

Efficient Resource Management for Infrastructure as Service Cloud Providers

Mehdi Keshavarz, Hyenyoung Yoon, Junseok Hwang

Technology Management, Economics and Policy Program Department of Industrial and Management Engineering College of Engineering, Seoul National University

Keshavarzm2005@gmail.com, hyyoon00@snu.ac.kr, junhwang@snu.ac.kr

Abstract— Cloud computing consists of a variety of computing services that are running on the service provider's infrastructure. Due to the restricted amount of computing resources and dramatic increase in cloud users, cloud providers may fail to provide service to potential cloud users. A cloud user who, at any given time, fails to receive the needed resources with the expected quality, may likely switch from one cloud provider to another. Cloud providers should offer high quality services, with a guarantee of not restricting the number of accepted requests, maximize their profit, utilize their resources, and dynamically increase their available resources using cloud federation strategy, which prompt cloud service providers to collaborate and trade computing resources. This study is based on different scenarios of standalone cloud and federated cloud using discrete event simulation, explores the impact of profit, and processing costs before and after cloud federation. The study presents policies for decisions whether cloud providers should outsource resources or use internal resources and the effect of outsourcing on the provider's benefits. Additionally, the study proposes a new resource allocation for cloud providers in the federated environment based on the quality matchmaking algorithm for selecting the suitable cloud provider.

Keywords— Cloud federation, cloud computing, Infrastructure-as-a-service, discrete-even simulation, resource allocation, service measurement index

I. INTRODUCTION

Computing power is transformed from traditional computing to the cloud computing approach, with the aim of providing computing services in a way that is comparable to public utilities such as water, electricity, gas, and telephony [1]. Cloud establishes dynamic and flexible application provisioning data centers [2]. Cloud enables "to access files, data, programs, and 3rd party services from a web browser via the Internet that are hosted by a 3rd party provider" [3] while paying only for services that are used.

Cloud federation assists providers to sell their idle resources and monetize them. Cloud users benefit from cloud federation by having extra computing resources at any requested time, having ubiquitous services, and a variety of resources with different pricing models.

Cloud computing works with the usage based cost model, where cloud customers are able to use resources and services in a pay-as-you-go approach. Therefore, if a cloud customer, at

any time, fails to receive resources with the expected quality, they may switch to another cloud provider.

One of the major issues in cloud computing is the lack of resources during peak demand. This motivates service providers to cooperate and trade their resources using cloud federation. Availability of resources in a single IaaS cloud provider is limited and data centers have restricted amount of resources. IaaS cloud providers aim to minimize their operational costs and attract more users by providing more services and resources to them. Because of the high competition in the IaaS cloud market, the cloud provider should provide resources at any time for users and satisfy the QoS of user requirement. Therefore, IaaS cloud providers should have an efficient and economical resource allocation and understand the effect of each resource allocation on their profit.

Giant cloud providers such as Google and Amazon, due to economies of scale, can potentially have more revenue through cost savings rather than smaller cloud providers. Therefore, cloud market should always be followed and controlled by giant providers, and in effect, small IaaS providers can never compete with giant providers and there is no market for smaller cloud providers. This situation can be the reason for a cloud provider to lock-in users and avoid standardization in the cloud market. According to Kim et al. (2014) [4], cloud federations allow IaaS cloud providers to reduce the cost of maintenance of its services and boost the resource utilization. Linthicum (2010) [5] identified the dangers of the influx of giant cloud monopolies wherein too much control is in the hand of a few and, businesses are left weak to shifting costs.

IaaS cloud providers provide their computing resources in the form of virtual machines to cloud users using the web portal. They aim to make profit by providing their available computing resources to handle many users, with the main goal of maximizing their profit while applying the guarantee of Quality of Service (QoS) of user requirements based on the agreed Service Level Agreement (SLA). To accomplish this objective, IaaS cloud providers need efficient and economical resource management strategies.

Nature of services is decomposable. Therefore, IaaS providers' resources such as computing power, memory, and network bandwidth cannot be stored for future selling and if not utilized carefully, due to depreciation cost, they may lose value in the future.

Since cloud providers should provide high QoS services, without having the limitation on the number of accepting end users, they must be capable of increasing their available resources at any requested time. One possible approach for getting additional VMs resources is renting idling resources from other external providers using cloud federation. Cloud participants in the cloud federation can rent part of their idle resources to other providers who need additional resources at that time. With the help of the federation mechanism, cloud service providers get fewer rejections of cloud users' requests and can boost the number of their users.

IaaS cloud providers consist of different resources such as, on-demand, on-spot using bidding strategy and reserved. On-demand resources allow cloud users to pay for resources by its hourly or daily usage and use pay as you go pricing model. On-spot virtual machine by using bidding strategy, allows cloud users to rent a resource by accepting the risk of cancellation due to new users willing to pay more for the same resource. Therefore, these types of resources are not permanent. Reserved resources allow cloud users to rent a resource for a specific duration of time that is usually yearly, which the price of the resource is lower than on-demand model. The challenge in whether a provider uses outsourcing or insourcing using different type of resources to maximize its profit is important. Additionally, analysing the profit of cloud provider in terms of outsourcing and insourcing is playing an important role.

IaaS cloud providers need to minimize their transaction costs using different resource management techniques and policies. Resource management, outsourcing, and trading resources among providers can be efficient for provider to invite more users and prevent market loss. Providers can propose different resources and prices with consideration of their user.

This study aims to consider policy makers of cloud providers, to achieve better understanding of resource allocation policies, outsourcing, service interoperability and cloud federation. We proposed a quality matchmaking algorithm for selecting the suitable outsourced cloud provider. We proposed three types of scenarios, based on the type of resources that are: on-demand, on-spot, and reserved. We analyze the profit of provider, and the effect of insourcing and outsourcing using three different types of resources using discrete event simulation. We consider two main scenarios: no internetworking approach, that the cloud provider is standalone and uses its in-house resources to handle cloud users; in effect, free competition, and internetworking approach, that the provider uses cloud federation to receive advantages from outsourcing.

II. RELATED WORK

There are many definitions and discussions of cloud computing; on what cloud computing is and is not. Berkeley defined cloud computing as *"the long-held dream of computing as a utility has the potential to transform a large part of the IT industry, making software even more attractive as a service"* [6].

Klems, Nimis, & Tai [7] defined cloud computing from an economic perspective view as, *"building on compute and storage virtualization technologies, and leveraging the modern*

Web, Cloud Computing provides scalable and affordable compute utilities as on-demand services with variable pricing schemes, enabling a new consumer mass market".

The definition published by the U.S. National Institute of Standards and Technology definition (NIST) seems to have captured the commonly agreed aspects of cloud computing. The NIST definition states that cloud computing is *"a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"* [8].

A. Federated Cloud Model Marketplace

Cloud federation allows several cloud service providers to share their resources and benefit of exchanging their idle resources. Additionally, cloud users can use resources and services hosted by different service providers in a seamless way. Cloud federation means to *"federate disparate data centers or cloud providers, including those owned by separate organizations to enable a seemingly infinite service computing utility"* [9]. Service owners can trade their services and resources through federation regulations. It requires one service provider to rent or share computing resources to another cloud provider. Thus, cloud providers benefit from cloud federation in areas of resource expansion, lessened customer attrition, disaster recovery, vendor lock-in problem avoidance, enhancement of scalability, and economies of scale.

Due to lack of resources, cloud providers aim to overcome resource constraints that can result to reject user requests, by outsourcing requests using brokering and federation strategies. Through cloud federation, providers can lease part of their resources to other members of the federation, usually at cheaper costs, in order to avoid wasting their non-storable resources [10].

Enhancement in provider's profit is achieved through properly used and implemented cloud federation [11]. Thus, the utilization of federation depends on the appropriate time that federation can be used.

OpenNebula [12] proposed an open-source framework to build a joined up cloud. OpenNebula framework manages the whole aspect of individual cloud providers and integrates them into a unified cloud. It supports external resource from Amazon EC2 in case of having many demands.

Reservoir [13] developed an open stack for IaaS cloud provider with the aim of solving complexity issues for cloud provider and proposes a model for federated cloud environment. Reservoir supports migration of resources across different geographical domains.

The Aneka-Federation [14] proposed a distributed IaaS federation system that aggregated several clouds using peer to peer technology to form a scalable and high-throughput computing resources. The objective of the Aneka-Federation system is to make efficient and scalable distributed cloud resource integration.

III. PROPOSED QUALITY MATCHMAKING ALGORITHM

In the cases of several foreign cloud providers existing in the cloud federation environment, selecting the best external cloud provider is a critical point and can affect the utility of users and profit of providers. One model for selecting is based on the SMI KPIs model which proposed by Cloud Service Measurement Index Consortium (CSMIC) [15]. The selection behaviour based on the SMI KPI model is depicted in Figure 1.

Each service provider in the cloud federation environment has SMI attributes for selecting the best foreigner cloud provider in case of the federation or resource migration.

This study proposes a quality matchmaker to select the best cloud service provider. The following steps explain the overall

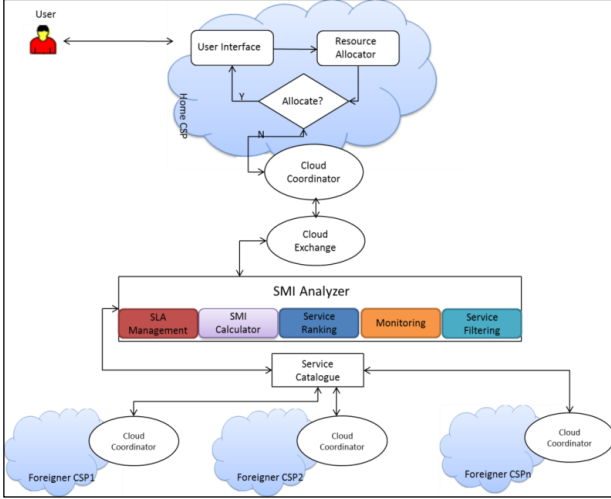


Figure 1 Proposed model for quality matchmaking

calculation for finding the appropriate cloud in the federation environment, federate for the request of users, or migrate for the resources of the main provider.

The quality matchmaking algorithm determines which cloud service provider csp_i from $CSP = \{csp_1, csp_2 \dots csp_n\}$, is selected based on user's QoS specifications.

A. Construct SMI attributes for CSP

Each datacenter consists of several attributes such as, number of hosts, virtual machines, and bandwidth. Additionally, each datacenter six measurements related to SMI KPI model are considered. These measurements are availability, reliability, costs (VM cost and traffic cost), throughput, service response time, and accuracy.

The SMI_{Attr} matrix is created, where n represents the total number of the Cloud Service Provider (CSP), and k represents the total number of SMI attributes.

$$SMI_{Attr} = \begin{bmatrix} q_{1,1} & q_{1,2} & \dots & q_{1,k} \\ q_{2,1} & q_{2,2} & \dots & q_{2,k} \\ \dots & \dots & \dots & \dots \\ q_{n,1} & q_{n,2} & \dots & q_{n,k} \end{bmatrix} \quad (1)$$

Each row in the matrix represents a service provider csp_i , while each column represents one of the SMI attributes.

B. Normalize each SMI attributes of CSP

Each SMI attributes are using different levels of measurement unit to compensate between different units; the value of each attribute should be normalized in the range $[0, 1]$. We use the following equations to normalize them:

$$Normalize_{min}(SMI_{Attr_{i,v}}) = \frac{SMI_{Attr_{i,v} \max} - SMI_{Attr_{i,v}}}{SMI_{Attr_{i,v} \max} - SMI_{Attr_{i,v} \min}} \quad (2)$$

$$Normalize_{max}(SMI_{Attr_{i,v}}) = \frac{SMI_{Attr_{i,v}} - SMI_{Attr_{i,v} \min}}{SMI_{Attr_{i,v} \max} - SMI_{Attr_{i,v} \min}} \quad (3)$$

Where $SMI_{Attr_{i,v}}$ is the attribute that needs to be normalized by minimization using equation 2 or maximization using equation 3. For example, response time needs to be normalized by minimization using equation 2 while availability needs to be normalized by maximization using equation 3.

C. Calculate the Euclidean distance

Euclidean distance measurement is used to evaluate the similarity of distance between two vectors $v_i = (v_{i1}, v_{i2} \dots v_{ik})$ and $v_j = (v_{j1}, v_{j2} \dots v_{jk})$, using the following equation:

$$Euclidean_{Distance}(v_i, v_j) = \sqrt{\sum_{h=1}^k (v_{ih} - v_{jh})^2} \quad (4)$$

By Euclidean distance we compute the distance between user specification requirement and providers' SMI attribute.

IV. SIMULATION SCENARIO

A. On-demand resources

When an IaaS provider receives a request, three types of scenarios can happen:

1) Insourcing (no federation): The provider acts as a standalone provider that does not need of federation. If the provider cannot execute the request, it has to be terminated.

2) Federation outsourcing: The provider is in the federated environment. If there is no available resource for processing, the provider shall communicate with the cloud exchange so it can transfer the request to another provider. In this case, outsourced cloud provider is selected by quality matchmaking algorithm.

3) Migrate idle on-demand resources: The provider checks the availability of idle on-demand resources and compares the profit of migrating the resources with outsourced the request.

B. On-spot and On-demand resources

1) Insourcing: If the provider cannot execute the request, the on-spot resource with lower bid is terminated; otherwise, the request shall be terminated. This policy analyses the profit of the IaaS cloud provider using insourcing.

2) Outsourcing: If there is no available resource for processing the incoming request, the provider communicates with cloud exchange for transferring the request to another provider. In this case, the provider uses the proposed matchmaking algorithm for selecting the suitable cloud provider. If outsourcing is not possible, then the availability of on-spot resources is considered; otherwise, the request shall be terminated. This policy analyses the effect of outsourcing on the provider's profit.

3) Migration of on-spot resources: When provider does not have enough resources, and does not want to terminate the on-spot resources, it migrates the available on-spot into another provider. This policy analyses the effect of resource migration on provider's profit.

C. Reserved and on-demand resources

1) Insourcing of reserved request (no federation): When the provider gets reserved resources, if the provider cannot execute the request and assign the resource to it, then terminate the resource.

2) Outsourcing of reserved request: The provider is in the federated environment and if there is not an available resource for processing, then the provider communicates with cloud exchange for migrating the request to another provider. In this case, outsourced cloud provider is selected by the proposed matchmaking algorithm.

3) Migration of reserved resources: In this policy, decisions are made based on the migrated reserved resources that have low usage.

V. EXPERIMENTS AND RESULTS

The Figure. 2. illustrates up to 2,500 cloud service providers. The execution time of the proposed quality matchmaking algorithm is about 10 seconds, which clearly indicates that our proposed approach can be used for online selection of cloud service providers.

The Figure 3 shows the impact of provider's profit on the three types of resource allocation. As shown according to the number of requests that is assigned to the provider, migrating idle resource allocation and federation, in the case of on-demand resources are more efficient compared to standalone cloud. Because in the migration of idle on-demand resource, it compares the profit of migrating idle resource with the given user's request and if the user's request is more profitable for the provider, it is better to migrate the idle resource to another foreign provider and execute the given process internally. Also, the cloud provider can earn more profit by selling the idle resources with other external providers when joining with the federated environment.

In the case of scenario 2, it shows that for the smallest quantity of on-spot virtual machine requests, handling of federation and migration for outsourcing help cloud providers boost profit. However, after high ratio of on-spot requests, providers were not able to increase profits because on-spot VM price is the beneficial factor on the cloud provider profit. Additionally, when a large number of allocating and running the virtual machine in the cloud data provider belongs to on-spot requests, providers can terminate on-spot virtual machines in order to run on-demand requests since on-demand requests are more profitable. Figure 4 shows the impact of percentage of

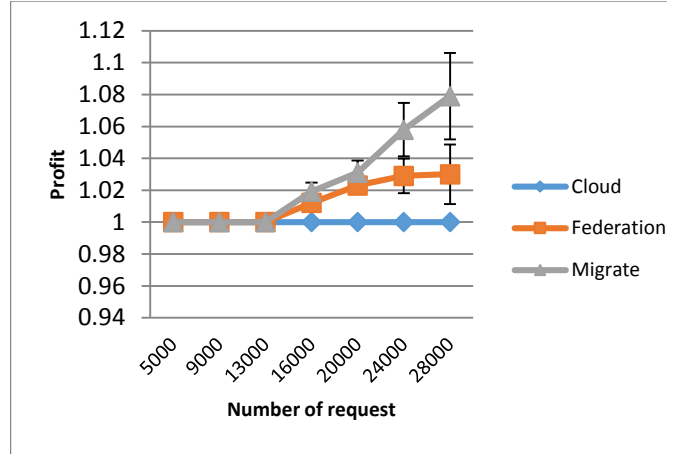


Figure 3 Effect of provider's profit on number of request on-spot requests on the provider's profit.

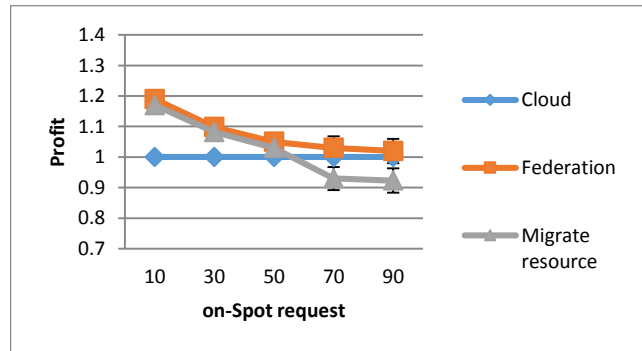


Figure 4 Effect of on-spot requests on provider's profit

In the case of scenario 3, which considers the reserved requests, since the provider cannot cancel the reserved request, in-sourcing can happen only if the provider has enough resources. Otherwise, outsourcing and migration is needed to reserve resources. Figure 5 shows the impact of percentage of reserved requests on the provider's profit. In the case of using migration or federation strategy, the profit of provider almost equal. However, if the amount of reserved requests increases,

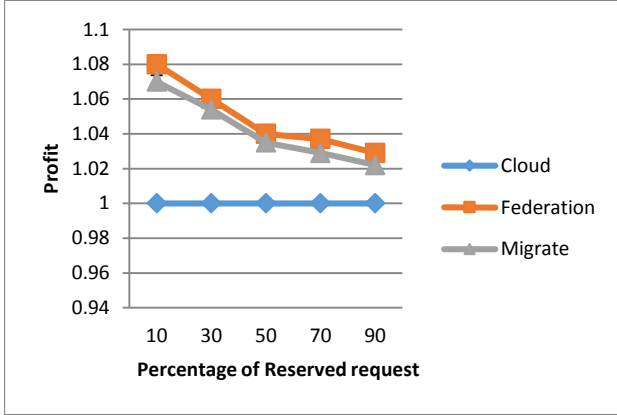


Figure 5 Effect of reserved requests on provider's profit.

the provider should migrate some of the idle reserved request.

VI. CONCLUSION AND DISCUSSION

The Cloud market industry is a competitive industry, wherein there are many providers already in the market and many providers competing to deliver similar services in a highly price-competitive environment. This study proposed efficient resource allocation which is profitable for cloud provider, and analyses the effect of provider's profit before and after cloud federation.

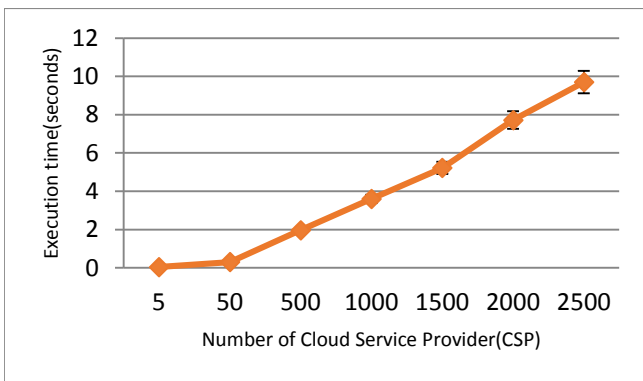


Figure 2 Execution time of quality matchmaking algorithm

With federation, even small service providers can come to the market and offer their infrastructure. Therefore, small cloud

providers can trade their resources with each other, and enable them to compete with big cloud providers.

Due to high competitive market of cloud provider if a cloud provider cannot guarantee or satisfy the user requirements, the user will lose the trust and switch to another cloud provider. Therefore, IaaS providers require certain levels trust to support reliability and to minimize latency.

The efficient management of resource can affect the profit of providers. Federation and migration is based on the type of the provider's resources. Providers that support different types of resources such as on-demand, on-spot and reserved virtual machines can terminate less profitable resources for accepting more profitable requests. Outsourcing enables cloud providers to trade their requests and idle resources to other external cloud providers.

Appropriate quality matchmaking using SMI factors for selecting providers in the cloud federation can reduce the SLA violation and provide a mechanism for cloud coordinators to identify the best cloud service provider.

One of the major constraints in the cloud computing and the federation is the lack of standard and protocol. In cloud federation, the resource trading can successfully happen through standardization where all providers have a similar standard cloud coordinator format that can match with each other. Therefore, cloud federation can be an efficient mechanism for boosting the standardization of the cloud market.

Despite the contributions of the current study for analysing the profit of cloud providers, there are still some research challenges in considering negotiation mechanism using brokering and negotiations in the federated environment. Additionally, security and compatibility are the challenges that can be considered as one of the SMI factors in quality matchmaking algorithm.

REFERENCES

- [1] L. R. Yin. "The Big Switch: Rewiring the World, from Edison to Google." *Journal of Information Privacy and Security* 5(2):67–68. 65, 2009.
- [2] A. Weiss, "Computing in the Clouds." *networker Journal* 11(4):16–25, 2007.
- [3] W. Kim, "Cloud Computing: Today and Tomorrow." *Journal of Object Technology* 8(1):65–72, 2009
- [4] K. Kim, S. Kang, and J. Altmann. "Cloud Goliath versus a Federation of Cloud Davids." in *Proceedings of the 11th International Conference on Economics of Grids, Clouds, Systems, and Services*, 2014
- [5] D. Linthicum, "The Danger of the Coming 'Big Cloud' Monopolies." *InfoWorld*. Retrieved November 15, 2015 (<http://www.infoworld.com/article/2624508/cloud-computing/the-danger-of-the-coming-big-cloud-monopolies.html>), 2010
- [6] M. Armbrust, et al. Above the Clouds: A Berkeley View of Cloud Computing, *Commun. ACM Journal*, 2009
- [7] M. Klems, N. Jens, and T. Stefan. "Do Clouds Compute? A Framework for Estimating the Value of Cloud Computing." *7th Workshop on E-Business, WEB 2008, Paris, France*, Pp. 110–23., 2009
- [8] P. Mell, and G. Tim, "The NIST Definition of Cloud Computing." National Institute of Standards & Technology, 2011

- [9] M. Armbrust et al. "A View of Cloud Computing." *Communications of the ACM* 53(4):50, 2010
- [10] M. Mihailescu, and M.T. Yong. "Dynamic Resource Pricing on Federated Clouds." Pp. 513–17 in *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*. IEEE, 2010
- [11] E.R. Gomes, B.V. Quoc, and K. Ryszard. "Pure Exchange Markets for Resource Sharing in Federated Clouds." *Concurrency and Computation: Practice and Experience* 24(9):977–91, 2012.
- [12] B. Sotomayor, S.M. Rubén, M. L Ignacio, and F. Ian. "Resource Leasing and the Art of Suspending Virtual Machines." Pp. 59–68 in *2009 11th IEEE International Conference on High Performance Computing and Communications*. IEEE, 2009
- [13] B. Rochwerger et al. "Reservoir - When One Cloud Is Not Enough." *Computer* 44(3):44–51, 2011
- [14] R. Ranjan, B. Rajkumar. "Decentralized Overlay for Federation of Enterprise Clouds." Pp. 191–217 in *Handbook of Research on Scalable Computing Technologies*. IGI Global. , 2010
- [15] J. Siegel, P. Jeff. "Cloud Services Measures for Global Use: The Service Measurement Index (SMI)." Pp. 411–15 in *2012 Annual SRII Global Conference*. IEEE, 2012.

analysis. On this topic of research, he is actively working for Digitalogy, and Technology and Humanities Convergence study



Dr. Mehdi Keshavarz (keshavarzm2005@gmail.com) is a PhD . at the Technology Management, Economics, and Policy Program of the College of Engineering, Seoul National University. His research interest is Cloud computing, cloud federation, and simulation.



Dr. Hyenyoung Yoon (hyyoon00@snu.ac.kr) is a professor for Technology Management, Economics and Policy Program at Seoul National University in Korea. Prior to this, she worked as a senior Researcher in a Korea Communications Agency and LG Electronics. She received her Ph.D. in Technology Management Economics and Policy Program from Seoul National University, a Master's degree in Information and telecommunication engineering from Ewha Woman's University. Her current research interests are internet of things, personal information protection, and resource management for cloud computing and spectrum management policy



Prof. Junseok Hwang (junhwang@snu.ac.kr) is Professor of Information Science and Technology at Technology Management, Economics and Policy Program (TEMEP), Seoul National University and Director of International Technology Policy Program (ITPP). Prior to this, he was an Assistant Professor in the School of Information Studies at Syracuse University. He received his Ph.D. in Information Science and Telecommunications from the University of Pittsburgh, a Master's degree in Telecommunications from the University of Colorado. His current research focuses on economics of information and network s, management and policy of convergence technologies, social impact study and forecasting of emerging technologies, knowledge management, patent research and policy