

## 1

## Fog Computing and Fogonomics

Yang Yang<sup>1</sup>, Jianwei Huang<sup>2</sup>, Tao Zhang<sup>3</sup>, and Joe Weinman<sup>4</sup>

<sup>1</sup>Shanghai Institute of Fog Computing Technology (SHIFT), ShanghaiTech University, Shanghai, China

<sup>2</sup>School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China

<sup>3</sup>National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

<sup>4</sup>XFORMA LLC, Flanders, NJ, USA

As a new computing paradigm, fog computing serves as the bridge that connects centralized clouds and distributed edges of the network and plays the crucial role in managing and coordinating multitier computing resources at the cloud, in the network, at the edge, and on the things (devices). In other words, fog computing provides a new architecture that spans along the cloud-to-things continuum, thus effectively pooling dispersive computing resources at global, regional, local, and device levels to quickly meet various service requirements. Together with the edge, fog computing ensures timely data processing, situation analysis, and decision-making at the locations close to where the data are generated and should be used. Together with the cloud, fog computing supports more intelligent applications and sophisticated services in different industrial verticals and scenarios, such as cross-domain data analysis, pattern recognition, and behavior prediction. Some infrastructure challenges and constraints in communication bandwidth, network connectivity, and service latency can be successfully addressed by fog computing, since it makes computing resources in any networks more accessible, flexible, efficient, and cost-effective. It is no doubt that fog computing will not only empower end users by enabling intelligent services in their neighborhoods but also more importantly, deliver a broad variety of benefits to business, consumers, governments, and societies. This book aims at providing a state-of-the-art review and analysis of key opportunities and challenges of fog computing in different application scenarios and business models.

The following three chapters address different technical and economic issues in collaborative fog and cloud scenarios. Specifically, Chapter 2 introduces the hybrid fog–cloud scenario that combines the whole set of resources from the edge up to the cloud, describing the challenges that need to be addressed to

enable realistic management solutions, as well as a review of the current efforts. The authors propose an architectural solution called Fog-to-Cloud (F2C) as a candidate to efficiently manage the set of resources in the IoT-fog-cloud stack. Such an architectural solution is conceptually supported by a service and technology agnostic software solution, which is discussed thoroughly in this chapter in comparison to other existing initiatives. The proposed F2C architecture has two key advantages: (i) it is open and secure by design, easily adoptable by any system environments through distinct software suites and (ii) it has an inherent collaborative model that supports multiple users to optimize resources utilization and services execution. Finally, the authors analyze main challenges for building a stable, scalable, and optimized solution, from both the resource and service perspectives, with special attention to how data must be managed.

In Chapter 3, the authors give an overview of fog computing and highlight the challenges due to the tremendous growth of various Internet of Things (IoT) systems and applications in recent years. They propose a mechanism to efficiently allocate the computing resources in the cloud and fog to different IoT users, in order to maximize their quality of experience (QoE), i.e. less energy consumption and computation delay. The competition among multiple users is modeled as a potential game to determine the computation offloading decisions. The existence of a pure Nash equilibrium (NE) is proven for this game, and it is shown that the equilibrium efficiency loss due to the strategic behavior of users is bounded. A best response strategy algorithm is then developed to obtain an NE of the computation offloading game. Numerical results reveal that the proposed mechanism significantly enhances the overall QoE, and in particular, 18% more users can benefit from computing services than the existing offloading mechanism. The results also demonstrate that the proposed mechanism is promising to enable low-latency computing services for delay-sensitive IoT applications.

In Chapter 4, the authors examine the pricing and performance trade-offs in data analytics. First, they introduce different types of computing devices employed in fog and cloud scenarios, review the current pricing techniques in use, and discuss their implications for performance criteria like accuracy and latency. Then, a data analytics case is studied under a testbed of temperature sensors, where the temperature readings can be analyzed either at local Raspberry Pis or on a cloud server. Local analysis can reduce the communication overhead as raw data are no longer sent to the cloud server, but it lengthens the computation time as Raspberry Pis have less computing capacity than cloud servers. Thus, it is not immediately clear whether fog-based or cloud-based analysis leads to a lower overall completion time; indeed, a hybrid algorithm that can utilize both types of resources in parallel will likely minimize the completion time. However, the choice between a fog-based, cloud-based, or hybrid algorithm also induces different monetary costs (including both computation and data transmission costs) and may lead to different levels of accuracy, since

the local analysis involves analyzing only subsets of data and later combining their results due to the Raspberry Pis' limited computing capacity. The authors examine these trade-offs for a simple linear regression scenario and show that there is a threshold number of samples above which a hybrid algorithm is preferred to the cloud-based one.

In Chapter 5, the authors outline a number of qualitative and quantitative arguments and frameworks to help rationally assess the economic benefits and trade-offs between different approaches. For example, resource consolidation tends to increase latency to and from distributed edge and fog services. On the other hand, it tends to reduce latency to cloud-based data and services. The statistics of independent, identically distributed workload demands can benefit from aggregation: multiple independent varying workloads tend to “cancel” each other out, leading to a precisely quantifiable smoothing effect that boosts utilization for a given resource level, which in turn reduces the weighted unit cost of resources. In short, there are many quantifiable characteristics of the fog, which can be evaluated in light of alternative architectures. Ultimately this illustrates that there is no “perfect” solution, as trade-offs need to be quantified and assessed in light of specific application requirements.

In Chapter 6, the authors analyze the design challenges of incentive mechanisms for encouraging user engagements in user-provided infrastructures (UPIs). Motivated by novel business models in network sharing solutions, they focus on mobile UPIs, where the energy consumption and data usage costs are critical, while storage and computation resources are limited. Hence, these parameters have large impacts on users' decisions for requesting/offering their resources from/to UPIs. This chapter reviews a set of incentive schemes that have been proposed for such UPIs, leveraging cooperative game theory, bargaining theory, and auctions. The authors shed light on the attained equilibriums, and study the efficiency and sensitivity on various system parameters. Furthermore, the impact of the network graph on the collaboration benefits in UPI systems is modeled and analyzed, and whether local user interactions achieve system-wide efficient sharing equilibriums is explored. Finally, key bottleneck issues are discussed in order to unleash the full potential of UPIs in fog computing.

In Chapter 7, the authors introduce a Fog-based Service Enablement Architecture (FogSEA), which is a light-weight, decentralized service enablement model. It supports fog services sharing at network edges by adopting a hierarchical management strategy and underpins cross-domain IoT applications based on a semantic-based overlay network. They also propose the Semantic Data Dependency Overlay Network (SeDDON) network, which maintains the semantic information about available microservices. SeDDON aims to reduce the traffic cost and the response time during service discovery. FogSEA produces less traffic and takes less time to return an execution result, comparing to the baseline approach. Generally, traffic increases as more microservices

join the network. SeDDON creation needs to send less messages at varying connectivity densities and microservice numbers. The main reason is that SeDDON allows microservices to advertise their services only once when they join the network, and only the microservice that detects the new node as a reverse-dependence neighbor needs to reply.

In Chapter 8, the authors firstly discuss new characteristics and open challenges of realizing fog orchestration for IoT services before summarizing the fundamental requirements. Then, they propose a software-defined orchestration architecture that decouples software-based control policies from the dependencies and operations of heterogeneous hardware. This design can intelligently compose and orchestrate thousands of heterogeneous fog appliances. Specifically, a resource filteringbased resource assignment mechanism is developed to optimize the resource utilization and fair resource sharing among multitenant IoT applications. Additionally, a component selection and placement mechanism is adopted for containerized IoT microservices to minimize the latency by harnessing the network uncertainty and security while considering different application requirements and appliance capabilities. Finally, a fog simulation platform is presented to evaluate the aforementioned procedures by modeling the entities, their attributes, and actions. The practical experiences show that the proposed parallelized orchestrator can reduce the execution time by 50% with at least 30% higher orchestration quality.

In Chapter 9, the authors focus on the problem of reliable Quality of Service (QoS) – aware service choreography within a fog environment where service providers may be unreliable. A distributed QoS optimized adaptive system is proposed to help users in selecting the best available service based on its reputation and to monitor the run-time performance of the service according to the predetermined Service Level Agreement (SLA). A service adaptation model is described to keep the system up with an expected run-time QoS when the SLA is violated. In addition, a performance validation mechanism is developed for the fog environment, which adopts a monitoring and negotiation component to enable the reputation system.

In Chapter 10, the authors propose a typical fog network consisting of multiple fog nodes (FNs), wherein some task nodes (TNs) have heavy computation tasks, while some helper nodes (HNs) have spare resources for sharing with their neighboring nodes. To minimize the delay of every task in such a fog network, a noncooperative game is formulated and investigated to model the competition among TNs for the communication resources and computation capabilities of HNs. Then, a comprehensive analytical model that considers circuit, computation, offloading energy consumptions is developed for accurately evaluating the overall energy efficiency. With this model, the trade-off relationship between performance gains and energy costs in collaborative task offloading is investigated. A novel delay energy balanced task scheduling (DEBTS) algorithm is proposed to minimize the overall energy

consumption while reducing average service delay and delay jitter. Further, extensive simulation results show DEBTS can offer much better delay-energy performance in task scheduling challenges.

In Chapter 11, the authors explore both noncooperative and cooperative perspectives of resource sharing issues in multiuser fog networks. On one hand, for the noncooperative distributed computation offloading scenario, the authors develop a game theoretic mechanism with fast convergence property and good performance guarantee. On the other hand, for the cooperation-based centralized computation offloading scenario, the authors devise a holistic dynamic scheduling framework for collaborative computation offloading, by taking into account a variety of system factors including resource heterogeneity and energy efficiency. Extensive performance evaluations demonstrate that the proposed competitive and cooperative computation offloading schemes can achieve superior performance gains over the existing approaches.

In Chapter 12, the authors design and implement an elastic fog storage solution that is fully client-centric, allowing to handle variable availability and possible untrustworthiness at different remote storage locations. The availability, security, and storage efficiency are ensured by employing data deduplication and erasure coding to guarantee a user's ability to access his or her files. By using the FUSE library, a prototype with proper POSIX interfaces is developed and implemented to study the feasibility and practicality issues, such as reusing file statistic information in order to avoid the metadata management overhead from a database system. The proposed method is evaluated by Amazon S3 as a cloud server and five edge/thing resources, and our solution outperforms cloud-only solutions and is robust to edge node failures, seamlessly integrating multiple types of resources to store data. Other fog-based applications can take advantage of this service as a data storage platform.

In Chapter 13, the authors propose a system design of Virtual Local-Hub (VLH) to effectively communicate with ubiquitous wearable devices, thus extending connection ranges and reducing response time. The proposed system deploys wearable services at edge devices and modifies the system behavior of wearable devices. Consequently, wearable devices can be served at the edge of the network without data traveling via the Internet. Most importantly, the system modifications on wearable devices are transparent to both users and application developers, so that the existing applications can fit into the system naturally without any modifications. Due to the limited computing capacity of edge devices, the execution environment needs to be light-weight. Thus, the system enables remote sharing of common and native function modules on edge devices. By using off-the-shelf hardware, a testbed is developed to conduct extensive experiments. The results show that the execution time of wearable services can be reduced by up to 60% with a low system overhead.

In Chapter 14, the authors present an overview of the primary security and privacy issues in fog computing and survey the state-of-the-art solutions

that deal with the corresponding challenges. Then, they discuss major attacks in fog-based IoT applications and provide a side-by-side comparison of the state-of-the-art methods toward secure and privacy-preserving fog-based IoT applications. This chapter summarizes all up-to-date research contributions and outlines future research directions that researcher can follow in order to address different security and privacy preservation challenges in fog computing.

We hope you enjoy reading this book on both technical and economic issues of fog computing. More importantly, we will be very happy if some chapters could inspire you to generate new ideas, solutions, and contributions to this exciting research area.