**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**

**Relator: MSc. Felipe Estrada Solano**

**Co-relator: PhD. Oscar Mauricio Caicedo Rendón**

**Protocolante: MSc. Héctor Fabio Bermúdez O. estudiante de Doctorado**

**Coordinador: PhD. Juan Carlos Corrales**

**Fecha:** 13de mayo de 2016

**Hora inicio:** 10:13 a.m.

**Hora fin:** 11:05 a.m.

**Lugar:** Salón de posgrado, FIET, Universidad del Cauca, Popayán

**Asistentes:**

PhD. Juan Carlos Corrales - Coordinador

PhD. Oscar Mauricio Caicedo Rendón, profesor invitado

MSc. Felipe Estrada Solano - Relator.

MSc. Héctor Fabio Bermúdez, Protocolante

Estudiantes de Maestría y Doctorado en Ingeniería Telemática

**Orden del día:**

1. Presentación a cargo del relator
2. Intervención del co-relator
3. Discusión
4. Conclusiones

**Desarrollo**

1. **Presentación a cargo del Relator**

The magister Felipe Estrada present the research proposal: "Traffic Engineering with Prediction based on Big Data Technologies for Configuring Software-Defined Networks". The advisor is the Professor Oscar Caicedo, and co-advisor the Professor Nelson Fonseca, from the Computing Institute of the University of Campinas. The outline is as follows:

1. Background
2. Motivation
3. Problem Statement
4. Hypothesis
5. Goals
6. Advance

**Background**

The Mag. Estrada preset three concepts: traffic prediction, Software-Defined Networking, and Big Data.

*Traffic prediction*: it 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.

*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. This architecture allows a simpler network operation by providing four major advantages. 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.

*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.

**Motivation**

For the motivating scenario, the mag. Estrada present the following sceneries:

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. The figure 1 describe this situation.

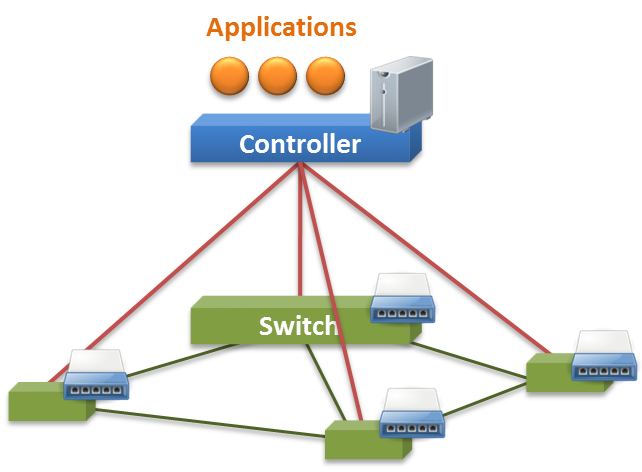
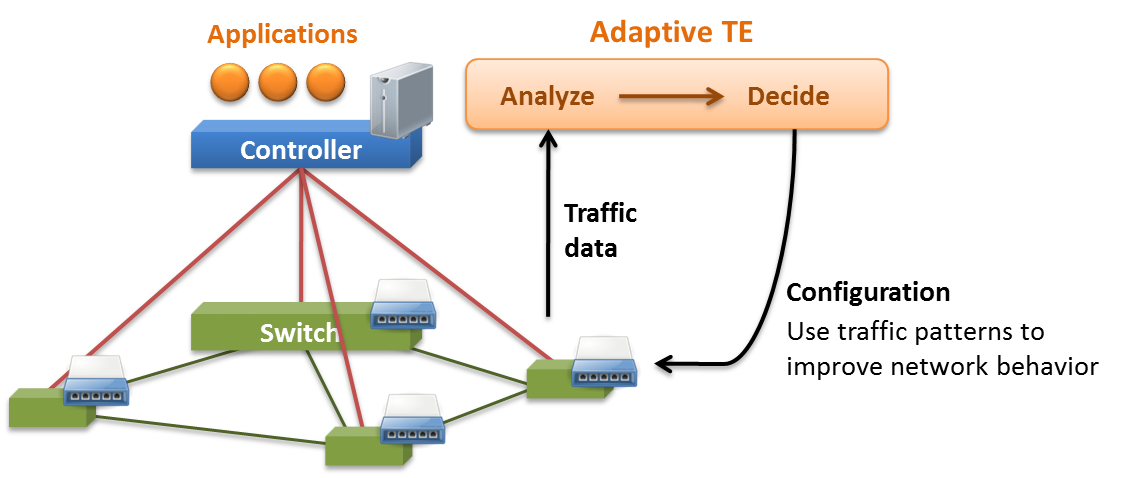


Figure 1. Implement more efficient and intelligent traffic management techniques

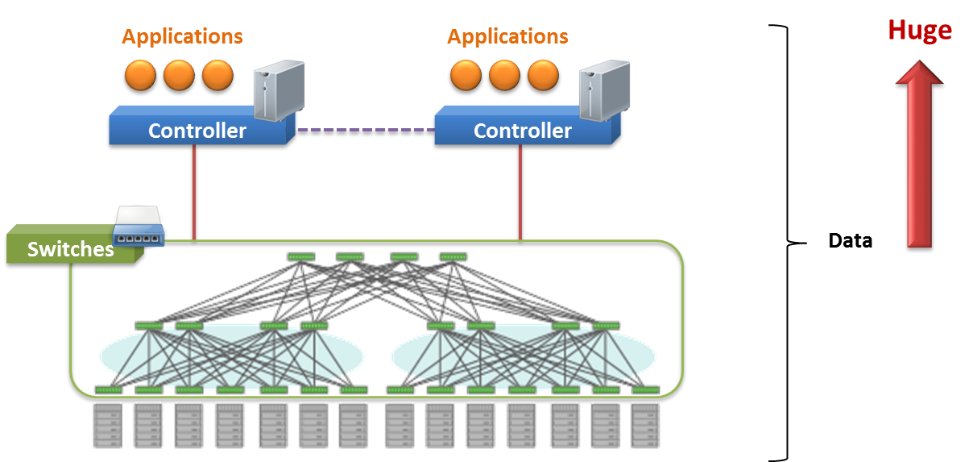
Traffic Engineering (TE)

The mag. Estrada present an example, Adaptive Traffic Engineering represents a promising approach for regulating traffic in networks. 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 your proposal focuses on using Big Data technologies.



**Problem Statement**

Based on the above information, for the problem statement the mag. Estrada 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.

**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.

**Goals**

Main goal

Propose a mechanism based on Big Data technologies for implementing traffic engineering with prediction for configuring SDN-based networks in a data-intensive environment.

specific goals:

*(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.

*(iii)* evaluate the effectiveness of the framework through an use case in an emulated data-intensive SDN environment.

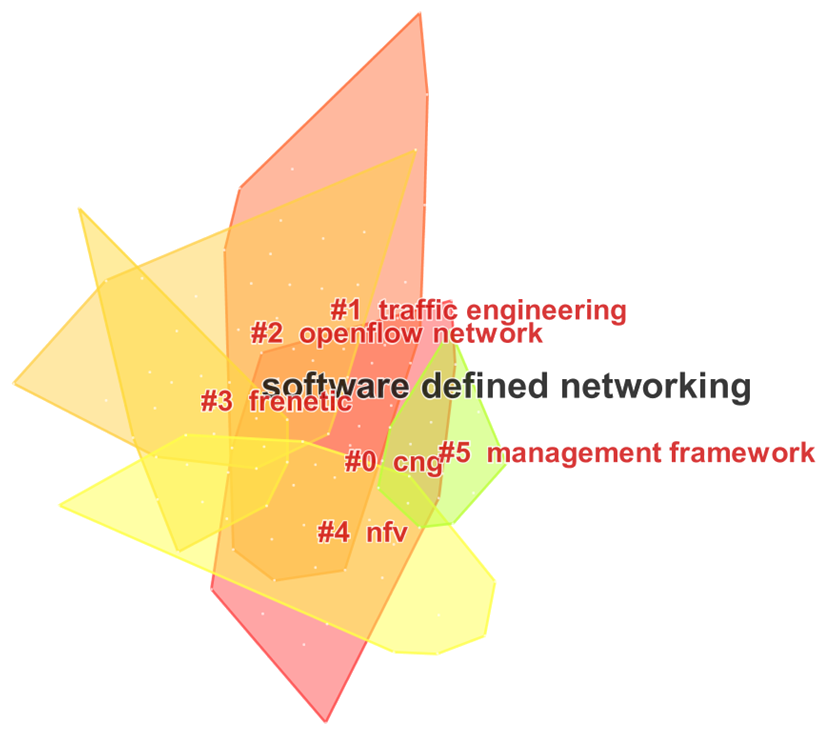
**Advance**

As an advance, the Mag. Estrada 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, he 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.

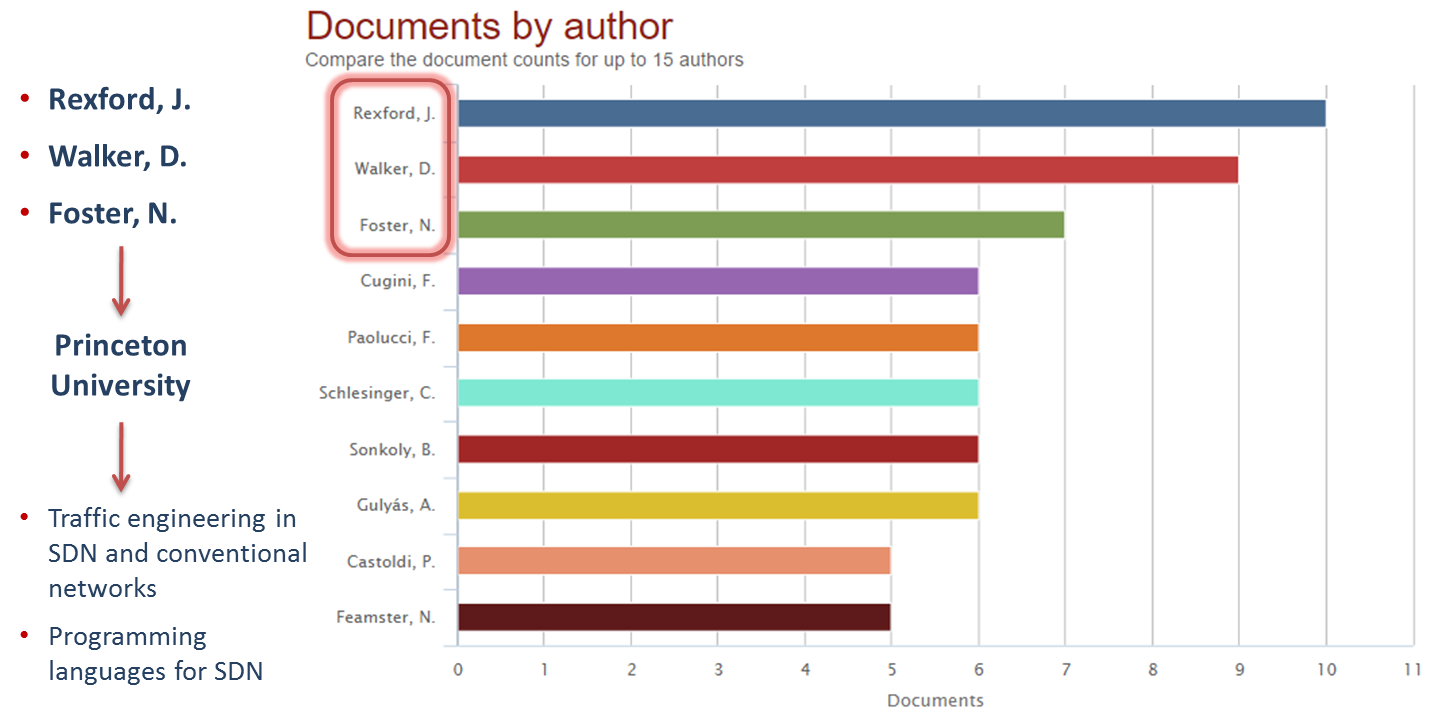
In the case of the FCAPS+P, the number of papers per functional area also maintains a similar relation, where both performance and fault present a major number of papers, accounting is still relegated from the investigation perspective, and configuration, programming, and security are constantly drawing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



To analyze this configuration in SDN, the Mag. Estrada 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.



The Mag. Estrada also used the author analysis tool from SCOPUS, 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.



From this systematic mapping, the mag. Estrada took the two most interesting papers about traffic prediction for SDN. First, the work "Network Traffic Prediction Model based on Training Data" from 2015. This paper 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". 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 this proposal, the mag. Estrada concluded that they only perform short-term traffic forecasting, while this 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 this idea is to deploy the prediction framework in, at least, an emulated SDN environment.

As near future steps, the Mag. Estrada plans 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. Finally, perform a review about Big Data for managing SDN for constructing the survey paper.

1. **Intervención del co-relator**

The Ph.D. Caicedo said that today the progress of the Felipe work is presents, highlights that the co-tutor is the Ph.D. Nelson Fonseca is IEEE senior member and Raul Hurtado who is IEEE senior members too, they are experts in SDN, they will support the work of Felipe.

1. **Discusión**

* Un estudiante da algunas recomendaciones sobre la pronunciación.
* Con respecto al primer objetivo específico se afirma que corresponde más a actividades, se recomienda revisarlo.
* Se pregunta qué tanto es el alcance del caso de uso que se usará como prueba, ya que un caso de uso puede ser muy grande o muy específico. Al respecto el Mag. Estrada responde que inicialmente se definen dos casos de uso, es posible que en el anteproyecto se cierre un poco, pero todo depende del avance alcanzado, pero sí es necesario definir más los casos de uso ya que se están dejando muy abiertos. Al respecto el PhD. Caicedo responde que para el caso de uso la idea es trabajar en un centro de datos en donde haya más de un controlador y poder sacar la información de tráfico de esa red, pero es necesario definir de una manera específica esos casos de uso, se está en ese proceso. El otro caso de uso puede ser en *carrier network*, ya que en este tipo de redes se ha utilizado el concepto de SDN pero es necesario definir muy bien las métricas usadas.
* Con respecto a los objetivos, se pregunta si en ingles se deben de realizar en infinitivo.
* El Ph.D. Corrales pregunta cómo se va a predecir. El Mag. Estrada afirma que el concepto de Big Data como concepto ha evolucionado y no solamente habla de la gestión de datos, también habla de la parte de Análisis de datos y es aquí en donde se comienzan a utilizar algoritmos o mecanismos de inteligencia artificial o de aprendizaje de máquinas, la idea es aplicar este tipo de mecanismos en un entorno de Big Data.
* El PhD. Corrales pregunta cuál es el punto de partida para poder predecir algo, ya que se debe de entender muy bien el dominio y la característica del dato, lo primero que uno se imagina es un grafo y sobre eso se empieza a predecir. Al respecto el PhD. Caicedo manifiesta que el co-tutor Nelson Fonseca manifestó que lo primero que se debe de hacer es definir la red y como se componen los nodos y las aristas de ese grafo que van a representar, ya que esa es la forma como se debe modelar la red para después poder predecir. El PhD. Corrales recalca que es muy importante definir la característica del dato. Además recomienda quitar la parte de predicción, dejarlo más abierto, si se alcanza a hacer predicción sería un plus.

**4. Conclusiones**

El coordinador da por terminada la sesión y agradece los comentarios.

Siendo las 11:05 am. Se termina la sesión.