CLARIFYING FOG COMPUTING AND NETWORKING: 10 QUESTIONS AND ANSWERS

BY MUNG CHIANG, SANGTAE HA, CHIH-LIN I, FULVIO RISSO, AND TAO ZHANG

1. WHAT IS FOG COMPUTING AND HOW IS IT DIFFERENT FROM EDGE COMPUTING?

Fog computing is an end-to-end horizontal architecture that distributes computing, storage, control, and networking functions closer to users along the cloud-to-thing continuum.

The word "edge" may carry different meanings. A common usage of the term refers to the edge network as opposed to the core network, with equipment such as edge routers, base stations, and home gateways. In that sense, there are several differences between fog and edge.

First, fog is inclusive of cloud, core, metro, edge, clients, and things. The fog architecture will further enable pooling, orchestrating, managing, and securing the resources and functions distributed in the cloud, anywhere along the cloud-to-thing continuum, and on the things to support end-to-end services and applications. Second, fog seeks to realize a seamless continuum of computing services from the cloud to the things rather than treating the network edges as isolated computing platforms. Third, fog envisions a horizontal platform that will support the common fog computing functions for multiple industries and application domains, including but not limited to traditional telco services. Fourth, a dominant part of edge is mobile edge, whereas the fog computing architecture will be flexible enough to work over wireline as well as wireless networks.

2. IS FOG JUST A SMALLER CLOUD?

First, the size of the fog is flexible — it can range from a single small fog node to large fog systems comparable to existing clouds, depending on the application needs.

While fog will bring many cloud-like services closer to end users and can have smaller footprints than the cloud, it has a different vision from that of smaller or mini-clouds. Mini-clouds tend to be designed as isolated computing platforms. Fog envisions a seamlessly integrated cloud-fog-thing architecture to enable computing anywhere along the cloud-to-things continuum. Fog-to-cloud and fog-to-fog interactions will therefore be a focus of an end-to-end fog computing architecture to distribute computing functions, and then manage, pool, orchestrate, and secure the distributed resources and functions. Fog may form a hierarchical architecture between the cloud and the things, with fog nodes at different architectural levels collaborating with each other to support end-to-end applications.

Fog also concerns the control of cyber-physical systems and D2D communication, in addition to computation and storage in clouds, big or small.

3. IS FOG EQUIVALENT TO IOT?

Fog is an architecture. The Internet of Things (IoT) often refers to a set of services and applications.

An architecture decides the allocation of functionalities. It formulates and answers questions such as "who does what, and at what timescale and location?" An architecture supports many applications, some in existence today and others more futuristic.

For example, TCP/IP represents an Internet architecture. It includes several key principles, such as addressing, and allocation of functionalities, such as congestion-independent, hop-by-hop routing, and congestion-dependent, end-to-end session control. Applications that leverage TCP/IP have come from a wide and increasing range: from the web to emails and from P2P to video streaming.

The relationship between fog and IoT is similar to that between the Internet architecture and the web applications. Fog also supports other areas of applications, such as those in fifth generation (5G) cellular or embedded artificial intelligence.

4. IS FOG FOR COMPUTATION, OR COMMUNICATION, OR CONTROL?

Fog is an umbrella term that includes an architecture for computation, an architecture for communication, an architecture for storage, and an architecture for control (both control of the network itself and networked control in cyber-physical systems).

For example, fog computing explores new ways to decompose a computational task so as to match an underlying computation substrate that is heterogeneous (in hardware and software capabilities), volatile (in availability, mobility, and security), and constrained (by bandwidth or battery). Fog communication explores how devices may talk to each other despite intermittent global connectivity. Fog control explores how clients might crowd-sense network conditions and self-configure, and how to leverage small and almost deterministic latency to enable feedback control loops.

5. WHAT ARE THE UNIQUE ADVANTAGES OFFERED BY FOG?

Unique advantages that are potentially offered by fog can be summarized with an acronym: "SCALE." These advantages in turn enable new services and business models, and may help broaden revenues, reduce cost, or accelerate product rollouts.

Security: While fog faces unique security challenges, it also offers certain advantages. In particular, by reducing the distance that information needs to traverse, there is less chance of eavesdropping. By leveraging proximity-based authentication challenges, identity verification can be strengthened.

Cognition: Awareness of client-centric objectives. A fog architecture, aware of customer requirements, can best determine where to carry out the computing, storage, and control functions along the cloud-to-thing continuum. Fog applications, being close to the end users, can be built to be better aware of and closely reflect customer requirements.

Agility: Rapid innovation and affordable scaling. It is usually much faster and cheaper to experiment with client and edge devices rather than waiting for vendors of large network and cloud boxes to initiate or adopt an innovation. Fog will make it easier to create an open marketplace for individuals and small teams to use open application programming interfaces (APIs), open software development kits (SDKs), and the proliferation of mobile devices to innovate, develop, deploy, and operate new services.

Latency: Real-time processing and cyber-physical system control. Fog enables data analytics at the network edge and can support time-sensitive control functions for local cyber-physical systems. This is essential for not only commercial applications but also for the Tactile Internet vision to enable embedded AI applications with millisecond reaction times.

Efficiency: Pooling resources along the cloud-to-thing continuum. Fog can distribute computing, storage, and control functions anywhere between the cloud and the endpoint to take full advantage of the resources available along this continuum. It can also allow applications to leverage otherwise idle computing, storage, and networking resources abundantly available on network edge and end-user devices such as tablets, laptops, smart home appliances, connected vehicles and trains, and network edge routers. Fog's closer proximity to the endpoints will enable it to be more closely integrated with end-user systems to enhance overall system efficiency and per-

formance. This is especially important for performance-critical cyber-physical systems.

6. Is Fog Good or Bad for Security and Privacy?

Fog systems and applications will often be distributed and operated remotely. Some fog systems can also be resource-constrained. Compared to centralized clouds, such distributed, remote, and resource-constrained fog systems pose additional security challenges often encountered in distributed systems. On the other hand, fog can bring more processing resources closer to the endpoints to help better protect the vast population of diverse endpoints that often do not have sufficient resources to adequately protect themselves. In other words, fog systems can provide a wide range of local security services to make the IoT as a whole more secure. For example, fog systems can perform local security monitoring, local threat detection, and local threat protection functions on behalf of the endpoints. Fog nodes can also serve as proxies of the endpoints to help manage and update the security credentials and software on the endpoints, eliminating the often impractical needs for all the endpoints to directly communicate with the remote cloud for such functions.

7. Will the Need for Fog Diminish as Network Capacity and Delay Improve over Time?

While it is true that a primary benefit of fog computing is its ability to reduce latency and delay, the drivers for fog go far beyond pure latency issues to include a variety of operational, regulatory, business, and reliability issues.

For example, instead of the traditional way of adding new applications by adding dedicated new local servers and networking gear, fog can provide a common end-to-end platform for all services provided to each customer. This can provide a unified platform to support life cycle management, networking, and security for all applications, which will reduce system complexity and costs and also allow applications from different providers to better interact with each other rather than stay siloed on their dedicated hardware and software platforms. Fog can enable critical services to be operated autonomously or managed from the cloud, the perimeter, or a variety of points in the network. Fog is equally advantageous for areas where network connectivity can be unreliable due to weather or other conditions. It can also significantly reduce network bandwidth loads through its proximity to where the data is generated. With fog, local operational and business policies can be applied to enable more efficient local data processing and analytics on premises.

As another example in cellular networks, cloud RAN, with centralized or distributed network architecture, has the advantage of being physically close to the end users and the capability of utilizing the network resources at the edge. Consequently, the cloud radio access network (C-RAN) will be an integral part of the solution to meet stringent network delay requirements that the traditional RAN network may fail to meet. Consequently, the Third Generation Partnership Project (3GPP) is now discussing a RAN architecture that contains both the central units and the distributed units. Extending prior notions in C-RAN, fog network is unique in the sense that the end user computing and storage resource is considered as an integral part of the whole network, by forming ad hoc subnets among end nodes. Careful exploitation of such features will bring unique values for fog networks.

8. What New Technologies and Standards, If Any, Do We Need to Develop for Fog?

Fog systems will need to interact with each other, with the clouds, and with a diverse range of user end devices. Therefore, the success and wide adoption of fog computing will rely on standards. While fog computing can benefit from many existing standards, new standards may also be required, for example, in the following areas:

Building unified fog-cloud platforms: Interfaces and protocols for the fog and the cloud to interact with each other to enable unified cloud-fog service platform and applications, move computing functions and applications between the cloud and the fog, pool resources distributed in the cloud and the fog, and manage the life cycle of the fog systems and applications.

Support distributed and hierarchical fog systems over possibly heterogeneous, volatile, and constrained physical resources: interfaces and protocols for different hierarchical levels in a fog system to interact with each other, and for different fog systems at the same hierarchical level to collaborate with each other to serve as each other's backup.

Access to fog services: A fog system, bringing resources closer to end users, can enable a wide range of new fog-based services. Standards will be required for users and their devices to interact with the fog system to discover, request, and receive fog services. So will automatic and lightweight bidding mechanisms for access to fog resources and services to reinforce the economic sustainability of the fog computing model, and enabling economic transactions.

Data management: Local processing and management of data is one of the important drivers for fog computing. Data, however, comes from an increasingly wide range of sources. Data management also imposes widely diverse requirements from industry to industry. New standards may be required to manage the diverse data, such as storing, accessing, and securing the data distributed in the fog and cloud.

Security and privacy: A distributed and remotely operated fog system can pose new security challenges not present in centralized systems. Addressing these new challenges may require new standards. For example, fog computing will need to run a diverse set of local hardware platforms. Therefore, new interfaces may be required for fog software to interact with the various hardware platforms, which may be provided by different vendors, to ensure a trusted computing environment. New interfaces and protocols may be required for automatic detection of security compromises in a distributed and remote fog system, and also for remote and automatic responses to security compromises.

Furthermore, although standards may exist for some fog computing needs, additional requirements in fog computing environments (e.g., low-latency, large number of resource-constrained devices) may necessitate new standards that are more suitable for fog computing environments.

9. What New Research Challenges Do We Have to Address to Enable Fog?

Research challenges in fog span a wide range: from computation decomposition over heterogeneous and constrained nodes to cloud-fog interface definition, from state consistency in dispersive computing to elastic storage over volatile substrate, from pricing for economic incentives to scalable security measures. Fundamental to these topics is the intrinsic trade-off between "local" and "global" and between "brick" and "click" as we slide between cloud and things in deciding where to allocate a function and how to glue them back together.

For example, fog computing enables a complex service to possibly be delivered through a set of elementary software elements that operate on heterogeneous nodes, such as end user terminals and local servers, but also network elements and data centers. The problem of the orchestration of

the above complex services is definitely an important challenge, complicated by the highly dynamic environment, the many fog-enabled applications installed on end user devices and things, and the necessity to support different administrative domains, to adapt the service to the extreme heterogeneity of the infrastructure, and to adapt the service (and the orchestration algorithms) to the external environment. For instance, fog applications cannot always count on the availability of powerful computing devices that can execute complex orchestration algorithms; in critical conditions (e.g., broken infrastructure as in the case of an earthquake), the fog infrastructure has to be able to orchestrate services even in the presence of limited computing capabilities, with an intrinsic degree of resiliency. Along this line, another important challenge is the capability to create self-adapting applications, which are able to automatically adapt their behavior based on the surrounding environment, for example, in case some required services (e.g., high-capacity storage or a high-precision sensor) cannot be reached, while still being able to deliver the service the user is expecting, albeit with some degradation.

10. What Commercial Opportunities Will Fog Bring?

Fog computing will bring many new commercial opportunities and will disrupt the existing industry landscapes and business models, disrupting the balance of power along the industry food chain.

For example, networking functions (e.g., routing and switching), application servers, and storage functions are already converging into integrated "fog nodes": edge devices that integrate edge router and local application server and storage functions are already commercially available. The emerging fog systems will empower the cloud to do what it cannot effectively do today by, for example, acting as proxies to connect and then provide cloud services to the many devices that cannot be practically connected to the cloud directly. A growing range of innovative fog-based services, including fog systems and services as a service, will emerge. The cloud and the fog will converge into unified end-to-end platforms and provide integrated services and applications, creating opportunities for fundamental disruptions to the existing cloud computing business models. Players of all sizes will be able to deploy fog systems and operate fog services. And the list goes on.