

SOFTWARE-DEFINED WIDE-AREA NETWORKS IN THE
COEVOLUTION OF NETWORK AND COMPUTATIONAL
BANDWIDTH

Iliyan Nenov

Received on September 16, 2022

Presented by S. Margenov, Corresponding Member of BAS, on October 25, 2022

Abstract

Software-Defined Wide-Area Networks (SD-WAN) are a technology that utilizes programmatic configuration to maximize network throughput. The main problem to achieve efficient use of network bandwidth with SD-WAN is the network delay (latency) in the communication. A network could be further slowed down by not enough memory or CPU resources. Here we conduct experiments, investigating the capability of an SD-WAN to maximize the network throughput at increasing levels of latency. We find that to a limited extent the computational optimization at the software side could offset the network limitations. At higher network latency the SD-WAN adds little value. Further increase of the capacity can only be achieved by hardware upgrades of the network bandwidth. Such solutions may be considered drivers of a coevolutional relationship between computational and network bandwidths, whereby the network use becomes more efficient.

Key words: network, throughput, bandwidth, coevolution, software-defined wide-area networks

Introduction. The amount of data delivered over a network is defined as throughput. It depends on the network's maximum possible data transfer rate (bandwidth) and on the delays of data transfers (latency). As a result, the throughput is always less than the bandwidth and low-latency networks use the bandwidth more efficiently. Network latency depends on many factors – traffic load, configurations, routing, distance and others. But these delays could be minimized to acceptable levels with the so-called software-defined wide-area

networks (SD-WAN). It is an approach to network management that enables dynamic programmatically efficient network configuration to achieve better network throughput [1].

Over the last decades network bandwidths have followed exponential growth, summarized by Edholm's law [2] – similar to Moore's law in microelectronics. Furthermore, the improvements in those two bandwidths have been a factor for each other's growth, i.e. they are in coevolution.

In biology, coevolution is the process by which two or more species evolve in tandem by playing a role in each other's evolution. The same principles are studied outside biology and the field of technology is no exception [3]. Popular examples for coevolution in technology are hardware and software.

Like hardware and software, network and computational bandwidths are in coevolutional relationship [4]. Upgrading the network bandwidth brings more data which pressures the computational bandwidth to upgrade too. Over time, we observe repeating patterns – while data swings searching the optimal destination for computation, the dominating software architecture changes accordingly: mainframes (centralized), client-server (distributed), cloud/mobile (centralized), edge-computing (distributed).

In this paper, we show how the network bandwidth can be used efficiently with SD-WAN solution and the limitations for achieving this with different types of software. Such technology could balance properly the supply and demand of data among computational nodes. These findings are important because having a better understanding about the limits of effective use of the network bandwidth will allow to construct better models for the future developments of network and computational bandwidths and the factors they play for each other's evolution.

One of the widely used methods for scaling fibre-optic network bandwidth is wavelength-division multiplexing (WDM) [5]. In general, the method relies on different wavelengths (i.e. light colours) to multiplex many optical carrier signals over a single optical fibre. Inside the fibre, the signal travels at 0.7 times the speed to light in vacuum. This does not leave much potential to scale with speed improvements thus the bandwidth upgrade depends mainly on the multiplexing.

Bottlenecks in the bandwidths and the role of SD-WAN solutions.

On data-centre level, at network bandwidth of 100 Gbit/s, DRAM latency appears as a major bottleneck in the system [4]. In recent years, transistors have reached their physical limits in regard to size. As a result, we observe mass adoption of multi-core architectures and especially GPU-based computations [6]. This drives the software to adopt parallel and distributed computing paradigms to meet the demands of the data and the users. Multi-core and GPU-based computations are now a standard and they are used to parallel the computation and thus to resolve the bottlenecks at the computational side. In parallel computing, one could observe scale limiting factors defined by Amdahl's law [7], yet the distributed architectures could outweigh most of these.

Most data-centres today are connected with 20 Gbit/s to 40 Gbit/s network bandwidth. About 10% of the data-centres are connected with 400 Gbit/s ports and some of the big vendors for hyperscale computing are considering the adoption of 800 Gbit/s Ethernet devices. These devices are built on top of the already adopted IEEE 802.3bs standard for 400 Gbit/s Ethernet. If such network bandwidth is widely enabled, there are three options for overcoming DRAM bottleneck: (1) increasing the number of physical cores (not scalable), (2) increasing the average size of internet packet (not generalizable), or (3) building specialized hardware (high cost), e.g. architecture that merges DRAM and CPU [4]. Therefore, WAN optimization techniques seems to provide a solution that gives time to both computational and network bandwidths to adapt and scale in response to the other.

On the edge of the network the challenges are quite different. The average bandwidth looks much smaller at 20–40 Mbit/s for mobile networks and 50–75 Mbit/s for broadband internet connections. In addition, for some networks the latency is too high or unpredictable reaching peaks around 200 ms. Inspired by the digitalization and Industry 4.0, we see trends for significant bandwidth upgrades at the edge networks. For example, the 5G networks are expected to be used between 75% and 90% of the population for some of the most developed countries [8]. Therefore, on the edge side, SD-WAN role could be even more important because more data traffic is expected to happen here.

A network communication experiment. The goal of our study was to identify the limits for reaching an effective use of network bandwidth with SD-WAN solution. We identified three core factors that determine the efficient use of the network bandwidth: the type of the software, the physical distance and the network latency. All other factors come down to these three. In our experiments we tested these core factors to the extreme in order to identify the limits of SD-WAN capabilities.

Software is an obvious factor because it determines the demands for network bandwidth. The distance in communication cannot be avoided as it is the path that has to be travelled in order to establish the communication. Lastly, all the bottlenecks in the system design, malfunctions and sub-optimal configurations lead to latency.

In our experiments, we used four types of internet applications that test the limitations of the SD-WAN solution in different areas. The applications we used are developed by a large software vendor of enterprise applications and used by thousands for companies:

1. **Web application** – tests the limits of the SD-WAN solution to apply smart caching, prefetching and the like.
2. **Multi-view application** – tests the limits to balance different strategies at once to maximize the throughput for many web-applications visualized in a dashboard.

3. **Data-analytics** – tests the limits for managing smart routing and random latency peaks.
4. **Content management system** – tests the limits of the SD-WAN solution to deal with various protocols in parallel.

With these four types of applications we executed two experiments with SD-WAN solution: the first we named “marathon”, the second “sprints”. In the first, we measured the performance gains with SD-WAN solution on stable and big-distance connection. In the second, we did the same measurements on a connection with simulation of increasing latency.

For our first experiment, we used a real long distance communication (approx. 10 250 km) between two data-centres, in Germany and Singapore, with no direct connection between them. The connectivity had less than 5% packet drops and the average latency, during 10-hours experimental load, was 167 ms – typical for this route and such distance. With this experiment we wanted to test how stable the results are with SD-WAN solution over extensive period of time, i.e. a marathon-run experiment.

For the second experiment, we used specialized equipment to simulate various network conditions like latency, packet drops, and connection failures. The effective use of network bandwidth depends on many factors, e.g. security policies, software problems, packet drops, equipment limitations and the like. Usually all these factors lead to an increase in the latency at the end. Therefore, it became one of the key factors for our experiments. In our second experiment, we simulated different latency levels: 100 ms, 150 ms, 200 ms, 250 ms, 300 ms, 350 ms, 500 ms to see the limits of the SD-WAN solution in achieving effective use of network bandwidth, i.e. sprint-runs experiment.

The execution of the experiments was fully automated with LoadRunner software to guarantee reproducible results. This allowed to distribute the load generators on multiple machines and to eliminate the bottlenecks in our experimental setup.

Here, throughput is measured as the number of requests per unit of time. In this way it is represented as relative to the response time of the application and is abstracted from network configuration details that could be changed by the SD-WAN. With such experimental setup it is possible to test the SD-WAN in various networks and to compare the results.

For example, if we achieve a response time of 1000 ms this means that the throughput is 60 requests per minute. Every request triggers data traffic over the network that could be measured in Mbit/s with LoadRunner. Accumulating the whole traffic, the LoadRunner tool gives data about the throughput in Mbit/s for the whole duration of the experiment. This is the same throughput but in different units of measurement. In this way, it can be calculated as a percentage of the network bandwidth.

Achieving more effective use of the bandwidth means higher throughput (i.e. closer to 100% of the bandwidth) and lower response time (i.e. more requests per unit of time). Hence in our experiments response time and throughput are the same and they can be used interchangeably. We measured the bandwidth through special software that was installed on both sides of the communication channel.

SD-WAN effect over long distance communication. For the long-distance experiment – the marathon – we did 10 hours long test executions, one with each of the four application types. The LAN response time is over the LAN network in the data-centre, i.e. without the overhead of the WAN. Figure 1 shows the response time for the web application software with all users ramped up. The measurements of the other three types of software were very similar. As expected, SD-WAN brings significant performance improvements to this test with stable network latency.

Table 1 summarizes the response times of the 10-hours load test. Comparing the mean values, the response time with SD-WAN is 2.91 times lower than the case with no SD-WAN. This is the ratio between the mean response time with and without SD-WAN after subtracting the LAN response time from both, thus leaving only the network effect. This means that the SD-WAN solution is able to increase the throughput about three times and to reach 93% utilization of the network bandwidth.

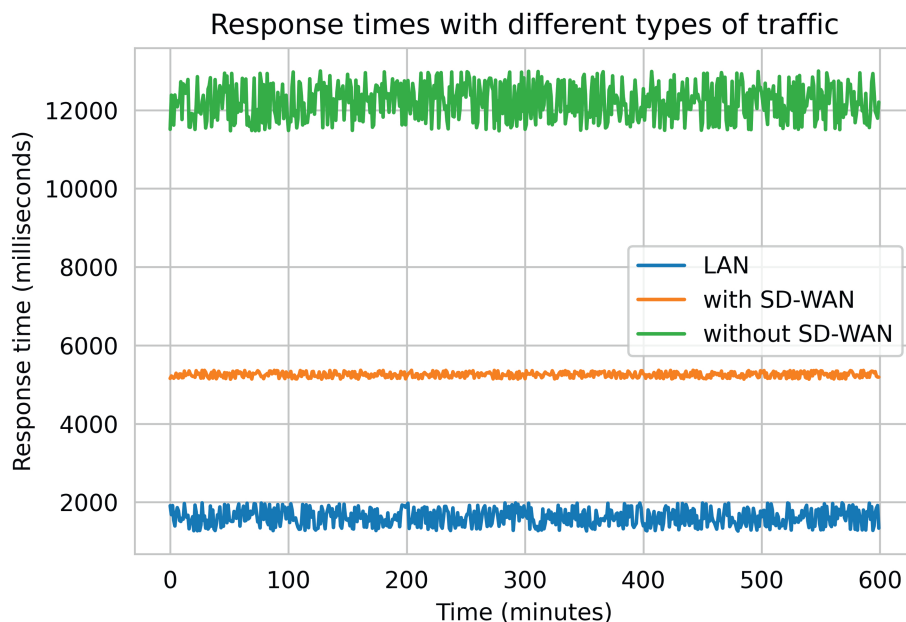


Fig. 1. Response times for web application with three types of communication (1) LAN at data-centre (base-line), (2) with SD-WAN between Frankfurt and Singapore, (3) without SD-WAN between Frankfurt and Singapore

T a b l e 1

Statistics of the response times on the web-application

	Mean response time, ms	Std. deviation ms	Average throughput Mbit/s	Bandwidth utilization, %
LAN (base-line)	1611	210	–	–
Without SD-WAN (GER-SG)	12 219	444	23.4	31.60
With SD-WAN (GER-SG)	5252	65	68.8	93.00

In addition, SD-WAN solution decreased the standard deviation in the response times compared with the LAN case. It was able to save traffic and thus to reduce the load on the server. As a result, we measured 8% less average CPU consumption in this case.

Network throughput with different latency levels. Using network equipment we simulated seven levels of network latencies: 100 ms, 150 ms, 200 ms, 250 ms, 300 ms, 350 ms, 500 ms. We tested the efficient use of network bandwidth in seven 1-hour long experimental runs – the sprints. For each latency, we loaded all four types of applications with and without SD-WAN. At the end, we charted

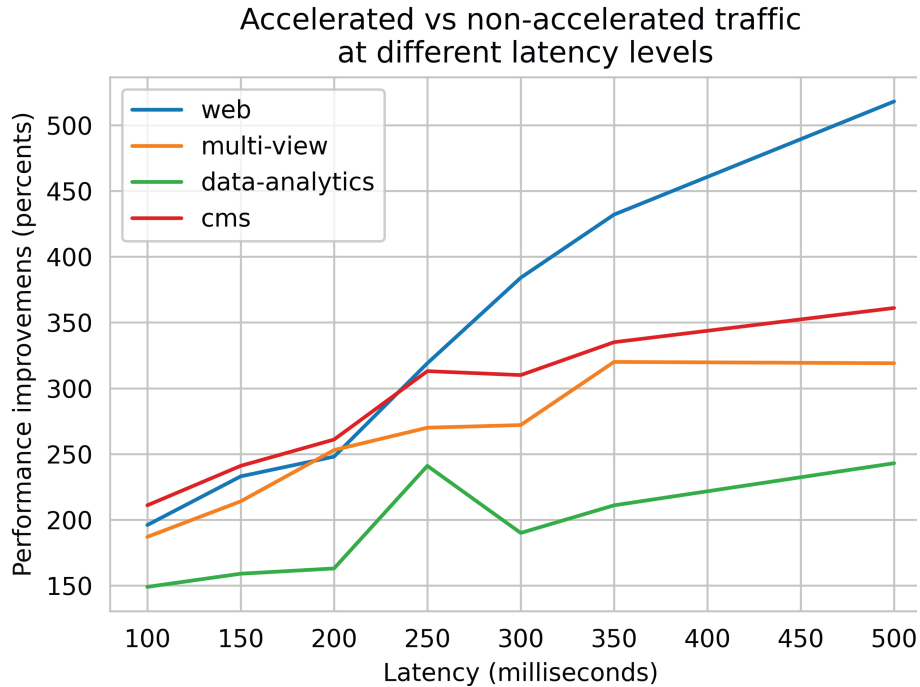


Fig. 2. Performance improvements of the response time with SD-WAN solution at different network latency

the improvement of the response time (in percentage). Again, we used fully automated setup with the LoadRunner software and the same types of applications as in the “marathon”.

There could be various causes for the network latency but from computational perspective two factors are of special interest: insufficient memory and heavy application traffic. The former is a sign of deficient computational bandwidth while the second is a sign for deficient network bandwidth. Latency bigger than 250 ms is considered high and could be often observed with 2G/3G mobile networks and satellite connections.

The results (see Fig. 2) show that all four applications benefit from the optimization techniques and achieve performance improvement on every latency level. At lower delays the performance improvements grow linearly with the latency increase. The only exception here is the data-analytics application. Between 250 ms and 350 ms we observe a chasm after which the performance improvements grow at sub-linear pace for all tested applications.

In our marathon experiment we observed that at latency in the range of 150–200 ms the SD-WAN solution achieves throughput that is 93% of the network bandwidth. Therefore, at latency levels above 250 ms there is not much gain in efficient use of network bandwidth. As a result, the performance improvements start to grow at sub-linear scale requiring more network and computational bandwidth to scale further.

Conclusion. The efficient use of the network bandwidth allows more data to be transferred, which in turn stimulates the upgrade of both network and computational resources. In this study we showed that through SD-WAN it is possible to bring the network throughput closer to the bandwidth limits. Thus the only way to scale up performance is by coevolutional development of network and computational bandwidth.

REFERENCES

- [1] XIE J. et al. (2019) A survey of machine learning techniques applied to Software Defined Networking (SDN): Research issues and challenges, *IEEE Commun. Surv. Tutorials*, **21**(1), 393–430, doi: 10.1109/COMST.2018.2866942.
- [2] CHERRY S. (2004) Edholm’s law of bandwidth, *IEEE Spectr.*, **41**(7), 58–60, doi: 10.1109/MSPEC.2004.1309810.
- [3] SEN S., E. AYDOGAN, A.I. AYSAN (2018) Coevolution of mobile malware and anti-malware, *IEEE Trans. Inform. Forensic Secur.*, **13**(10), 2563–2574, doi: 10.1109/TIFS.2018.2824250.
- [4] THOMAS S., R. MCGUINNESS, G.M. VOELKER, G. PORTER (2018) Dark packets and the end of network scaling. In: *Proc. 2018 Symp. Architectures for Networking and Communications Systems*, Ithaca, New York, 1–14, doi: 10.1145/3230718.3230727.

- [5] PATKI P. G. et al. (2020) Recent progress on optical regeneration of wavelength-division-multiplexed data, *IEEE J. Select. Topics Quantum Electron.*, **27**(2), 1–12, doi: 10.1109/JSTQE.2020.3025482.
- [6] FANG J., C. HUANG, T. TANG, Z. WANG (2020) Parallel programming models for heterogeneous many-cores: a comprehensive survey, *CCF Trans. HPC*, **2**(4), 382–400, doi: 10.1007/s42514-020-00039-4.
- [7] RICHINS D., T. AHMED, R. CLAPP, V. JANAPA REDDI (2018) Amdahl’s law in big data analytics: Alive and kicking in TPCx-BB (BigBench). In: 2018 IEEE Int. Symp. High Performance Computer Architecture (HPCA), Vienna, 630–642, doi: 10.1109/HPCA.2018.00060.
- [8] JAHNG J. H., S. K. PARK (2020) Simulation-based prediction for 5G mobile adoption, *ICT Express*, **6**(2), 109–112, doi: 10.1016/j.ict.2019.10.002.

Faculty of Economics and Business Administration
Sofia University “St. Kliment Ohridski”
125 Tsarigradsko Shosse Blvd, Bl. 3
1113 Sofia, Bulgaria
e-mail: iliyann@uni-sofia.bg