forward with lower energy costs, better economical, secure and reliable. The data centers should not be considered as an IT asset but these data centers are ultimately business decisions. The strategies taken in data centers impact and transform to the change in business strategies.

When it comes to security, data centers are adopting series of possible fundamentals securities to make efficient and reliable. The journey from physical premises to virtual environment of onward network connectivity is improving as per required. The protection of concrete structure may ideally be protected by numerous disaster recovery prevention and backup strategies at the time of earthquake, power, fire and other natural disasters. In addition to access in data centers the strict procedure and policies must follow while accessing the data centers. Firewalls in data center are uppermost defence border to handle security, these firewalls are hold client's private information and online transactions, predefined written policies to allow and deny the packets. The recognized favourite packets are allowed and rest strange requests are discarded. Generally traditional firewalls hold fewer ports to attach with servers, afterward if one of the port can constantly hit, then the firewall be hanged and slow down responding. Attackers can have more chance to pass through the network.

Besides, a general architecture regarding flow of traffic in data centers is presented in Fig. 1. Conceptually, the flow of traffic is categorized in three directions. The global data center traffic by destination report shows the bifurcation of traffic flow in Fig. 2 [11], [12], [13].

II. Network Traffic in Data Center Networks

The classification of global traffic and the volume of traffic flow from source to destination is discussed below. Following are some key points to discuss the traffic among data centers:

1) *East-West Traffic*: Maximum and noticeable volume of traffic is produced within data centers. This heavily flow pushed from east-west to north-south compartment. East-West traffic considers to move horizontally and it also facilitates to communicate with application servers i.e. server to server and VM to VM communication. Usually, this type of traffic is consumed by social networking where clients updates their status by one click. The volume of east-west traffic is greater than all other traffics and mostly remains within the boundaries of data centers.
2) *North-South*: The North-South traffic usually moves vertically and travels between data centers to users in order to invoke video steaming and games. In addition to it, north-south traffic circulates between data centers to data centers.
3) *Internal Traffic*: The volume of this sort of traffic is slow in rising and the flow pushed by upper bound of section (north-sound). The traffic flow travels between data centers to data centers, likewise between public to private and vice-versa.

A. Framework for Forecasting and Plannings

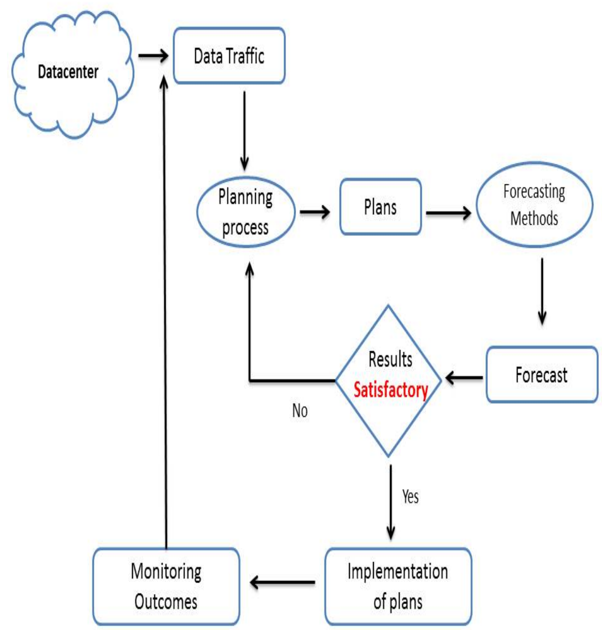

Fig. 3. Framework for forecasting and planning [14].

Prior to discuss planning and forecasting in detail, it is instructive to consider generalized framework for forecasting in terms of planning and decision. A general architecture is showed in Fig. 2. Conceptually, the architecture comprises of the following key components for taking management decisions:

- *Data traffic*: The process begins with the collection of data. The Data Traffic component extract the network traffic from multiple servers and devices to data bank, likewise this raw format data sorted in a database.
- *Planning process*: This layer can be further divided in two components (a) Plan and (b) forecasting methods. The gathered information from data bank is ready to

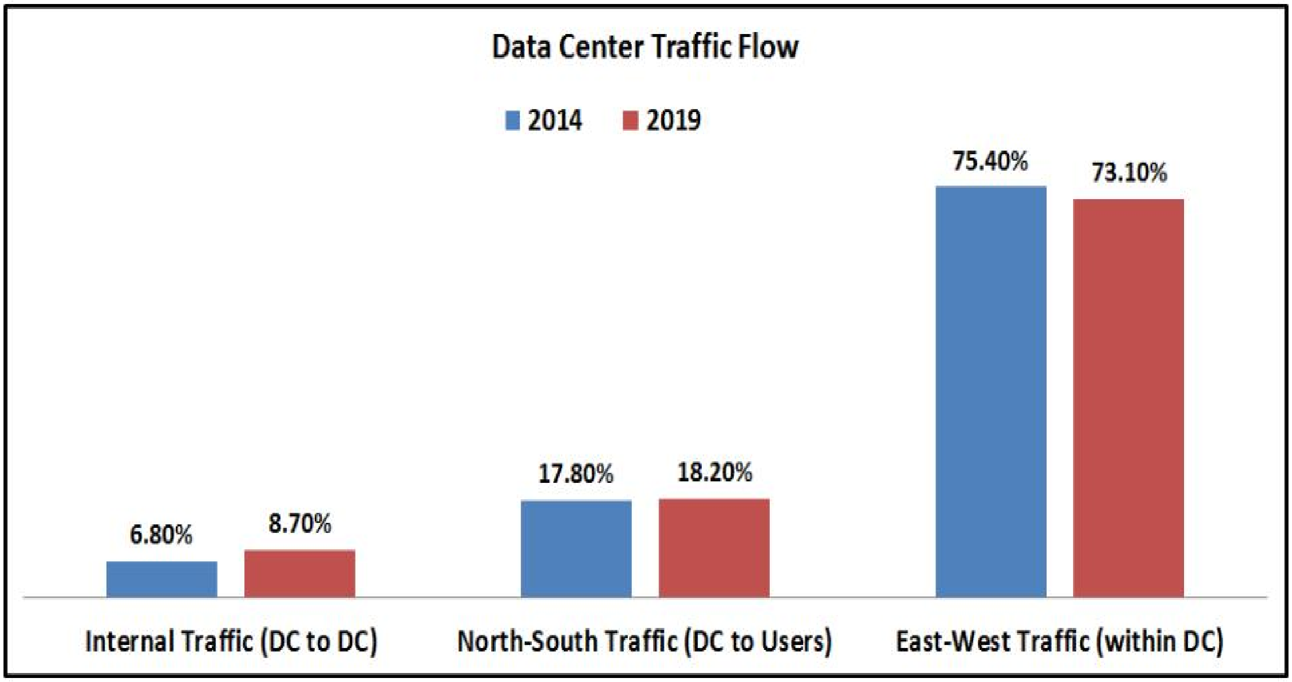

Fig. 1. Worldwide data centers and volume of traffic consumption [10].

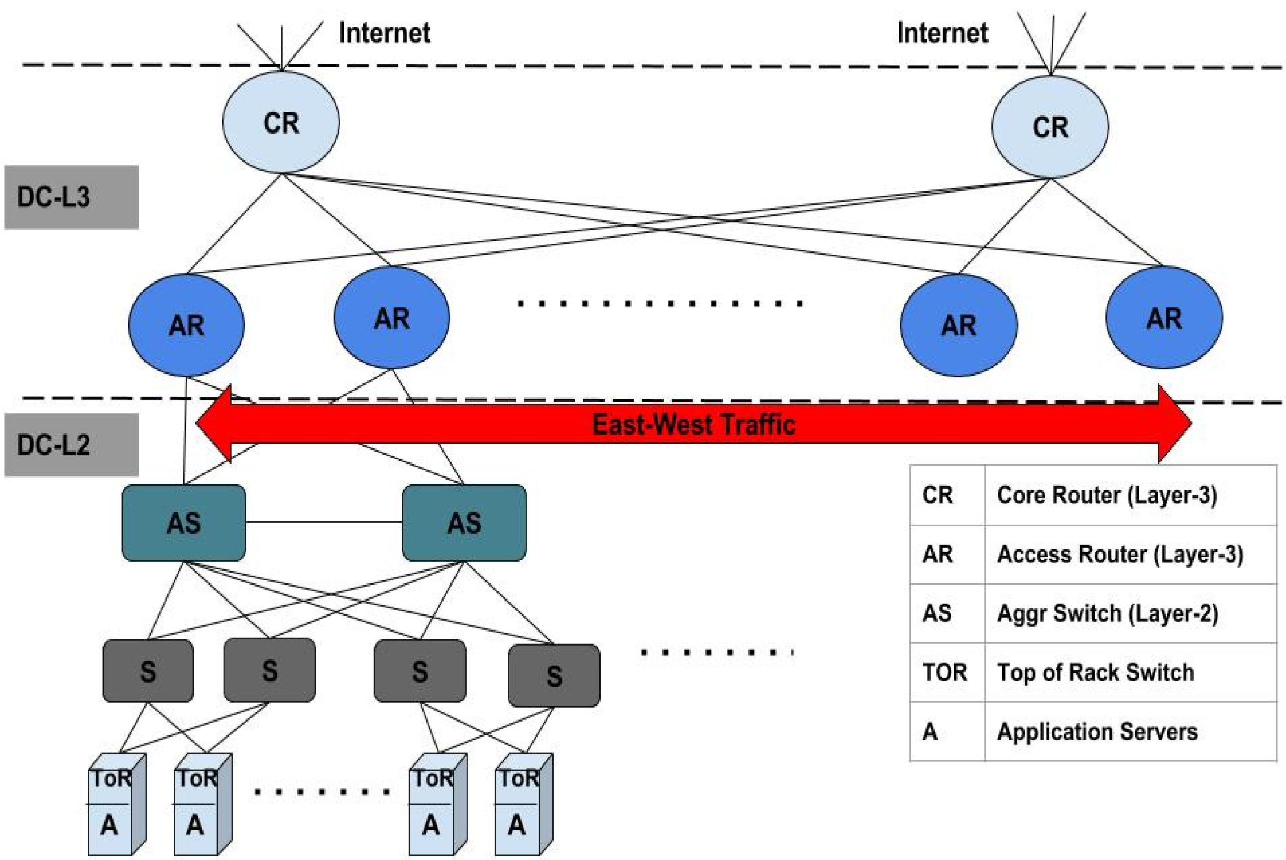

Fig. 2. Typical data center topology.

develop the objectives. It also determine the needed resources for implementing the further tasks. However, this will help to take the decision for planning process to choose appropriate plan for forecasting method.

- *Forecast*: The multiple forecasting methods can be used for prediction; qualitative methods are specialized in this subjective nature. The Forecast approach also rely on personal experiences and market trends for estimation of making opinions. The second approach contains quantitative methods; unlike qualitative this approach needs mathematical computation and algorithms. However, the quantitative methods are more appropriate for neural networks for time series to predict. These comparative models will be more discussed in chapter three and four.
- *Results*: The data center operators wait to attain outcomes for possible plans. If an appropriate outcomes are satisfactory then go ahead for further decisions. Moreover, in condition of not satisfactory will go through and revise the plan as a loop till satisfied.
  While taking futuristic movements for decisions, the management has better strategy to choose the best alternative plans for future. Forecasting methods help out to change the behavior of management till satisfactory outcomes and also repeatedly assists to make rational decisions. As shown in Fig. 3 [The framework for forecasting and planning [14] to predict the network traffic].
  The process starts with the collection of network data extracted in a data bank. Moreover, the data bank unit pushes towards the planning process in order to choose an appropriate plan for forecasting method. Data center operators wait to attain outcomes for possible plans. If the appropriate outcomes are satisfactory then go ahead for management decisions, otherwise in condition for not satisfactory this will revise the plan until satisfied.

B. Forecasting Models

In general, the forecasting models are employed for estimation, projection and forthcoming activities. However, some of the models entirely work for general purpose while others are for technological forecasting. The general purpose models are exclusively to predict economical activities, i.e. executive opinion, market survey, sales and Delphi method. On the other hand, the technological forecasting models are beneficial for accommodating, identifying expansions, operational decisions and policy making of predictable technological innovation (see Fig. 4).

C. Choosing Forecasting Method

Since our collected network traffic is in historical regular time intervals, such as the generated traffic data in days, weeks and months so we decided to choose our forecasting with Autoregressive Integrated Moving Average models (ARIMA). Moreover the observations were required intensive change and time interval for accuracy and get prompt decisions.

ARIMA is an extraordinary technique for modeling which provides diversity of tools to choose best fitting model. Time

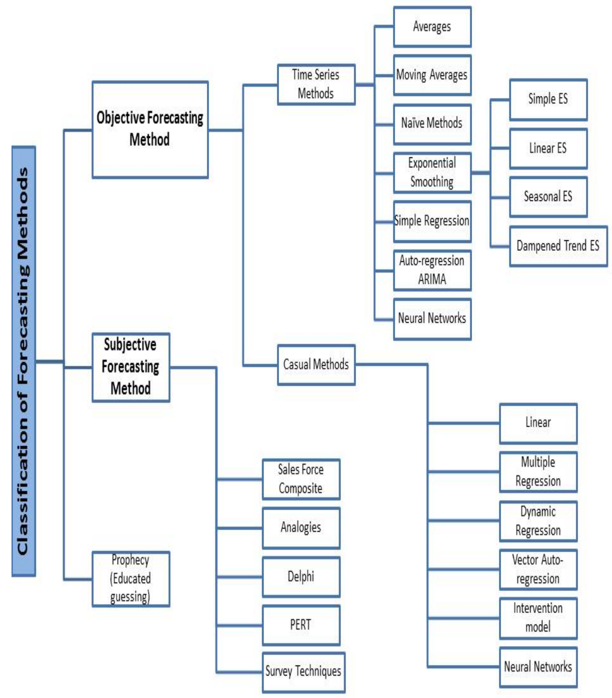

Fig. 4. Classification of forecasting models [15].

series is more compatible and being used successfully for forecasting across multiple domains of research. In order to take its advantage, we also choose these sophisticated techniques for modeling to predict the forthcoming network traffic. The common patterns within time series components are identical, in "Trend" component in which the data of long term fixed and fitted easily. While in "Seasonal" type of patterns the series of lower period of data pattern (week and days) accommodated. However, the "Cyclic" having bursty data or all of sudden and unexpected data pattern dealt with [16].

All three components (trend, seasonal and cyclic) are decomposed with time series in (1) and (2), where "y" illustrates time series, "t" shows time period, "S" displays seasonal, "T" represents Trend cycle whereas "E" is error component [17].

$$y_t = S_t + T_t + E_t \tag{1}$$

$$y_t = S_t * T_t * E_t \tag{2}$$

D. Forecast Accuracy Assessment

While evaluating the accuracy of forecast, the MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) are most suitable and common factors to apply. However, MAPE (Mean Absolute Percentage Error) is used for accessing accuracy an MAD (Mean Absolute Error) measures the size of errors in unit [18].

These are generalized standards for measuring the prediction accuracy of forecasting. These equations can be mathematically written as below:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} = \left[\frac{actual_t - forecast_t}{actual_t}\right] * 100 \quad (3)$$

$$MAD = \frac{1}{n}\sum = [Actual_t - Forecast_t] \quad (4)$$

To measure errors in accuracy is the foremost part while inspecting precision of forecasting. It is substantial to calculate the accuracy of forecast. In simple difference between actual and forecast (shown in (3)-(4)) can be considered is an error. Furthermore, the MAP appears as a better choice to analyze volume of error in terms of "percentage", whereas MAD calculates the error in "units" [19]. The consideration of errors are at the same scale unit, for example the experimental data "yt" is in Kbps and then error output "et" is also in kbps. We will demonstrate further in supplementary chapters after getting the results while forecasting.

III. Methodology

A. Design of Data Center

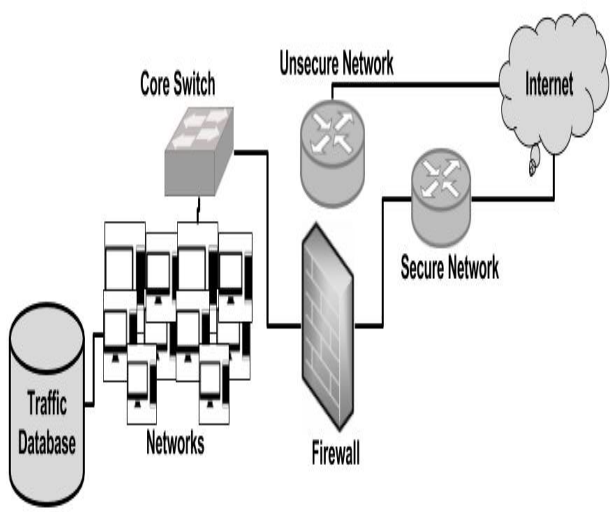

Fig. 5. Standard data center.

Since various optional network topologies to choose while creating data center, amongst several architectural models, we select an ordinary and typical three-tier architecture [20] which contains layer two to four in series of devices. For example, gateway router associated with aggregation router, and aggregation router linked with access switches. The chain of connected networks relay with each other and start from higher layers to lower layers. Gateway router are distribution point with aggregation layer and access switches at lower. Upper layers devices are directly connected with the outside traffic called the Internet, while at middle layer router collect the traffic from upper layer and distribute to lower layer switches which are relate to servers and end-user computers.

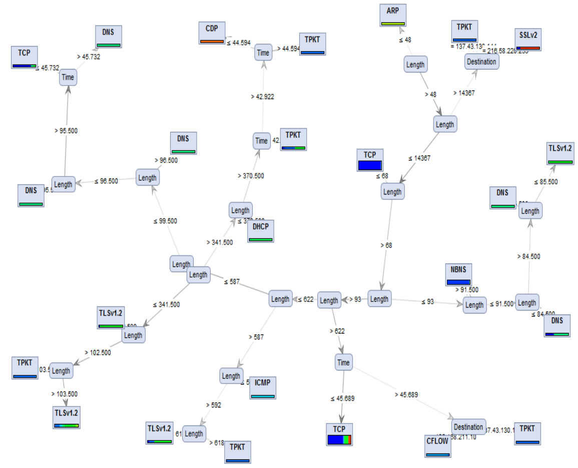

Fig. 6. Example of protocols shown in shape of tree.

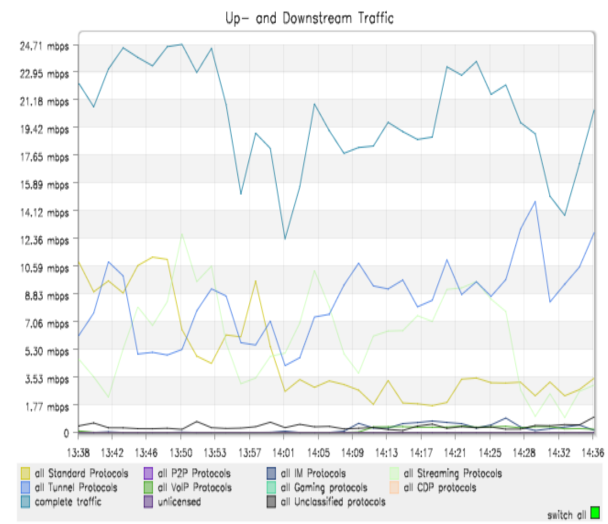

Fig. 7. Up down traffic of last hour.

B. Traffic Observation

During the simulation phase of a dynamic infrastructure, the traffic patterns of the applications are generally not known by the network architect and are difficult to predict.

References

[1] P. P. R. Bureau, "World population data sheet." 2003.

[2] R. Pingdom, "Internet 2012 in numbers," *Available at royal. pingdom. com*, 2013.

[3] Y. Takahashi, K. Ishibashi, M. Tsujino, N. Kamiyama, K. Shiomoto, T. Otoshi, Y. Ohsita, and M. Murata, "Separating predictable and unpredictable flows via dynamic flow mining for effective traffic engineering," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–7.

[4] W. W. W. Consortium *et al.*, "Internet live stats."

[5] C. G. C. I. Cisco, "Forecast and methodology, 2013–2018 (2014)."

[6] P. Bodík, I. Menache, M. Chowdhury, P. Mani, D. A. Maltz, and I. Stoica, "Surviving failures in bandwidth-constrained datacenters," in *Proceedings of the ACM SIGCOMM 2012 conference on Applications,*

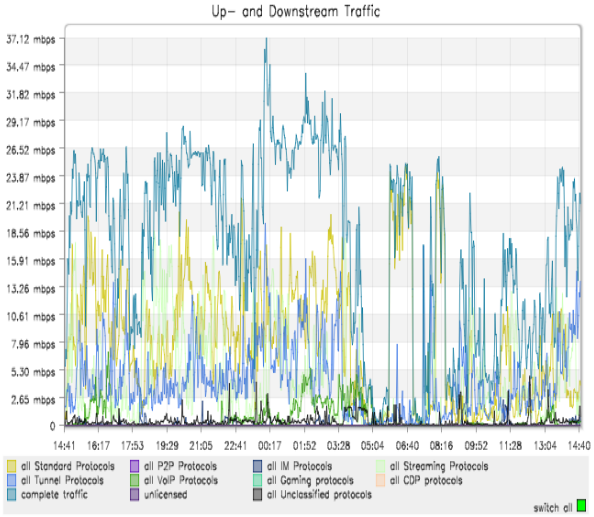

Fig. 8. Up down traffic of last day.

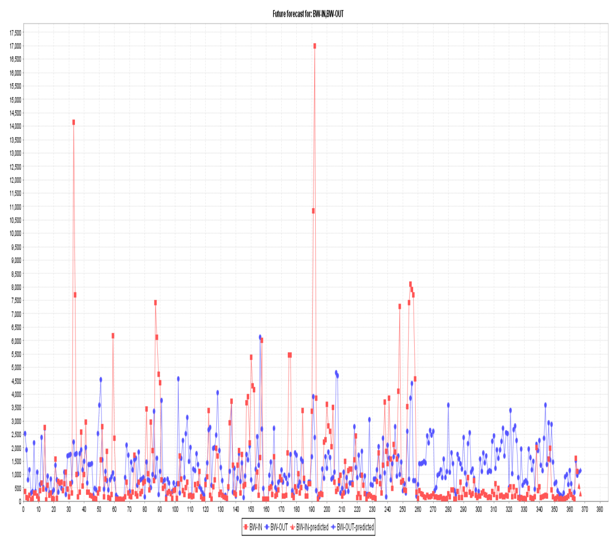

Fig. 9. Extracted traffic for bandwidth-in-out and predicted.

technologies, architectures, and protocols for computer communication. ACM, 2012, pp. 431–442.

[7] A. M. De Franceschi, L. F. Kormann, and C. B. Westphall, "Performance evaluation for proactive network management," in *Communications, 1996. ICC'96, Conference Record, Converging Technologies for Tomorrow's Applications. 1996 IEEE International Conference on*, vol. 1. IEEE, 1996, pp. 22–26.

[8] P. K. Upadhyay and S. Prakriya, "Performance of analog network coding with asymmetric traffic requirements," *IEEE Communications Letters*, vol. 15, no. 6, pp. 647–649, 2011.

[9] M. G. Mikurak, "Technology sharing during demand and supply planning in a network-based supply chain environment," Oct. 31 2006, uS Patent 7,130,807.

[10] C. T. Z. Era—Trends and Analysis. White paper. [Online]. Available: Whitepaper:TheZettabyteEra\OT1\textemdashTrendsandAnalysis

[11] C. V. Networking, "Cisco global cloud index: Forecast and methodology, 2012-2017,(white paper)," 2013.

[12] M. P. of Vulnerability and Failures. The cisco blog on vulnerabilities causes failures. [Online]. Available: Cisco-Blog:trends-in-data-center

[13] C. V. Forecast and M. 2015-2020. White paper. [Online]. Available: Whitepaper:CiscoVNIForecastandMethodology,2015-2020

[14] J. S. Armstrong, *Principles of forecasting: a handbook for researchers and practitioners*. Springer Science & Business Media, 2001, vol. 30.

[15] T. on Neural Network Forecasting. neural-forecasting.com. [Online]. Available: http://www.neural-forecasting.com/tutorials

[16] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*. John Wiley & Sons, 2008.

[17] R. J. Hyndman, G. Athanasopoulos, S. Razbash, D. Schmidt, Z. Zhou, Y. Khan, and C. Bergmeir, "forecast: Forecasting functions for time series and linear models," *R package version*, vol. 5, 2014.

[18] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, no. 4, pp. 527–529, 1993.

[19] E. Stellwagen, "Forecasting 1 1: A guide to forecast error measurement statistics and how to use them," 2015.

[20] D. Kliazovich, P. Bouvry, and S. U. Khan, "Greencloud: a packet-level simulator of energy-aware cloud computing data centers," *The Journal of Supercomputing*, vol. 62, no. 3, pp. 1263–1283, 2012.