Overturning the Eight Fallacies of Distributed Computing with the Octopus Edge Network

(invited paper)

Sébastien Bigo
Nokia Bell Labs, Route de Villejust, 91620 Nozay
sebastien.bigo@nokia-bell-labs.com

Abstract: Named after the mollusk nervous system, the Octopus network is a sequel of low-latency ultra-reliable edge networks. Its dynamic and deterministic characteristics open a new era for distributed computing by breaking the notorious eight fallacies. © 2020 Nokia

1. Introduction

In a metaphoric animal representation of converged Information Technologies (IT) and telecom infrastructures, the nervous system (the infrastructure which performs a function) makes decisions after massive cooperation of neurons (servers). It interacts with the physical world via its sensors (connected "things", e.g. robots, cars, connected objects, interfaces with users). Most of the neurons are located inside the brain (the central cloud), which performs the most complex decisions, but is remote and therefore not capable of instantaneous responsiveness. Another set of neurons in the spine (servers in the edge cloud) are located close to the sensors for reflex reactions (simpler, time-sensitive applications). The axons of the neurons (optical fibers) transport neural influx between neurons and sensors and across neurons. The synapses (optical and electrical switches) transfer or not the neural influx to the other neurons. Overall, the nervous system is flooded with neural pulses predominantly generated across the neurons of the brain (i.e. server-to-server traffic largely dominates inter-user or user-to-server communications). Today IT and telecom infrastructures are very much like the primitive nervous system of a lobster, with a (limited) brain and no spine for time-sensitive reactions. In ten years from now, we conjecture that it will mutate into that of one of the longest-lasting creatures on earth: the octopus. An octopus has eight smaller brains (edge data-centers) supervised by a ninth brain (central cloud). These smaller brains are distributed and located close to the sensors, inside the eight tentacles. The small brain off-loads the central brain by reacting fast to local stimuli (performing local time-sensitive applications), while the central brain orchestrates the more complex decisions. The overall flatness of the octopus cognitive system ensures remarkable cooperation and synchronization with superior resilience; the local brains continue to operate after cut off from the rest of the system, and the system can recover its integrity by growing (pay-as-you-grow) a new tentacle.

If organized like the octopus nervous system, the distributed information technology (IT) infrastructure of edge-clouds of the upcoming 5G networks will undoubtedly attract digital service providers into new business opportunities. One can easily predict that the largest potential for revenue creation will leverage its unmatched responsiveness. Large enterprises planning for the digital transformation of their manufacturing processes, e.g. car makers, have been among the first to foresee opportunities there for virtualizing the IT of their factory floor or for creating a digital twin to act on their factory floor through virtual-reality glasses, but they made clear that they expect latency to be strictly guaranteed and constant (i.e. with negligible jitter, 1-10µs [1]) for security reasons, not just across the interconnection network, but end-to-end[2]. All these scenarios pave the way to a world where the precedence of today's user-to-user or user-to-machine data traffic is taken over by machine-machine data traffic, where the word machine designates any connected "thing", but predominantly a server. In this world, we conjecture that mass market, bandwidth-hungry, best-effort applications, e.g. video streaming, coexist with time-critical, ultra-reliable applications over the same IT infrastructure. Regrettably, distributing IT over could steer the profitability of cloud businesses down because serves are more poorly utilized in average, unless machines are capable of efficiently cooperating across multiple locations. Cloud Operators are already working on this challenge for the type of applications hosted in their DCs [3].

In this paper, we particularly address the case of time-sensitive applications mentioned above, e.g. control of machines in a factory floor. For those applications, ultra-reliability and strictly guarantees of end-to-end delivery time and bandwidth (no timing jitter) are prerequisites. We show that an edge network consisting of a duly orchestrated concatenation of time-division multiplexed network technologies can successfully meet the end-to-end performance challenges for interconnecting time-sensitive machines, while simultaneously hosting mainstream best effort applications. We call this new type of network dynamic deterministic network (DDN). [4] While in short-term use cases of DDN, the word "machines" refers primarily to moving robots, DDN also opens considerable opportunities when the word "machines" designates servers. More specifically, 25 years ago, L. Peter Deutsch popularized a set of false assumptions that junior programmers invariably make when programming distributed applications and that are now widely known as the eight fallacies of distributed computing [5]. We show that DDN has the potential to shake and even break them, opening avenues for resurrecting alternative forms of real-time computing, where real time does not mean short in average but short always.

2. Dynamic deterministic networks

The hard challenge for DDN is to offer strict performance guarantees end-to-end, i.e. down to each application flow, to the relatively the small portion (typ. <10% of total traffic) of priority traffic which is time sensitive, while ensuring that the cost to carry the dominant best-effort traffic is weakly or not impacted. No congestion is allowed for priority flows. Congestion creates jitter, and ultimately packet loss. Hence, today's split of technologies, namely circuit switching for transport and packet switching for (statistical) multiplexing and aggregation at the edges is being challenged. Over-dimensioning switching capacity, i.e. performing statistical multiplexing at low load, as exemplified in today's central DCs, alleviates the probability of congestion and decreases jitter but cannot contain it when the number of priority flows competing for the same output exceeds a small amount, e.g. a few tens [4]. Reserving bandwidth over some circuit container, as in today's optical transport networks, seems like a natural work-around to process priority flows but unlike today's optical networks, it must be end-to-end and for any input to any output of the network. This requirement can drive costs very high unless the circuit containers is reconfigured nearly as quickly as flows come and go. Overall, the new distributed edge cloud dream network should provide the jitter of circuit switching with the dynamics of statistical multiplexing. While service turn-up times in optical transport networks have been drastically decreasing by the introduction of various innovations over time, it is interesting to notice that they have reached a point where the next disruption could allow for similar turn-up times for transport services as for IT services, so that service dynamics could be efficiently managed end-to-end for the first time in history [6].

When analyzing all the options on the table, we drew the conclusion that the common denominator for the new distributed edge cloud network should be scheduled, slotted and synchronous/isochronous (S^3), across all network segments, e.g. access, wireless, optical transport and DC. Some segments are already compliant, others require updates or hardening up to the servers themselves, while the overall combination of them should support end-to-end scheduling at scale. Slotted networks use time division multiplexing to temporally interleave data containers of fixed duration (slots), thereby allowing for synchronous or isochronous ultra-reliable delivery of time-critical applications. Trains of slots are like temporarily allocated circuits which can be reworked to host best effort (e.g. video) traffic opportunistically as easily as in today's IP/Ethernet.

3. Experimental demonstration

We recently demonstrated a generic distributed computing network, with low latency (~µs), allowing deterministic data delivery in time (~ns jitter) and capacity (no packet loss for conforming traffic), slicing, dynamic reconfiguration (~ms time), between devices (antennas, robots) and data centers or across data centers, through a transport network.

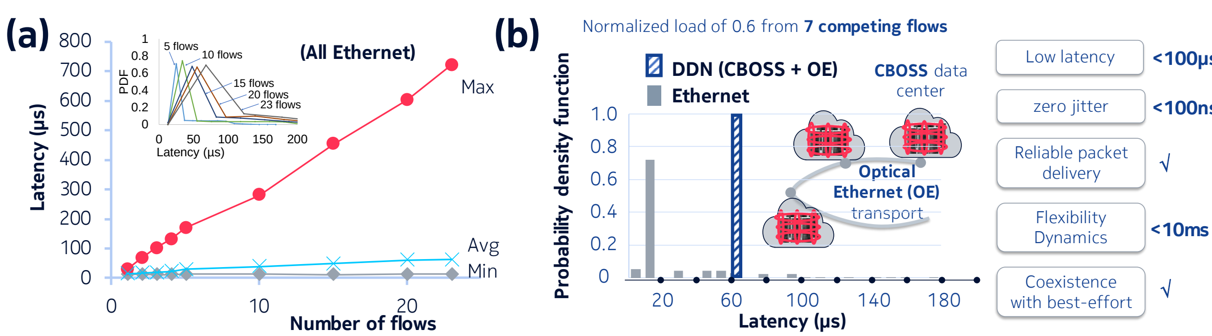

Fig. 1. (a) Latency over legacy Ethernet-based data center versus number of flows. Inset = corresponding probability density functions (b) Experimental characterization of DDN versus Ethernet distributed edge cloud prototype

We leverage two DDN technologies optimized for two network segments: Cloud Burst Optical slot switching (CBOSS) for interconnecting servers inside data center, and Optical Ethernet (OE) for transporting across data centers. CBOSS is a time and wavelength division multiplexed optical slot switching ring (or torus, for scalability) fabric. OE is a time slotted bus where data are regenerated electrically at each node, but where the most time- and energy-hungry processing (such as FEC) is done only at in/egress nodes. Slots may be dynamically reserved end-to-end through orchestration of any mix of OE and CBOSS, and transition between segments is always performed in the electronic domain. In each segment, non-reserved slots are left to best effort traffic. Time sensitive flows can therefore be physically isolated (hard slicing) and carried across the network without interaction with best effort traffic or between themselves. In the test-bed, we connect 2 servers (emulating an industrial robot and remote server) and 2 Edge DCs: one 3- nodes/3 wavelengths CBOSS network inter-connecting 8 servers, and a single server (to emulate the second DC) with a transport 2-node OE bus. Fig 1 (b) reports histograms of the measured latency at the output of the network where seven 1Gb/s priority flows are competing. We compare DDN and Ethernet. CBOSS and OE achieve deterministic (jitter<100ns) latency of 70µs per flow, while Ethernet yields an average latency of 34µs with large tails up to 200 µs. We also measured that, at constant load, the maximum latency scales linearly with the number of priority flows over best-effort technologies like-Ethernet (reaching. 1 ms with just thirty flows, Fig 1 (a) but remains constant with CBOSS+OE [4]. These measurements confirm the potential of DDN to deliver true real time applications in a distributed edge cloud.

4. Breaking the eight fallacies of distributed computing

We believe that the unique advantages of DDN can disrupt established habits to the point that they could revive the old dream of distributed computing, a dream overshadowed by the success of central clouds, where IT jobs are performed inside a single location. L Peter Deutsch discussed its specific challenges in his famous list of eight fallacies in 1994. Interestingly, all of them would be shaken by the introduction of DDN technologies (Table 1).

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
The Eight Fallacies	Network is **reliable**	**Latency** is zero	**Bandwidth** is infinite	Network is **secure**	**Topology** does not change	There is **one administrator**	**Transport cost** is zero	Network is **homogeneous**
Over legacy network	~10^{-3} data loss ratio	~100ms latency ~10µs jitter	100Gb/s per port	Soft slicing by VLAN	Upon changes, discontinuities in service	Multiple Controls per technology domain	Transports and switching costs add up	Circuit + packet side by side with specific controls
Over DDN	<<10^{-9} data loss ratio (measured)	~100µs latency ~10ns jitter (measured)	>1Tbit/s per port	Hard slicing per flow	Upon changes, hot and hitless reconfigurations	End-to-end orchestration	Transport and switching are one equipment	Slotted, Synchronous, Scheduled technology all over edge
Evolution	**>10^6 gain**	**10^3 gain**	**>10 gain**	**Security is extended end-to-end**	**Network is built to sustain changes**	**Someone owns and runs it all**	**The cost of transport is diluted**	**There is one common technology**

Table 1. Deutsch's eight fallacies of distributed computing (top line) and why DDN overturns them

With all the above, we conjecture that DDN could be to fog computing, what GPU has been to multiple CPUs. But exactly like GPU is a Formula 1 car fit for racing circuits whereas CPU is a four-wheel drive utility vehicle designed for any type of road, DDN will bring unique benefits for true real-time, scheduled jobs, whereas it hosts mainstream jobs no better but as flexibly as today's best effort networking technologies, i.e. subject to Deutsch's fallacies.

5. Conclusion

In the prospects of a connected world where machine-to-machine traffic dominates, we reviewed the benefits of the Octopus network, a platform to organize the end-to-end exchange of time-critical flows across distributed edge clouds. We showed that dynamic deterministic networking (DDN) technologies are good candidates to turn the octopus network into a reality, while unleashing opportunities for distributed computing.

6. References

[1] 5G-PPP, "5G and the factories of the future," 2015, white paper. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf

[2] E. Grossman, "Deterministic networking use cases," Jun. 2018, IETF draft. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-ietf-detnet-use-cases

[3] M. Filer, J. Gaudette, Y. Yin, D. Billor, Z. Bakhtiari, and J. L. Cox, "Low Margin Optical Networking at Cloud Scale", J. Opt. Commun. Netw., Vol. 11, Issue 10, pp. C94-C108 (2019)

[4] N. Benzaoui, M. Szczerban Gonzalez, M. V. Rivera, J. M. Estaran, H. Mardoyan, W. Lautenschlaeger, U. Gebhard, L. Dembeck, Y. Pointurier, and S. Bigo, "DDN: Dynamic deterministic networks," in Proc. European Confe. on Optical Communication (ECOC 2018), Th3B.6.

[5] Arnon Rotem-Gal-Oz, "Fallacies of Distributed Computing Explained", https://www.rgoarchitects.com/Files/fallacies.pdf

[6] S. Bigo, "A Post-5G Network to Break the Eight Fallacies of Distributed Computing", in proc. Asia Communications and Photonics Conference, paper (ACP 2019), Nov 2019, Chengdu, paper M3C.3.