Finding your Way in the Fog: Towards a Comprehensive Definition of Fog Computing
Luis M Vaquero, Luis. Rodero-Merino

HP Laboratories
HPL-2014-60

Keyword(s):
for computing

Abstract:
The cloud is migrating to the edge of the network, where routers themselves may become the virtualisation infrastructure, in an evolution labelled as "the fog". However, many other complementary technologies are reaching a high level of maturity. Their interplay may dramatically shift the information and communication technology landscape in the following years, bringing separate technologies into a common ground. This paper offers a comprehensive definition "the fog", comprehending technologies as diverse as cloud, sensor networks, peer-to-peer networks, network virtualisation functions or configuration management techniques. We highlight the main challenges faced by this potentially break-through technology amalgamation.
Finding your Way in the Fog: Towards a Comprehensive Definition of Fog Computing

Luis M. Vaquero
Hewlett-Packard Labs
Bristol, United Kingdom
luis.vaquero@hp.com

Luis Rodero-Merino
Gradiant
Vigo, Spain
lrodero@gradiant.org

ABSTRACT
The cloud is migrating to the edge of the network, where routers themselves may become the virtualisation infrastructure, in an evolution labelled as “the fog”. However, many other complementary technologies are reaching a high level of maturity. Their interplay may dramatically shift the information and communication technology landscape in the following years, bringing separate technologies into a common ground. This paper offers a comprehensive definition “the fog”, comprehending technologies as diverse as cloud, sensor networks, peer-to-peer networks, network virtualisation functions or configuration management techniques. We highlight the main challenges faced by this potentially breakthrough technology amalgamation.

Categories and Subject Descriptors
C.2 [Computer Communication Networks]: [Distributed Systems - Network Operating Systems]

General Terms
Computing Theory

Keywords
Fog computing, Network Function Virtualisation (NFV), peer-to-peer (P2P), Internet of Things (IoT), Sensor networks, Cloud computing, Configuration Management

1. INTRODUCTION
The information and communication technologies (ICT) community typically takes time to agree on the real meaning, reach and context of the new terms that appear associated to new technology trends and their associate buzz/hype. Web services, cloud computing, big data are a few examples of hyped terms that were confusion when first coined.

The term fog computing is reaching this initial state of confusion now. Unlike the examples above, ‘the fog’ is not constrained to a particular technological area. As a result, we can expect the initial confusion about ‘what the fog is?’ to reach unprecedented levels.

As it often happens with new technologies, a consensus definition needs to be agreed on by the community to mitigate hype and confusion. The very first definitions tend to focus on just a few aspects, like scalability in the cloud or interoperability in web services. The fact that the fog agglutinates many converging technological trends makes this problem even more severe. In fact, looking at any of the technologies related to the fog from a single angle may offer the false view that there is little new to it. For instance, recent definition attempts have presented it as just an evolution to our current cloud model. See, for instance, Cisco’s view of the fog [6].

In this paper, we offer a broader and integrative view of the fog. We present it as the result of several emerging trends on technology usage patterns on the one side, and the advances on enabling technologies on the other side. From the analysis of both aspects, we propose a definition of fog computing that encompasses its features and impact. Also, this work introduces the obstacles that will have to be overcome so that fog computing can mature and unfold its entire potential.

This paper is structured as follows. Section 2 discusses devices ubiquity as the main factor that will bring the fog, along with a brief overview of the main works that address the demands for smaller and more capable devices. Section 3 deals with the challenges on services and networks management that fog applications will introduce, while Section 4 summarises the advances proposed at several levels to provide connectivity to the billions of devices that will be the norm in the fog. Section 5 explains how privacy demands by users will be another propeller of the technological changes that will shape the fog. With all those ingredients taken into account, Section 6 presents our definition of the fog, and Section 7 lists the open challenges that will have to be solved in the future to make the fog a reality. Finally, Section 8 summarises the conclusions of this work.

2. DEVICE UBIQUITY
There is a huge increase in the number of devices getting connected to the network. This increase is driven by 2 sources: user devices and sensors/actuators. Cisco conservatively estimates that there will be 50 billion connected devices by 2020 [9]. This explosion in the number of devices per person is explained by the proliferation of mobile devices (e.g. mobile phones and tablets, specially in developing countries). But these impressive numbers will soon be overpassed by the myriad of sensing/acting devices placed virtually everywhere (the so called Internet of Things, IoT, and pervasive sensor networks). Wearable computing devices (smart watches, glasses, etc.), smart-cities [12], smart metering devices deployed by energy suppliers to analyse consumption at the home level [13], self-driving vehicles, Wearable computing devices (smart watches, glasses, etc.), smart-cities [12], smart metering devices deployed by energy suppliers to analyse consumption at the home level [13], self-driving vehicles,
sensor networks and the like will be major drivers to the ubiquity of connected devices.

All these applications are fostering the presence of devices everywhere around us. Thus ubiquity has prompted intensive research, leading to a new breed of technical achievements is under way to solve today’s limitations in device size and battery lifespan (see subsection 2.1). This may itself ease the deployment of more devices, creating a virtuous circle.

2.1 Size and Battery Lifespan

Cost is a major factor driving devices to be as small as possible. This also increases device portability and reduces power consumption, which may be crucial in some context (e.g., portable phones or long-lasting fire sensors in a remote forest). Packaging and power management technologies aim to create smaller and more autonomous devices that can run way longer at a minimum price.

System on Chip (SoC) technologies embed components such as CPU, memory (e.g., HP’s memristor [8]), timers and external interfaces in a single chip. They require less room and consume less power than typical multi-chip systems. System in Package (SiP) is a solution somewhere in between SoCs and multi-chip systems: it ensembles circuits in a single unit or ‘package’, and is used today for small devices such as smart phones.

Even when better packaging may improve power consumption, this alone may not be enough for it to last longer. The IoT is calling for long life sensors which sometimes will not be able to connect to any power supply. Today’s lithium-ion batteries (LiB) are used for portable devices of all kinds; solid-state LiB solutions are expected to replace them in the medium term, increasing up to three times today’s energy density. Still, batteries based on chemical power sources can become a limiting factor in future developments: higher power requirements in a tiny fraction of the size of current batteries.

Research efforts are focused on 3D microbatteries. ‘3D’ is a term that encompasses the efforts to arrange the anode and cathode of batteries in 3D layouts (beyond the typical 2D formations), to enhance both its energy and power density. Using those 3D structures at microscopic scale is resulting in batteries of tiny size and big power. Also, we have to watch the evolution of RF-powered computing [10], which poses that energy can be harvested from ambient radiofrequency signals (such as TV, cellular) to power low-end devices that sense, compute and communicate. Also renewable energy fed devices are already available.

3. SERVICE/NETWORK MANAGEMENT

Having many devices can be very helpful to improve our processes at all levels (from our home to the planet as a whole) and help us understand them better. These devices need to be configured and maintained once they get deployed (e.g., a future phone hosting a service sold to a third party user or a remote sensor at the bottom of the sea). Managing networks of billions of heterogeneous devices that run one or more services is incredibly challenging and complex. Several fog technologies have been evolving to help tame this complexity: “softwareisation” of network and service management for better flexibility; asymptotic/declarative techniques for scaling management; “small” edge clouds to host services close to the endpoints (or at the endpoints themselves); and peer-to-peer (P2P)- and sensor network-like approaches for auto-coordination of applications.

3.1 “Softwareisation” of Network Management

Configuring and keeping updated and secure fog networks, services and devices is done separately (e.g., routers, servers, services and devices are separately managed by different people). These tasks are labour intensive and error prone. For instance, well-known Internet companies claim a single admin handles thousands of machines running a single service type. Configuring and maintaining many different types of services running on billions of heterogeneous devices will only exacerbate our current management problems. The fog needs heterogeneous devices and their running services to be handled in a more homogeneous manner; ideally fully automated by software.

Network Function Virtualisation (NFV) is arguably the most remarkable technology in this regard. NFV is the reaction of telco operators to their lack of agility and constant need for reliable infrastructures. NFV tries to provide the ability of dynamically deploying on-demand network services (e.g., a firewall, a router or a WAN accelerator, a new LAN or a VPN) or user-services (e.g., a database) where and when needed. Software Defined Networks SDN are one if the pillars needed for NFV, since some network services (e.g., creating new “virtual” networks on top of the physical infrastructure) can be done by software only. For instance, some gateways can be deployed as virtual machines and their traffic tightly controlled thanks to SDN capabilities in a local edge cloud, see Figure 1. The “softwareisation” of a classically hardware-driven business build around routers and servers where services got deployed will result in cheaper and more agile operations.

A complementary approximation is proposed by Cisco with its first software-only version of the IOS wrapped in with a Linux distribution (IOx)3. The router itself becomes an SDN-enabled virtualisation infrastructure where NFV and application services are deployed close to the place where they are actually going to be used. But IOx’s computing capabilities will still be limited (edge routers are not carrier grade after all).

Figure 1: An SDN-enhanced cloud at the edge of the network as a cornerstone for NFV

There are recent NFV proof-of-concepts [4, 5], but NFV capabilities do not reach end user devices or sensors yet. In addition, NFV and IOx offer telco operator’s and vendor’s point of view. Network gear equipment is only a tiny fraction of the devices of the fog. Billions of user handheld devices and potentially trillions of sensors need to have a similar automation process that can cope with the required scale.

2Fog services also run on end user devices, not just on well-controlled “central” servers

3.2 Asymptotic Techniques
At fog scale, only declarative and asymptotic techniques seem feasible [14]. These techniques engage components in their own management tasks so that: 1) the admin only specifies the final desired state (declarative) as opposed to individual commands; and 2) she assumes the configuration may never take place because by the time it is being deployed the system may have changed (e.g. fog nodes are gone or new nodes show up). As an example of these techniques, see work on declarative and asymptotic management done by HP Labs in the past [14,7]. Other vendors are also starting to use declarative systems to tame scale and complexity, for instance see Cisco’s approach at managing OpFlex (a kind of Cisco’s OpenFlow supported by IBM and Midokura) SDNs⁴.

3.3 Clouds at the Edge
Mini-clouds are getting deployed closer to the edge (to the user) via private clouds. Telcos and gear vendors are moving on that direction too. Long Term Evolution (LTE)’s Enhanced Packet Core (EPC) can easily be expanded to include their own mini clouds. Having a small cloud at the EPC can help to deliver services close to users (at the edge) and confine traffic there while reducing “trombone routes” with the help of SDNs. Also, IOx is just an evolution of the current cloud model in which routers can become the virtualisation infrastructure given their ubiquity and hierarchical placement help to achieve locality. The fog enables user devices to become the virtualisation platform themselves. As such, they can lease some computing/storage capacity for applications to run on them.

Figure 2: Edge clouds as entry points for IoT and virtualised sensor networks

In the fog, both the network and the services running on top of it can be deployed on demand in a fog of edge devices. Service delivery to specific locations in the network is greatly simplified. For example [16] gives an example of storage functions being dynamically deployed in different mini clouds in selected network locations so that bulky data transfers are sped up.

3.4 Distributed Management
The management techniques discussed so far rely on a provider (e.g. the telco operator) as the sole responsible of network and service operation. But there are also P2P-and sensor network-like approaches that allow endpoints to cooperate in order to achieve similar results, but can scale better. P2P technologies have been around for a while and they are mature enough to help deliver the vision of the fog. They can exploit locality while removing the need for a central management point. Applications like Popcorn Time⁵ have shown the benefits of a P2P model to deliver global services at scale. Many of the ideas of P2P content distribution networks (CDNs) are applicable to the fog too; a fog application could be seen as a CDN where some sort of data is exchanged between peers.

Thus, in the fog a subset of network and user device/sensor elements can behave as mini-clouds. As a result the fog becomes an environment where applications and data are no longer to necessarily stay in centralised data centres. This improves scalability and empowers users to retain control and ownership of their own data/apps. Applications will then be implemented by using droplets or tiny pieces of code that can securely run in devices at the edge with minimum interaction with central/coordinating elements, reducing unnecessary/undesired uploads of data to central servers in corporate data centres.

4. CONNECTIVITY AT FOG SCALE
The presence of (potentially tiny) devices everywhere is only one of the ingredients of the fog. As mentioned above, all these devices need to be connected. The sheer volume of devices (50 billion handheld user devices in 2020) together with many more sensing/acting devices of the IoT (working 24/7) will likely dwarf present connectivity and bandwidth problems. A special report in The Economist titled “Augmented Business” described how cows will be monitored to ensure healthier, more plentiful supply of meat for people to consume. On average, each cow generates about 200 MB of information a year⁶.

4.1 Physical Connectivity
A consequence of having hundreds of billions of devices consuming and producing data at the edge of the network is that these networks become a huge bottleneck [1]. Network operators have been intensively investing in a variety of new wireless access technologies to cope with the abrupt increase in devices per user, but these Wide/Metropolitan Area networks (WAN and MAN), Local Area Networks (LAN) and Personal Network (PN) investments may fall short in an IoT world.

Most efforts in WAN/MAN are focused on LTE; LTEv12 will be the first specification that fulfils all the requirements of the Intl. Telecom. Union to be labelled 4G⁷. 4G

---

⁵http://venturebeat.com/2014/03/15/movie-streaming-app-popcorn-time-is-coming-back-from-the-dead-thanks-to-a-torrent-site/
⁶https://github.com/Yify/popcorn-app
⁷The USDA estimates there are around 1.2 billion cattle on the planet. That would represent around 224 PB of data per year globally for cow monitoring only
⁸The term 5G is starting to be used in many fora, however there are not official specifications yet that define what 5G networks will be. Today 5G is mostly a term that encompasses efforts to describe the features that networks will
LTE/EPC is supposed to be fully rolled out by 2017 [1] and it will expand the available bandwidth of edge networks [3]. LAN technology has improved to reduce congestion and increase the available bandwidth at lower power consumption, see for instance the latest Wi-Fi specification, 802.11ac. Finally, there have been huge improvements in PNs. These short range technologies require nodes to organise themselves, as no central access point may be available. Bluetooth Low Energy, ANT+, ZigBee and RF4CE are the most remarkable.

Figure 3 summarises the evolution in download bandwidth capacity brought by new wireless technologies.

4.2 Network Connectivity

Beyond improvements on wireless networks, other developments are needed to enable communication in scenarios where having all endpoints connected to some WAN/LAN is not feasible (due to costs, lack of enough ‘link’ points such as base stations, etc.)

In the fog, each node must be able to act as a router for its neighbours and must be resilient to node churn (nodes entering and leaving the network) and mobility. Mobile Ad-hoc Networks (MANET), which have been an important research topic for several years now [17], could be the basis for future fog networks as they will enable the formation of densely populated networks without requiring fixed and costly infrastructures to be available beforehand. In fact, Bluetooth LE, ANT+, ZigBee and RF4CE all allow the construction of MANETs at least up to local range. However, most work is still to be done to enable MANET in MAN and WAN networks. Wireless Mesh Networks (WMN) are a solution close to MANETs. A WMN can use mesh routers at its core, which have little to no mobility. Nodes use those routers to get connectivity, or other nodes if no direct link with the routers can be established. Routers enable access to other networks such as cellular, Wi-Fi, etc. There is still an intensive research activity on MANET and WMNs.

On top of MANET/WMNs (or right on top of the wireless network if feasible) we find the protocols that have been developed for the IoT, like MQTT, CoAP and DDW. All are designed with two goals in mind: low resource consumption and resilience to failures; they tend to follow a publish/subscribe (pub/sub) communication model.

Both network and IoT protocols can benefit from data locality: they no longer need to send all the data around the world all the time. Only aggregates may be sent or a pub/sub model can be enforced that can hugely alleviate our connectivity needs, confining potential congestion problems at the edge of the network (more so with the advent of edge router/handledd/sensor enabled mini clouds). In addition to confining traffic at the edge, this has a very positive impact on privacy.

5. PRIVACY

Today, we constantly leak personal information by using different products, platforms and services. Albrecht et al. picture a blunt, but honest, reality: “we may think were in charge of our shopper cards and our mobile apps and our smart fridges, but lets not fool ourselves. The information is not ours. It belongs to Google, and IBM, and Cisco Systems and the global Mega-Corp that owns your local supermarket. If you dont believe us, just try removing your data from their databases” [2].

Users are becoming increasingly concerned about the risk of having their private data exposed. As a result, besides the technical challenges introduced by the ubiquity of devices, there is another trend that will push for a fog scenario where data is not sent to a few centralised services, but it is instead kept ‘in the network’ for better privacy. Data ownership will be a very important cornerstone of the fog, where some applications will be able to use the network to run applications and manage data without relying on centralised services.

Storing encrypted sensitive data in traditional clouds is an alternative to keep privacy. However, this makes it really hard to perform any processing over such data. There is important research work on this topic, for example using crypto-processors or applying special encryption functions that cipher while keeping some of its original properties, thus allowing to perform certain limited tasks on it [15]. Still, such options have limited applicability.

As a result, users will demand innovative ways to preserve their privacy from any potential big-brother-like entity. This will be a great incentive to adopt fog technologies, as they will enable the network to replace centralised services.

6. A DEFINITION OF THE FOG

Previous sections have introduced the set of challenges envisioned and technologies devised that will shape the fog, which are summarised in Figure 4.

Figure 4: Fog Challenges and Enabling Technologies
following definition of the fog:

Fog computing is a scenario where a huge number of heterogeneous (wireless and sometimes autonomous) ubiquitous and decentralised devices communicate and potentially cooperate among them and with the network to perform storage and processing tasks without the intervention of third-parties. These tasks can be for supporting basic network functions or new services and applications that run in a sandboxed environment. Users leasing part of their devices to host these services get incentives for doing so.

This definition encompasses the features which we deem will be key ingredients of the fog: ubiquity, improved network capabilities as a hosting environment, and better support for cooperation among devices.

If only because the partial overlap of the terms, the differences between fog and cloud computing could be hard to grasp for some users. Some could consider the fog just an “extension” of the cloud. Table 1 compares the features of both cloud and fog computing to clarify how they differ.

7. CHALLENGES AHEAD

Although the research efforts and user trends described in previous sections are pushing to bring the fog, the path is far from paved. There are many open problems that will have to be addressed to make the fog a reality. It is necessary to clearly identify them so future research works have these problems into account. The set of open challenges for the fog to become a reality is: 1) Discovery/Sync Applications running on devices may need either some agreed ‘centralised’ point (e.g. establish an “upstream” backup if there are too few peers in our storage application); 2) Compute/Storage limitation Current trends are improving this fact with smaller, more energy-efficient and more powerful devices (e.g. one of today’s phones is more powerful than many high end desktops from 15 years ago). Still new improvements are granted for non consumer devices; 3) Management In addition to setting up the communication routes across end nodes, IoT/ubiquitous computing nodes and applications running on top need to be properly setup and configured to operate as desired. Having potentially billions of small devices to be configured, the fog will heavily rely on decentralised (scalable) management mechanisms that are yet to be tested at this unprecedented scale. One thing that can be predicted with certain degree of confidence is that there will be no full control of the whole fog and asymptotic declarative configuration techniques will become more common; 4) Security The same security concerns that apply to current virtualised environments can be foreseen to affect fog devices hosting applications. The presence of secure sandboxes for the execution of droplet applications poses new interesting challenges: Trust and Privacy. Before using other devices or mini-clouds in the network to run some software, isolation and sandboxing mechanisms must be in place to ensure bidirectional trust among cooperating parties. The fog will allow applications to process users data in third-party’s hardware/software. This of course introduces strong concerns about data privacy and its visibility for those third-parties; 5) Standardisation Today no standardised mechanisms are available so each member of the network (terminal, edge point...) can announce its availability to host others’ software components, and for others to sent it their software to be run; 6) Accountability/Monetisation Having users able to share they spare resources to host applications is crucial to enable new business models around the concept of the fog. A proper system of incentives needs to be created. The incentives can be financial or otherwise (e.g. unlimited free data rates). On the other hand the lack of central controlling entity in the fog makes it difficult to assert if a given device is indeed hosting a component (droplet) or not; 7) Programmability Controlling application lifecycle is already a challenge in cloud environments [18]. The presence of small functional units (droplets) in more locations (devices) calls for the right abstractions to be in place, so that programmers do not need to deal with these difficult issues [11]. Easy to use APIs for programmers will heavily rely on simple Management mechanisms that provide them with the right abstractions to hide the massive complexity of the fog. Some vendors like Microsoft have already taken some steps in positioning themselves in this space9. Table 2 discusses which of these challenges are inherently new and which ones have been inherited by the fog from one of its “parent” technologies.

8. CONCLUSIONS

The fog is nothing but the convergence of a set of technologies that have been developing and maturing in an independent manner for quite some time. The integration of these into a single IT scenario is an answer to the new requirements introduced by device ubiquity and demands for agile network and service management and data privacy. As a result the fog will dramatically shift many of our current practices at almost every layer of the IT stack, like apps development, network traffic management, network/service provision, accounting, apps collaboration mechanisms, etc. This article has provided a broad overview of this convergence and what are the common points that link all these technologies together, creating a new paradigm that some have already named as “fog” computing.

9. ACKNOWLEDGEMENTS

The authors want to thanks Julio Guijarro, Amip Shah, Suxsant Sae-Lor and Rocio Arroyo for their useful suggestions and constructive criticism on previous versions of this manuscript.

10. REFERENCES


9http://www.siliconrepublic.com/enterprise/item/36359-microsoft-is-preparing-to-d


