**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

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

**Seminario de Investigación**

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

**Carlos Felipe Estrada Solano**

Ph.D. Student

April 22nd, 2016

1. **Introduction**

For this session, we modified the title of the PhD proposal to "Traffic Prediction based on Big Data Technologies for Configuring Software-Defined Networks". The outline is as follows: the background, the motivating scenario, our initial approach, the systematic mapping, some conclusions, and the Ph.D. study plan.

1. **Background**

The background presents three concepts: Software-Defined Networking, traffic engineering, and Big Data.

The Software-Defined Networking paradigm, or SDN, defines a new architecture for future networks that overcomes the difficult management of conventional networks. In a conventional network, both the Data Plane, responsible of packet forwarding, and the Control Plane, in charge of decision policies, are in the same network device, such as routers and switches. SDN separates the Data and the Control planes, leaving the Data Plane in network devices and migrating the Control Plane to a logically centralized software program, also known as controller, allowing a simpler network operation. 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).

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 deploys the network infrastructure made up of interconnected Network Devices for packet forwarding. In the middle, the Control Plane deploys a Network Operative System, or Controller, that compiles the network logic and enforces decision policies on the Data Plane through Southbound Interfaces, or SBI. OpenFlow is the most well-known and used open standard SBI. 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 custom Network Applications. Therefore, at the top, the Application Plane contains Network Applications that execute high-level networking functions. In addition, recent investigations have considered a Management Plane in the SDN architecture that interacts with the other three horizontal planes for conducting an integrated management of the whole SDN environment (Wickboldt et al., 2015) (ONF, 2014). For the Master's degree project (Estrada-Solano et al., 2016) (Caicedo Rendon et. al., 2016, 2013), we extended this Management Plane by proposing a high-level approach based on the Open System Interconnection network management model, which divides into four submodels: Information, Organizational, Communication, and Functional. Then, we focused on defining the Information Model by using the Common Information Model specification to represent the SDN environment from a management perspective and, also, designed a mashup-based and event-driven framework that allows Network Administrators to easily build and execute SDN management tools.

**Figure 1. SDN architecture**

Next, traffic engineering, which defines mechanisms that dynamically analyze, recognize, classify, predict, and regulate traffic to improve network management (Otoshi et al., 2015). For example, a specific mechanism may classify various traffic types from distinct applications to provide the most suitable network service for each traffic type in a very short time period. Therefore, traffic engineering represents an interesting opportunity to improve management techniques for every SDN management functional area, which we defined in our previous work as Fault, Configuration, Accounting, Performance, Security, and Programming (Estrada-Solano et al., 2016). This is the FCAPS+P model. In addition, it is worth to mention that traffic engineering with prediction is a promising approach to accommodate time-varying traffic without frequent route changes, allowing, for example, to avoid congestion on the basis of the predicted traffic (Akyildiz et al., 2014).

Following, the concept of Big Data has evolved as a solution to tackle the challenges in data management, describing six dimensions, also known as the 6 Vs (Gandomi and Haider, 2015) (Marr, 2014) (Laney, 2001). Volume, that refers to the vast amounts of data. Velocity, that is related to the rate at which data are generated and should be analyzed. Variety, that makes reference to the different types of data. Veracity, that refers to the integrity of the data. Variability, that is related to the different data flow rates. And value, that makes reference to obtaining significant results from data. Big Data has been applied in different domains, including networking, where experts highlight that is going to have short and long term benefits in the future Internet (Mayer-Schönberger and Cukier, 2013) (Oracle, 2013) (Cloud Computing Consortium, 2012). In short words, Big Data allows to collect and analyze huge amounts of data to obtain significant results for predicting events and improving decision-making.

This extraction of significant results with Big Data can be broken into two groups (Gandomi and Haider, 2015) (Nguyen, 2014) (Labrinidis and Jagadish, 2012) (Jiang, 2012): first, Data Management, which refers to the storage in Big Data and is composed of engineering mechanisms for collecting, storing, and preparing data; this is closely related to the Volume, Velocity, Variety, and Variability dimensions. Second, Data Analytics, which refers to the analytics in Big Data and is made up of techniques that analyze and get intelligence from data, such as Data Mining and Machine Learning; this is related to the Veracity and Value dimensions.

1. **Motivation**

For the motivating scenario, we present the four major advantages of the SDN architecture for operating networks. 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, specially 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.

Then, from an SDN deployment, a management component can collect and store data in order to perform specific analysis techniques and take decisions based on the obtained results. These decisions allow applying SDN configurations that optimize the network performance. For example, updating the packet forwarding by using predicted traffic patterns. Therefore, the knowledge of relationships between network status and network configuration may help network to decide the best parameters according to real performance feedback.

Nevertheless, there are some SDN deployments, such as datacenters and wide area networks, that generate huge amounts of data. Then, the question is how to apply traffic engineering in a data-intensive SDN environment? One possible solution is to integrate Big Data technologies.

1. **Initial Approach**

So, our initial approach is to address traffic engineering for configuring SDNs, working along with the network core, with Big Data approaches.

In the Computer Networks Group, guided by Ph.D. Caicedo, we address this approach as following: the undergraduate students cope with traffic collection in SDN to support the work of graduate students; the master student tackles traffic classification or recognition in SDN by using the collected data, providing support to the Ph.D. project; finally, this proposal focus on traffic prediction or estimation in SDN by relying on the aforementioned works. The three approaches are based on Big Data technologies.

1. **Systematic Mapping**

To start with this proposal, we are carrying out a systematic mapping to explore the status of traffic engineering in SDN. So far, this systematic mapping consists of the research questions, the search process, the selection process, and some initial results (Kitchenman et al., 2009) (Petersen et al., 2008).

The research questions that we pretend to solve through this systematic mapping are, first, what are the different solutions for implementing traffic engineering in SDN?, second, what research topics about traffic engineering in SDN are being addressed?; and third, what are the limitations of current investigation about traffic engineering in SDN?

For the search process, we constructed the query by using the keywords "software-defined networking" and "traffic engineering", both with quotation. Then, by adding and discarding associated terms, we got the following search query:

("software-defined networking" OR openflow OR "software defined networking" OR "software defined network" OR "software defined networks“) AND ("traffic engineering" OR "traffic management" OR "traffic analysis" OR "traffic monitoring" OR "traffic classification" OR "traffic prediction" OR "traffic steering")

With this query, we performed the search on ACM, IEEE, ScienceDirect, and SpringerLink. The first three are recommended by (Brereton et al., 2007) along with other four ones, which were discarded because they didn't provide significant results or most of their results are included in the above digital libraries. In addition, we used SpringerLink because it provides interesting papers working on this topic. At the end, the search query returned a total of 602 papers.

Figure 2 shows the number of papers per source and per year. The figure clearly depicts that, from 2011 to 2015, there was an increment in the number of papers related to traffic engineering in SDN. Only ACM presented a small decrease in 2015 compared with the number in 2014.

**Figure 2. Number of papers per source and per year**

In general, the total number of papers per year corroborate the increase of research related to traffic engineering in SDN during the last years (see Figure 3).

**Figure 3. Total number of papers per year**

To reduce the number of papers, we applied a selection process based on the following criteria. Include papers that propose traffic engineering solutions for managing SDN, and exclude review papers, such as systematic reviews, surveys, and tutorials, and papers not subject to peer review.

First, we selected the candidate papers by reviewing the title, abstract, and keywords. If these components presented poor information, then we reviewed the conclusions. As a result, we excluded almost 200 papers, for a total of 407 candidate papers.

Then, we continued with the selected papers, a selection that is currently in progress. This selection is based on the review of the introduction and conclusions of each paper. If this components lack of relevant information, then review the whole paper. So far, we have selected 86 papers from 2015.

**Figure 4. Traffic engineering in SDN regarding the FCAPS+P model**

From this partial selection, we present some initial results. First, the classification of papers regarding the FCAPS+P model. Figure 4 shows that (i) most of the work focuses on addressing performance issues, specially load-balancing and congestion-avoidance, (ii) an important number of papers work on fault challenges, particularly fault tolerance and anomaly detection, (iii) it is very uncommon the research on accounting problems, where we find one that copes with usage capabilities; and (iv) configuration, programming, and security has not been completed addressed, representing an interesting research opportunity. Current works in configuration focus on policy update and time-based configurations; in programming, papers mainly perform checking invariants and debugging errors; and in security, works concentrate on denial of service and malicious bots. It is noteworthy that addressing configuration issues may support other functional areas, such as performance and fault, and that recent trends in configuration propose an Intent-Based Networking (IBN).

Regarding the mechanisms for traffic engineering in SDN (see Figure 5), most of the work focus on traffic monitoring without performing additional actions. The figure shows that traffic collection, classification, steering, and prediction are interesting research opportunities. It is important to mention that traffic steering is more related to Network Function Virtualization, or NFV, which can represent an interesting use case for SDN.

**Figure 5. Traffic engineering mechanisms in SDN**

Through this data extraction, we found that it will be also important to classify the papers regarding the network context, the data context, and the outcome, in order to focus our proposal. For example, campus networks might be an small SDN environment for Big Data requirements, unlike datacenters and Wide Area Networks.

1. **Conclusions**

As conclusions, first, that applying traffic engineering for solving SDN configuration issues represents an interesting research topic. Second, that traffic prediction in SDN presents an attractive research opportunity. And third, that complementing the systematic mapping will provide better insight about traffic engineering in SDN, allowing to solve the following questions: which network context should be addressed in this project?, which are the specific techniques for conducting traffic collection, classification, and prediction in SDN?; and how many papers use Big Data for applying traffic engineering in SDN?

1. **Ph.D. Study Plan**

The Ph.D. study plan is as follows. For 2016, *(i)* submit a survey paper about Big Data for managing SDN to the journal IEEE Surveys and Tutorials, *(ii)* complete the courses Research Seminary I and Thesis III (Ph.D. Thesis Proposal), *(iii)* perform the research internship at the University of Waterloo under the supervision of Professor Raouf Boutaba, a distinguished researcher in the area of network management, *(iv)* submit a conference paper with the initial results of the research internship, perhaps to COMPSAC 2017; and *(v)* optionally, submit the systematic mapping study through a paper for a journal or conference with higher acceptance rate. For 2017, submit one journal paper and one conference paper, complete the course Research Seminary II, and start Teaching Practice. For 2018, submit one journal paper and one conference paper, complete the credits of Research Seminary III and Teaching Practice, and, at least, submit the Ph.D. thesis.

**References**

Agrawal, D.; Bernstein, P.; Bertino, E.; Davidson, S.; Dayal, U.; M., F. …, Widom, J. (2012), 'Challenges and Opportunities with Big Data', Technical report, Computing Community Consortium, A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. http://cra.org/ccc/resources/ccc-led-whitepapers/.

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.

Brereton, P.; Kitchenham, B. A.; Budgen, D.; Turner, M. & Khalil, M. (2007), 'Lessons from Applying the Systematic Literature Review Process Within the Software Engineering Domain', *J. Syst. Softw.* **80**(4), 571--583.

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.

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.

Jiang, J.Aggarwal, C. C. & Zhai, C., ed., (2012), *Mining Text Data*, Springer US, Boston, MA, chapter Information Extraction from Text, pp. 11--41.

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'(C11-728045-03), Technical report, Cisco.

Kitchenham, B.; Pearl Brereton, O.; Budgen, D.; Turner, M.; Bailey, J. & Linkman, S. (2009), 'Systematic literature reviews in software engineering – A systematic literature review', *Information and Software Technology* **51**(1), 7--15.

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.

Labrinidis, A. & Jagadish, H. V. (2012), 'Challenges and Opportunities with Big Data', *Proc. VLDB Endow.* **5**(12), 2032--2033.

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

Marr, B. (2014), 'Big Data: The 5 Vs Everyone Must Know', [Online]. Available: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know.

Mayer-Schönberger, V. & Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt.

Nguyen, H. (2014), 'Data Science vs Data Engineering', [Online]. Available: http://insightdataengineering.com/blog/Data\_Science\_vs\_Data\_Engineering.html.

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

ONF (2012), 'Software-Defined Networking: The New Norm for Networks', Technical report, Open Network Foundation.

Oracle (2013), 'Ideas Economy: Finding Value in Big Data', *The Economist*, 2--19.

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.

Petersen, K.; Feldt, R.; Mujtaba, S. & Mattsson, M. (2008), Systematic Mapping Studies in Software Engineering, *in* 'Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering', British Computer Society, Swinton, UK, UK, pp. 68--77.

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.

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.