Fog Computing: Focusing on Mobile Users at the Edge

Tom H. Luan, Longxiang Gao, Yang Xiang
School of Information Technology
Deakin University
Burwood Melbourne, VIC 3125, Australia
Email: {tom.luan,longxiang.gao,yang.xiang}@deakin.edu.au

Zhi Li, Limin Sun
Institute of Information Engineering
Chinese Academy of Sciences
Beijing, China
Email: {lizhi,sunimin}@iie.ac.cn

Abstract—With smartphones becoming our everyday companions, high-quality mobile applications have become an important integral of people’s lives. The intensive and ubiquitous use of mobile applications have also led to the explosive growth of mobile data traffics. Therefore, to accommodate the surge mobile traffic yet providing the guaranteed service quality to mobile users represent a key issue of the next generation mobile networks. This motivates the emergence of Fog computing as a promising, practical and efficient solution tailored to serving mobile traffics. Fog computing deploys highly virtualized computing and communication facilities at the proximity of mobile users. Dedicated to serving the mobile users, Fog computing explores the predictable service demand patterns of mobile users and typically provides desirable localized services accordingly. Stitching above features, Fog computing can provide mobile users with the demanded services via low-latency and short-distance local connections. In this article, we outline the main features of Fog computing and describe its concept, architecture and design goals. Lastly, we discuss on the potential research issues from the networking’s perspective.

I. INTRODUCTION

Our networking today is shaped by two salient trends.

Cloud-based Architecture: the Internet has shifted to the cloud-based architecture. Notably, cloud computing has already evolved as the key computing infrastructure for Internet services with full-fledged service models and applications. As reported in Cisco Cloud Index (2013-2018) since 2008, most Internet traffic has originated or terminated in a data center. By 2016, it is predicted that nearly two-thirds of total workloads in traditional IT space will be processed in the cloud.

Surge Mobile Traffics: in 2011, the smartphone shipment worldwide overtook that of PCs for the first time in history. The smartphones have brought extensive computing and communication capability to the palm of people’s hand. As a result, a multitude of innovative mobile applications, such as social networking, navigation, and internet-of-things applications including Ehealthcare, smart home and community, etc., have resulted in the fundamental changes in the pattern we live. The intensive use of mobile applications finally leads to the surge growth of the mobile data traffic. As reported in Cisco Visual Networking Index (2013-2018), by 2018 the global mobile data traffic will increase to 11-fold of that in 2013. Therefore, from the nation’s perspective, to accommodate the surge mobile data traffic in a scalable and efficient approach represents an urgent goal. From the user’s perspective, acquiring desirable mobile applications with guaranteed service quality at anytime anywhere is an important determinant of life quality.

With cloud becoming the overarching approach for centralized information storage, retrieval and management, and mobile devices becoming the major destination of information, the successful integration of cloud computing and mobile applications therefore represents an important task. However, due to the far distance between the cloud and mobile devices, the enjoyable high-rate data exchange between cloud and mobile user is still hurdled by fundamental engineering challenges. This makes high-quality cloud-based mobile service applications still a far end to achieve, and motivates the emergence of Fog computing as an effective solution towards the smooth convergence of cloud and mobile applications.

The term “Fog computing” was first proposed by Cisco in 2012 [1]. Similar systems typically known as edge computing, such as Cyber Foraging [2], Cloudlets [3], can date back to early 2000. However, the research on Fog computing and related systems still remains at a very early stage. The goal of this article is to investigate on the key features of Fog computing and identify its main design goals and open research issues accordingly from the networking’s perspective. In the rest part of the article, we unfold our journey by first describing the basic system architecture of Fog computing and showcase exemplary applications. After that, we discuss on the fundamental motivation behind the Fog computing and its comparisons with existing related networking systems. Lastly, we discuss on the potential research directions and conclude this article.

II. SYSTEM ARCHITECTURE

Fog computing extends the cloud-based Internet by introducing an intermediate layer between mobile devices and cloud, aiming at the smooth, low-latency service delivery from the cloud to mobile. This accordingly leads to a three hierarchy Mobile-Fog-Cloud architecture as depicted in Fig. 1.

The intermediate Fog layer is composed of geo-distributed Fog servers which are deployed at the edge of networks, e.g., parks, bus terminals, shopping centers, etc. Each Fog server is a highly virtualized computing system, similar to a lightweight cloud server, and is equipped with the on-board large-
Fog computing is dedicated to serving the mobile users with the local short-distance high-rate connections. This overcomes the drawback of cloud which is far to mobile users with elongated service delays. Therefore, the fog is interpreted as “the cloud close to the ground” [1].

III. EXEMPLARY IMPLEMENTATIONS

The fog thus behaves as a surrogate of cloud or a private cloud at the user’s premises. This enables Fog servers to be more efficient to handle the localized computation requests. Therefore, Fog computing targets to deliver the localized and location-based service applications to mobile users. In what follows, we showcase some examples of Fog computing implementation from this perspective:

a) Shopping Center: Assuming that a number of Fog servers are deployed inside a multi-floor shopping center, which collectively form an integrated localized information system. The Fog servers at different floors can pre-cache floor-related contents, such as the layout and ads of stores on a particular floor. The Fog servers can deliver engaged services including indoor navigation, ads distribution and feedback collections to mobile users through WiFi.

b) Parkland: The Fog computing system can be deployed in the parkland to provide localized travel services. For instance, Fog servers can be deployed at the entrance and other important locations of the park. The Fog server at the entrance can pre-cache information including park map, travel guide and local accommodations; other Fog servers at different locations inside the park can be incorporated with sensor networks for environment monitoring and provide navigation to travellers. By connecting the Fog servers to the park administration office and cloud, the Fog servers can be used as an information gateway to send timely alerts and notifications to travellers.

c) Inter-state Bus: Greyhound has launched “BLUE” [4], an on-board Fog computing system over inter-state buses for entertainment services. As an example illustrated in Fig. 2, a Fog server can be deployed inside the bus and provides on-board video streaming, gaming and social networking services to travellers using WiFi. The on-board Fog server connects to the cloud through cellular networks to refresh the pre-cached contents and update application services. Using its computing facility, the Fog server can also collect and process user’s data, such as number of travellers and their feedbacks, and reports to cloud.

d) Vehicular Fog Computing Networks: Luan et. al. [5] present the application of Fog computing as an integrated large-scale network for localized content disseminations. Fig. 3 shows a motivating scenario. Assuming that a store installs a Fog server at its parking lot with the purpose to distribute the store flyer. In step 1, the store uploads flyers to the Fog server via wireless connections, and the Fog server distributes the flyers wirelessly to vehicles driving through its coverage using wireless communications. With the vehicle moving to different locations, it can further disseminate the cached flyers to other vehicles using wireless communications, as depicted in step 2. In step 3, the flyers can also be retrieved and cached at other Fog servers deployed at different locations, e.g., bus stop, and further propagated in the network.

IV. COMPARISON TO CLOUD COMPUTING

Based on the examples presented, the Fog computing has following three features:

- Wireless: Fog computing is dedicated to serving the mobile users. Each Fog server typically has limited
Wireless coverage, e.g., 200m using WiFi, and directly interacts with mobile users using the single-hop wireless connections.

- **Local Services**: Fog computing is dedicated to serving the localized information and providing location-based service applications. For example, the Fog computing system deployed in a specific park only provides the navigation services associated with the park.

- **Distributed Management**: A Fog computing system may typically be deployed and managed by the local business, with the purpose to deliver designated contents and services to specific user groups.

Using the example in Fig. 3 for illustration, the Fog servers provide localized content distribution using wireless communications, which matches the first two features. The Fog server deployed nearby the store may be installed and managed by the store owner for the distribution of store flyers; the Fog server at the bus stop may be managed by the bus company for the distribution of bus information, e.g., bus schedules, safety manual, etc., to mobile users waiting for the bus. The Fog computing system in Fig. 3 is therefore distributedly constructed with Fog servers distributed installed and managed by different entities to serve their own purposes, which matches the third feature.

Table I summarizes the differences between Fog computing and Cloud computing.

By targeting to different user groups at different locations, Fog computing extends cloud computing to better serve local mobile traffics. As the system architecture shown in Fig. 1, the Fog servers deployed at different locations would be used to deliver engaged services specified by their owners. The Fog servers at different locations can connect to the same cloud and form an integrated Fog computing system in a wide region.

V. WHY FOG COMPUTING?

The above features of Fog computing attribute to its motivation as an enabling technology for the convergence of mobile applications and cloud. To show the rationale of Fog computing, in what follows we take a retrospect study by revisiting the design of cloud-based Internet and service requirements of mobile users.

A. Global and Local Information

Cloud computing is an efficient centralized solution for information management and distribution. It is efficient to serve the information requests from the traditional desktop Internet users. To be specific, the desktop users, typically accessing the Internet at homes and offices are often interested in the information which is irrelevant to their locations, such as the world news, stock market at different cities or countries, to name a few. We refer to such information as the global information. As a contrast, we refer to the location-based service information related to the location of users as the local information.

Note that cloud computing is an optimized approach for serving such global information services. In specific, cloud computing is a scalable and efficient approach to store and manage the information originated at different locations of the world. In the meantime, using a reliable public IP, cloud computing provides a convenient way to distribute the cached global information from a central server for public accesses. The convenient and reliable cloud computing framework, together with fast-rate cable connections at homes/offices, therefore favor the desktop users to meet their requests on global information.

The mobile users, however, have distinguished service requirements from the desktop users. This requires the current cloud-based Internet to be modified accordingly to cater to the specific service requirements of mobile users. In specific, differ from desktop users, the mobile users, particularly smartphones, are typically in the outdoor environments. This makes their service requirements closely related to their current locations. In other words, mobile users are more interested in the local information around them. For example, a mobile user in a...
**TABLE I**

**COMPARISON OF FOG COMPUTING AND CLOUD COMPUTING**

|                      | Fog Computing                                                                 | Cloud Computing                                      |
|----------------------|-------------------------------------------------------------------------------|-----------------------------------------------------|
| **Target User**      | Mobile users                                                                  | General Internet users.                              |
| **Service Type**     | Limited localized information services related to specific deployment locations| Global information collected from worldwide          |
| **Hardware**         | Limited storage, compute power and wireless interface                         | Ample and scalable storage space and compute power   |
| **Distance to Users**| In the physical proximity and communicate through single-hop wireless connection| Faraway from users and communicate through IP networks|
| **Working Environment** | Outdoor (streets, parklands, etc.) or indoor (restaurants, shopping malls, etc.) | Warehouse-size building with air conditioning systems |
| **Deployment**       | Centralized or distributed in reginal areas by local business (local telecommunication vendor, shopping mall retailer, etc.) | Centralized and maintained by Amazon, Google, etc. |

**B. Physical and Communication Distance**

The cloud-based Internet can be inefficient to serve the local information desired by mobile users. As a motivating example shown in Fig. 4(a), assuming that a mobile user inside a shopping center intends to retrieve flyers of stores within the shopping center. To do this using the cloud-based Internet, the stores may need to first upload their flyers to a remote cloud server over Internet, and then direct mobile users to retrieve the desired information from the remote cloud server. In other words, although the physical distance between the mobile user (destination) and stores (original source) is short, using the remote cloud as the information depot, the actual communication distance can be far, e.g., from the cloud server to mobile user in this example.

The Fog computing paradigm represents a practical and efficient solution to resolve the mismatch between physical and communication distances. As a remedy shown in Fig. 4(b) a Fog server can be deployed inside the shopping center and to distribute the local store flayers to mobile users. As such, the physical distance is equal to the communication distance and users can acquire low-latency desirable services.

By minimizing the communication distance, the Fog computing therefore brings the following two advantages:

- **To mobile users**: compared to cloud, the Fog computing can provide enhanced service quality with much...
increased data rate and reduced latency and response time. Moreover, by reducing the bandwidth cost of data transmission in the backbone, the users can also be benefited from the reduced service cost.

- **To network:** by avoiding the duplicated back and forth traffic between cloud and mobile user, not only the backbone bandwidth can be significantly saved, the energy consumption of core networks can also be greatly reduced, which contributes to the sustainable development of networking.

VI. STORAGE, COMPUTE AND COMMUNICATION

A Fog server is a generic virtualized device equipped with the storage, computing and communication capability. Dedicated to serving the localized and location-based applications, a Fog server manages its on-board resources to fully explore the location information and predictable user demand with the following functions.

A. Storage

In a predefined service area, a Fog server predicts the mobile user’s demand on information and pre-cache the desirable information accordingly using a proactive way in its storage. Such information can be either retrieved from the cloud or uploaded by its owner. For example, the Fog server installed at a restaurant can pre-cache the menu of the restaurant and dish recipes to serve the mobile users inside the restaurant. In another example, the Fog servers deployed in the airport can pre-cache the flight and local transportation information which is desirable to travellers in the airport. Therefore, the key design issue of Fog computing is to predict the user’s demand and proactively select the contents to cache in the geo-distributed Fog servers based on the specific locations.

**Compared to Related Systems:** The Content Delivery Network (CDN) [7] represents the most mature catch networks and extensively pursued in both academic and industry. CDN is the Internet-based cache network by deploying cache servers at the edge of Internet to reduce the download delay of contents from remote sites. CDN mainly targets to serve traditional desktop Internet users, which have much broader interests and blur service demands that are more difficult to predict than those of mobile users. With precise service region, Fog computing has more clear target users and provides the local information service to mobile users. The Fog computing is a generic platform for edge computing and focuses on the localized service and computing requests. The prototype and techniques in [10], [11] can be incorporated in Fog computing overview.

B. Compute

A salient feature that differentiates Fog computing from the traditional cache networks is that Fog servers are intelligent compute system. This allow a Fog server to autonomously and independently serve local computation and data processing requests from mobile users. [10] shows the applications of Fog computing in the cognitive applications. In another example, a Fog server inside the shopping mall or parkland can maintain an on-board geographic information system, and provide the real-time navigation and video streaming to connected mobile users.

Bridging the mobile and cloud, a Fog server can also be conveniently used to collect the environmental data or demographic data from mobile users at the deployed spot, and transport the collected big data to cloud for in-depth analysis; the results can be provided to third party for strategic and valuable insights on business and government event planning, execution and measurement.

**Compared to Related Systems:** Despite of the high computing power, the cloud is faraway from mobile users and can hardly support real-time computing intensive applications due to the bandwidth-constrained IP networks. The demand of real-time resource-intensive mobile applications, e.g., cognitive and internet-of-things applications, motivates the design of ubiquitous edge computing system [10], [11]. Cloudlets [3], [10] adopt the same framework of Fog computing, in which a Cloudlet server, similar to the Fog server, is deployed in the proximity of mobile users and processes the computing requests of mobile devices at real-time for video streaming and data processing. A comparison of processing delays using Cloudlets and Amazon clouds is shown in http://elijah.cs.cmu.edu/demo.html. Transparent computing [11] is a highly virtualized system, which targets to develop the computing system transparent to users with cross-platform and cross-application support.

The Fog computing is a generic platform for edge computing and focuses on the localized service applications and computation requests. The prototype and techniques in [10], [11] can be incorporated in Fog computing framework.

C. Communication

Fog server can equip different wireless interfaces, e.g., WiFi, Bluetooth and visible light communications [12] according to the specific application scenarios. The Fog computing...
differs from traditional radio access networks, e.g., WiFi and Femtocell networks, in two important ways.

**Cross-layer Design:** Unlike traditional WiFi access points, the Fog server manages an autonomous, all-inclusive network by providing both service applications and wireless communications to mobile users in the coverage. Therefore, a Fog server can work without Internet connections as stated in [5]. Note that the Fog computing tailors its applications based on the specific deployment location and environment, and therefore is highly service-oriented. To this end, a Fog server can manage all the communication layers and effectively enables the cross-layer design [13] to provide the best service quality to users. For example, as in “BLUE” [4], a Fog server can cache a number of videos and deliver Youtube-like video streaming services to mobile users in the proximity. In this case, based on the context, wireless channel and video popularity information, the video services can be conveniently adapted towards the optimal performance via cross-layer adjustments.

**Predictable Location-based Service:** The key of Fog computing is to provide the localized network and information applications to mobile users, whereas the traditional radio access networks focus on the provision of Internet applications and global information. With this distinguished feature, the design of Fog computing communications needs to consider the specific deployment environment and the features of mobile users in the considered scenario. For example, a Fog computing system deployed in the shopping mall needs to address the diverse mobilities of users, whereas the similar system deployed in the inter-state bus [4] only needs to consider static on-board passengers.

**VII. Future Research Topics**

Based on the Mobile-Fog-Cloud hierarchy shown in Fig. 5, we envision potential research directions from the communication efficiency’s viewpoint as follows.

**A. Communications between Mobile and Fog**

Note that a Fog server manages three-dimensional resources including storage, computing and communication. The service quality acquired by users relies on the collective performance of resource utilization from all the three dimensions. Moreover, as Fog computing typically provides pre-defined application services and targets to specific user groups, the service-oriented resource allocation customized to the specific deployment environments is thus necessary. For example, considering the on-board Fog computing system inside the inter-state bus as in Fig. 2 three types of traffics may coexist including video streaming, gaming and web surfing delivered through the same Fog server. As such, a cross-layer MAC design at the Fog server can be devised based on the application’s information. Considering that Fog servers have limited storage and deliver limited localized services only, another key design issue is how to optimally select the desirable information contents to cache at each Fog server and determine the appropriate service applications which cause the least service failure rates to mobile users. The solution needs to consider the predictable pattern of mobile service requests, available storage and compute power of a Fog server.

The Fog computing can also be incorporated with the 5G cellular networks. In this case, by making the cellular base station a Fog server with on-board storage and compute facility, the entire Fog system can provide greater coverage and dedicated services to cellular users.

**B. Communications between Fog and Cloud**

The cloud performs two roles in the integrated Fog computing system. First, the cloud is the central controller of Fog servers deployed at different locations. With each Fog server focusing on the service delivery to mobile users at specific locations, the cloud manages and coordinates the geo-distributed Fog server clusters at different regions. Second, the cloud is the central information depot. The Fog servers at different locations select the information contents from the cloud and then deliver the copied contents from its cache to the mobile users. With above two roles, the design goal of the communications between fog and cloud can be two-fold: 1. how to enable the reliable and scalable control of Fog servers at the cloud; and 2. how to develop the scalable data routing scheme from cloud to Fog server for content updates.

Note that the dual functions of cloud as stated above well match the architecture of a software-defined networking (SDN) [14], which decouples the traffic routing to the control plane and data plane. It is thus promising to apply the SDN scheme for the control of Fog computing.

**C. Communications between Fogs**

As summarized in Table 1 and illustrated in Fig. 1, the Fog servers at different locations may be deployed by different entities for different commercial usages. As a result, they may conform to different policies defined by owners and are therefore independent to each other. However, as the Fog servers belonging to different owners may locate at close distance to each other and serve the similar mobile users, it is thus possible for separate Fog server clusters to collaborate for joint service delivery. For instance, as the example in Fig. 3 the Fog servers across the city collaborate for the distribution of similar contents to vehicles. In addition, since Fog servers...
co-located in the same region may be connected to the Internet through the same Internet service provider with the high-rate low-cost connections, to enable efficient collaborations among nearby Fog servers can alleviate the traffic between cloud and Fog servers, and improve the system performance with saved bandwidth cost and enhanced data rate.

D. Internet-of-Things Applications

As Fog servers are deployed at the physical spot close to mobile users and can be equipped with sensors, it is convenient to incorporate the Fog computing with the Internet-of-things applications. [1], [15] present the examples of adopting Fog computing in the applications of smartgrid, vehicular networks and sensor networks.

VIII. Conclusion

This article presents Fog computing, a new networking frontier dedicated to serving mobile users. By deploying reserved compute and communication resources at the edge, Fog computing absorbs the intensive mobile traffic using local fast-rate connections and relieves the long back and forth data transmissions among cloud and mobile devices. This significantly improves the service quality perceived by mobile users and, more importantly, greatly save both the bandwidth cost and energy consumptions inside the Internet backbone. Therefore, Fog computing represents a scalable, sustainable and efficient solution to enable the convergence of cloud-based Internet and the mobile computing. The purpose of this article is to investigate on the major motivation and design goals of Fog computing from the networking’s perspective. We emphasis that the emergence of Fog computing is motivated by the predictable service demands of mobile users, and Fog computing is thus mainly used to deliver the localized service requests. As a Fog server possesses hardware resources in three-dimensions (storage, compute and communications), the three-dimensional service-oriented resource allocations are therefore the key of Fog computing. Moreover, with the three-tier Mobile-Fog-Cloud architecture and rich potential applications in both mobile networking and Internet-of-things, the Fog computing also opens broad research issues on network management, traffic engineering, big data and novel service delivery. Therefore, we envision a bright future of Fog computing.

REFERENCES

[1] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog Computing and Its Role in the Internet of Things,” in Proceedings of ACM MCC, 2012, pp. 13–16.
[2] R. Balan, J. Flinn, M. Satyanarayanan, S. Sinnamohideen, and H. I. Yang, “The Case for Cyber Foraging,” in Proceedings of ACM SIGOPS, 2002, pp. 87–92.
[3] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, “The Case for VM-Based Cloudlets in Mobile Computing,” IEEE Pervasive Computing, vol. 8, no. 4, pp. 14–23, 2009.
[4] G. M. Relations, “Greyhound Launches “BLUE”, an Exclusive Wi-Fi Enabled Onboard Entertainment System,” https://www.greyhound.com/en/newsroom/viewrelease.aspx?id=528&year=2013, Accessed on Dec. 2014.
[5] T. H. Luan, L. X. Cai, J. Chen, X. Shen, and F. Bai, “Vtube: Towards the Media Rich City Life with Autonomous Vehicular Content Distribution,” in Proceedings of IEEE SECON, 2011, pp. 359–367.
[6] “Location-based Mobile Services Have Huge Untapped Potential Worldwide,” 2012. [Online]. Available: http://thenexthome.com/mobile/2012/04/24/report-location-based-mobile-services-have-huge-untapped-potential-worldwide/
[7] G. Peng, “CDN: Content Distribution Network,” arXiv preprint cs/0411069, 2004.
[8] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A Survey of Information-Centric Networking,” IEEE Communications Magazine, vol. 50, no. 7, pp. 26–36, 2012.
[9] E. Baytuğ, M. Bentiss, and M. Debbah, “Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks,” IEEE Communications Magazine, vol. 52, pp. 82 – 89, 2014.
[10] M. Satyanarayanan, Z. Chen, K. Ha, W. Hu, W. Richter, and P. Pillai, “Cloudlets: at the Leading Edge of Mobile-Cloud Convergence,” in Proceedings of MobiCASE, 2014.
[11] Y. Zhang and Y. Zhou, “Transparent Computing: a New Paradigm for Pervasive Computing,” in Ubiquitous Intelligence and Computing, pp. 1–11.
[12] T. Komine and M. Nakagawa, “Fundamental Analysis for Visible-light Communication System using LED Lights,” IEEE Transactions on Consumer Electronics, vol. 50, no. 1, pp. 100–107, 2004.
[13] C. X. Lin, X. Shen, J. W. Mark, L. Cai, and Y. Xiao, “Voice Capacity Analysis of WLAN with Unbalanced Traffic,” IEEE Transactions on Vehicular Technology, vol. 55, no. 3, pp. 752–761, 2006.
[14] H. Kim and N. Faeemster, “Improving Network Management with Software Defined Networking,” IEEE Communications Magazine, vol. 51, no. 2, pp. 114–119, February 2013.
[15] I. Stojmenovic and S. Wen, “The Fog Computing Paradigm: Scenarios and Security Issues,” in Proceedings of FedCSIS, 2014.