

Information and Communication Technology Series • T. Russell Hsing and Vincent Lou, Series Editors



FOG AND FOGONOMICS

CHALLENGES AND PRACTICES OF FOG
COMPUTING, COMMUNICATION, NETWORKING,
STRATEGY, AND ECONOMICS

EDITED BY
YANG YANG | JIANWEI HUANG
TAO ZHANG | JOE WEINMAN



WILEY

Fog and Fogonomics

Fog and Fogonomics

Challenges and Practices of Fog Computing,
Communication, Networking, Strategy, and Economics

Edited by

Yang Yang

Shanghai Institute of Fog Computing Technology (SHIFT)
ShanghaiTech University
Shanghai, China

Jianwei Huang

The Chinese University of Hong Kong
Shenzhen, China

Tao Zhang

National Institute of Standards and Technology (NIST)
Gaithersburg, MD, USA

Joe Weinman

XFORMA LLC
Flanders, NJ, USA

WILEY

This edition first published 2020
© 2020 John Wiley & Sons, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Yang Yang, Jianwei Huang, Tao Zhang, and Joe Weinman to be identified as the authors of this work has been asserted in accordance with law.

Registered Office

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication data applied for

ISBN: 9781119501091

Cover Design: Wiley

Cover Images: Cloud computing © Just_Super/Getty Images,
Network and city background © Busakorn Pongparnit/Getty Images

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To our families.

– Yang, Jianwei, Tao, and Joe

Contents

List of Contributors *xvii*

Preface *xxi*

- 1 Fog Computing and Fogonomics** *1*
Yang Yang, Jianwei Huang, Tao Zhang, and Joe Weinman

- 2 Collaborative Mechanism for Hybrid Fog-Cloud Scenarios** *7*
Xavi Masip, Eva Marín, Jordi Garcia, and Sergi Sànchez
 - 2.1 The Collaborative Scenario *7*
 - 2.1.1 The F2C Model *11*
 - 2.1.1.1 The Layering Architecture *13*
 - 2.1.1.2 The Fog Node *14*
 - 2.1.1.3 F2C as a Service *16*
 - 2.1.2 The F2C Control Architecture *19*
 - 2.1.2.1 Hierarchical Architecture *20*
 - 2.1.2.2 Main Functional Blocks *24*
 - 2.1.2.3 Managing Control Data *25*
 - 2.1.2.4 Sharing Resources *26*
 - 2.2 Benefits and Applicability *28*
 - 2.3 The Challenges *29*
 - 2.3.1 Research Challenges *30*
 - 2.3.1.1 What a Resource is *30*
 - 2.3.1.2 Categorization *30*
 - 2.3.1.3 Identification *31*
 - 2.3.1.4 Clustering *33*
 - 2.3.1.5 Resources Discovery *33*
 - 2.3.1.6 Resource Allocation *34*
 - 2.3.1.7 Reliability *35*
 - 2.3.1.8 QoS *36*
 - 2.3.1.9 Security *36*
 - 2.3.2 Industry Challenges *37*

2.3.2.1	What an F2C Provider Should Be?	38
2.3.2.2	Shall Cloud/Fog Providers Communicate with Each Other	38
2.3.2.3	How Multifog/Cloud Access Is Managed	39
2.3.3	Business Challenges	40
2.4	Ongoing Efforts	41
2.4.1	ECC	41
2.4.2	mF2C	42
2.4.3	MEC	42
2.4.4	OEC	44
2.4.5	OFC	44
2.5	Handling Data in Coordinated Scenarios	45
2.5.1	The New Data	46
2.5.2	The Life Cycle of Data	48
2.5.3	F2C Data Management	49
2.5.3.1	Data Collection	49
2.5.3.2	Data Storage	51
2.5.3.3	Data Processing	52
2.6	The Coming Future	52
	Acknowledgments	54
	References	54
3	Computation Offloading Game for Fog-Cloud Scenario	61
	<i>Hamed Shah-Mansouri and Vincent W.S. Wong</i>	
3.1	Internet of Things	61
3.2	Fog Computing	63
3.2.1	Overview of Fog Computing	63
3.2.2	Computation Offloading	64
3.2.2.1	Evaluation Criteria	65
3.2.2.2	Literature Review	66
3.3	A Computation Task Offloading Game for Hybrid Fog-Cloud Computing	67
3.3.1	System Model	67
3.3.1.1	Hybrid Fog-Cloud Computing	68
3.3.1.2	Computation Task Models	68
3.3.1.3	Quality of Experience	71
3.3.2	Computation Offloading Game	71
3.3.2.1	Game Formulation	71
3.3.2.2	Algorithm Development	74
3.3.2.3	Price of Anarchy	74
3.3.2.4	Performance Evaluation	75
3.4	Conclusion	80
	References	80

4	Pricing Tradeoffs for Data Analytics in Fog-Cloud Scenarios	83
	<i>Yichen Ruan, Liang Zheng, Maria Gorlatova, Mung Chiang, and Carlee Joe-Wong</i>	
4.1	Introduction: Economics and Fog Computing	83
4.1.1	Fog Application Pricing	85
4.1.2	Incentivizing Fog Resources	86
4.1.3	A Fogonomics Research Agenda	86
4.2	Fog Pricing Today	87
4.2.1	Pricing Network Resources	87
4.2.2	Pricing Computing Resources	89
4.2.3	Pricing and Architecture Trade-offs	89
4.3	Typical Fog Architectures	90
4.3.1	Fog Applications	90
4.3.2	The Cloud-to-Things Continuum	90
4.4	A Case Study: Distributed Data Processing	92
4.4.1	A Temperature Sensor Testbed	92
4.4.2	Latency, Cost, and Risk	95
4.4.3	System Trade-off: Fog or Cloud	98
4.5	Future Research Directions	101
4.6	Conclusion	102
	Acknowledgments	102
	References	103
5	Quantitative and Qualitative Economic Benefits of Fog	107
	<i>Joe Weinman</i>	
5.1	Characteristics of Fog Computing Solutions	108
5.2	Strategic Value	109
5.2.1	Information Excellence	110
5.2.2	Solution Leadership	110
5.2.3	Collective Intimacy	110
5.2.4	Accelerated Innovation	111
5.3	Bandwidth, Latency, and Response Time	111
5.3.1	Network Latency	113
5.3.2	Server Latency	114
5.3.3	Balancing Consolidation and Dispersion to Minimize Total Latency	114
5.3.4	Data Traffic Volume	115
5.3.5	Nodes and Interconnections	116
5.4	Capacity, Utilization, Cost, and Resource Allocation	117
5.4.1	Capacity Requirements	117
5.4.2	Capacity Utilization	118
5.4.3	Unit Cost of Delivered Resources	119

5.4.4	Resource Allocation, Sharing, and Scheduling	120
5.5	Information Value and Service Quality	120
5.5.1	Precision and Accuracy	120
5.5.2	Survivability, Availability, and Reliability	122
5.6	Sovereignty, Privacy, Security, Interoperability, and Management	123
5.6.1	Data Sovereignty	123
5.6.2	Privacy and Security	123
5.6.3	Heterogeneity and Interoperability	124
5.6.4	Monitoring, Orchestration, and Management	124
5.7	Trade-Offs	125
5.8	Conclusion	126
	References	126
6	Incentive Schemes for User-Provided Fog Infrastructure	129
	<i>George Iosifidis, Lin Gao, Jianwei Huang, and Leandros Tassiulas</i>	
6.1	Introduction	129
6.2	Technology and Economic Issues in UPIs	132
6.2.1	Overview of UPI models for Network Connectivity	132
6.2.2	Technical Challenges of Resource Allocation	134
6.2.3	Incentive Issues	135
6.3	Incentive Mechanisms for Autonomous Mobile UPIs	137
6.4	Incentive Mechanisms for Provider-assisted Mobile UPIs	140
6.5	Incentive Mechanisms for Large-Scale Systems	143
6.6	Open Challenges in Mobile UPI Incentive Mechanisms	145
6.6.1	Autonomous Mobile UPIs	145
6.6.1.1	Consensus of the Service Provider	145
6.6.1.2	Dynamic Setting	146
6.6.2	Provider-assisted Mobile UPIs	146
6.6.2.1	Modeling the Users	146
6.6.2.2	Incomplete Market Information	147
6.7	Conclusions	147
	References	148
7	Fog-Based Service Enablement Architecture	151
	<i>Nanxi Chen, Siobhán Clarke, and Shu Chen</i>	
7.1	Introduction	151
7.1.1	Objectives and Challenges	152
7.2	Ongoing Effort on FogSEA	153
7.2.1	FogSEA Service Description	156

7.2.2	Semantic Data Dependency Overlay Network	158
7.2.2.1	Creation and Maintenance	159
7.2.2.2	Semantic-Based Service Matchmarking	161
7.3	Early Results	164
7.3.1	Service Composition	165
7.3.1.1	SeDDON Creation in FogSEA	167
7.3.2	Related Work	168
7.3.2.1	Semantic-Based Service Overlays	169
7.3.2.2	Goal-Driven Planning	170
7.3.2.3	Service Discovery	171
7.3.3	Open Issue and Future Work	172
	References	174
8	Software-Defined Fog Orchestration for IoT Services	179
	<i>Renyu Yang, Zhenyu Wen, David McKee, Tao Lin, Jie Xu, and Peter Garraghan</i>	
8.1	Introduction	179
8.2	Scenario and Application	182
8.2.1	Concept Definition	182
8.2.2	Fog-enabled IoT Application	184
8.2.3	Characteristics and Open Challenges	185
8.2.4	Orchestration Requirements	187
8.3	Architecture: A Software-Defined Perspective	188
8.3.1	Solution Overview	188
8.3.2	Software-Defined Architecture	189
8.4	Orchestration	191
8.4.1	Resource Filtering and Assignment	192
8.4.2	Component Selection and Placement	194
8.4.3	Dynamic Orchestration with Runtime QoS	195
8.4.4	Systematic Data-Driven Optimization	196
8.4.5	Machine-Learning for Orchestration	197
8.5	Fog Simulation	198
8.5.1	Overview	198
8.5.2	Simulation for IoT Application in Fog	199
8.5.3	Simulation for Fog Orchestration	201
8.6	Early Experience	202
8.6.1	Simulation-Based Orchestration	202
8.6.2	Orchestration in Container-Based Systems	206
8.7	Discussion	207
8.8	Conclusion	208
	Acknowledgment	208
	References	208

9	A Decentralized Adaptation System for QoS Optimization	213
	<i>Nanxi Chen, Fan Li, Gary White, Siobhán Clarke, and Yang Yang</i>	
9.1	Introduction	213
9.2	State of the Art	217
9.2.1	QoS-aware Service Composition	217
9.2.2	SLA (Re-)negotiation	219
9.2.3	Service Monitoring	221
9.3	Fog Service Delivery Model and AdaptFog	224
9.3.1	AdaptFog Architecture	224
9.3.2	Service Performance Validation	227
9.3.3	Runtime QoS Monitoring	232
9.3.4	Fog-to-Fog Service Level Renegotiation	235
9.4	Conclusion and Open Issues	240
	References	240
10	Efficient Task Scheduling for Performance Optimization	249
	<i>Yang Yang, Shuang Zhao, Kunlun Wang, and Zening Liu</i>	
10.1	Introduction	249
10.2	Individual Delay-minimization Task Scheduling	251
10.2.1	System Model	251
10.2.2	Problem Formulation	251
10.2.3	POMT Algorithm	253
10.3	Energy-efficient Task Scheduling	255
10.3.1	Fog Computing Network	255
10.3.2	Medium Access Protocol	257
10.3.3	Energy Efficiency	257
10.3.4	Problem Properties	258
10.3.5	Optimal Task Scheduling Strategy	259
10.4	Delay Energy Balanced Task Scheduling	260
10.4.1	Overview of Homogeneous Fog Network Model	260
10.4.2	Problem Formulation and Analytical Framework	261
10.4.3	Delay Energy Balanced Task Offloading	262
10.4.4	Performance Analysis	262
10.5	Open Challenges in Task Scheduling	265
10.5.1	Heterogeneity of Mobile Nodes	265
10.5.2	Mobility of Mobile Nodes	265
10.5.3	Joint Task and Traffic Scheduling	265
10.6	Conclusion	266
	References	266

11	Noncooperative and Cooperative Computation Offloading	269
	<i>Xu Chen and Zhi Zhou</i>	
11.1	Introduction	269
11.2	Related Works	271
11.3	Noncooperative Computation Offloading	272
11.3.1	System Model	272
11.3.1.1	Communication Model	272
11.3.1.2	Computation Model	273
11.3.2	Decentralized Computation Offloading Game	275
11.3.2.1	Game Formulation	275
11.3.2.2	Game Property	276
11.3.3	Decentralized Computation Offloading Mechanism	280
11.3.3.1	Mechanism Design	280
11.3.3.2	Performance Analysis	282
11.4	Cooperative Computation Offloading	283
11.4.1	HyFog Framework Model	283
11.4.1.1	Resource Model	283
11.4.1.2	Task Execution Model	284
11.4.2	Inadequacy of Bipartite Matching–Based Task Offloading	285
11.4.3	Three-Layer Graph Matching Based Task Offloading	287
11.5	Discussions	289
11.5.1	Incentive Mechanisms for Collaboration	290
11.5.2	Coping with System Dynamics	290
11.5.3	Hybrid Centralized–Decentralized Implementation	291
11.6	Conclusion	291
	References	292
12	A Highly Available Storage System for Elastic Fog	295
	<i>Jaeyoon Chung, Carlee Joe-Wong, and Sangtae Ha</i>	
12.1	Introduction	295
12.1.1	Fog Versus Cloud Services	296
12.1.2	A Fog Storage Service	297
12.2	Design	299
12.2.1	Design Considerations	299
12.2.2	Architecture	300
12.2.3	File Operations	301
12.3	Fault Tolerant Data Access and Share Placement	303
12.3.1	Data Encoding and Placement Scheme	303
12.3.2	Robust and Exact Share Requests	304
12.3.3	Clustering Storage Nodes	305

12.3.4	Storage Selection	306
12.3.4.1	File Download Times	307
12.3.4.2	Optimizing Share Locations	307
12.4	Implementation	309
12.4.1	Metadata	310
12.4.2	Access Counting	311
12.4.3	NAT Traversal	312
12.5	Evaluation	312
12.6	Discussion and Open Questions	318
12.7	Related Work	319
12.8	Conclusion	320
	Acknowledgments	320
	References	320
13	Development of Wearable Services with Edge Devices	325
	<i>Yuan-Yao Shih, Ai-Chun Pang, and Yuan-Yao Lou</i>	
13.1	Introduction	325
13.2	Related Works	328
13.2.1	Without Developer's Effort	329
13.2.2	Require Developer's Effort	330
13.3	Problem Description	331
13.4	System Architecture	332
13.4.1	End Device	332
13.4.2	Fog Node	333
13.4.3	Controller	333
13.5	Methodology	333
13.5.1	End Device	334
13.5.1.1	Localization	334
13.5.1.2	Speech Recognition	335
13.5.1.3	Retrieving Google Calendar Information	336
13.5.2	Fog Node	337
13.5.3	Controller	338
13.6	Performance Evaluation	339
13.6.1	Experiment Setup	339
13.6.2	Different Computation Loads	340
13.6.3	Different Types of Applications	342
13.6.4	Remote Wearable Services Provision	344
13.6.5	Estimation of Power Consumption	346
13.7	Discussion	348
13.8	Conclusion	349
	References	350

14	Security and Privacy Issues and Solutions for Fog	353
	<i>Mithun Mukherjee, Mohamed Amine Ferrag, Leandros Maglaras, Abdelouahid Derhab, and Mohammad Aazam</i>	
14.1	Introduction	353
14.1.1	Major Limitations in Traditional Cloud Computing	353
14.1.2	Fog Computing: An Edge Computing Paradigm	354
14.1.3	A Three-Tier Fog Computing Architecture	357
14.2	Security and Privacy Challenges Posed by Fog Computing	360
14.3	Existing Research on Security and Privacy Issues in Fog Computing	361
14.3.1	Privacy-preserving	361
14.3.2	Authentication	363
14.3.3	Access Control	363
14.3.4	Malicious attacks	364
14.4	Open Questions and Research Challenges	366
14.4.1	Trust	367
14.4.2	Privacy preservation	367
14.4.3	Authentication	367
14.4.4	Malicious Attacks and Intrusion Detection	368
14.4.5	Cross-border Issues and Fog Forensic	369
14.5	Summary	369
	Exercises	370
	References	370
	Index	375

List of Contributors

Mohammad Aazam

Carnegie Mellon University (CMU)
USA

Nanxi Chen

Chinese Academy of Sciences
Bio-vision Systems Laboratory
SIMIT
865 Changning Road
200050
Shanghai
China

Shu Chen

IBM Ireland
Watson Client Solution
Dublin
Ireland

Xu Chen

School of Data and Computer Science
Sun Yat-sen University
Guangzhou
China

Mung Chiang

Department of Electrical and
Computer Engineering
Purdue University
West Lafayette, IN
USA

Jaeyoon Chung

Myota Inc.
Malvern, PA
USA

Carnegie Mellon University
University of Colorado Boulder
Boulder, CO
USA

Siobhán Clarke

The University of Dublin
Distributed Systems Group
SCSS
Trinity College Dublin
College Green
Dublin 2
Dublin
Ireland

Abdelouahid Derhab

Center of Excellence in Information
Assurance (CoEIA)
King Saud University
Saudi Arabia

Mohamed Amine Ferrag

LabSTIC Laboratory
Department of Computer Science
Guelma University
Guelma
Algeria

Lin Gao

Department of Electronic and
Information Engineering
Harbin Institute of Technology
Shenzhen
China

Jordi Garcia

Advanced Network Architectures Lab
(CRAAX)
Universitat Politècnica de Catalunya
(UPC)
Vilanova i la Geltrú
Barcelona
Spain

Peter Garraghan

School of Computing and
Communications
Lancaster University
Lancaster
UK

Maria Gorlatova

Department of Electrical Engineering
Princeton University
Princeton, NJ
USA

Sangtae Ha

Department of Computer Science
University of Colorado Boulder
Boulder, CO
USA

Jianwei Huang

School of Science and Engineering
The Chinese University of Hong Kong
Shenzhen
China

Carlee Joe-Wong

Department of Electrical and
Computer Engineering
Carnegie Mellon University (CMU)
Pittsburgh, PA
USA

Fan Li

The University of Dublin
Distributed Systems Group
SCSS
Trinity College Dublin
College Green
Dublin 2, Dublin
Ireland

Tao Lin

School of Computer and
Communication Sciences
École Polytechnique Fédérale de
Lausanne
Lausanne
Switzerland

Zening Liu

School of Information Science and
Technology
ShanghaiTech University
Shanghai
China

George Iosifidis

School of Computer Science and
Statistics
Trinity College Dublin
University of Dublin
Ireland

Yuan-Yao Lou

Graduate Institute of Networking and
Multimedia and Department of
Computer Science and Information
Engineering
National Taiwan University
Taipei City
Taiwan

Leandros Maglaras

School of Computer Science and Informatics
Cyber Technology Institute
De Montfort University
Leicester
UK

Eva Marín

Advanced Network Architectures Lab (CRAAX)
Universitat Politècnica de Catalunya (UPC)
Vilanova i la Geltrú
Barcelona
Spain

Xavi Masip

Advanced Network Architectures Lab (CRAAX)
Universitat Politècnica de Catalunya (UPC)
Vilanova i la Geltrú
Barcelona
Spain

David McKee

School of Computing
University of Leeds
Leeds
UK

Mithun Mukherjee

Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis
Guangdong University of Petrochemical Technology
Maoming
China

Ai-Chun Pang

Graduate Institute of Networking and Multimedia and Department of Computer Science and Information Engineering
National Taiwan University
Taipei City
Taiwan

Yichen Ruan

Department of Electrical and Computer Engineering
Carnegie Mellon University (CMU)
Moffett Field, CA
USA

Sergi Sánchez

Advanced Network Architectures Lab (CRAAX)
Universitat Politècnica de Catalunya (UPC)
Vilanova i la Geltrú
Barcelona
Spain

Hamed Shah-Mansouri

Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver
Canada

Yuan-Yao Shih

Department of Communications Engineering
National Chung Cheng University
Taipei City
Taiwan

Leandros Tassioulas

Department of Electrical Engineering, and Institute for Network Science
Yale University
New Haven, CT
USA

Kunlun Wang

School of Information Science and
Technology
ShanghaiTech University
Shanghai
China

Joe Weinman

XFORMA LLC
Flanders, NJ
USA

Zhenyu Wen

School of Computing
Newcastle University upon Tyne
Newcastle
UK

Gary White

Distributed Systems Group, SCSS,
Trinity College Dublin
The University of Dublin
College Green
Dublin
Dublin 2
Ireland

Vincent W.S. Wong

Department of Electrical and
Computer Engineering
The University of British Columbia
Vancouver
Canada

Jie Xu

School of Computing
University of Leeds
UK

Beijing Advanced Innovation Center
for Big Data and Brain Computing
(BDBC)
Beihang University
Beijing
China

Renyu Yang

School of Computing
University of Leeds
UK

Beijing Advanced Innovation Center
for Big Data and Brain Computing
(BDBC)
Beihang University
Beijing
China

Yang Yang

Shanghai Institute of Fog Computing
Technology (SHIFT)
ShanghaiTech University
Shanghai
China

Tao Zhang

National Institute of Standards and
Technology (NIST)
Gaithersburg, MD
USA

Shuang Zhao

Shanghai Institute of Microsystem
and Information Technology (SIMIT)
Chinese Academy of Sciences
China

Liang Zheng

Department of Electrical Engineering
Princeton University
Princeton, NJ
USA

Zhi Zhou

School of Data and Computer Science
Sun Yat-sen University
Guangzhou
China

Preface

In the eternal dance driven by the evolution of technology and its applications, computing infrastructure has evolved through numerous waves, from the mainframe, to the minicomputer, to the personal computer, client-server, the smartphone, the cloud, and the edge. Whereas the cloud typically is viewed as pooled, centralized resources and the edge comprises the distributed resources that connect to endpoint devices and things, the fog, which is the latest wave, spans the cloud to device continuum.

To understand the fog, it helps to first understand the cloud. Cloud computing has a variety of definitions, ranging from those of standards bodies, to axiomatic and theoretical frameworks, to various vendor and analyst marketing and positioning statements. It typically is viewed as processing, storage, network, platform, software, and services resources that are available to multiple customers and various workload types. These resources are available “for rent” under a variety of pricing models, such as by the hour, by the minute, by the transaction, by the user, and so forth. Further variations include freemium models, discounts for advance reservation and purchase, for sustained flat use, and dynamic pricing. While some analysts define the cloud as having these resources accessed over the (public) Internet, there is no reason that other networking technologies cannot be used as well, ranging from cellular wireless radio access networks to interconnection facilities to dense wave division multiplexing and a variety of other public and private networks.

In any event, the reality of the cloud is that the major cloud providers have each built dozens of large hyper-scale facilities packed with thousands, or even hundreds of thousands of servers, whose capacity and services are accessible on demand and with pay-per-use charging by a wide variety of customers. This “short-term rental” consumption and business model exists in many other industries beyond cloud computing, e.g. overnight stays in hotels for a per-night fee; cars rentals with a daily-rate; airline, train, and bus ticket for each usage; dining at restaurants and cafés. It even exists in places that we do not normally consider: a bank loan is a means of renting capital by the day or month, where the pay-per-use fee is called the interest rate.

Cloud computing use is still growing at astronomical rates, due to the many advantages that it offers. Clouds gain their strength in large part through their consolidation into large masses of resources. This enables cost-effective dynamic allocation of resources to customers on demand and with a pay-per-use charging model. Large hotels can offer rooms for rent at attractive rates because when one convention leaves, another one begins checking in, and the remaining breakage is rented out to other people. Rental car agencies have thousands of customers; when some are returning cars, others are driving them, and still others are arriving at the counters to begin their rentals. In addition to economies of scale, these demand smoothing effects through statistical multiplexing of multiple diverse customer workloads help generate a compelling customer value proposition. They enable elasticity for many workloads, and smoothing enables higher utilization than if the varying workloads were partitioned into smaller silos. Higher utilization reduces wasted resources, lowering the unit cost of each resource.

However, this main advantage of the cloud – consolidated resources – is also its main weakness. Hyper-scale size and centralized pooled resources mean that computing and storage are located far from their actual use in factories, automobiles, smartphones, wearables, irrigation sensors, and the like. Moreover, in stark contrast to the days when computers were housed in temples and only acolytes could tend to them, computing has become pervasive, ubiquitous, low power, and cheap. Rather than the alleged prognostication from decades ago that there was a world market for “maybe five computers,” there are tens of billions of intelligent devices distributed in the physical world. It is clear that sooner or later, we will have hundreds of billions – or even a trillion – smart, connected, digital devices. It is an easy calculation to make. There are seven billion people in the world, so it only takes 15 devices per person, on average, to reach 100 billion globally. In the developed world, it is not unusual for an individual to have 4 or 5 video surveillance cameras, a few smart speakers, a laptop, a desktop, a tablet, a smartphone, some smart TVs, a fitness tracker, and a few Wi-Fi lightbulbs or outlets. To this basic observation one can add three main insights.

First, the global economy is developing even as the price of technology is plummeting, suggesting that every individual will be able to own multiple such devices.

Second, ever more devices are becoming smart and connected. For example, the smart voice-activated microwave has been introduced by Amazon; soon it will be virtually impossible to buy an object that is not smart and connected.

Third, these calculations often undercount the number of devices out there. Because in addition to consumer devices with dedicated ownership by an individual or household, there will be additional tens and hundreds of billions of devices such as manufacturing robots and traffic lights and retail point-of-sale systems and hospital wheelchair tracking systems and autonomous delivery

vehicles. A trillion connected devices can be deployed if every individual has 60 or seventy devices – not unlikely once you start adding in light bulbs and outlets and nonconsumer device counts make up the other half-trillion.

These massive resource-limited devices with various functionalities and capabilities, when they are deployed and connected, contribute to the future Internet of Things (IoT) to enable different intelligent applications and services, such as environment monitoring, autonomous driving, city management, and medicine and health care. Moreover, emerging wireless capabilities, as embodied in 5G, reduce latency from tens of milliseconds to single digits. To fully take advantage of these capabilities requires processing and storage resources in proximity to the device. There is absolutely no way that the optimal system architecture in such a situation would be to interconnect all these devices across a dumb wide area network to a remote consolidated facility, i.e. the cloud. Instead, multiple layers of processing and storage are needed to bring order, collaboration, intelligence, and solutions out of what otherwise would be a random chaos of devices.

This is the fog.

A number of synonyms and related concepts with nuanced differences exist, such as edge computing, mobile edge computing, osmotic computing, pervasive computing, ubiquitous computing, mini-cloud, cloudlets, and so on and so forth.

And, various bodies have proposed various definitions. The OpenFog Consortium defines fog computing as “a system-level horizontal architecture that distributes resources and services of computing, storage, control and networking anywhere along the continuum from Cloud to Things.” The US National Institute of Standards and Technology similarly defines it as a “horizontal, physical or virtual resource paradigm that resides between smart end-devices and traditional cloud or data centers. This paradigm supports vertically-isolated, latency-sensitive applications by providing ubiquitous, scalable, layered, federated, and distributed computing, storage, and network connectivity.”

In other words, the fog is simply multiple interconnected layers of computing along the continuum from cloud to endpoints such as user devices and things. This may include racks or microcells in server closets, residential gateways, factory control systems, and the like.

Whereas clouds are hyper-scale, fog nodes may be intermediate size, or even miniature. Whereas clouds rely on multiple customers and workloads, fog nodes may be dedicated to one customer, and even one use. Whereas clouds have state of the art power distribution architectures including multiple grids with diverse access, generators and/or fuel cells, or hydrothermal energy, fog nodes may be powered by batteries or even energy scavenging. Whereas clouds use advanced thermal management strategies including hot-cold aisles, water cooling, airflow simulation and optimization, fog nodes may be cooled by the environmental ambient. Whereas clouds are built in walled data centers, fog

nodes may be in homes, factories, agricultural fields, or vineyards. Whereas clouds have fixed street addresses, fog nodes may be mobile. Whereas clouds are engineered for uptime and five nines connectivity, fog nodes may be only intermittently powered, available, within a coverage area, or functional. Whereas clouds are offered by a specific vendor, fog solutions are inherently heterogeneous ecosystems.

Perhaps this is why fog is likely to have an impact across many domains – the economy, technology, standards, market disruption, society and culture, and innovation – on par with cloud computing’s impact.

Of course, similar to how cloud’s advantages are their weaknesses, fog’s advantages can also be its weaknesses. The strength of mobility can lead to intermittent connectivity, which increases the challenges of reliable message passing. Low latency to endpoints means high latency for massive databases, which can be in the cloud. Small footprints can mean an inability to process massive compute jobs. Heterogeneity can create robustness by largely eliminating systemic failures due to design flaws; it can also create a nightmare for monitoring, management, and root cause analysis. This book will document, explore, and quantify many of these challenges and identify and propose solutions and promising directions for future research.

We, the editors, sincerely hope that this collection of insights from the world’s leading fog experts and researchers helps you in your journey to the fog.

Shanghai, China, 27 May 2019

Yang Yang
Jianwei Huang
Tao Zhang
Joe Weinman

1

Fog Computing and Fogonomics

Yang Yang¹, Jianwei Huang², Tao Zhang³, and Joe Weinman⁴

¹Shanghai Institute of Fog Computing Technology (SHIFT), ShanghaiTech University, Shanghai, China

²School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China

³National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA

⁴XFORMA 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