Cost-Efficient Resource Provision for Multiple Mobile Users in Fog Computing

Shuaibing Lu[a,b], Jie Wu[b], Yubin Duan[b], Ning Wang[b], and Juan Fang[a]

[a]Faculty of Information Technology, Beijing University of Technology
[b]Dept. of Computer and Info. Sciences, Temple University

Outline

1. Background and Motivation
2. Model and Formulation
3. Provision With Single Replication (PSR)
4. Provision With Multiple Replications (PMR)
5. Evaluations
6. Conclusions

1. Background and Motivation

Cloud Data Center Networks (DCNs)

- Supporting cloud-based applications for large enterprises

Fog Computing

- Providing computation, storage, and networking services between end devices and traditional cloud DCNs.

1. Background and Motivation

Motivation

- IoT devices generate data constantly and the analysis must be very rapid
- Meeting computation and communication demands
- Reduce the transmission latency and decrease monetary cost

Objective

- Find an assignment of the workload for the workload of users with minimum cost, and support all users' demands in both computation and communication resources constraints.

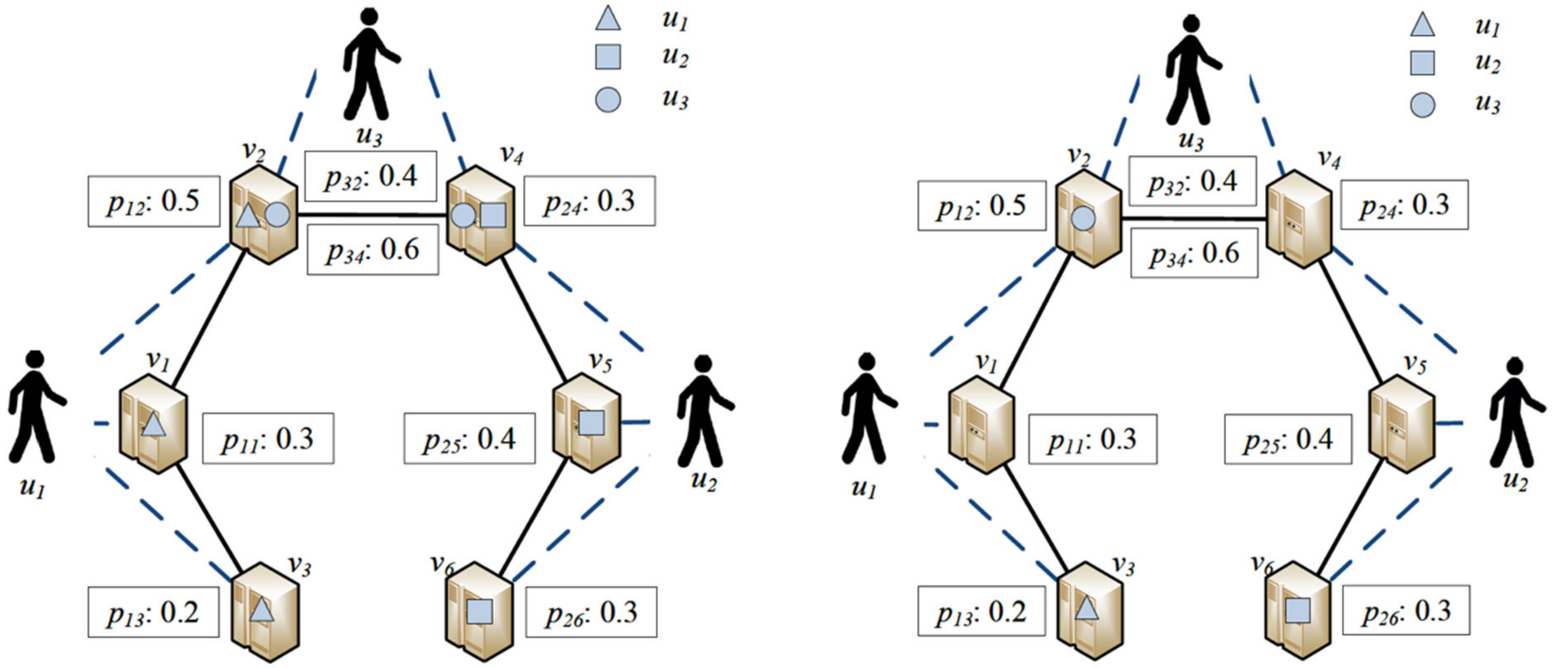
2. Model and Formulation

Formulation

- Find an appropriate scheme which takes the set of users' workloads as input and decides the amounts and locations of the replications for users' workloads in the fog computing system accordingly such that the total cost is minimized.

minimize	$\sum_{k \in U} f_C(u_k)$	(1)
subject to	$f_C(u_k) = f_R(u_k) + f_T(u_k)$	(2)
	$f_R(u_k) = \sum_{i \in X_k} r_{ki}$	(3)
	$f_T(u_k) = \sum_{j \in V_k / X_k} p_{ki} \cdot t_{ij}$	(4)
	$0 \le p_{ki} \le 1, \sum_{i \in V_k} p_{ki} = 1$	(5)

2. Model and Formulation



(a) Transmission cost minimization. (b) Replication cost minimization.

Fig 1. Illustration of two extreme assignments.

3. Provision With Single Replication (PSR)

Three Cases

- $\eta = 1$

1) Calculate the value of weight w_{ki} according to G_k;
2) Construct a bipartite graph;
3) Calculate the maximum weight matching by using the Kuhn-munkras algorithm;
4) Return the resource provision scheme X of U;

- $\eta > 1$

Optimal solution still can be calculated by applying the bipartite matching.

- $\eta = \infty$

Each user offloads his/her workload to the fog node with the smallest cost that contains the transmission and replication cost w_{ki}. (greedy approach)

4. Provision With Multiple Replications (PMR)

- Non-Adaptive Algorithm based on Submodular

Theorem 2: The reward function $f(Z_k) = W_k - C(Z_k)$ of the D-PMR problem is a submodular function.

Theorem 3: The reward function $f(Z_k) = W_k - C(Z_k)$ is non-monotone function.

Definition 1 (additive error): Let $X_k = V_k(1/2)$ denote a uniformly random subset of V_k of user u_k. For each element x, define $\omega(x) = E[f(X_k) \cup \{x\}] - f(X_k\{x\})$.

Non-Adaptive Algorithm based on Submodular

- **Step 1**
  - Use random sampling to find the estimated value $\widetilde{\omega}(v_j)$ for each fog node v_j in V_k;
- **Step 2**
  - Independently, sample a random set with probability $p_{ki} = 1/2$, where $\boldsymbol{X_k} = \boldsymbol{V_k(1/2)}$;
- **Step 3**
  - With probability 8/9, return $\boldsymbol{X_k}$;
  - With probability 1/9, return $X_k' = \{v_j \in V_k : \widetilde{\omega}(v_j) > 0\}$;

Theorem 4: The NA-D-PMR algorithm is bounded by $\frac{1}{3}(2W + OPT)$, and the time complexity is $O(|U| \cdot |V|)^6$.

Approximation Algorithm based on Local Search

- Step 1
  - $C(X \cup \{v_j|_{j\in Y}\}) - C(X) < 0$ ➡ $\Delta = C(X \cup \{v_j|_{j\in Y}\}) - C(X) < 0$
  - Update set $X = X \cup \{v_j|_{j\in Y}\}$;
- Step 2
  - $C(X \setminus \{v_j|_{j\in Y}\}) - C(X) < 0$ ➡ $\Delta = C(X \setminus \{v_j|_{j\in Y}\}) - C(X) < 0$
  - Update set $X = X \setminus \{v_j|_{j\in Y}\}$;
- Step 3
  - $C(X \setminus \{v_i|_{i\in X}\} \cup \{v_j|_{j\in Ys}\}) - C(X) < 0$ ➡ $\Delta = C(X \setminus \{v_i|_{i\in X}\} \cup \{v_j|_{j\in Y}\}) - C(X) < 0$
  - Update set $X = X \setminus \{v_i|_{i\in X}\} \cup \{v_j|_{j\in Y}\}$;

The LS-D-PMR is a $3 + \epsilon$ approximation algorithm and the time complexity is $\boldsymbol{O(|U| \cdot |V| \cdot M)}$.

Approximation Algorithm based on Local Search

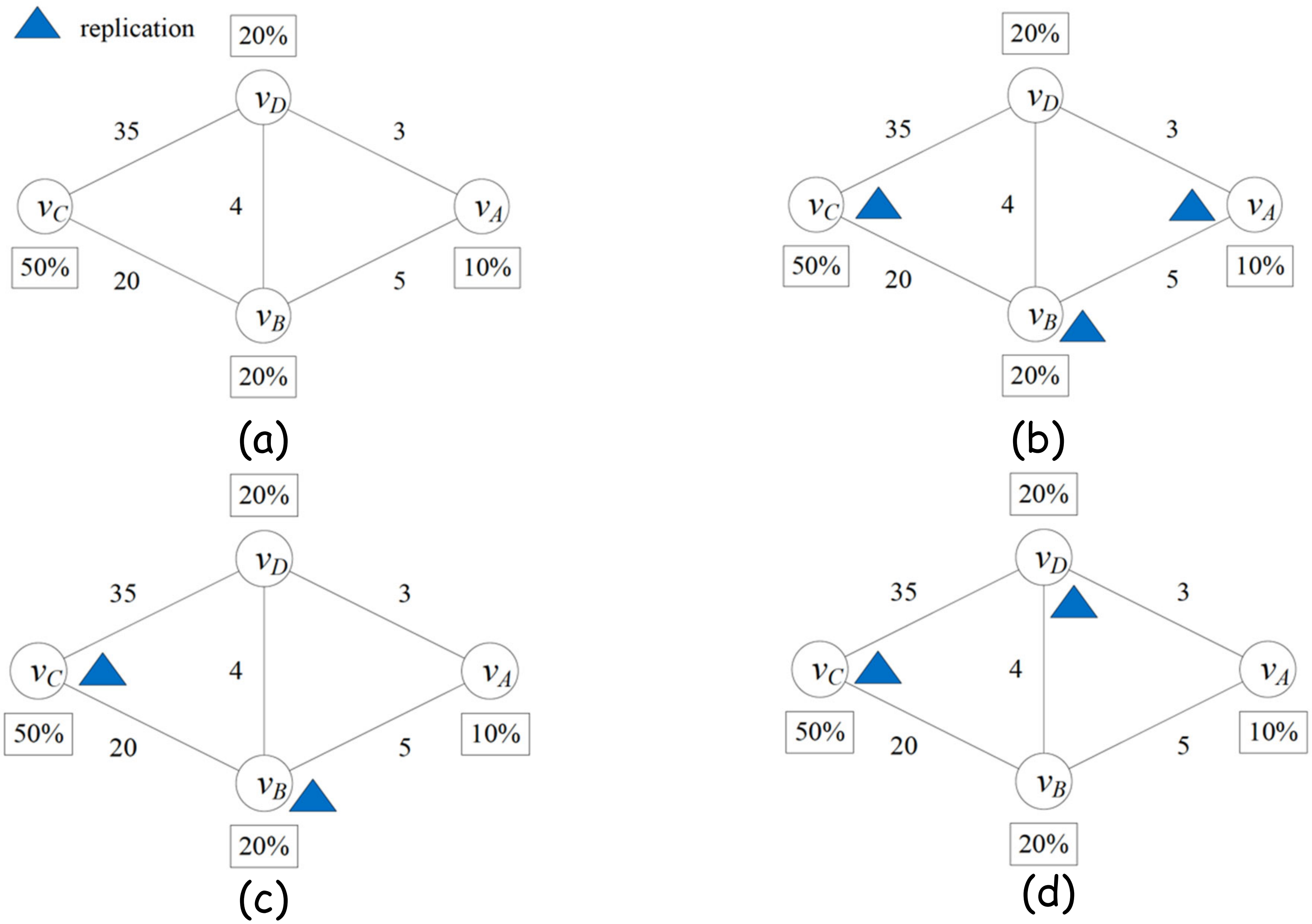


Fig 2. The motivation example.

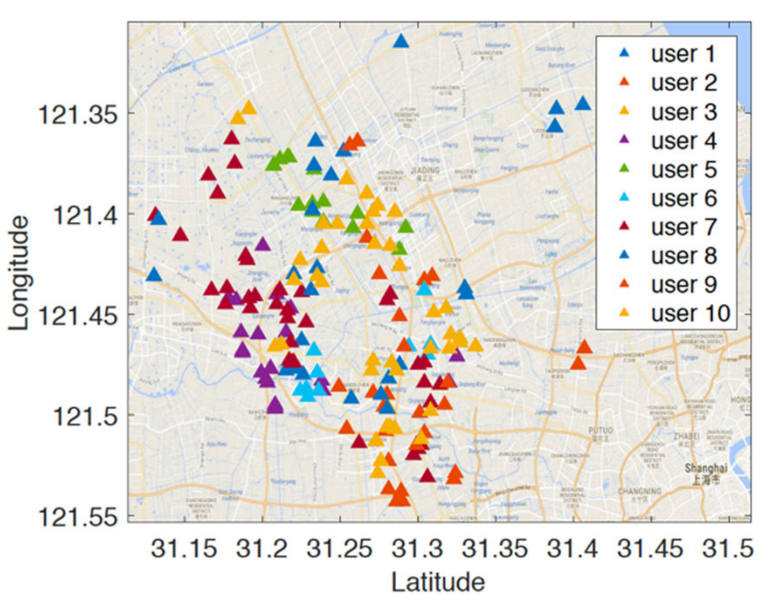
5. Evaluations

- **Basic Setting-Real Dataset**
  - Mobike Dataset:16680 users， 102361 records
  - Microsoft GPS trajectory dataset: 182 users
- **Three Comparison algorithms**
  - Random Adding algorithm (RA):the replications on fog nodes of each user are added randomly
  - Random Removing algorithm (RR): The replications on fog nodes of each user are removed randomly
  - Greedy Adding algorithm (GA): the replications on fog nodes of each user are greedily added by the probabilities.

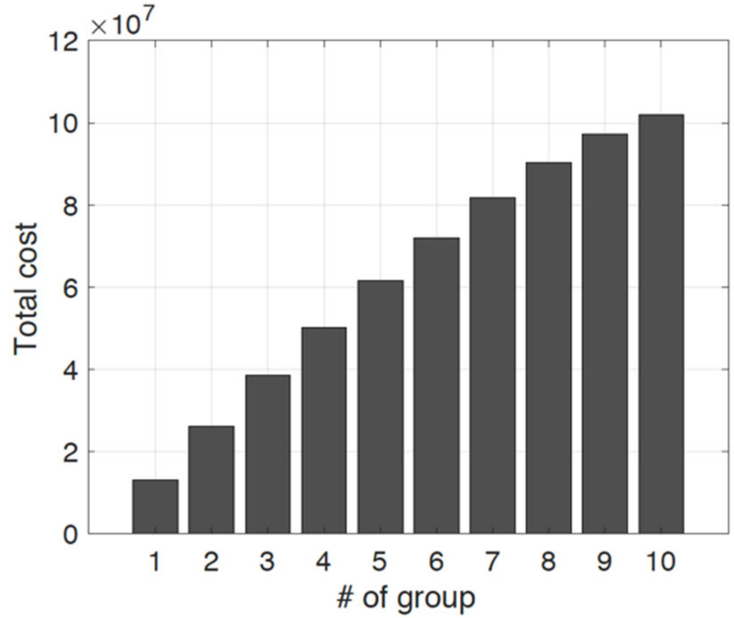
Experiment Results (PSR)

❑ Cost under the PSR (Mobike users')

- With the scaling number of users, the total cost increases.
- Although the increase in the number of users is linear, the increase in total cost is non-linear.



(a) Mobike users' interesting points.

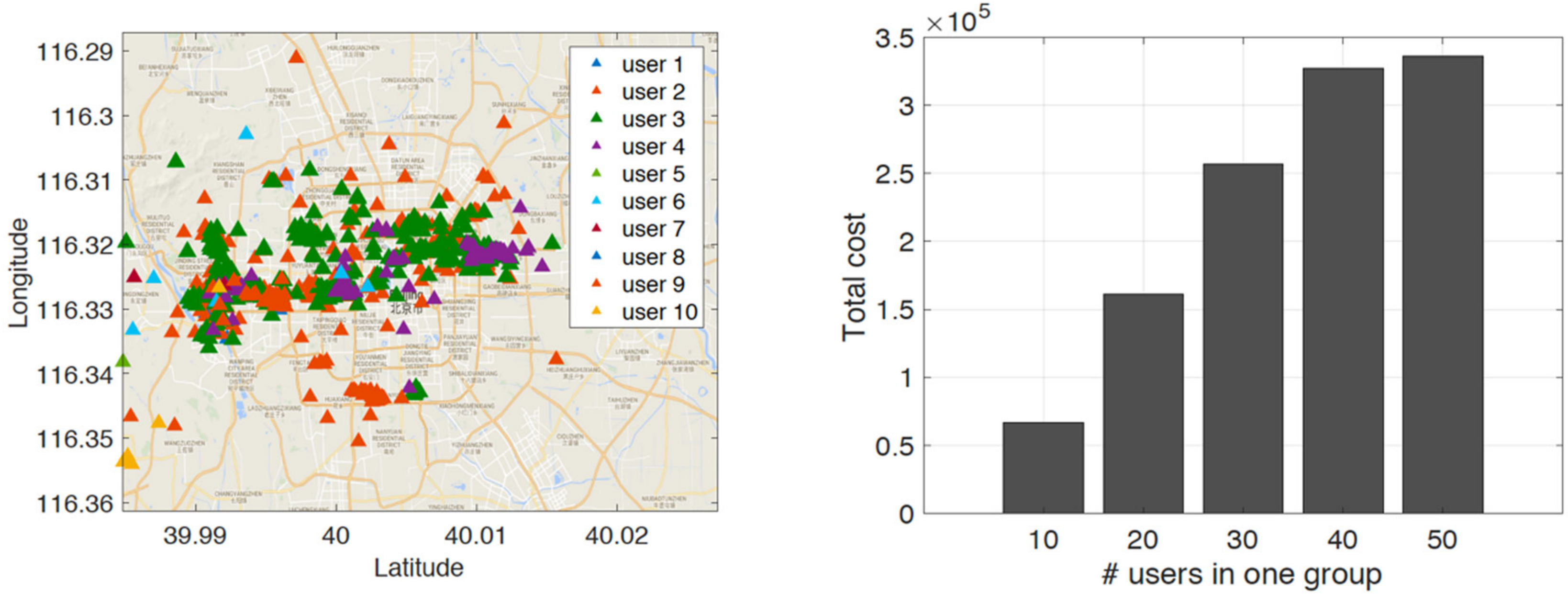


(b) Total cost with single replication.

Fig 3. Mobike users' interesting points and total cost with single replication.

Experiment Results (PSR)

- **Cost under the PSR (Phone users')**
  - Clustering the interesting points for 50 users based on the Microsoft GPS trajectory dataset.
  - Users whose trajectories cover more fog nodes will have a larger total cost.



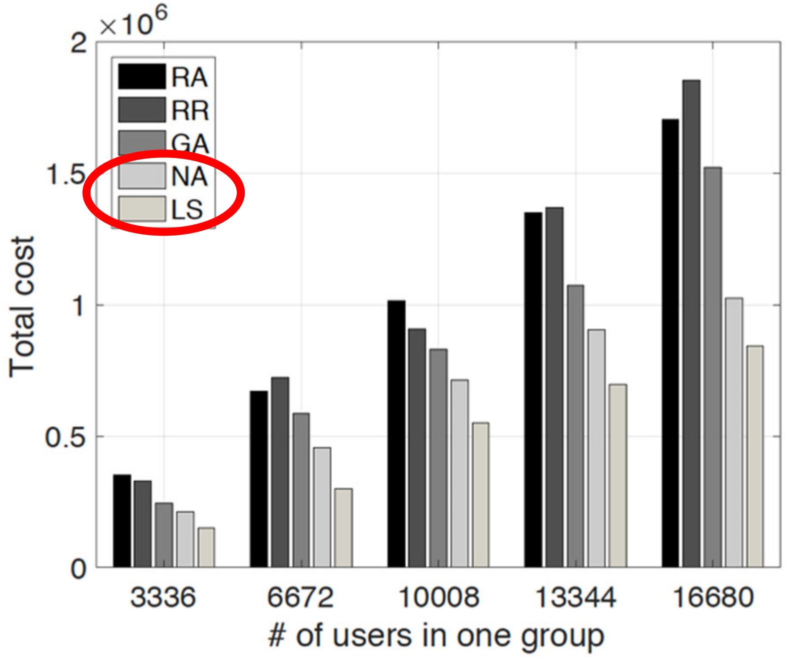
(a) Phone users' interesting points. (b) Total cost with single replication.

Fig 4. Phone users' interesting points and total cost with single replication.

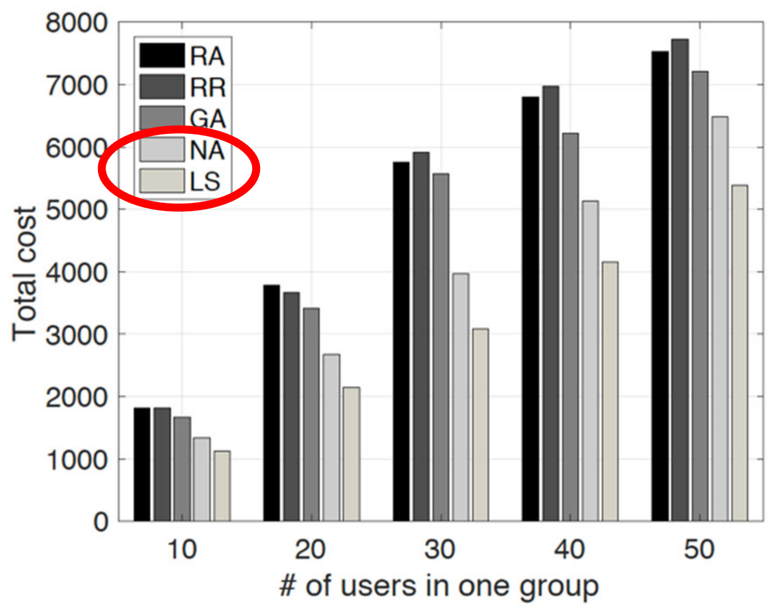
Experiment Results (PMR)

Cost under the PMR

- The total costs of the NA algorithm for some group of users are fluctuating
- LS can have better performance due to the reduction of the number of iterations.



(a) The fluctuation of 30 users.



(b) Average cost of users.

Fig 5. The cost under the LPFN case.

6. Conclusions

- **Objective**
  - Find an assignment of the workload for the workload of users with minimum cost, and support all users' demands in both computation and communication resources constraints.
- **RP with two cases**
  - **Provision With Single Replication (PSR)**
  - **Provision With Multiple Replications (PMR)**
- **Experiments**
  - Real Dataset

Q&A