

## Five reasons why the edge is becoming the center

Peter Van Roy  
Université catholique de Louvain

Feb. 7, 2019

The “edge” of the Internet consists of the numerous heterogeneous, loosely coupled nodes situated at the logical extreme of the Internet. The edge is increasing in relevance with respect to data centers. The current situation, in which data centers are effectively in control of the Internet, is becoming unstable. The Internet is evolving toward a new equilibrium in which the role of the edge will be much more important. One could say that “the edge is becoming the center”. Here are five points that illustrate this trend:

1. The edge is growing exponentially, in data generation, data storage, and computation abilities, and the cloud (i.e., data centers) does not have such growth. Many studies (easy to find by Googling) show that the number of Internet of Things devices, which are part of the edge, has been growing exponentially for the last decade and they predict that this will continue for the foreseeable future. While mobile phones are already flattening out now, at several billion (basically one phone per person on Earth), the number of edge devices continues to grow and is projected to reach hundreds of billions and more. A keynote talk at NetFutures 2017 predicted >1000 devices per person by 2027; this may be optimistic but the error is in the year and not in the number of devices! The bulk of the Internet's storage and computation ability will in a few years be at the edge, and data centers will become small in relative size to the edge. Eventually data centers will be too small to effectively control or manage the edge.
2. The Internet is a hotbed of fake news and manipulation. The reasons for this are both sociological and technical. One important technical reason is that today's social networks can quickly amplify any voice, no matter how small. We are bombarded with so much information that separating wheat from chaff is difficult. This can be addressed by means of *provenance*: a digitally secure trace of the origin and path of a piece of data. Provenance when combined with curation (i.e., the selection of credible news sources by reputed organizations) can solve the fake news problem. Since most new data comes from the edge, provenance must be created directly at the edge. Among other things, this requires some security infrastructure, such as enclaves, directly at the edge. See the discussion item presented at the 2017 Ethical Forum: <http://www.info.ucl.ac.be/~pvr/ProtectedInternet.pdf>
3. Today's large-scale applications are built on the idea that “truth” is at the center, usually stored in a database. This is straightforward for simple client/server systems, but becomes much more complex when the center is small and the edge is vastly bigger and ever growing, and when edge data gets ever closer to raw real-world data. Applications and developers need to spend much effort to make sure that the database never diverges too much from what is true in the real world. An alternative is to consider the truth to be at the edge (e.g., near the user, much closer to the real world), and to have the central database automatically converge to the truth. This simplifies the application architecture and reduces the bottleneck created by the database. See the article by Christopher Meiklejohn presented at the Salon des

Refusés workshop 2017: <http://www.info.ucl.ac.be/~pvr/tendency-article.pdf>

4. The GAFAM companies (Google, Amazon, Facebook, Apple, Microsoft) exploit user data without their consent. An alternative is Personal Information Management Systems (PIMS), in which all user information is controlled by the user and any use by others is negotiated. Each user has their own PIMS and the PIMS is located near the user, at the edge. The success of the European GDPR is giving new strength to this approach. See the talk by Serge Abiteboul: <https://hal.inria.fr/hal-01621687/file/17.Ethics.Data.Namur.pdf>
5. Personal healthcare systems are a killer app for the edge. They will have a network of sensors continuously monitoring a person's physiological conditions, and the enormous data generated by these sensors will never be sent to the cloud, but be analyzed near the person by a “personal data center”. Analysis will be done by a combination of big data and machine learning, and training will also be done at the edge. Training will be online (interleaved with inferencing) and there will be collaboration between individual personal healthcare systems, to make training more effective by combining (in suitably anonymized form) the relevant information from many people. The final result will be to simultaneously increase quality and decrease cost of healthcare. This is discussed in detail in the LightKone project Deliverable D2.1 (eventually to be made public at [lightkone.eu](http://lightkone.eu)).

The increasing importance of the edge as exemplified by these five points is well underway. Multiple consortia and platforms exist to help move computation out of the cloud and toward the edge, such as OpenFog, Microsoft Azure IoT, Amazon IoT Greengrass, and so on. But this is just the beginning. These platforms are far from exhausting the new opportunities created by the expansion of the edge. It is clear that new abilities will appear at the edge that were never inside a cloud. It is also clear that the current situation, in which a small set of powerful organizations effectively controls the Internet, is changing. The locus of control is moving from the physical data centers managed by these organizations (e.g., Google, Facebook, Amazon) to a logical infrastructure that is much more dispersed. There will be a large number of small logical centers of control, each managing its own authority, protected by strong cryptography, and all negotiating with each other. This form of decentralized control is becoming practical because the Internet continues to evolve toward improved reliability and connectivity, increased bandwidth, and reduced latency. The edge is becoming the center, indeed.