The draft for the cost minimization of a federated cloud participant

***Abstract*—We plan to consider the assignment of virtual machines under the dynamic request arrival scenario. The objective is to minimize the total cost.**

I. System Model

We study the problem that how an individual cloud provider in the federated clouds minimize its cost.

We consider a time slotted system. Let $t = 0, 1, 2, 3..., T$ be the time slots. Every time slot, there are new jobs arriving from tenants which are needed to be assigned to data centers.

We consider the cloud provides m types of instances. These instances are characterized by their sizes of CPU, memory and storage. The tenants submit their requests for instances in the form of m-tuples. Let $\mathcal{S} = (s_1, ..., s_m)$ represent the number of requested instances. A job is represented by the tuple $(\mathcal{S}, w)$, w is the required time of the instances. $0 \leq s_i \leq s_i^{max}$. $w^{min} \leq w \leq w^{max}$. Hence, there are in total $R = s_1^{max} \times s_2^{max} \times ... \times s_m^{max} \times (w^{max} - w^{min} + 1)$ types of jobs. Let $\mathcal{R}$ be the set of all job types.

Each job is associated with a traffic matrix T^r.

At time slot t, $0 \leq t \leq T$, $A^r(t)$ new type r jobs arrive at the provider. We assume the provider's price is fixed. The price for one type i instance one unit time is p_i.

The provider has a total of N homogeneous servers. Each server can host $n_i, 0 \leq i \leq m$ type i instances. The provider guarantees to serve the job request within a delay bound l. To avoid the violation of no service due to limit of servers, the cloud provider is federated with a set of other cloud providers, which is denoted by $\mathcal{C}$, $|\mathcal{C}| = C$. The federated peer cloud $c \in \mathcal{C}$'s price for instance i one unit time is p_i^c. The price for data transfer out is p_d^c per unit volume per time unit for federated peer cloud provider c.

A. Problem definition

We formulate the cost minimization problem for the cloud provider with federated peer cloud providers.

Let $\mu^r(t)$ be the number of new type r jobs served by the provider at time t. Hence, there are in total $\sum_{i=0}^{w^r-1} \mu^r(t - i)$ type r jobs is being served by the cloud. The vector of requested instances by type r jobs is $\mathcal{S}^r$. Let T_i^r represent the traffic matrix for job i, $1 \leq i \leq \sum_{i=0}^{w^r-1} \mu^r(t - i)$ in type r jobs.

(the variables include $\mu^r(t)$, i.e., the number of new type r jobs served at time t, still need to define the variables for the assignment of jobs in different clouds.)

(Objective: minimize Total cost = operational cost of local cloud + cost of running VMs in peer cloud + data transfer cost + migration cost).

(The next step is to model the assignment of the running jobs into different cloud providers.)