Revolutionising Computing Infrastructure For Citizen Empowerment

Noa Zilberman

University of Cambridge
noa.zilberman@cl.cam.ac.uk

Abstract

The world has dramatically changed over the last decade. Almost every aspect of our lives is being digitally monitored: from our social networks activity, through online shopping habits to healthcare and financial records. The emergence of Internet of things and the growing presence of cyber physical systems only increase citizens' exposure to digital monitoring by commercial enterprises. In order to maintain citizens' right for privacy while still encouraging an evolving digital economy, people should be given the right to choose where their data is stored and who holds it, a currently unattainable privilege. We propose that through the revolution of computing infrastructure, enabled by new computing architectures, a healthier competitive environment can thrive. In this environment, companies will compete for customers, offering privacy and information control as a service. Such competition, when supported by regulation, will empower citizens, allowing them to take back control of their data.

***Keywords*—** Computing Infrastructure; Data Control; Privacy; Citizen Empowerment.

1 Introduction

The most important piece of information we hold during our lifetime may be our personal information. Information on who we are, what we like and dislike, health and financial records as well as family ties and friends: our personal information defines us. For this reason, privacy is an important concern in today's digital era, where every aspect of our lives is collected and analysed.

Research exploring ways to take back control of our data is considerable (e.g. [14, 13]). The importance of data protection has not escaped the eyes of legislators, and regulation such as General Data Protection Regulation (GDPR) has been set in motion. While legislation covers aspects ranging from the handling of collected data to responsibility, accountability and consent, it does not cover the users' ability to choose where their data is stored.

While users today can choose not to use the services of giants such as Google, Facebook or Amazon, such a choice has implications on the daily lives of the users, and their ability to interact with others. It is rarely noted, however, that the growth of these powerful corporates was fuelled by a technological gap: the lack of hierarchical computing infrastructure.

In this paper, we assert that better control of our personal data can only be achieved through a revolution of computing infrastructure. We propose a model for scaling computing infrastructure, where more and more companies can become part of the infrastructure and where competition between these companies strives. Under such conditions, privacy and information control will become a service, and citizens will have the power to choose how their data is handled.

2 Motivation

Health records are becoming a dominant factor in the debate on progress versus privacy. Already a decade ago Google offered Google Health as a mean to collect, store, and manage medical records online [22]. While Google Health failed, DeepMind Health [9] is a more recent Google initiative, aiming to help clinicians provide better care. Other cloud service providers also offer health related services, such as Salefoce's HealthCloud [30] and Microsoft Health [23]. Amazon and Merck even initiated the Alexa Diabetes Challenge [20].

While using technology to improve our health is a blessing, the way that it is done needs to be questioned. For Example, in July 2017 the Information Commissioner Office (ICO) ruled the Royal Free NHS Foundation Trust failed to comply with the Data Protection Act when it provided patient details to Google DeepMind [17]. In this case, no less than 1.6 million identifiable patient records were shared with Google without explicit patient consent. While here data was shared with Google by a different party, in many cases users knowingly share their health information, e.g. using Internet of Things (IoT) devices for health monitoring. This data is shared with the application, which stores and processes it in the cloud. The physical location of the data, e.g. in which data centre the data is stored, is not a parameter a user can control.

The problem is not limited to health records: voice assistants, such as Apple's Siri, Amazon's Echo, Google Home and Microsoft Cortana record your voice and store some or all of the data in the cloud [24]. All around us, every piece of data that can be collected is stored in the cloud. This means that the few companies that physically store the data hold tremendous power over our lives [28]. While there is no immediate alternative, emerging computing architectures do offer better means to take back control of our data.

3 Computing as an infrastructure

Infrastructure is the basic physical and organizational structures and facilities needed for the operation of a society or enterprise [27]. Infrastructure has greatly evolved over history: from roads and canals in ancient history to more recent railways, electricity and water and sewage systems. Communication infrastructure, including telephony and Internet, are more recent additions.

While computing is at the base of digital economy, it is currently not an infrastructure [18]. Any type of infrastructure mentioned above is deployed at varying scales and for varying needs (e.g., a 4-lanes interchange on a motorway vs. a mini roundabout in a rural area). In contrast, computing is deployed only at the edges: the end-user and the cloud provider.

Today, data generated by users is streamed to the cloud, where it is processed and stored. It is a directly linked service, with no computation provided along the way. Similarly, computing equipment is designed to handle data only on these two extremes of scale: either user size (IoT, mobile devices and servers) or cloud-size (data centre). Anything in-between is a simple aggregation of equipment, e.g. a rack of servers, unsuitable to process scaling amounts of streamed information.

The state of computing infrastructure stands in a stark contrast to the complementing networking infrastructure. Networking infrastructure scales from low-end user equipment to high-end data centre networks, with different types of equipment and technologies at each scaling point. Networking infrastructure does not only scale in equipment: it also involves many service providers along the way, with known rules applying to the way traffic is treated [11]. A result of this construct is a large number of networking infrastructure players, and while some of them are very large, no single corporate dominates the market. Furthermore, the large number of competitors allows users to freely select their networking service providers from amongst a wide set of choices, and the providers compete to obtain new clients.

Not treating computing as an infrastructure has long lasting effects on users' experience, also in terms of resilience and political climate effects. The distributed denial of service attack on Dyn in 2016, a US based company, affected European services, such as those provided by the BBC or the Swedish Government [7, 35]. While not a direct example, this shows the worldwide dependency on services located in other countries. Political disputes or accidents (e.g. submarine optic fibre cut) can cut citizens from both essential and convenient services hosted in the cloud outside their country.

Legislation already puts limitations on cross border data flow [8], and the Parliament is concerned by the effects of Brexit on free data flow [15]. The solutions, however, focus on building data centres locally rather than on changing existing paradigms. The interests of users will never come first as long as data centre providers control the information, no matter if their data centre is situated locally or abroad.

4 Scaling computing infrastructure

The need for a scalable computing infrastructure has not been unnoticed. The introduction of 5G networks, in particular, has driven mobile providers to explore solutions to the increasing load and requirements from their networking infrastructure, combined with computing needs. Fog Computing [4], and Mobile Edge Computing [16] which advocate pushing computing to the edge, is a move at the same direction as networking. However, Fog Computing currently focuses on the Enterprise market (e.g. [26]) and IoT rather than the personal user, and does not provide the technological leverage required to transition the infrastructure.

The Computer As Network (CAN) architecture [36] proposes a new server-level computer architecture that attends specifically to the challenge of scalable computing infrastructure. It is based on the insight that scalable computing infrastructure cannot strive without attending to the increasing performance gap between networking and comput-

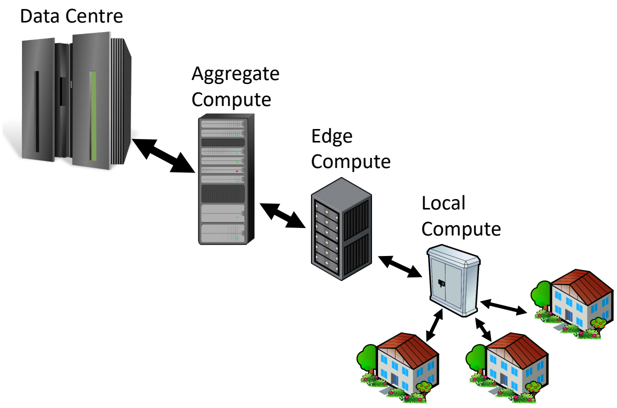

Figure 1: Scaling computing infrastructure: from the local compute resource, to the data centre.

ing [37, 36]. CAN bridges the performance gap between networking and computing by borrowing ideas and practices from the networking world, to address the challenges presented by computing. The architecture explicitly puts a networking-fabric at the core of the computing device, and treats every I/O transaction between elements in the system as a networking transaction.

The fundamental revolution proposed by CAN is the ability to build "Tiny Terabit Data Centres" in a box, at a cost that is at same scale as a standard server. This means that providing computing infrastructure will not require datacentre scale equipment anymore, nor the financial investment required for such. Compute equipment can be located as close to the home as a local network cabinet used in e.g. Fibre To The Cabinet (FTTC), with scaling capabilities at each point-of-presence in the hierarchy (as demonstrated in Figure 1).

Promoting such scalable computing equipment is mandatory. This can be done in a manner similar to the Open Compute Project [25] (promoting computing equipment for the cloud) or Telecom Infra Project [33] (promoting telecommunications equipment). The key aspect here will be the cost, both in terms of investment and maintenance. A low cost, low power solution will allow easy deployment and a minimal financial commitment and enable a large number of companies to become players in the market.

5 Citizen empowerment model

A large number of new competitors, whether small or medium size, providing computing infrastructure services opens a completely new set of possibilities for the personal user. It will revolutionize the way we treat cloud computing and applications: while today we choose the application, and that automatically implies the compute service provider, in the new model we will choose the compute service provider, and the provider will imply the available cloud-based applications.

To make an analogy, this is similar to selecting a television provider (e.g. BT, Sky or Virgin), where the selected provider sets the television channels available and the shows that one can watch, and the customer can choose to contract with more than one service provider to gain access to additional channels (e.g. Netflix, Amazon).

An example of such new infrastructure entity is a "neighbourhood cloud": a local compute resource, located in the user's vicinity and being able to collect information from tens to hundreds of thousands of devices, while maintaining privacy control per data stream and complemented by software solution (e.g. as the software platform proposed by [13]). With the increasing number of sensors and IoT devices used, such a resource will support hundreds to thousands of neighbours - a small number on a city scale. The implication of a neighbourhood cloud is that users will no longer be required to store e.g. their health monitoring information (from IoT devices, pace makers etc.) in an unknown location, as all the information can be stored and processed locally by a provider they choose.

A large number of service providers also means a healthier competition, and providers can propose different privacy policies for data control rights to the users in order to draw their attention. Beyond application packs, providing the users with a platform to install their own applications is not a technologically challenging task, and approaches to doing so securely can be adapted from the data centre model (e.g. [19, 3]).

It is not unexpected that new computing paradigms will also call for community or non-profit organisations led neighbourhood clouds. Where the cost of installing the equipment is low, and the trust between the user and the compute provider must be high, such initiatives are likely to gain the support of many people. Even when maintenance and support are taken into account, the success of community driven models, such as the Raspberry Pi foundation [29] is indicative.

Assuming that enabling a choice of data storage location and competition will empower citizens is not a myth. Companies such as Facebook and Google depend on the size of their user base, and the information gained from the users is key to their success [10]. By sharing less information with these giants, users create a healthier and safer environment for themselves, and take back control of their personal information.

It is the government's place to enable the competitive and regulated environment required for the success of computing infrastructure. The most likely model is the one used for communications, and it is expected that many companies providing communication infrastructure will want to offer computing services as well. Yet, as computing differs from communications, the two should not be treated as identical.

6 Discussion

Turning compute into an infrastructure was not practical until recently. In part, the change is contributed to the emergence of lower power and lower cost processors at the core of high performance servers [21, 31]. The economic model is a key factor here, as the vast number of systems that need to be deployed dictates a low cost solution to be a mandatory requirement.

Another element required is isolation of users and resources. In communication infrastructure this is already provided, guaranteeing quality of experience and as an enabler for shared wireless services. In computing, however, resource isolation is still the source of considerable study, especially where high performance systems are involved (e.g. [2, 1, 6]). For computing infrastructure to succeed we need to apply networking practices to compute [36] while maintaining easy management, low cost and a strong isolation between users.

One may question why computing as an infrastructure is required. Beyond current technological limits, why isn't home compute equipment sufficient. However, personal computing equipment sales (including desktops, notebooks and laptops) have been steadily declining over the past five years [12, 32] as people move to cheaper, lighter and more accessible mobile devices. It is unlikely that people will buy new compute equipment when they can get the same services for what seems to be free. The understanding that current services are not free, as you pay with your personal data, is growing but have not yet reached a critical mass [34].

Privacy is the priority requirement from computing infrastructure. While currently governments have limited control over cloud service providers and application designers, it is much easier to control privacy related aspects in a locally regulated environment. As an example, if it is forbidden by law to allow private data beyond the first compute unit, local companies will have an incentive not to share it further and the government will find it easier to enforce the law. This is contrary to the common model where companies provide services for free in exchange for personal data that they later sell [5], and where governments find it hard to deal with global conglomerates.

New advancements in security, such as homomorphic encryption, may further contribute to aspects of user privacy while running applications on a computing service. While new computing infrastructure will likely create new cyber security threats, it can also help addressing different current threats. As an example, a Distributed Denial of Service (DDoS) attack is not effective when the target services are physically distributed, reducing the effect of a single (or a few) points of failure.

Not all applications fit the proposed models: social networks, for example, are likely to continue running in the cloud, as they aggregate information from many users. Electronic mail services, on the other hand, can operate just as well when located close to the user. Sensor based applications and applications sensitive to latency are the leading examples of applications best placed close to the user.

7 Conclusions

The world is rapidly changing. More and more data that we consider to be personal is held these days by a third party, stored in locations unknown and uncontrolled by users. Leaving our most precious data in the hands of entities we cannot choose or control is of utmost concern. In this paper, we proposed a radical rethinking of data control paradigms, by allowing users to choose where their data is stored and who handles it. We assert that this paradigm can come true only by treating computing as an infrastructure. The vision is enabled by new computing architectures, proposing low cost, high performance platforms. Where computing becomes an infrastructure, it allows new competitors to enter the market and a wide offer of privacy-preserving cloud-compute services to the user. Citizens will be able to take back control of their data, as well as a foundation of our society: the right to choose.

8 Acknowledgements

We thank Andrew W Moore for his contribution to the CAN project. We acknowledge the support from the Leverhulme Trust (ECF-2016-289) and the Isaac Newton Trust.

References

[1] S. Angel, H. Ballani, T. Karagiannis, G. O'Shea, and E. Thereska. End-to-end performance isolation through virtual datacenters. In *OSDI*, pages 233–248, 2014.

[2] G. Banga, P. Druschel, and J. C. Mogul. Resource containers: A new facility for resource management in server systems. In *OSDI*, volume 99, pages 45–58, 1999.

[3] A. Baumann, M. Peinado, and G. Hunt. Shielding applications from an untrusted cloud with haven. *ACM Transactions on Computing System*, 33(3):8:1–8:26, Aug. 2015.

[4] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli. Fog computing and its role in the Internet of things. In *Workshop on Mobile Cloud Computing*, pages 13–16. ACM, 2012.

[5] J. Brustein. Start-ups seek to help users put a price on their personal data. *The New York Times*, 12:3, 2012.

[6] B. Burns, B. Grant, D. Oppenheimer, E. Brewer, and J. Wilkes. Borg, Omega, and Kubernetes. *Communications of ACM*, 59(5):50–57, Apr. 2016.

[7] E. Chiel. Here are the sites you can't access because someone took the Internet down, 2016. [`http://splinternews.com/here-are-the-sites-you-cant-access-because-someone-took-1793863079`, Online; accessed August 2017].

[8] N. Cory. Cross-border data flows: Where are the barriers, and what do they cost? *Information Technology and Innovation Foundation (ITIF)*, May 2017.

[9] DeepMind. DeepMind Health. https://deepmind.com/applied/deepmind-health/ [Online; accessed August 2017].

[10] A. Ezrachi and M. E. Stucke. Virtual competition, 2016.

[11] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, 2001.

[12] Gartner. Gartner says worldwide pc shipments declined 8.3 percent in fourth quarter of 2015, 2016. [Online; accessed August 2017].

[13] H. Haddadi, H. Howard, A. Chaudhry, J. Crowcroft, A. Madhavapeddy, and R. Mortier. Personal data: Thinking inside the box. In *Decennial Conference on Critical Alternatives*, 2015.

[14] W. Heath, D. Alexander, and P. Booth. Digital enlightenment, mydex, and restoring control over personal data to the individual. In *Digital Enlightenment Forum Yearbook 2013: The Value of Personal Data*, pages 253–269, 2013.

[15] House of Lords, European Union Committee. Brexit: trade in non-financial services, 18th Report of Session 2016–17, 2017.

[16] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young. Mobile edge computing a key technology towards 5G. *European Telecommunications Standards Institute (ETSI)*, (11), Septmeber 2015.

[17] Information Commissioner's Office. Royal Free - Google DeepMind trial failed to comply with data protection law, July 2017. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/.

[18] Infrastructure and Projects Authority. National infrastructure delivery plan 2016–2021, Mar. 2016. Policy Paper.

[19] A. N. Khan, M. M. Kiah, S. U. Khan, and S. A. Madani. Towards secure mobile cloud computing: A survey. *Future Generation Computer Systems*, 29(5):1278–1299, 2013.

[20] Luminary Labs. The Alexa diabetes challenge, Apr. 2017. http://www.alexadiabeteschallenge.com/.

[21] M. Malik and H. Homayoun. Big data on low power cores: Are low power embedded processors a good fit for the big data workloads? In *2015 33rd IEEE International Conference on Computer Design (ICCD)*, pages 379–382, Oct 2015.

[22] M. Mayer. Google Health, a first look, Feb. 2008. https://googleblog.blogspot.co.uk/2008/02/google-health-first-look.html [Online; accessed August 2017].

[23] Microsoft. Microsoft Health. https://www.microsoft.com/microsoft-health/ [Online; accessed August 2017].

[24] T. Moynihan. Alexa and Google Home record what you say. But what happens to that data? *Wired*, may 2016. https://www.wired.com/2016/12/alexa-and-google-record-your-voice/ [Online; accessed August 2017].

[25] Open Compute Project. http://www.opencompute.org/ [Online; accessed May 2017].

[26] OpenFog Consortium Architecture Working Group. Openfog architecture overview. *White Paper*, 2016.

[27] Oxford English Dictionary. *infrastructure, n.* Oxford University Press, June 2017. [Online; accessed August 2017].

[28] J. Powles and H. Hodson. Google deepmind and healthcare in an age of algorithms. *Health and Technology*, pages 1–17, 2017.

[29] Raspberry Pi. https://www.raspberrypi.org/. [Online; accessed August 2017].

[30] Salesforce. Salesforce Health Cloud:salesforce for healthcare - patient management software in the cloud.

[31] M. Shahrad and D. Wentzlaff. Towards deploying decommissioned mobile devices as cheap energy-efficient compute nodes. In *9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17)*, Santa Clara, CA, 2017. USENIX Association.

[32] Statistica. Shipment forecast of laptops, desktop pcs and tablets worldwide from 2010 to 2020 (in million units), 2017. https://www.statista.com/statistics/272595/global-shipments-forecast-for-tablets-laptops-and-desktop-pcs/ [Online; accessed August 2017].

[33] Telecom Infra Project. https://telecominfraproject.com/ [Online; accessed May 2017].

[34] J. Van Dijck. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2):197, 2014.

[35] J. Westerholm. Så sänktes twitter och regeringen.se i attacken, 2016. [`https://sverigesradio.se/sida/artikel.aspx?programid=83\&artikel=6547041`, Online; accessed August 2017].

[36] N. Zilberman, A. W. Moore, and J. A. Crowcroft. From photons to big-data applications: terminating terabits. *Royal Society Philosophical Transactions A*, 374(2062):20140445, 2016.

[37] N. Zilberman, P. M. Watts, C. Rotsos, and A. W. Moore. Reconfigurable network systems and software-defined networking. *Proceedings of the IEEE*, 103(7):1102–1124, 2015.