

Hybrid IP/SDN Routing for Inter-Data Center Communications

Vítor Pereira
Centro Algoritmi
Department of Informatics
University of Minho
Email: vitor.pereira@algoritmi.uminho.pt

Miguel Rocha
Centre of Biological Engineering
Department of Informatics
University of Minho
Email: mrocha@di.uminho.pt

Pedro Sousa
Centro Algoritmi
Department of Informatics
University of Minho
Email: pns@di.uminho.pt

Abstract—Internet Service Providers (ISPs) and dedicated inter-Data Center Wide Area Networks have been exploring Software-Defined Networking (SDN) features to achieve a high utilization of the available resources. This work proposes a scalable hybrid IP/SDN routing model, and optimization procedures fostered by Evolutionary Computation algorithms, to achieve near optimal network resources utilization under changing traffic requirements.

Keywords—Software-defined Networking, Hybrid IP/SDN networks, Network Resources Utilization, Evolutionary Algorithms.

I. INTRODUCTION

Software Defined Networking (SDN) offers many advantages over traditional networks, such as fine-grained control for data traffic flows and the enforcement of security policies. Wide Area Networks (WANs), that connect data centers, have been exploiting these advantages to achieve high-throughput with low-latency while making the infrastructures more resilient and adaptable to changes [1], [2]. However, a private WAN or an Internet Service Provider (ISP) usually handles millions of forwarding entries, and SDN-enabled switches can typically support only tens of thousands. Possible solutions to address this concern are to deploy a limited number of SDN-enabled devices alongside traditional network devices or leverage SDN with traditional routing. A network containing a mix of SDN and legacy network devices, or functionalities, such as the Google’s private SDN backbone network B4 [1], is commonly referred to as a hybrid IP/SDN network.

Hybrid IP/SDN networks can commonly be deployed in three different ways: 1) setup up SDN switches in a legacy network, that is, the network is built with two types of devices, legacy routers and SDN switches, 2) hybrid IP/SDN switches having both SDN switching and legacy switching functionality, and 3) deploying hybrid IP/SDN switches on legacy networks as a combination of the previous. The last ones have particular advantages that result from combining the time tested-maturity of Interior Gateway Protocols (IGP) with the flexibility of SDN. Hybrid IP/SDN switches add flexibility to traditional IGP by enabling arbitrary forwarding paths that overcome distributed routing protocols limitations. On the other hand, IGPs give SDN networks scalability and robustness. Enabling

IGPs destination based and SDN match/action forwarding strategies allows to dramatically reduce the number of switch entries that need to be managed by a controller.

II. ROUTING MODEL FOR HYBRID IP/SDN

The mathematical model of a data communications is defined as follows. Consider the capacitated directed graph $G = (N, A, c)$, where N and A denote the sets of nodes and arcs, representing routers and links, the last with a capacity constraint c_a , $a \in A$. Given a demand matrix D , with entries d_{st} that model the amount of traffic to be sent from a source s to a target t , the problem is to route these demands along network paths while minimizing congestion. The network congestion is evaluated using the piecewise convex continuous function Φ^* proposed by Fortz and Thorup [3]. Also used as optimization objective, Φ^* assigns to each link a , $a \in A$, a congestion cost Φ_a as a function of the utilization l_a/c_a . The goal is then to minimize the normalized sum of all link’s utilization costs, by setting adequate IGP link weights and SDN forwarding rules.

Load balancing mechanisms have an important role in improving network resources utilization and responding to unforeseen events such as traffic variation and link failure. We extend the Distributed Exponentially-weighted Flow Splitting (DEFT) [4] routing protocol to enable traffic forwarding along non-shortest paths and accommodate corrections on traffic load balancing. The shortest-path distance from a node u to a node t , when traffic is routed using a next-hop v , is here expressed as $w_{u,v} + d_v^t$, where d_v^t is the shortest-path distance from the next-hop v to t , and $w_{u,v}$ is the weight of the link (u, v) . The extra length of the path from u to t through v , when compared to the shortest path with the same origin and destination, d_u^t , can thus be expressed by Equation 1, and denoted as $h_{u,v}^t$. The exponential function $\Gamma(h_{u,v}^t)$, Equation 2, indicates the relative amount of traffic destined to t that node u will forward via outgoing link (u, v) , and decreases with the extra length $h_{u,v}^t$ of a path. The proportion of traffic with destination t to be forwarded at u using the outgoing link (u, v) is then computed by Equation 4.

$$h_{u,v}^t = d_v^t + w_{u,v} - d_u^t \quad (1)$$

$$\Gamma(h_{u,v}^t, p) = \begin{cases} f(h_{u,v}^t, p), & \text{if } d_v^t < d_u^t \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$f(h_{u,v}^t, p) = \begin{cases} 1, & \text{if } h = 0 \\ e^{-h_{u,v}^t \times p}, & \text{if } p \geq 0 \text{ and } h \neq 0 \\ 1 - \frac{p}{h_{u,v}^t}, & \text{if } p < 0 \text{ and } h \neq 0 \end{cases} \quad (3)$$

$$P(h_{u,v}^t) = \frac{\Gamma(h_{u,v}^t)}{\sum_{(u,i) \in A} \Gamma(h_{u,i}^t)}, \quad (4)$$

The manipulation of the p variable (node- p variable) in Equations 2 and 3 permits to change the amount of traffic forwarded on non-shortest-paths without any change to the installed IGP link weights and forwarding paths. A different p is assigned to each node which can be optimized to improve hop by hop traffic load balancing.

A. Routing Model

The proposed traffic load balancing scheme suggests a natural separation between shortest path and non-shortest path traffic forwarding. While all traffic that travels over edge-to-edge shortest paths is to be managed by the IGP, traffic that makes use of non-shortest paths is to be managed via SDN. When all nodes are hybrid IP/SDN switches, ingress edge routers classify incoming traffic as suitable to be routed by IP routing protocols or via SDN. A Local Management Entity (LME) identifies which flows are candidate to be routed via SDN using a match criterion (e.g., source/destination, type of service, protocol, in/out TCP port). If a flow is identified to be SDN routable, its first packet is sent to a controller who will select the forwarding path. The SDN controller ensures the traffic splitting ratios on non-shortest paths and install the required rules on the Ternary Content-Addressable Memory (TCAM) of all SDN enabled switches on the selected path. If the controller decides that a flow should be forwarded by the IGP, no rules are installed.

B. Incremental Deployment of Hybrid IP/SDN Networks

The positioning and number of SDN enabled nodes represents also, in itself, an optimization problem. The network topology and resources constraints, traffic requirements and budget are some conditions that need to be considered when selecting which and how many legacy nodes should be replaced by SDN enabled devices. In this context, we also propose an approach for incremental deployment of SDN enabled nodes. We simultaneously optimize the IGP configuration, OSPF or IS-IS link weights, and the SDN nodes placement. The idea is to improve legacy routing by performing non-equal load balancing on non-shortest path links, and thus improve the distribution of traffic in the network available resources. This new optimization problem is formulated as follows: given a defined number k of hybrid IP/SDN switches to be deployed, a network topology and traffic necessities, the aim is to find the location where SDN functionalities

TABLE I
COMPARISON OF CONGESTION AND MLU VALUES WITH OPTIMIZED CONFIGURATIONS.

Topology	Nodes	Links	OSPF/ECMP	Hybrid IP/SDN	MCF
<i>Rand30₂</i>	30	110	1.42 (72%)	1.38 (67%)	1.30 (67%)
<i>Rand30₄</i>	30	220	1.67 (88%)	1.64 (78%)	1.23 (67%)
<i>Rand50₂</i>	50	194	1.68 (94%)	1.61 (92%)	1.39 (67%)
<i>Rand50₄</i>	50	280	3.67 (112%)	2.38 (98%)	1.22 (67%)

should be installed as well as a link weights configuration that allow to minimize the network congestion. In incremental deployments of hybrid IP/SDN switches flows are classified to be routed by the IGP at each individual hybrid IP/SDN switch. This task can be performed solely by the LME without the need to send any packet to the controller. The role of the controller is to evaluate the network performance and implement load balancing corrections on non-shortest paths whenever necessary.

III. EXPERIMENTAL RESULTS

A. Full Deployment of hybrid IP/SDN Nodes

We devised a set of experiments to evaluate the optimization method considering distinct synthetic network topologies with links' capacity varying uniformly in the interval [1; 10] Gbits and whose characteristics are summarized in Table I. For each network we randomly generated a set M of 100 traffic matrices (TMs) which model heavy traffic requirements between each pair of source/destination nodes following the scheme presented in [5].

The first set of experiments, whose results are presented in Table I, compares the network resources when the networks are configured with optimized OSPF settings and with the proposed hybrid IP/SDN routing model, when considering traffic necessities modeled as a single TM randomly selected from the set M . The network resources utilization is evaluated using both the congestion metric Φ^* and the Maximum Link Utilization (MLU). We resort to the Non-dominated Sorting Genetic Algorithm (NSGAI) [6], to simultaneously optimize both objectives obtaining a set of solutions, a Pareto Front, each with a different trade-off between the objectives. The operators used by the EA to evolve solutions are: Random and Incremental/decremental mutations, Uniform and Two point crossovers. Solutions are encoded as a set of integers, link weights, taken from the reduce range [1,20], to narrow the size of the search space and favor the occurrence of ECMP. Also included in Table I are the congestion and MLU values of the Multi-Commodity Flow (MCF) problem.

The results, that average 30 experiments for each topology with a stopping criterion of 1500 iterations, show that the proposed hybrid IP/SDN routing model outperforms optimized OSPF configurations. By forwarding portions of traffic along non-shortest path links, the hybrid routing model is able to achieve a better utilization of network resources and, simultaneously, reducing the MLU. However, the hybrid model configurations do not perform as well as those obtained by solving the MCF problem. In the hybrid model, after each hop,

TABLE II
AVERAGE CONGESTION AND MLU VALUES WITH AND WITHOUT LBC.

Topology	<i>Rand30₂</i>	<i>Rand30₄</i>	<i>Rand50₂</i>	<i>Rand50₄</i>
One TM	3.44 (107%)	44.64 (138%)	6.94 (116%)	44.64 (138%)
One TM+LBC	2.10 (94%)	22.84 (125%)	3.93 (107%)	22.84 (125%)
Two TMs	1.91 (93%)	6.79 (119%)	5.55 (115%)	6.79 (119%)
Two TMs+LBC	1.69 (85%)	3.68 (105%)	3.58 (106%)	3.68 (105%)

traffic becomes closer to the destination when considering the shortest path distance. The LP configuration, however, allows traffic to be routed along longest path. But although the LP achieves a better utilization of network resources, it is not without disadvantages as it leads to an increasing latency.

The second set of experiments explores the traffic Load Balancing Correction (LBC) on non-shortest paths to adapt the hybrid IP/SDN configurations to distinct traffic requirements. In this context, we compare two initial configuration that consider, respectively, one and two TMs. In the first case, the configuration is optimized for a single TM, the average of M . The second use k-means clustering [7] to choose two representative TMs, D_1 and D_2 , from M . To avoid a poor clustering, the initial centroids are two opposite vertices of the higher-dimensional polyhedra which bounds the TMs space M . The initial configuration is then obtained by running a multi-objective EA which simultaneously minimizes Φ_1^* and Φ_2^* , the normalized congestion metric Φ when considering the TMs D_1 and D_2 respectively. The solution with the smallest tradeoff value $\alpha \times \Phi_1^* + (1 - \alpha) \times \Phi_2^*$ is selected from the Pareto optimal set, where α is the real value that reflects the proportion of elements in each of the two clusters.

The average congestion values of 10 experiments, for each network, are presented in Table II. Results show that initial configurations that consider two TMs are able to better accommodate changing traffic necessities as they are not tied to a particular TM. To further ameliorate the distribution of traffic, node-p values are also optimized (LBC). Although insufficient in certain cases, this approach always introduces improvements on congestion values.

B. Incremental Deployment of Hybrid IP/SDN Nodes

SDN is typically not fully deployed in networks due to several reasons, such as a limited budget and the fear of downtime during the transition to SDN. In this set of experiments we evaluate the incremental deployment of the proposed hybrid IP/SDN model varying in the number k of hybrid node (10, 20, 30 and 40% of the total number of topology nodes). We also include results where all nodes are legacy routers (0%) as well as results where all nodes are hybrid IP/SDN switches (100%). The results presented in Table III are the mean of 30 runs of the EA optimization process.

Results show that not all nodes need to implement SDN functionalities to attain congestion values equivalent to those obtained when all nodes are hybrid. A percentage of 40% of hybrid IP/SDN nodes, or less, is sufficient to achieve such congestion levels. With only 20% it is already possible to observe a significant improvement in congestion. The algo-

TABLE III
CONGESTION BY PERCENTAGE OF HYBRID IP/SDN NODES.

Topology	Nodes	Links	0%	10%	20%	30%	40%	100%
<i>Rand30₂</i>	30	110	1.42	1.40	1.39	1.38	1.38	1.38
<i>Rand30₄</i>	30	220	2.70	2.48	1.90	1.86	1.72	1.71
<i>Rand50₂</i>	50	194	1.68	1.66	1.64	1.63	1.62	1.61
<i>Rand50₄</i>	50	380	3.67	2.45	2.41	2.4	2.39	2.38

rithm tends to select as hybrid nodes the ones with the higher in/out degree and contribute more significantly to the Traffic Engineering tasks.

IV. CONCLUSIONS

Software Defined Networking provides the ability to move network management to a centralized controller allowing to exercise per-flow traffic engineering and better usage of networks' infrastructures. However, SDN has scalability issues and added complexity in managing network functions. Hybridizing SDN with a link state protocol solves the scalability problems and attenuates complexity. We proposed an hybrid IP/SDN routing model coupled with a traffic load balancing correction mechanism for inter-data center WANs. We showed the validity of the proposed hybrid IP/SDN routing model in full and incremental deployment contexts. In the last, the results suggest that 30% of hybrid nodes, or less, offer an utilization of network resources nearly equivalent to those where all nodes are hybrid. The optimization framework used in the experiments is publicly available at <http://www.bio.di.uminho.pt/netopt>.

ACKNOWLEDGMENTS

This work has been supported by national funds through FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019.

REFERENCES

- [1] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hözlze, S. Stuart, and A. Vahdat, "B4: Experience with a globally-deployed software defined wan," *SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 3–14, Aug. 2013.
- [2] C.-Y. Hong, S. Kandula, R. Mahajan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer, "Achieving high utilization with software-driven wan," in *Proceedings of the ACM SIGCOMM 2013 Conference*, ser. SIGCOMM '13. New York, NY, USA: ACM, 2013, pp. 15–26.
- [3] B. Fortz and M. Thorup, "Internet traffic engineering by optimizing ospf weights," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*. IEEE, 2000, pp. 519–528.
- [4] D. Xu, M. Chiang, and J. Rexford, "Deft: Distributed exponentially-weighted flow splitting," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 71–79.
- [5] V. Pereira, M. Rocha, and P. Sousa, "Optimizing segment routing using evolutionary computation," in *FNC/MobiSPC*, ser. Procedia Computer Science, vol. 110. Elsevier, 2017, pp. 312–319.
- [6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, July 2002.