

# Paying Less for More? Combo Plans for Edge-Computing Services

Shizhe Zang  
University of Sydney

Wei Bao  
University of Sydney

Phee Lep Yeoh  
University of Sydney

Branka Vucetic  
University of Sydney

Yonghui Li  
University of Sydney

## Abstract

Due to the strong potential of edge computing, service providers are aware of its excellent market opportunity. In this paper, we examine new pricing plans for edge-computing services that jointly consider communication and computing costs in mobile devices and edge servers. In our proposed model, users can freely choose to use: 1) a pay-as-you-go payment for communication and computing costs, 2) an upfront data plan for unlimited communication volume, 3) an upfront computing plan for unlimited computing workload, or 4) a combo plan for unlimited communication volume and computing workload. Based on this, we design a smart plan-purchasing scheme, namely smart online reservation (SOR) scheme for users and prove it achieves the best possible competitive ratio among all possible online deterministic schemes. With SOR adopted by users, we develop pricing plan guidelines for service providers to achieve profitability by motivating users to purchase the combo plan. We provide trace-driven simulations to verify the advantage of SOR and identify how different prices influence user behavior.

## 1 Introduction

Edge computing is a promising solution to accommodate the explosive growth in communications and computing demand by mobile devices and the Internet of Things [8]. Equipped with ubiquitously available wireless access points and computing nodes (e.g., Virtual Machine and containers), mobile data can be efficiently and promptly processed at the network edge instead of being processed at a remote and possibly congested central data center.

Studies show that edge computing represents an over \$80B market by 2021 [7]. Edge-computing service providers (ESPs) are keenly aware of this lucrative market opportunity. In 2018, AT&T is going to launch an

edge-computing test zone in Silicon Valley [3]. One important issue to be addressed by ESPs is how to develop attractive and cost-effective pricing plans for their customers. Obviously, the conventional approach of charging by transmitted data volume is no longer valid since the cost of data processing largely depends on its computing workload rather than data volume.

Similar to mobile plans in their childhood (early 3G era), where data usage and voice calls were charged separately [5], we envision that customers will be charged for their communication volume and computing workload separately in the early stages of edge computing. However, as edge-computing services reach maturity, combo plans that jointly charge for data transmission and data processing services will emerge. Empirical analysis of customer behavior for tiered mobile data plans was conducted in [12] without considering computing plans. In [9], an incentive-compatible auction mechanism was designed for real-time resource allocation in mobile edge computing without considering future requests. A key question we will address is: *How to design the price of edge-computing combo plans to attract users while guaranteeing the profit of ESPs without knowing users behavior?*

In this paper, we investigate the transition of the edge-computing service pricing design from the early stage to the mature stage. We consider that users are allowed to pay for their communication volume and computing workload in a pay-as-you-go (PAYG) fashion, or they can prepay upfront reservation fees of  $C_1$  for a data plan with unlimited communication volume and  $C_2$  for a computing plan with unlimited computing workload for a period of time. They can also choose the combo plan, which has an upfront reservation fee of  $C_3$  offering both unlimited communication volume and computing workload for a period of time.

Based on the above plans, we propose the smart online reservation (SOR) scheme for the users without any future knowledge of task arrivals to minimize

overall expense. We **prove** that the proposed SOR scheme achieves the competitive ratio of  $1 + \frac{C_1+C_2}{C_3}$ , which is the best possible competitive ratio among all possible online deterministic schemes. According to our analysis for the SOR Scheme, we identify a key pricing design guideline for ESP: by setting  $C_3 < \min\left(\sqrt{C_1(C_1+C_2)}, \sqrt{C_2(C_1+C_2)}\right)$ , users will select the combo plan over individual data and computing plans. Furthermore, we provide trace-driven simulations to verify the advantage of SOR over other benchmark schemes in terms of cost minimization and show how the prices of different plans influence user behavior. Due to the space limitation, the proof of the competitive ratio is in our technical report [13].

## 2 Problem Formulation

We consider the scenario where different edge-computing tasks labeled by  $i = 1, 2, 3, \dots$ , arrive from the user device and are served by the ESP. Task  $i$  is characterized by a 3-tuple  $(a_i, b_i, t_i)$ , where  $a_i$  is the communication volume to serve the task,  $b_i$  is the computing workload to serve the task, and  $t_i$  is the arrival time instance of the task. Without loss of generality, we assume  $t_1 < t_2 < \dots$ . We assume that the user *does not have the knowledge of the future*. In other words, the user only knows  $(a_i, b_i, t_i)$  at  $t_i$ , i.e., when the  $i$ th task arrives.

### 2.1 PAYG and Plan Pricing

The user is charged for his/her usage of the edge-computing service (data transmission and computing resource usage). Two payment pricing approaches are available: pay-as-you-go (PAYG) and all-you-can-use plans. PAYG lets users pay for the communication volume and computing workload without long-term commitments. A cost of  $p_1$  (resp.  $p_2$ ) is incurred for a unit communication volume (resp. computing workload). Here we focus on fine-grained pricing as it is widely advocated in practice (e.g., [14] and [2]). Besides, non fine-grained PAYG pricing is also available in real markets (e.g., [15] and [2]). That is,  $K$  dollars per  $M$  mega bytes is charged. If less than  $M$  mega bytes are used,  $K$  dollars are still fully charged. In our subsequent analysis, we found that the non fine-grained pricing does not affect our analysis much. Therefore, we only focus on fine-grained pricing.

For all-you-can-use plans, we consider three types:

*Data plan:* An upfront fee  $C_1$  is paid at  $t$ ; all communication volume in  $[t, t + T)$  will be free of charge.

*Computing plan:* An upfront fee  $C_2$  is paid at  $t$ ; all computing workload in  $[t, t + T)$  will be free of charge.

*Combo plan:* An upfront fee  $C_3$  is paid at  $t$ ; all communication volume and computing workload in  $[t, t + T)$  will be free of charge.

For each plan,  $T$  is defined as *reservation period*<sup>1</sup> and we assume  $C_3 \leq C_1 + C_2$  and  $C_3 \geq \max(C_1, C_2)$ . Otherwise, it is straightforward to show that the combo plan is always better or worse than reserving data and computing plans individually.

### 2.2 Plan Reservation Problem Formulation and Analysis

Upon the arrival of tasks, the user should determine whether to reserve a plan or pay for the task in a PAYG fashion. Purchasing a plan will involve a high reservation cost, but it may save money if there are many tasks in the upcoming period of  $T$ . However, the user does not have any knowledge about the future tasks: he/she has to employ an *online* reservation policy, which only depends on the *historical and current information* to determine how to reserve the plans.

Let  $\mathbf{S} = \{(a_1, b_1, t_1), (a_2, b_2, t_2), \dots, (a_N, b_N, t_N)\}$  denote all tasks. Given  $\mathbf{S}$ , the offline minimization problem is to determine the optimal *offline* strategy, which tells when to purchase each of the plans to minimize the overall expense. A special case of the offline reservation problem (all tasks only have communication volume but zero computing workload) is equivalent to the offline ski-rental problem [10, 16] that is known to be NP-complete. Therefore, the offline minimization problem is NP-complete. However, an online solution to our plan reservation problem is substantially more complicated compared with online ski-rental problem. This is because the user has an extra combo plan option which covers both the communication volume and computing workload. When purchasing the combo plan, the user needs to compare it with individual plans with unknown future information.

To quantify the performance of an online strategy, we adopt the conventional competitive analysis [4]. An online reservation algorithm  $A$  is  $c$ -competitive ( $c$  is a constant) if for all possible  $\mathbf{S}$ , we have

$$C_A(\mathbf{S}) \leq c \cdot C_{OPT}(\mathbf{S}), \quad (1)$$

where  $C_A(\mathbf{S})$  is the outcome of algorithm  $A$  and  $C_{OPT}(\mathbf{S})$  is obtained by solving the problem optimally offline.

## 3 Problem Solution: SOR Scheme

In this section, we present our SOR scheme for users to choose the most suitable pricing plans to minimize the

<sup>1</sup>The plan reservation period  $T$  is an arbitrary value and determines the frequency of a user to change plans.

cost. SOR employs different algorithms under different  $C_1$ ,  $C_2$ , and  $C_3$  values and achieves a minimum possible competitive ratio. We further discuss how ESPs should design the prices so that the combo plans are more attractive compared with individual plans.

*Case 1. Inexpensive Combo Plan:* Let  $\sqrt{C_1(C_1+C_2)}$  denote the geometric average between the data plan fee and the sum of data and computing plan fees, while  $\sqrt{C_2(C_1+C_2)}$  denotes the geometric average between the computing plan fee and the sum of data and computing plan fees. In Case 1, we consider  $C_3 \leq \min\left(\sqrt{C_1(C_1+C_2)}, \sqrt{C_2(C_1+C_2)}\right)$ , which means the combo plan fee is less than both geometric averages. In this case, Algorithm 1 is employed: It only purchases combo plans whenever there are opportunities.

The algorithm starts whenever there is a new task arrival (Line 1). If there is no on-going plan to cover the task, the algorithm decides whether the task is satisfied by PAYG, or a new plan is purchased to cover this task (as well as potential tasks). In order to make such a decision, we formally define the typical data and communication costs as follows. At time  $t$ , let  $E_1$ ,  $E_2$ , and  $E_3$  denote the expiration time instant of last data, computing and combo plans respectively. The typical data cost is the sum of PAYG data costs during  $(\max(t-T, E_1, E_3), t]$ . The typical computing cost is the sum of PAYG data costs during  $(\max(t-T, E_2, E_3), t]$ . If  $x$  and  $y$  (typical cost) satisfies the conditions in Line 4, a combo plan is reserved to cover the current arrival as well as all other task arrivals in the following period of  $T$ . Otherwise, the new task is satisfied by a PAYG payment.

---

**Algorithm 1** SOR Case 1: Inexpensive Combo Plan

---

- 1: Proceed if there is a new task.
  - 2: **if** there is no on-going combo plan **then**
  - 3: | Calculate  $x$  and  $y$  as the typical data cost and computing cost.
  - 4: | **if**  $x + y \geq C_3$  **or**  $y \geq C_2$  **or**  $x \geq C_1$  **then**
  - 5: | | Reserve a combo plan.
  - 6: | **else**
  - 7: | | Satisfy the task by PAYG.
  - 8: | **end if**
  - 9: **end if**
  - 10: Repeat from step 1.
- 

*Case 2. Expensive Combo Plan:* We consider that the combo plan fee is greater than both of the geometric averages,  $C_3 > \max\left(\sqrt{C_1(C_1+C_2)}, \sqrt{C_2(C_1+C_2)}\right)$  in case 2. Here, we propose Algorithm 2 preferring to buy the computing and data plans separately for the purpose of flexibility. The key difference between Algorithm 1 and Algorithm 2 is that there are exclusive and non-exclusive data and computing plans. Once an exclusive

data/computing plan is on going, the user can no longer reserve a combo plan. A non-exclusive data/computing plan does not enforce this prohibition.

In Algorithm 2, if  $x \geq C_1$  (Line 11) or  $y \geq C_2$  (Line 13), the user should reserve an exclusive data plan or exclusive computing plan. When  $y \geq C_3 - C_1$  (Line 6) and the exclusive data plan is still valid, the user should reserve a non-exclusive computing plan. If  $x \geq C_3 - C_1$  and the exclusive computing plan is still valid, the user should reserve a non-exclusive data plan. If we have  $x + y \geq C_3$ ,  $x \leq C_1$  and  $y \leq C_2$ , then a combo plan is estimated to be beneficial. Otherwise, the task should be satisfied by a PAYG payment.

---

**Algorithm 2** SOR Case 2: Expensive Combo Plan

---

- 1: Proceed if there is a new task.
  - 2: **if** there is no plan fully covering the task **then**
  - 3: | Calculate  $x$  and  $y$  as the typical data and computing costs.
  - 4: | **if** there is an on-going exclusive computing plan **and**  $x \geq C_3 - C_2$  **then**
  - 5: | | Reserve a non-exclusive data plan.
  - 6: | **else if** there is an on-going exclusive data plan **and**  $y \geq C_3 - C_1$  **then**
  - 7: | | Reserve a non-exclusive computing plan.
  - 8: | **else**
  - 9: | | **if**  $x + y \geq C_3$  **and**  $x \leq C_1$  **and**  $y \leq C_2$  **then**
  - 10: | | | Reserve a combo plan.
  - 11: | | **else if**  $x \geq C_1$  **then**
  - 12: | | | Reserve an exclusive data plan.
  - 13: | | **else if**  $y \geq C_2$  **then**
  - 14: | | | Reserve an exclusive computing plan.
  - 15: | | **else**
  - 16: | | | Satisfy the task by PAYG.
  - 17: | | **end if**
  - 18: | **end if**
  - 19: **end if**
  - 20: Repeat from step 1.
- 

*Case 3. Moderate Combo Plan Price:* In Case 3, we consider that the combo plan fee is moderate and in the middle between the geometric averages,  $\sqrt{C_1(C_1+C_2)} < C_3 \leq \sqrt{C_2(C_1+C_2)}$ .

In Algorithm 3, when  $x \geq C_1$ , it has the same strategy as Algorithm 2. Since the cost of computing plan is greater than the cost of data plan, it is beneficial for the user to buy a combo plan when  $y \geq C_2$  (Line 7). If we have  $x + y \geq C_3$ ,  $x \leq C_1$  and  $y \leq C_2$ , then the combo plan is estimated to be beneficial. Otherwise, the new task should be satisfied by a PAYG payment.

Please note that the case  $\sqrt{C_2(C_1+C_2)} < C_3 \leq \sqrt{C_1(C_1+C_2)}$  is symmetric to Case 3. A separate Algorithm 4 to handle this case is omitted here due to the

---

**Algorithm 3** SOR Scheme: Moderate Combo Plan

---

```
1: Proceed if there is a new task
2: if there is no plan fully covering the task then
3:   | Calculate  $x$  and  $y$  as the typical data and computing costs
4:   | if there is an on-going exclusive data plan and
   |  $y \geq C_3 - C_1$  then
5:     | Reserve a non-exclusive computing plan.
6:   else
7:     | if  $(x + y \geq C_3$  and  $x \leq C_1)$  or  $y \geq C_2$  then
8:       | Reserve a combo plan.
9:     | else if  $x \geq C_1$  then
10:      | Reserve an exclusive data plan.
11:    | else
12:      | Satisfy the task by PAYG.
13:    | end if
14:  | end if
15: end if
16: Repeat from step 1.
```

---

space limitation.

In the following theorems, we analyze the performance of SOR scheme and highlight that it achieves the minimum possible competitive ratio. Due to the space limitation, the proofs of the following theorems are omitted here and provided in our technical report [13].

**Theorem 1** *The competitive ratio of SOR scheme with fine-grained PAYG charging is  $1 + \frac{C_1 + C_2}{C_3}$ .*

**Theorem 2** *The competitive ratio of SOR scheme is minimum among all deterministic online algorithms.*

If the pricing for PAYG cost is non fine-grained ( $K_1$  dollars per  $M$  units of communication volume and  $K_2$  dollars per  $N$  units of computing workload), the competitive ratio will be  $1 + \frac{C_1 + C_2 + K_1 + K_2}{C_3}$ .  $K_1$  and  $K_2$  are most likely small compared to  $C_3$  (After all,  $C_3$  is the upfront fee for a combo plan which covers unlimited usage).

Given that the SOR scheme achieves the minimum competitive ratio, in the following discussion we assume that all rational users will adopt the SOR scheme when they reserve their edge-computing service plan to avoid overcharge. As such, we can highlight the following important design guideline for the combo plan pricing: The user will stick to the combo plan by adopting Algorithm 1 in Case 1 where  $C_3$  is relatively inexpensive,  $C_3 < \min\left(\sqrt{C_1(C_1 + C_2)}, \sqrt{C_2(C_1 + C_2)}\right)$ . This is the critical price point as the combo plan is more attractive compared with individual plans. When the price of the combo plan is moderate, the user starts choosing the individual computing and data plans according to Algorithms 3. When the price of combo plan is expensive,

according to Algorithm 2, the user will choose individual plans in most cases due to its flexibility.

## 4 Evaluation

In our simulation, we apply our SOR scheme to an edge-computing scenario based on the publicly open mobile record trace of an actual user at the Czech Technical University [6]. The trace shows the time instances when the user uses the mobile phone to implement different applications. For our edge-computing scenario, each application is associated with typical communication volume and computing workload, as shown in Table 1 [11]. As a result, the arrival time instance, communication workload, and computing workload are employed as the input data to generate different traces representing behaviors of different users.

The computing prices are inferred from Amazon EC2 t2.xlarge instance. PAYG rate is set to  $p_2 = \$0.0464$  per vCPU-hour<sup>2</sup> and plan fee is set to  $C_2 = \$18$  for a week<sup>3</sup>. The communication prices are inferred from AT&T with PAYG rate  $p_1 = \$0.01$  per MB. We consider two data plan prices  $C_1 = \$15$  and  $C_1 = \$21$  for a week (different QoS levels in data transmission are considered).

Table 1: Description of Mobile Edge Computing Tasks

Tasks	MB	vCPU - hour	Probability of Arrivals
Browsing	1	0	20%
Video Streaming	16	0	20%
Machine Learning	10	0.2185	20%
Virtual Reality	20	1.1360	20%
Scientific Programming	0.2	3.4486	20%

We evaluate the performance of our proposed SOR scheme by comparing it with four benchmark strategies.

*PAYG Only (P)*: All tasks are satisfied by PAYG.

*Combo Only (C)*: A combo plan is purchased when a new task arrives and previous combo plan expires.

*Individual Plan (I)*: A communication (resp. computing) plan is purchased when new communication volume (resp. computing workload) arrives and previous plan expires.

*Offline (O)*: The user knows all task arrivals in advance. Whenever a task arrives, the user decides to purchase a plan to make sure that the sum of further PAYG cost it will cover is greater than the plan fee.

In Figs. 1a and 1b, we show the overall expenses of our proposed SOR scheme and the four benchmark schemes. In Figs. 1c and 1d, we compare the number of tasks covered by PAYG and plans under different schemes. In all

---

<sup>2</sup>vCPU-hour is defined as the computing workload that uses a virtual CPU (a hyper-thread of an Intel Xeon processor with up to 3.3GHz clock speed) for an hour [1].

<sup>3</sup>Since the trace only spans for three months, we scale to period to 1 week.



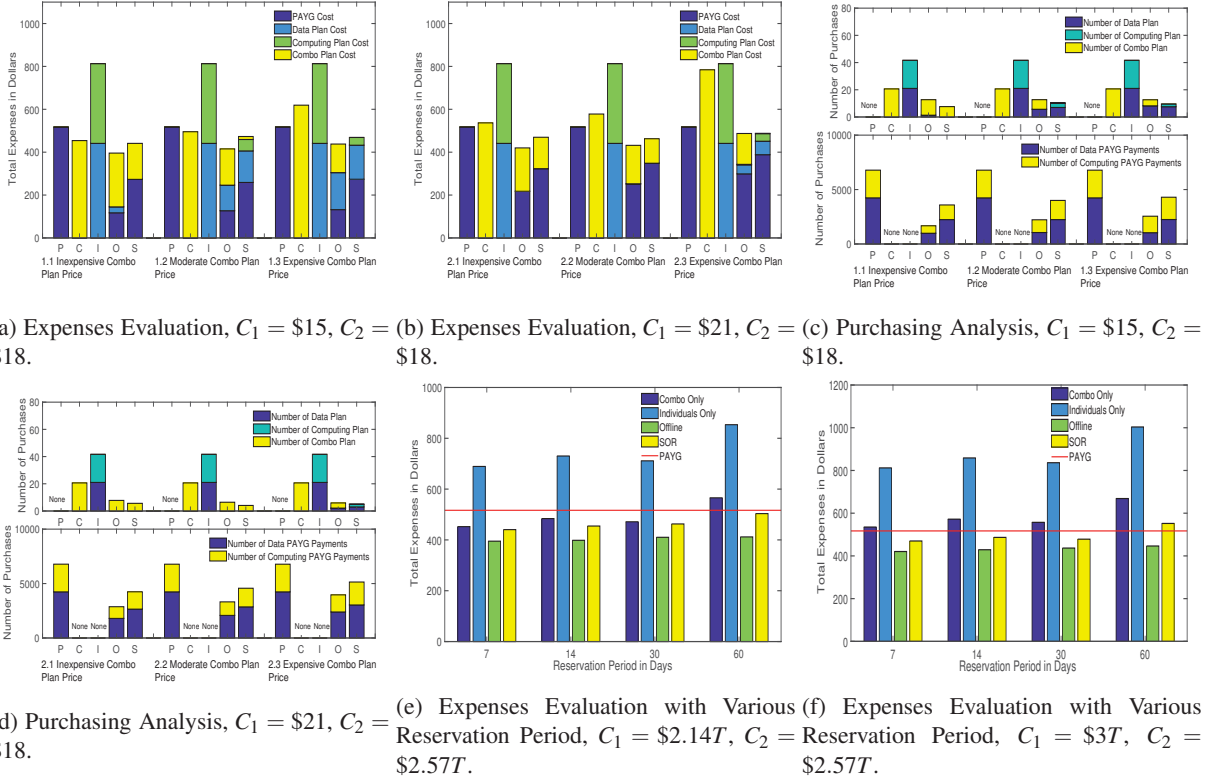


Figure 1: Performance evaluation and analysis of SOR Scheme.

of the four figures, we set  $T = 7$  days. In Figs. 1a and 1c, we set  $(C_1, C_2) = (\$15, \$18)$  while  $C_3$  is set to  $\$22$  (inexpensive),  $\$24$  (moderate), and  $\$30$  (expensive) in Cases 1.1, 1.2, and 1.3 respectively. In Figs. 1b and 1d, we set  $(C_1, C_2) = (\$21, \$18)$  while  $C_3$  is set to  $\$26$ ,  $\$28$ , and  $\$30$  in Cases 2.1, 2.2, and 2.3. The labels P, C, I, O, and S indicate the PAYG only, Combo Only, Individual Plan, Offline, and SOR schemes respectively.

It can be noted from Figs. 1a and 1b that our proposed SOR achieves the lowest total expenses compared to all online benchmark schemes in overall expenses, illustrating that SOR is a smart scheme for users without future information. Moreover, the cost of our SOR is very close to that of the offline scheme especially when  $C_3$  is high, even though the offline scheme has the information of all task arrivals in advance.

From Figs. 1c and 1d, we see that for the inexpensive combo plan price, our SOR and the offline benchmark scheme purchases fewer individual computing and data plans, which matches the implementation of Algorithm 1. In case 2.2, even individual data and computing plans are allowed, combo plan still dominates the user's choice. When  $C_3$  is expensive, SOR purchases individual data and computing plans to replace most combo plans. The results recommend that ESPs should set  $C_3$  in the inexpensive range to make the combo plan more attractive

to users.

In Figs. 1e and 1f, we show the overall expenses of our SOR and four benchmark schemes with various reservation periods  $T$  of 7, 14, 30 and 60 days. The price of all the plans are set proportional to the length of  $T$ . It can be noted from Figs. 1e and 1f that our SOR is better than any online benchmark schemes in most situations. As  $T$  increases, the expenses of SOR increase but more slowly compared with other online schemes. This is because the plan purchasing is less flexible and the user is more likely to be in a long-term plan even if there is no task. This result suggests that if ESP sets large reservation period  $T$ , it will push users to use PAYG.

## 5 Conclusion

This paper studied the mobile edge-computing service pricing design from perspectives of users and ESP. A smart online reservation(SOR) scheme was proposed for the user and was **proved** to achieve the best possible competitive ratio  $(1 + \frac{C_1 + C_2}{C_3})$  among all online deterministic schemes. Important plan pricing design guidelines were provided for the ESP. Trace-driven simulation results verified the cost advantage of SOR scheme and identify how different prices influence user behavior.

## References

- [1] AMAZON. Amazon EC2 instance types, 2018. Available at <https://aws.amazon.com/ec2/instance-types/#burst>
- [2] AMAZON. Amazon EC2 on-demand pricing, 2018. Available at <https://aws.amazon.com/ec2/pricing/on-demand/>
- [3] AT&T. AT&T to launch edge computing test zone in silicon valley in early 2018, Nov. 2017. Available at [http://about.att.com/story/launch\\_edge\\_computing\\_test\\_zone.html](http://about.att.com/story/launch_edge_computing_test_zone.html)
- [4] BORODIN, A., AND EL-YANIV, R. *Online Computation and Competitive Analysis*. Cambridge University Press, New York, NY, USA, 1998.
- [5] COLON, A. iPhone 4s voice and data plans: AT&T vs. Sprint vs. Verizon, Oct. 2011. Available at <https://www.pcmag.com/article2/0,2817,2394224,00.asp>
- [6] FICEK, M. Crawdad dataset ctu/personal (v. 20120315). Tech. rep., Czech Technical University, 2012. Available at <https://crawdad.org/ctu/personal/20120315>
- [7] GLOBE NEWSWIRE. \$80 billion mobile edge computing opportunity: Market assessment and forecasts 2016 - 2021, Jan. 2017.
- [8] HU, Y. C., PATEL, M., SABELLA, D., SPRECHER, N., AND YOUNG, V. Mobile edge computing—a key technology towards 5G. *ETSI white paper 11*, 11 (2015), 1–16.
- [9] JIN, A. L., SONG, W., AND ZHUANG, W. Auction-based resource allocation for sharing cloudlets in mobile cloud computing. *IEEE Transactions on Emerging Topics in Computing* (Jan 2018), 45–57.
- [10] KARLIN, A. R., KENYON, C., AND RANDALL, D. Dynamic tcp acknowledgement and other stories about  $e/(e-1)$ . In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2001), STOC '01, ACM, pp. 502–509.
- [11] MIETTINEN, A. P., AND NURMINEN, J. K. Energy efficiency of mobile clients in cloud computing. *HotCloud 10* (2010), 4–4.
- [12] SUGIYAMA, K., URAKAWA, J., TAYA, M., YAMADA, A., KOBAYASHI, A., AND TAGAMI, A. Empirical analysis of customer behavior for tiered data plans in mobile market. In *2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (April 2016), pp. 389–394.
- [13] S.ZANG, W.BAO, P.YEOH, B.VUCETIC, AND Y.LI. Technical report: Paying less for more? combo plans for edge-computing services. Tech. rep., University of Sydney, 2018. Available at <http://rp-www.cs.usyd.edu.au/~weibao/TechReport/HotEdge18TR.pdf>
- [14] TELSTRA. Telstra pre-paid mobile offers & rates, 2018. Available at <https://www.telstra.com.au/mobile-phones/prepaid-mobiles/40-prepaid-multi-fit-sim-starter-kit?#long-life-plus>
- [15] VERIZON. Add a data boost - my verizon app, 2018. Available at <https://www.verizonwireless.com/support/knowledge-base-205749/>
- [16] WANG, W., LIANG, B., AND LI, B. Optimal online multi-instance acquisition in iaas clouds. *IEEE Transactions on Parallel and Distributed Systems* 26, 12 (Dec 2015), 3407–3419.