Towards Power Consumption-Delay Tradeoff by Workload Allocation in Cloud-Fog Computing

Ruilong Deng*, Rongxing Lu*, Chengzhe Lai† and Tom H. Luan‡

* School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798
† National Engineering Lab. for Wireless Security, Xi'an University of Posts and Telecommunications, Xi'an, China 710121
† State Key Laboratory of Information Security, Chinese Academy of Sciences, Beijing, China 100093
‡ School of Information Technology, Deakin University, Burwood, Victoria 3125 Australia
Email: {rldeng, rxlu}@ntu.edu.sg, lcz.xupt@gmail.com, tom.luan@deakin.edu.au

***Abstract*—Fog computing, characterized by extending cloud computing to the edge of the network, has recently received considerable attention. The fog is not a substitute but a powerful complement to the cloud. It is worthy of studying the interplay and cooperation between the edge (fog) and the core (cloud). To address this issue, we study the tradeoff between power consumption and delay in a cloud-fog computing system. Specifically, we first mathematically formulate the workload allocation problem. After that, we develop an approximate solution to decompose the primal problem into three subproblems of corresponding subsystems, which can be independently solved. Finally, based on extensive simulations and numerical results, we show that by sacrificing modest computation resources to save communication bandwidth and reduce transmission latency, fog computing can significantly improve the performance of cloud computing.**

I. Introduction

The Internet has shifted to the cloud based structure. Cloud computing provides the outsourced computation and storage capabilities to end users such that they are not bothered to manage their own infrastructures but enjoy the flexible and convenient services [1]. However, with the surging traffic generated in recent years, the transmission of the extraordinarily huge-volume data to the cloud has not only posed a heavy burden on communication bandwidth, but also resulted in unbearable transmission latency and degraded service to end users [2]. In addition to real-time interaction and low latency, with mobile users and traffic becoming dominant nowadays, the support of mobility and geo-distribution is also critical [3]. Therefore it is desirable to revisit the design of computing paradigms to better satisfy the service requirements for the mobile big data era.

To address the above challenges, Cisco has delivered the vision of fog computing in January 2014, which aims to process in part workload and services locally on fog devices (such as hardened routers, switches, IP video cameras, and etc.), rather than being transmitted to the cloud [4]. Conceptually, "the fog is a cloud close to the groud", just as fog computing extends cloud computing to the edge of the network. Fog computing, that is specially featured by "proximity to end users", "densely and geographically distributed", and "support for mobile computing", can provide the same services (computation, networking, and storage) as the cloud. For example, with the fog, some of the processing can be taken place at routers rather than transmitting all data to the cloud. From this view of point, fog computing is not aimed to substitute cloud computing but to complement it, which can ease bandwidth burden and reduce transmission latency. In addition, fog computing can support and facilitate applications that do not fit well with the cloud: (i) applications that require very low and predictable latency, such as online gaming and video conferencing; (ii) geographically distributed applications such as pipeline monitoring and sensor networks; (iii) fast mobile applications such as smart connected vehicles; and (iv) large-scale distributed control systems such as smart energy distribution and smart traffic lights [5].

While the fog provides localization, i.e., enabling real-time interaction and low latency at the network edge, the cloud provides centralization, which arouses applications that require the interplay and cooperation between the edge (fog) and the core (cloud), particularly for the Internet of Things and big data analytics [6]. In this paper, we consider a cloud-fog computing system. On one hand, with the huge-volume and ever-increasing service requests, the power consumption on powering up (and cooling) cloud servers is soaring. It is thus important and desirable to consider the energy management in the cloud-fog computing system [7]. For service providers, on the other hand, it is equally crucial to guarantee the quality of service (e.g., latency requirements) of end users. The reason is that the unbearable response latency leads to revenue loss since users will subscribe to other vendors with better service [8]. To this end, we systematically investigate the fundamental tradeoff between the power consumption and the end-user delay in the cloud-fog computing system.

In this paper, we firstly model the power consumption and delay functions of each part of the cloud-fog computing system, and formulate the workload allocation problem. We then develop an approximate solution to the primal problem through decomposition, and formulate three subproblems of corresponding subsystems respectively. These subproblems can be independently solved via existing optimization techniques. Finally, based on extensive simulations and numerical results, we show that fog computing can significantly improve the performance of cloud computing in terms of reducing

978-1-4673-6432-4/15/$31.00 ©2015 IEEE

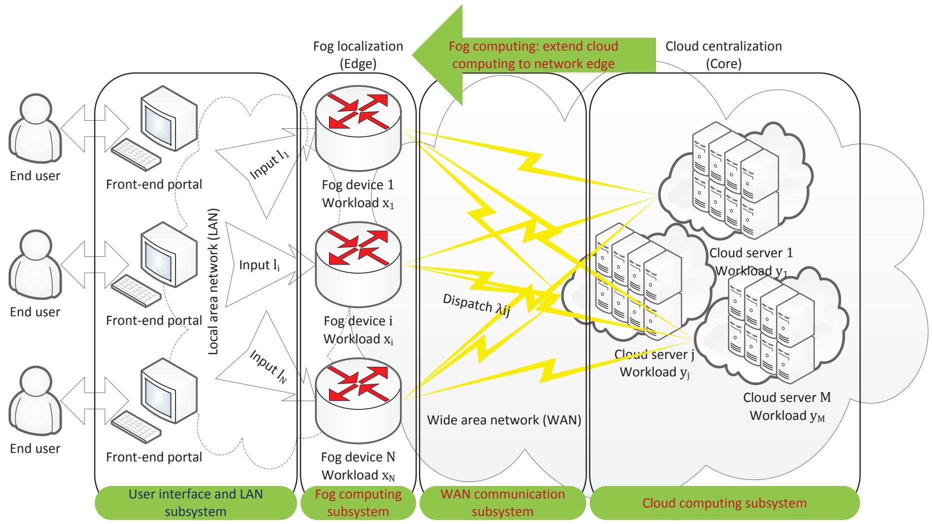


Fig. 1. An overall architecture of a cloud-fog computing system with four subsystems and their interconnections/interactions.

communication latency. To the best of our knowledge, this is the first effort towards providing a systematic framework of computation and communication co-design in the cloud-fog computing system. We hope that this pioneering work can throw light on how the fog can extend and complement the cloud. Specifically, the original contributions of this paper are summarized in the following three folds:

1) We are the first to mathematically formulate and investigate the power consumption-delay tradeoff problem by workload allocation in the cloud-fog computing system.
2) We have developed an approximate solution to decompose the primal problem into three subproblems of corresponding subsystems, which can be independently solved.
3) We have conducted extensive simulations to validate the performance of our scheme and demonstrated that the fog can significantly complement the cloud with much reduced communication latency.

The rest of this paper is organized as follows. We describe the model of the cloud-fog computing system and formulate the power consumption-delay tradeoff problem in Section II. In Section III, we approximately decompose the primal problem into three subproblems of corresponding subsystems. Simulations are conducted in Section IV with numerical results, and we draw concluding remarks in Section V.

II. System Model and Problem Formulation

We illustrate the overall architecture of the cloud-fog computing system in Fig. 1, which has been divided into four subsystems. The front-end portals act as user interfaces that receive service requests from end users. These requests are separately input to a set $\mathcal{N}$ of fog devices through a local area network (LAN). Since fog devices are generally located in the vicinity of end users, thus the LAN communication delay could be omitted (compared to WAN). Fog computing can process some latency-sensitive requests and forward other delay-tolerant requests to cloud computing. There is a set $\mathcal{M}$ of cloud servers, each of which hosts a number of homogeneous computing machines. The delay-tolerant requests are dispatched from each fog device to each cloud server through a wide area network (WAN). Since WAN covers a large geographical area from the edge throughout to the core, the communication delay and constrained bandwidth should be taken into account. In the following, we mainly consider the power consumption and computation/communication delay of the latter three subsystems (i.e., fog computing, WAN communication, and cloud computing).

A. System Model

1) Power Consumption of Fog Device: For the fog device i, the computation power consumption can be modelled by a function of the computation amount x_i, which is a monotonic increasing and strictly convex function. The piece-wise linear function and the quadratic function are two alternatives [9]. In fact, the fog computing devices can accommodate any form of power consumption functions as long as they satisfy the following two properties: (i) the computation power consumption always increases when the computation amount increases; (ii) the marginal power consumption for each fog device is increasing. For simplicity but without loss of generation, we can express the power consumption P_i^{fog} of the fog device i by the following function of the computation amount x_i:

$$P_i^{\text{fog}} \triangleq a_i x_i^2 + b_i x_i + c_i,$$

where $a_i > 0$ and $b_i, c_i \geq 0$ are pre-determined parameters.

2) Computation Delay of Fog Device: For simplicity but without loss of generality, for the fog device i with the traffic arrival rate x_i and the service rate v_i, the average computation delay D_i^{fog} is

$$D_i^{\text{fog}} \triangleq \frac{x_i}{v_i}.$$

3) Power Consumption of Cloud Server: Each cloud server hosts a number of homogeneous computing machines. The configurations (e.g., CPU frequency) are assumed to be equal for all machines at each server. Thus, each machine at the same server has the same power consumption profile. We approximate the power consumption value of each machine at the cloud server j by a function of the machine CPU frequency f_j: $A_j f_j^p + B_j$, where A_j and B_j are positive constants, and p varies from 2.5 to 3 [8].

When the allocated workload increases, more cloud servers are powered on; when it decreases, excess ones are turned off for energy saving. Let a binary variable σ_j denote the on/off state of the cloud server j, where 1 means that the server is on and 0 means off. Besides, let an integer variable n_j denote the number of turned on machines at the cloud server j. Thus, the power consumption P_j^{cloud} of the cloud server j can be obtained by multiplying the on/off state, the on-state machine number, with each machine power consumption value [7]:

$$P_j^{\text{cloud}} \triangleq \sigma_j n_j \left(A_j f_j^p + B_j \right).$$

4) Computation Delay of Cloud Server: The M/M/n queueing model is employed to describe each machine at the cloud server j. In this model, the average computation delay (waiting time plus service time) is $[1/(n\mu - \lambda)]$, where n is the number of working machines, λ and μ are the traffic arrival and service rate respectively. At the cloud server j, let each machine has the same service rate μ_j. We can generally convert μ_j to f_j by $\mu_j = f_j/K$, where K is in terms of (# of cycles)/request.

From the above, for the cloud server j with n_j turned on machines, when each machine has the traffic arrival and service rate y_j and f_j/K respectively, then the average computation delay D_j^{cloud} is given by [8]

$$D_j^{\text{cloud}} \triangleq \frac{\sigma_j}{\frac{n_j f_j}{K} - y_j}.$$

5) Communication Delay for Dispatch: Let d_{ij} denote the average delay of the WAN transmission path from the fog device i to the cloud server j. Thus, when the traffic rate dispatched from the fog device i to the cloud server j is λ_{ij}, the corresponding communication delay D_{ij}^{comm} is

$$D_{ij}^{\text{comm}} \triangleq d_{ij} \lambda_{ij}.$$

B. Constraints

1) Workload Balance Constraint: We denote the total request input from all front-end portals as L. The traffic arrival rate from all front-end portals to the fog device i is denoted by l_i. Thus, we have

$$L \triangleq \sum_{i \in \mathcal{N}} l_i.$$

Besides, let X and Y denote the allocated workload for fog computing and cloud computing, respectively. Then, we have

$$\begin{cases} X \triangleq \sum\limits_{i \in \mathcal{N}} x_i \\ Y \triangleq \sum\limits_{j \in \mathcal{M}} y_j. \end{cases}$$

We describe the workload balance constraint on the traffic rate dispatched from each fog device to each cloud server. The end-user requests are either handled by a fog device, or forwarded to a cloud server to be processed. The corresponding relationships between the workload and traffic rate are listed as (i) workload balance constraint for each fog device:

$$l_i - x_i = \sum_{j \in \mathcal{M}} \lambda_{ij} \quad \forall i \in \mathcal{N}, \tag{1}$$

(ii) workload balance constraint for each cloud server:

$$\sum_{i \in \mathcal{N}} \lambda_{ij} = y_j \quad \forall j \in \mathcal{M}, \tag{2}$$

From (i) and (ii) we can easily obtain (iii) workload balance constraint for the holistic cloud-fog computing system:

$$L = X + Y.$$

2) Fog Device Constraint: For the fog device i, there exists a limit on the processing ability due to physical constraints. Let X_i denote the computation capacity of the fog device i. In addition, the workload x_i allocated to the fog device i should be no more than the traffic arrival rate l_i to that device. From the above, we have

$$0 \le x_i \le \min\{X_i, l_i\} \quad \forall i \in \mathcal{N}. \tag{3}$$

3) Cloud Server Constraint: For the cloud server j, there exists a limit on the computation rate of each machine due to physical constraints. Let $f_j^{\min}$ and $f_j^{\max}$ denote the lower and upper bound on the machine working frequency, respectively:

$$f_j^{\min} \le f_j \le f_j^{\max} \quad \forall j \in \mathcal{M}. \tag{4}$$

In addition, for the cloud server j, the number of working machines n_j has an upper bound N_j. Thus, for the integer variable n_j, we have

$$n_j \in \{0, 1, 2, \ldots, N_j\} \quad \forall j \in \mathcal{M}. \tag{5}$$

The binary variable σ_j denote the on/off state of the cloud server j. When σ_j equals 1, it means that the cloud server j is on; when σ_j equals 0, it means that the cloud server j is off, and meanwhile the number of on-state machines equals 0. Thus, we have

$$\sigma_j \in \{0, 1\} \quad \forall j \in \mathcal{M}. \tag{6}$$

4) WAN Communication Bandwidth Constraint: For simplicity but without loss of generality, the traffic rate λ_{ij} is assumed to be dispatched from the fog device i to the cloud server j through one transmission path. Furthermore, these transmission paths do not overlap with each other. There is a limitation C_{ij} on the bandwidth capacity of each path. Thus, the bandwidth constraint of the WAN communication is

$$0 \le \lambda_{ij} \le C_{ij} \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M}. \tag{7}$$

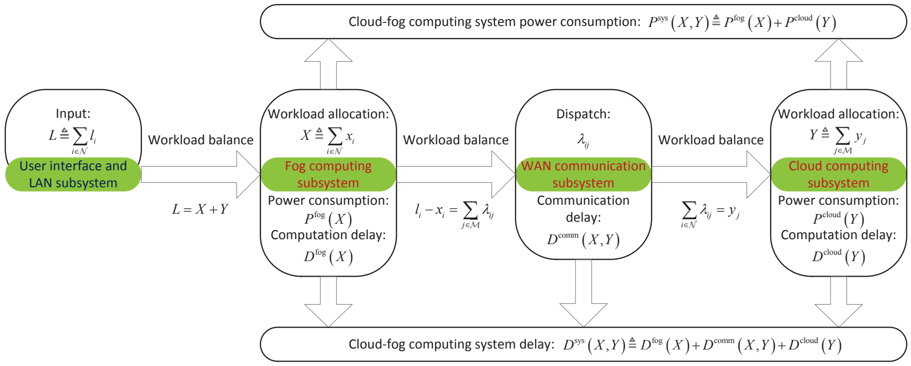


Fig. 2. An overall framework of power consumption-delay tradeoff by workload allocation in a cloud-fog computing system.

C. Problem Formulation

Towards power consumption-delay tradeoff in cloud-fog computing, it is important to minimize the aggregated power consumption of all fog devices and cloud servers. The power consumption function of the cloud-fog computing system is defined as

$$P^{\text{sys}} \triangleq \sum_{i \in \mathcal{N}} P_i^{\text{fog}} + \sum_{j \in \mathcal{M}} P_j^{\text{cloud}}.$$

On the other hand, it is equally crucial to guarantee the quality of service (e.g., latency requirements) of end users. The end-user experienced delay consists of the computation (including queueing) delay and the communication delay. Therefore, the delay function of the cloud-fog computing system is defined as

$$D^{\text{sys}} \triangleq \sum_{i \in \mathcal{N}} D_i^{\text{fog}} + \sum_{j \in \mathcal{M}} D_j^{\text{cloud}} + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} D_{ij}^{\text{comm}}.$$

We consider the problem of minimizing the power consumption of the cloud-fog computing system while guaranteeing the required delay constraint $\overline{D}$ for end users. For simplicity but without loss of generation, we take the average end-user experienced delay as a constraint to approximate the worst (maximum) delay for an individual task. That is, we have the Primal Problem (**PP**):

$$\min_{x_i, y_j, \lambda_{ij}, f_j, n_j, \sigma_j} \quad P^{\text{sys}}$$
$$\text{s.t.} \quad \begin{cases} D^{\text{sys}} \leq \overline{D} \\ (1) - (7). \end{cases}$$

The decision variables are the workload x_i assigned to the fog device i, the workload y_j assigned to the cloud server j, the traffic rate λ_{ij} dispatched from the fog device i to the cloud server j, the machine CPU frequency f_j, the machine number n_j, and the on/off state σ_j at the cloud server j. The objective of workload allocation in the cloud-fog computing system is to tradeoff between the system power consumption and the end-user experienced delay.

III. Decomposition and Solution

Note that in **PP**, the decision variables come from different subsystems and are tightly coupled with each other, which makes the relationship between the workload allocation and the system power consumption-delay tradeoff not clear. To this end, we develop an approximate solution to decompose **PP** into three subproblems of corresponding subsystems, which can be independently solved via existing optimization techniques. We illustrate the decomposition and each subproblem/subsystem interactions in Fig. 2, which in general provides the overall framework of power consumption-delay tradeoff by workload allocation in the cloud-fog computing system.

A. Power Consumption-Delay Tradeoff for Fog Computing

We firstly consider to tradeoff between the power consumption and computation delay in the fog computing subsystem. That is, we have the Subproblem One (**SP1**):

$$\min_{x_i} \quad \sum_{i \in \mathcal{N}} \left(a_i x_i^2 + b_i x_i + c_i + \eta_i \frac{x_i}{v_i} \right)$$
$$\text{s.t.} \quad \begin{cases} \sum\limits_{i \in \mathcal{N}} x_i = X \\ (3). \end{cases}$$

where the adjustable parameter η_i is the weight to tradeoff between the power consumption and computation delay at the fog device i.

Given the allocated workload X for the fog computing subsystem, **SP1** is a convex problem with linear constraints. This problem can be easily solved using convex optimization techniques [10], [11]. After we obtain the optimal workload x_i^* assigned to the fog device i, we can calculate the power consumption and computation delay in the fog computing subsystem, respectively as

$$\begin{cases} P^{\text{fog}}(X) = \sum\limits_{i \in \mathcal{N}} \left[a_i \left(x_i^* \right)^2 + b_i x_i^* + c_i \right] \\ D^{\text{fog}}(X) = \sum\limits_{i \in \mathcal{N}} \frac{x_i^*}{v_i}. \end{cases}$$

B. Power Consumption-Delay Tradeoff for Cloud Computing

At the cloud server j, for the delay-sensitive requests, their average response delay should be bounded by a certain threshold that is specified as the service level agreement, since the agreement violation would result in loss of business revenue. We assume that the average response delay should be smaller than the adjustable parameter $\overline{D}_j$, which can be regarded as a threshold that identifies the revenue/penalty region at the cloud server j. That is

$$D_j^{\text{cloud}} \leq \overline{D}_j.$$

After mathematical transformations, we have the following computation delay constraint for the cloud server j [8]:

$$\frac{n_j f_j}{K} - \frac{\sigma_j}{\overline{D}_j} \geq y_j.$$

We now consider to tradeoff between the power consumption and computation delay in the cloud computing subsystem. That is, we have the Subproblem Two (**SP2**):

$$\begin{array}{ll} \min\limits_{f_j, n_j, \sigma_j} & \sum\limits_{j \in \mathcal{M}} \sigma_j n_j \left(A_j f_j^p + B_j \right) \\ \text{s.t.} & \left\{ \begin{array}{l} \sum\limits_{j \in \mathcal{M}} \left(\frac{n_j f_j}{K} - \frac{\sigma_j}{\overline{D}_j} \right) \geq Y \\ (4) - (6). \end{array} \right. \end{array}$$

Given the allocated workload Y for the cloud computing subsystem, **SP2** is a mixed integer nonlinear programming (MINLP) problem. Since the generalized Benders decomposition (GBD) is an effective method to solve this problem with guaranteed optimality, we design the GBD algorithm [12, Ch. 13]. Due to the space limitation, we omit the detail. After we obtain the optimal solution f_j^*, n_j^* and σ_j^*, we can calculate the optimal workload assigned to the cloud server j by

$$y_j^* = \left\{ \begin{array}{ll} \frac{n_j^* f_j^*}{K} - \frac{1}{\overline{D}_j}, & \text{if } \sigma_j^* = 1 \\ 0, & \text{if } \sigma_j^* = 0, \end{array} \right.$$

together with the power consumption and computation delay in the cloud computing subsystem, respectively as

$$\left\{ \begin{array}{l} P^{\text{cloud}}(Y) = \sum\limits_{j \in \mathcal{M}} \sigma_j^* n_j^* \left[A_j \left(f_j^* \right)^p + B_j \right] \\ D^{\text{cloud}}(Y) = \sum\limits_{j \in \mathcal{M}} \frac{\sigma_j^*}{\frac{n_j^* f_j^*}{K} - y_j^*} = \sum\limits_{j \in \mathcal{M}} \sigma_j^* \overline{D}_j. \end{array} \right.$$

C. Communication Delay Minimization for Dispatch

We finally consider how to choose the traffic dispatch rate λ_{ij} to minimize the communication delay in the WAN subsystem. That is, we have the Subproblem Three (**SP3**):

$$\begin{array}{ll} \min\limits_{\lambda_{ij}} & \sum\limits_{i \in \mathcal{N}} \sum\limits_{j \in \mathcal{M}} d_{ij} \lambda_{ij} \\ \text{s.t.} & (1)(2)(7). \end{array}$$

From Section III-A and III-B, given the allocated workload X for fog computing and Y for cloud computing, we can obtain the optimal workload x_i^* assigned to the fog device i and y_j^* assigned to the cloud server j. Given x_i and y_j, **SP3** is regarded as an assignment problem. This problem can be efficiently solved using existing algorithms such as the Hungarian method [13]. After we obtain the optimal traffic rate λ_{ij}^* dispatched from the fog device i to the cloud server j, we can calculate the communication delay in the WAN subsystem as

$$D^{\text{comm}}(X, Y) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} d_{ij} \lambda_{ij}^*.$$

D. Putting It All Together

Based on the approximate decomposition and with the solution to the above three subproblems, on one hand, the power consumption function of the cloud-fog computing system is rewritten as

$$P^{\text{sys}}(X, Y) \triangleq P^{\text{fog}}(X) + P^{\text{cloud}}(Y),$$

which means that the system power consumption comes from the fog devices and cloud servers. On the other hand, the delay function of the cloud-fog computing system is rewritten as

$$D^{\text{sys}}(X, Y) \triangleq D^{\text{fog}}(X) + D^{\text{cloud}}(Y) + D^{\text{comm}}(X, Y),$$

which means that the system delay comes from the computation delay of the fog devices and cloud servers, as well as the communication delay of the WAN.

After solving the above three subproblems, we can approximately solve **PP** by considering the following approximate problem named **PP-approx**:

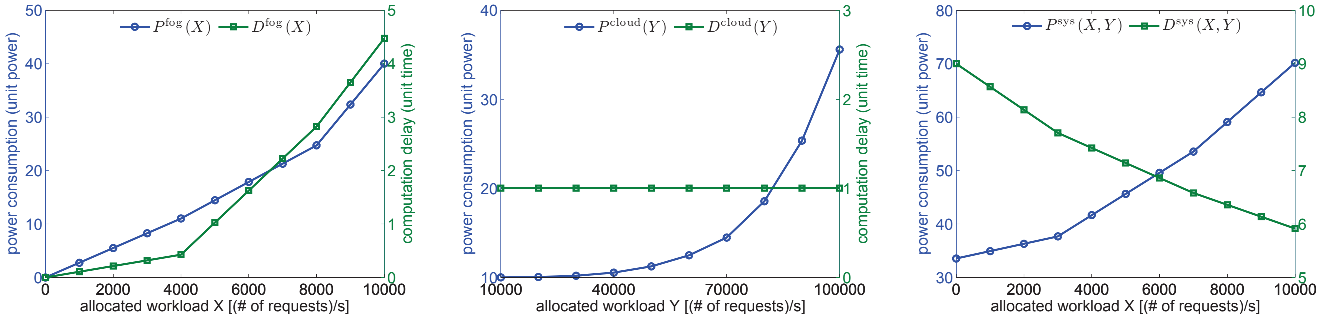
$$\begin{array}{ll} \min\limits_{X, Y} & P^{\text{sys}}(X, Y) \\ \text{s.t.} & \left\{ \begin{array}{l} D^{\text{sys}}(X, Y) \leq \overline{D} \\ X + Y = L, \end{array} \right. \end{array}$$

which can be iteratively solved. The approximation ratio is dependent on the choice of two adjustable parameters η_i and $\overline{D}_j$. If these parameters could be appropriately chosen, then the solution to **PP-approx** would be the optimal solution to **PP**. How to evaluate the approximation ratio of the proposed decomposition is left as our future work.

IV. Numerical Results

Simulation results are presented in this section to validate the power consumption-delay tradeoff by workload allocation to fog computing and cloud computing. For simplicity but without loss of generality, we consider the scenario with five fog devices and three cloud servers (Internet data centers) in the cloud-fog computing system. Some important parameters used in the simulation are summarized in TABLE I, referring to [7], [8]. All the following results are obtained by MATLAB.

Firstly, we vary the allocated workload X for fog computing from 0 to 10^4, to evaluate how they affect the power consumption $P^{\text{fog}}(X)$ and computation delay $D^{\text{fog}}(X)$ in the subsystem. Under different values of X, we solve **SP1** and obtain the optimal workload x_i^* assigned to the fog device i. Based on this we calculate $P^{\text{fog}}(X)$ and $D^{\text{fog}}(X)$, and draw their curves in Fig. 3(a). It is seen that both the



(a) Fog computing subsystem

(b) Cloud computing subsystem

(c) Cloud-fog computing system

Fig. 3. An illustration of power consumption-delay tradeoff by workload allocation in a cloud-fog computing system.

TABLE I
PARAMETER SETUP

Param	Value	Param	Value
l_i	$[30\ 15\ 15\ 20\ 20] \times 10^3$	$f_j^{\min}$	1.0
A_j	[3.206 4.485 2.370]	$f_j^{\max}$	[3.4 2.4 3.0]
B_j	[68 53 70]	N_j	$[30\ 60\ 25] \times 10^3$
p, K	3, 1	$\overline{D}_j$	unit time

power consumption and computation delay increase with the allocated workload for fog computing.

Next, we vary the allocated workload Y for cloud computing from 10^4 to 10^5, to evaluate how they affect the power consumption $P^{\text{cloud}}(Y)$ and computation delay $D^{\text{cloud}}(Y)$ in the subsystem. Under different values of Y, we solve **SP2** and obtain the optimal workload y_j^* assigned to the cloud server j. Based on this we calculate $P^{\text{cloud}}(Y)$ and $D^{\text{cloud}}(Y)$, and draw their curves in Fig. 3(b). The result shows that the computation delay stays steady and the power consumption increases with the allocated workload for cloud computing.

Finally, from the above x_i^* and y_j^*, we further solve **SP3** and obtain the communication delay $D^{\text{comm}}(X,Y)$ in the WAN subsystem. Based on these we calculate the system power consumption $P^{\text{sys}}(X,Y)$ and delay $D^{\text{sys}}(X,Y)$, and draw their curves in Fig. 3(c). The figure illustrates that, when some workload is allocated for fog computing, the system delay decreases while the power consumption increases. This is because in the cloud-fog computing system, cloud computing is more powerful and energy-efficient than fog computing; while the fog, with the advantage of proximity to end users, can sacrifice modest computation resources to save WAN bandwidth and reduce communication latency, in such a way to significantly improve the performance of the cloud.

V. Conclusion

In this paper, we have introduced the vision of fog computing, a newly emerged paradigm that extends cloud computing to the edge of the network. Concretely, we develop a systematic framework to investigate the power consumption-delay tradeoff issue in the cloud-fog computing system. We formulate the workload allocation problem and approximately decompose the primal problem into three subproblems, which can be independently solved within corresponding subsystems. Extensive simulation and numerical results are presented to verify our theoretical analysis. We hope that this pioneering work can provide guidance on studying the interaction and cooperation between the fog and cloud.

VI. Acknowledgment

The authors would like to thank the support from Nanyang Technological University under Grant NTU-SUG (M4081196) and MOE Tier 1 (M4011177). R. Deng is supported in part by EEE Cybersecurity Research Program, NTU.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, 2014.

[3] 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 *Proc. IEEE SECON*, 2011, pp. 359–367.

[4] Cisco delivers vision of fog computing to accelerate value from billions of connected devices. [Online]. Available: http://newsroom.cisco.com/release/1334100

[5] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for internet of things and analytics," in *Big Data and Internet of Things: A Roadmap for Smart Environments*. Springer, 2014, pp. 169–186.

[6] I. Stojmenovic and S. Wen, "The fog computing paradigm: Scenarios and security issues," in *Proc. FedCSIS*, 2014, pp. 1–8.

[7] L. Rao, X. Liu, L. Xie, and W. Liu, "Coordinated energy cost management of distributed internet data centers in smart grid," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 50–58, 2012.

[8] L. Rao, X. Liu, M. D. Ilic, and J. Liu, "Distributed coordination of internet data centers under multiregional electricity markets," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 269–282, 2012.

[9] R. Deng, J. Chen, X. Cao, Y. Zhang, S. Maharjan, and S. Gjessing, "Sensing-performance tradeoff in cognitive radio enabled smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 302–310, 2013.

[10] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[11] J. He, P. Cheng, L. Shi, and J. Chen, "Time synchronization in WSNs: A maximum value based consensus approach," in *Proc. IEEE CDC-ECC*, 2011, pp. 7882–7887.

[12] D. Li and X. Sun, *Nonlinear integer programming*. Springer, 2006.

[13] H. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics*, vol. 52, no. 1, pp. 7–21, 2005.