**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Programas de Maestría y Doctorado en Ingeniería Telemática**

**Seminario de Investigación**

**Traffic Engineering with Prediction based on Big Data Technologies for Configuring Software-Defined Networks**

**Carlos Felipe Estrada Solano**

Ph.D. Student

May 13th, 2016

1. **Introduction**

For this session, we modified the title of the PhD proposal to " Traffic Engineering with Prediction based on Big Data Technologies for Configuring Software-Defined Networks". The advisor is Professor Oscar Caicedo. Also, recently, Professor Nelson Fonseca, from the Computing Institute of the University of Campinas, − and who was here three (3) weeks ago giving the conference Big Data for Networking − has accepted to co-advise my work.

The outline is as follows: the background, the motivating scenario, the problem statement, the hypothesis, the goals, and some advances.

1. **Background**

The background presents three concepts: traffic prediction, Software-Defined Networking (SDN), and Big Data.

Traffic prediction refers to collect and analyze traffic information from the network to predict future traffic behavior. The prediction methods can be applied for different time scales, such as milliseconds, seconds, minutes, hours, days, or even months. This predicted traffic behavior is useful for conducting different management techniques, such as anomaly detection, admission control, capacity planning, and traffic engineering. Usually, these management techniques lead to configure the network for optimizing its operation. In particular, traffic engineering establishes that the knowledge of relationships between network behavior and network configuration may help to decide the best parameters according to real performance feedback. In short words, traffic engineering aims to optimize network behavior regarding traffic information. Furthermore, traffic engineering with prediction is a promising approach to accommodate time-varying traffic without frequent route changes, for example, to avoid congestion on the basis of the predicted traffic (Otoshi et al., 2015) (Akyildiz et al., 2014).

The Software-Defined Networking paradigm, or SDN, defines a new architecture for future networks that overcomes the difficult management of conventional networks. SDN separates the Data Plane, in charge of packet forwarding, and the Control Plane, responsible of decision policies, leaving the Data Plane in network devices, such as router or switches, and migrating the Control Plane to a logically centralized software program, also known as Controller. The Control Plane communicates with the Data Plane through an SDN protocol (Feamster et al., 2013) (Kim and Feamster, 2013) (Anwer and Feamster, 2010) (Shimonishi and Ishi, 2010). This architecture allows a simpler network operation by providing four major advantages (Akyildiz et al., 2014). First, a centralized global view about the network state, such as resource capabilities and dynamic status, and about the deployed applications, such as quality requirements and service agreements. Second, a dynamic programmability of multiple forwarding devices, for example, for allocating resources to prevent congestion and improve performance. Third, open interfaces for handling the Data Plane, such as OpenFlow, and for developing the Application Plane, like APIs based on protocols and programming languages. Fourth, a flexible flow management, especially in OpenFlow because of its multiple flow tables. These unique features lead the SDN architecture to emerge as a promising scenario for efficiently and intelligently implementing management techniques, particularly traffic engineering.

**Figure 1. SDN architecture**

Figure 1 depicts a typical SDN architecture composed of three horizontal planes and three interfaces (Efremova and Andrushko, 2015) (Kiran and Kinghorn, 2015) (Kreutz et al., 2015) (Casado et al., 2014) (Rijsman and Singla, 2013) (ONF, 2012). At the bottom, the Data Plane for packet forwarding. In the middle, the Control Plane deploys the Network Operative System, or Controller, for compiling network logic and enforcing decisions on the Data Plane through the Southbound Interfaces, which the most well-know is OpenFlow. This Control Plane also defines East/Westbound Interfaces to deploy distributed Controllers and includes Network Slicers that virtualize the underlying Data Plane, allowing multiple Controllers running over the same network infrastructure. Such Controller provides generic services and Northbound Interfaces to the Application Plane, facilitating to develop and integrate high-level networking functions as Network Applications. Recent investigations have considered a Management Plane in the SDN for conducting an integrated management of the whole SDN environment (Wickboldt et al., 2015) (ONF, 2014). In our previous works (Estrada-Solano et al., 2016) (Caicedo Rendon et. al., 2016, 2013), we extended this Management Plane to define a reference architecture composed of four sub-models: Information, Organizational, Communication, and Functional. Then, we defined a CIM-based Information Model to characterize the SDN environment, and designed a mashup-based and event-driven framework for easily building and executing management tasks.

The concept of Big Data was born from coping with the challenges in Data Management, mainly focused in collecting, storing, and preparing data, presenting three dimensions: Volume, that refers to the amount of data, ranging from Terabytes to Zettabytes and beyond; Velocity, that is related to the rate at which data are generated, including batch and streaming; and Variety, that makes reference to the different types of data, such as structured and unstructured data. Over time, other dimensions has been added to the concept of Big Data, taking into account the challenges about analyzing and getting intelligence from data, also known as Data Analytics. These dimensions are: Veracity, that refers to the integrity of data, ranging from reliable and unreliable; Variable, that is related to the changes of data interpretation over time, such as constant or irregular; Valence, that makes reference to the inter-relatedness of data, from fully correlated to completely disconnected; and Value, that is about obtaining results from data, recognizing trivial and significant results. In short words, Big Data allows to manage and analyze huge amounts of data to obtain significant results for predicting events and improving decision-making (Gandomi and Haider, 2015) (Jagadish et al., 2014) (Siewert, 2013) (Laney, 2001).

1. **Motivation**

For the motivating scenario, SDN, as conventional networks, requires to cope with traffic management, including time-varying traffic and different traffic types. Using traditional solutions as redundant resource capacity presents critical drawbacks: requires overly large capacity to cope with traffic variations, causes low resource utilization, and, hence, provides a costly and poor scalability as the traffic increases. Therefore, there is the need to address traffic management in SDN with limited resources, avoiding traffic congestion and optimizing resource utilization. Fortunately, as aforementioned, SDN provides unique features that allow to implement more efficient and intelligent traffic management techniques, such as traffic engineering.

For example, Adaptive Traffic Engineering represents a promising approach for regulating traffic in networks (Otoshi et al., 2015). In SDN, this Adaptive Traffic Engineering can be performed in the Controller as a Network Application that periodically measures and analyzes the traffic in the network for dynamically deciding if performing configurations to improve network behavior, for example, in the case of network congestion. However, this adaptive approach only mitigates the observed traffic conditions, not future traffic issues. Furthermore, high time-varying traffic may cause heavy load at the controller, generate high traffic between the controller and the underlying network devices, and affect the performance of the network due to frequent complex configurations. In this sense, prediction for traffic engineering emerges as an interesting approach to tackle these drawbacks.

Then, from an SDN deployment, a management component can collect and store traffic-related data in order to perform specific analysis techniques and predict future traffic regarding the obtained results. The predicted traffic patterns allow applying SDN configurations for improving network behavior and reducing adaptive modifications during a control period. Here, the challenges are *(i)* defining the time-scale of the prediction methods to establish the control period of the network configuration; and *(ii)* performing a network configuration that will demand only few and simple adaptive configurations that help to optimize network performance.

Moreover, there are some SDN deployments, such as datacenters and wide area networks, that generate huge amounts of data. Then, the aforementioned challenges become more difficult for traffic engineering with prediction in a data-intensive SDN environment. It is important to highlight that our proposal focuses on using Big Data technologies.

1. **Problem Statement**

Based on the above information, for the problem statement, this proposal establishes that due to the high variability of network traffic and the limited network resource capacity, performing traffic engineering for configuring SDN-based networks in a data-intensive environment remains inefficient in terms of time and traffic. Therefore, the research question is as follows: how to carry out an effective approach that performs traffic engineering for configuring data-intensive SDNs? It is to remark that, so far, this word effective is considered as a broad term that encompasses metrics about time and traffic, including the prediction time-scale for the control period, the time for analyzing data, the optimization of traffic in the network, and the traffic generated when conducting configuration.

1. **Hypothesis**

To address the research question, this approach proposes the following hypothesis: carrying out traffic prediction supported by Big Data technologies – working along the network core – would lead to perform an effective traffic engineering for configuring SDN-based networks in a data-intensive environment. Recall that traffic engineering aims to optimize network behavior regarding traffic related information.

1. **Goals**

Based on the hypothesis, the main goal is to propose a mechanism based on Big Data technologies for implementing traffic engineering with prediction for configuring SDN-based networks in a data-intensive environment. The specific goals are: *(i)* characterize the relationships between network traffic and network configuration for SDN in a data-intensive environment, *(ii)* design a framework for traffic prediction supported by Big Data technologies for conducting traffic engineering in a data-intensive SDN; and *(iii)* evaluate the effectiveness of the framework through an use case in an emulated data-intensive SDN environment.

1. **Advance**

As an advance, we continued with the data extraction process of the systematic mapping about traffic engineering in SDN, updating the corresponding plots. In the case of traffic engineering mechanisms in SDN (see Figure 2), we observed that the number of papers per each mechanism maintains a similar relation, where monitoring concentrates the major number of papers, both classification and steering represent an attractive opportunity for current research, and traffic prediction is starting to take off. Recall that this proposal focuses on traffic prediction.

**Figure 2. Traffic engineering mechanisms in SDN**

In the case of the FCAPS+P (see Figure 3), the number of papers per functional area also maintains a similar relation, where both performance and fault present a major number of papers, accounting is still relegated from the investigation perspective, and configuration, programming, and security are constantly drawing more attention from the research community. It is to remind that this work concentrates in the configuration functional area.

**Figure 3. Traffic engineering in SDN regarding the FCAPS+P model**

To analyze this configuration in SDN, we used SCOPUS and CiteSpace. The analysis results show that traffic engineering stands out as one of the major areas of research in SDN configuration. In addition, these results provided interesting associated terms, such as datacenter network and wide area network, which give us an idea about the network context for our approach, and scalability and capacity planning, which indicate some network challenges to take into account in traffic engineering for SDN configuration.

We also used the author analysis tool from SCOPUS (see Figure 4), obtaining the three main authors in SDN configuration: Jennifer Rexford, David Walker, and Nate Foster, the three from the Princeton University. These authors are actively working in a research area of our interest: traffic engineering for SDN. Nevertheless, most of their work in this area focuses on the performance management area. These authors also work in traffic engineering for conventional networks and in programming languages for SDN. The latter is more related with the programming functional area.



**Figure 5. Author analysis tool from SCOPUS**

From this systematic mapping, two interesting papers about traffic prediction for SDN stand out. First, the work "Network Traffic Prediction Model based on Training Data" (Park et al., 2015), which proposes a prediction model based on the K Nearest Neighbors algorithm, frequently used in the area of road traffic prediction. The simulation results of the algorithm presented a sub-optimal performance. The authors state that these results were caused by the lack of historical data for performing extensive simulations. This work used a dataset from the Lincoln Laboratory, which the more recent one data from 2000. This dataset was built for intrusion detection systems and the network wasn't an SDN deployment. As future work, the authors plan to rebuild the model for high precision, evaluate such model with real-time data, and compare their approach with other prediction models.

The other work is "EMD-based Multi-model Prediction for Network Traffic in Software-Defined Networks" (Dai et al., 2014). This paper presents an interesting concept about "appropriate" traffic matrix, which refers to a result from a prediction algorithm that reflects the traffic change trend of the future. This work proposes a short-term forecasting algorithm that decomposes traffic series by using the Empirical Mode Decomposition and that performs Multi-Model Prediction. This algorithm demonstrated better results than similar algorithms, such as Auto-Regresive and Moving Average and Support Vector Regression. The authors used a dataset from Internet2, which data from August 2008 and whose network wasn't an SDN deployment.

Comparing these works with the present proposal, we can conclude that they only perform short-term traffic forecasting, while our work is more likely to focus on long-term prediction. Also, these works did not perform configuration of an SDN-based network, they only adopted SDN as a technology for collecting data. In addition, they didn't consider a data-intensive environment, instead, they assumed that the data fits in a single controller. Furthermore, the used datasets aren't from an SDN deployment, although it would be interesting to further review the Internet2 dataset. And finally, such works only perform simulations, while our idea is to deploy the prediction framework in, at least, an emulated SDN environment.

As near future steps, this proposal plan to include works from 2016 about traffic prediction in SDN. Also, review traffic prediction approaches in other networks different of SDN, which can give us a clearer idea about traffic prediction solutions. In addition, describe two use cases for traffic engineering with prediction in a data-intensive SDN in order to be more specific in our research proposal. And finally, perform a review about Big Data for managing SDN for constructing the survey paper.

**References**

Akyildiz, I. F., Lee, A., Wang, P., Luo, M. & Chou, W. (2014). A Roadmap for Traffic Engineering in SDN-OpenFlow Networks. Comput. Netw. 71, 1--30.

Anwer, M. B. & Feamster, N. (2010). Building a Fast, Virtualized Data Plane with Programmable Hardware. SIGCOMM Comput. Commun. Rev. 40 (1), 75--82.

Caicedo Rendon, O. M., Estrada-Solano, F. & Granville, L. Z. (2013). A Mashup-Based Approach for Virtual SDN Management. In Computer Software and Applications Conference (COMPSAC), 2013 IEEE 37th Annual (pp. 143-152).

Caicedo Rendon, O. M., Estrada-Solano, F., Guimarães, V., Rockenbach Tarouco, L. M. & Granville, L. Z. (2016). Rich dynamic mashments: An approach for network management based on mashups and situation management. Computer Networks 94, 285 - 306.

Casado, M., Foster, N. & Guha, A. (2014). Abstractions for Software-defined Networks. Commun. ACM 57 (10), 86--95.

Dai, L., Yang, W., Gao, S., Xia, Y., Zhu, M. & Ji, Z. (2014). EMD-Based Multi-Model Prediction for Network Traffic in Software-Defined Networks. In 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems (pp. 539-544).

Efremova, L. & Andrushko, D. (2015). What's in OpenDaylight?. [Online]. Available: https://www.mirantis.com/blog/whats-opendaylight/

Estrada-Solano, F., Ordonez, A., Granville, L. Z. & Caicedo Rendon, O. M. (2016). A CIM-based Information Model for Heterogeneous SDN Management . Computer Communications (Submitted to Computer Communications)

Feamster, N., Rexford, J. & Zegura, E. (2013). The Road to SDN. Queue 11 (12), 20:20--20:40.

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics . International Journal of Information Management 35 (2), 137 - 144.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R. & Shahabi, C. (2014). Big Data and Its Technical Challenges. Commun. ACM 57 (7), 86--94.

Kim, H. & Feamster, N. (2013). Improving network management with software defined networking. Communications Magazine, IEEE 51 (2), 114-119.

Kiran, S. & Kinghorn, G. (2015). Cisco Open Network Environment: Bring the Network Closer to Applications (White Paper C11-728045-03). Cisco.

Kreutz, D., Ramos, F. M. V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S. & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. Proceedings of the IEEE 103 (1), 14--76.

Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety (Research Report 949). META Group.

ONF (2014). SDN architecture v1.0 (Technical Reference TR-502). Open Network Foundation.

ONF (2012). Software-Defined Networking: The New Norm for Networks (White Paper ). Open Network Foundation.

Otoshi, T., Ohsita, Y., Murata, M., Takahashi, Y., Ishibashi, K. & Shiomoto, K. (2015). Traffic Prediction for Dynamic Traffic Engineering. Comput. Netw. 85 (C), 36--50.

Park, J., Raza, S. M., Thorat, P., Kim, D. S. & Choo, H. (2015). Computational Science and Its Applications – ICCSA 2015: 15th International Conference, Banff, AB, Canada, June 22-25, 2015, Proceedings, Part IV. In O. Gervasi, B. Murgante, S. Misra, L. M. Gavrilova, C. A. M. A. Rocha, C. Torre, D. Taniar & O. B. Apduhan (ed.), (pp. 117--127). Springer International Publishing. (ISBN: 978-3-319-21410-8.)

Rijsman, B. & Singla, A. (2013). Day One: Understanding OpenContrail Architecture. Juniper Networks Books.

Shimonishi, H. & Ishii, S. (2010). Virtualized network infrastructure using OpenFlow. In Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP (pp. 74--79).

Siewert, S. B. (2013). Big data in the cloud Data velocity, volume, variety, veracity (Technical Report ). IBM developerWorks.

Wickboldt, J., De Jesus, W., Isolani, P., Both, C., Rochol, J. & Granville, L. (2015). Software-defined networking: management requirements and challenges. Communications Magazine, IEEE 53 (1), 278--285.