ISSN 2383-6318(Print) / ISSN 2383-6326(Online)
KIISE Transactions on Computing Practices, Vol. 24, No. 7, pp. 345-351, 2018. 7
https://doi.org/10.5626/KTCP.2018.24.7.345

자원 제한적인 기기를 위한 Q-Learning 보완, 응답 기반 크라우드센싱 프레임워크

(Q-Learning Supplemented Response Based Crowdsensing Framework for Resource Constrained Devices)

샤시 라즈 판디 † 사바 수 하일 † 문 승 일 † 홍 충 선 ††
(Shashi Raj Pandey) (Sabah Suhail) (Seung Il Moon) (Choong Seon Hong)

요 약 모바일 크라우드 센싱에서 가장 중요한 과제는 스마트 디바이스가 다양한 목표 지향적 응용프로그램을 위한 다양한 센싱 작업을 수행하도록 동기를 부여하는 것이다. 이는 작업 소유자와 스마트 디바이스 간의 상호 작용으로 스마트 디바이스가 작업 소유자로부터의 작업 수용 여부를 결정하는 데에 영향을 줄 수 있으며, 기존의 연구에서는 다양한 인센티브 기법과 기술을 사용하였다. 하지만 이 외에도 참여 디바이스의 에너지 제한 문제나 알려지지 않은 상호 작용 환경에서 간과되어 왔던 기능을 기반으로 하는 작업을 할당하는 문제 역시 해결해야 하는 주요 문제들이다. 본 논문에서는 이러한 문제의 해결을 위하여 작업 할당을 위한 노드들의 사용율 최대화를 위해 최적의 작업 할당 알고리즘을 제한하였고, 참여 노드에 대한 누적 보상을 향상시키기 위한 크라우드 센싱의 분산 형 Q-Learning 프레임워크를 모델링하였다. 그리고 시뮬레이션 결과를 통해 제안된 알고리즘의 성능을 입증하였다.

키워드: 모바일 크라우드센싱, 자원 제약적 기기, 사물인터넷, Q-학습, 유틸리티 모델

Abstract In mobile crowdsensing, the most significant challenge is to enable smart devices to perform various sensing tasks for diverse goal-oriented applications. This can be accomplished by the interaction of task owners with smart devices via a specific platform (application interface) to influence their acceptance for task completion, employing various incentive schemes and techniques mentioned in the existing literatures. However, it becomes critical to handle distinct energy restrictions of participating devices and appropriately assign task loads based upon their capabilities that have mostly been overlooked, even more so in an unknown interaction environment. In this paper we address this issue first by evaluating an optimal task-load assignment that maximizes a participating resource constraint node's utility at a resourceful node (broker), and then modeling a distributed Q-learning framework of crowdsensing to improve the cumulative reward for participating nodes. Simulation results show that the proposed algorithm converges quickly for the designed framework, and is very efficient to employ.

Keywords: mobile crowdsensing, resource-constrained devices, internet of things, Q-learning, utility model

· 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2015-0-00557, IoT 기기의 물리적 속성, 관계, 역할 기반 Resilient/Fault-Tolerant 자율 네트워킹 기술 연구)
· 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 Grand ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2018-2015-0-00742)
· 이 논문은 제44회 한국소프트웨어종합학술대회에서 'Q-learning Supplemented Crowdsensing Framework for Resource Constrained Devices'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 경희대학교 컴퓨터공학
shashiraj@khu.ac.kr
sabah@khu.ac.kr
moons85@khu.ac.kr
†† 종신회원 : 경희대학교 컴퓨터공학 교수(Kyung Hee Univ.)
cshong@khu.ac.kr
(Corresponding author임)

논문접수 : 2018년 3월 16일
(Received 16 March 2018)
논문수정 : 2018년 5월 18일
(Revised 18 May 2018)
심사완료 : 2018년 5월 21일
(Accepted 21 May 2018)

정보과학회 컴퓨팅의 실제 논문지 제24권 제7호(2018. 7)

1. Introduction

The significant increase of high end mobile devices with higher capabilities, and the notion Internet of Things (IoT) has enabled a cooperative working environment of mobile crowdsensing (MCS) to achieve better quality of sensing data.

It is considered as an emerging framework that leverages varied location based services and applications spanning from traffic monitoring, localization, environmental monitoring and even daily activities [1,2].

The key enabling idea behind mobile crowdsensing is device's participation for a particular sensing task. For this purpose authors [3], have formulated a platform based campaigning using user profiling to motivate users for a specific task. In literature not many work discuss about 'inconvenience' metrics of participating nodes such as energy constraints. Our earlier work [4] incorporates this issue while formulating an optimization framework for improving user's participation.

For resource constrained participants such as IoT nodes, the story is little different, and this situation becomes very critical that needs to be addressed with detailed scrutiny. The interaction environment between participating nodes (agents) and platform becomes dynamic in nature as the agents are unaware to evaluate the maximal utility point related to task load offers. Furthermore, they can not decide the cumulative reward function related to current action of acceptance or rejection to the participation. Thus, the challenge appears to integrate these situations for the improvement in participation of resources constrained devices.

To narrate this scenario, in this paper we have proposed a distributed Q-learning framework of crowdsensing by evaluating an appropriate task load allocation scheme that optimizes the utility of participating nodes. In our model, a localized set of possible participating nodes are assisted by a resourceful broker to instantiate an optimal task load. The agents then interact with task owner via a light application interface to collect and complete the sensing tasks employing proposed Q-learning approach. The optimal task load is considered as a terminal point, or the maximal utility point for the agents to explore the environment and improve the cumulative reward.

This research work is extended from our previous paper [5]. The rest of the paper is organized as follows: Section 2 discusses about background literature and related works about mobile crowdsensing. Following this, in Section 3 we present our system model with the interaction between participating agents and crowdsensing platform via a resourceful broker. We formulate our problem and discuss about the proposed algorithm in Section 4. In Section 5 we discuss simulations results of the proposed algorithm. Finally, we conclude this paper in Section 6.

2. Background and Related Works

The expected revenue from location based service market would be crossing $43.3 billion worth by 2019 [6], which indicates massive growth of things connected to the internet: Internet of Things (IoT). With the increasing demand for sophisticated services corresponding to the growth in number of connected heterogeneous devices, a critical observation is to collect and transmit required data for the crowdsensing platform. In this regards, improved number of participations affects the quality of received data, and to motive active devices for participation in the crowdsensing framework, a well incentive scheme is expected.

In the literature, a number of incentive mechanisms have been formulated to motivate users for participation in crowdsensing framework. In [7], [8] authors have also discussed about the potential of enhancing quality of received data because of strong commitments from users on appropriate incentive plan. Most of the related works, however, formulate the problem as an individual utility maximization game [7] [9] where the usual interaction would be in terms of bids for a particular sensing task. Also, the related works largely overlook critical aspects of user's participation in the crowdsensing framework such as inconvenience measures (in terms of energy constraints, privacy, time of participation and so on.). On one hand, the heterogeneity of devices has to be appropriately addressed while on the other, the resource restriction on each device needs to be considered for improving participation, and eventually

the improved quality of data.

Notably, a large number of resource constraint devices will come online, that means the aforementioned issues have to be incorporated while designing crowdsensing framework. Motivated by different incentive mechanisms and spaces for improving the requirement for resource constraint nodes to participate in crowdsensing, this research work exploits a learning interaction framework between participating nodes (agents) and crowdsensing platform that improves the overall quality of data as reflected with maximal platform's favorable condition in our optimization problem.

3. System Model

The system model is shown in Fig. 1 where we consider a platform that communicates as a task owner to the number of participating resource constraints nodes (agents) via an application interface. We have considered a resourceful broker that serves as an intermediate moderator for sensory message exchanges of the allocated task. It also assists to the dynamic interaction environment for appropriate number of task load allocation based upon the energy constraints of the participating nodes.

When the task owner has some localized sensing related application service to complete, it evaluates the incentive strategy to improve participations. The pricing scheme is exposed via the application interface along with task description. In such case, the broker acts as a moderator to finalize optimum task load allocation for each individual participants that maximizes will maximize their utilities considering inconvenience metric (energy constraints). The participating nodes will accept or reject to the individual task load offers to improve their cumulative reward while achieving the target of maximal utility point. A maximal utility point can obtained from the utility model computed by the broker considering available resources of participating nodes, and incentives offered by the task owner for task loads.

4. Problem Formulation and Algorithm

We consider a set of participating nodes (considered agents) $u \in U$, $U=\{1,2,3,...,M\}$ localized under a broker that interacts with the task requestor via an application interface. In our formulation we consider the states for an agent u as a set $S_u = \{1,2,3,\cdots,N\}$ that is defined by the tuple of task load allocated ($l \in L$, $L=\{1,2,3,\cdots,N\}$) and the incentive value (p) : [l,p] .The action set is defined as $x_u = ['accept','reject']$ for the sequence of given task loads at time t. The broker facilitates dynamic interaction environment by responding terminal task load allocation to the agents for maximal utility point based upon their individual energy restriction modeled by a concave utility function. The utility of an agent u is

$$U=\begin{cases} p_u - c_u, & \text{if } u \in U \\ 0, & otherwise. \end{cases}$$

which can be modeled as in equation (1).

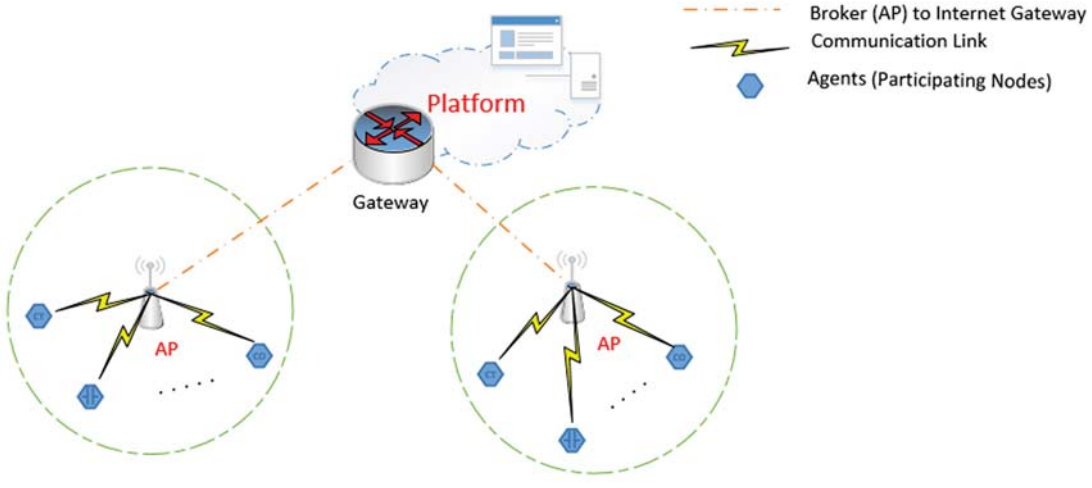

Fig. 1 System Model

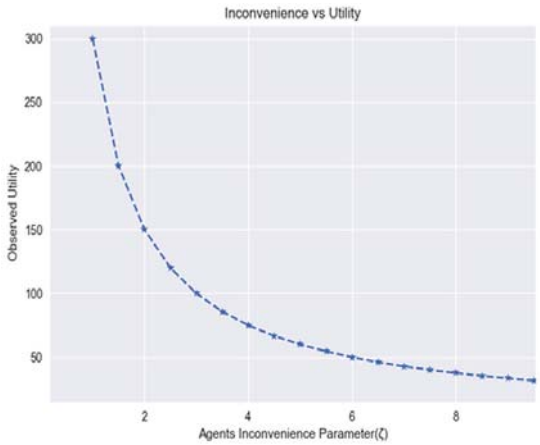

Fig. 2 Inconvenience versus Utility

$$U_{ul} = p_{\{ul\}} E_u l - \zeta_u l^2 \quad (1)$$

Here,

E_u is energy profile, ζ_u is inconvenience parameter and $p_{\{ul\}}$ is incentive offered for task load l to agent u respectively. Fig. 2 illustrates the dependency of inconvenience metrics (energy constraints) of agents and their immediate utility to better represent the scenario.

To define the optimal utility task load for resource constraint agents, we refer to our earlier work [5,10], where we analyze the historical responses by participating agents to the crowdsensing framework over the task load allocation and incentive plan, and design a profile based pricing scheme. The motivation behind this is to improve agent's participation for the crowdsensing framework. We consider platform's adverse impact factor ($\psi < 1, \in [0,1]$) that indicates the favorable situation for agents to participate. The individual agent's inconvenience parameter as defined in the utility framework is bounded as ($\zeta, \zeta > 0$), with individual agent's bias response defined as (β , $\in [0,1]$).

The objective of participation maximization in equation (2) improving platform's favorable condition is solved by the resourceful broker.

$$\max \sum_{u \in U} \sum_{l \in L} (1 - \psi \hat{\zeta}) \sigma\left([\Phi_{ut} \Phi_{up}]^T [l_t p_{ul}] \right) \quad (2)$$

subject to:

$$0 \le p_{ul} \le p_{\max}, l \in L, u \in U \quad (3)$$

$$\sum_{l \in L} p_{ul} = C_l \quad (4)$$

Constraint (3) considers the incentive limitation for each user on a particular task, and constraint (4) guarantees budget constraint for the platform while designing incentive plan for the participating agents, $\hat{\zeta}$ is normalized inconvenience metric feedback by the profile of set of agents for a particular task $\sum_{u \in U} \zeta / u$.

Our proposed sigmoid optimization problem is NP-hard in nature. It can be easily shown; as one can reduce a well know integer linear NP-hard problem (Karp 1972) [11],

$$\begin{aligned} &\text{find} && x \\ &\text{subject to:} && Ax = b \\ & && x \in \{0,1\}^n \end{aligned}$$

, to sigmoid programming

$$\begin{aligned} &\max && \sum_{i=1}^{n} f(x_i) = x_i(x_i - 1) \\ &\textit{subject to:} && Ax = b \\ & && 0 \le x_i \le 1, \ i = 1, \cdots, n \end{aligned}$$

where, $f(x)$ is a chosen function to enforce a penalty on non-integral solutions. Then, we can get solution the sigmoid problem as 0 if and only if there exists an integral solution to the first constraint $Ax = b$. We refer to the solution approach by Udell M, Boyd S (2014) [12] to solve this optimization problem. In accordance with the optimal incentive plan $p'_{\{ul\}}$ for each agent, it is very critical to incorporate energy constraints of participating resource deficient agents and offer optimal utility - task load which maximizes their utility and improves participation [10]. The optimal task load can be analyzed for a given incentive plan by setting first order derivative of equation (1) as zero. i.e., at $p'_{\{ul\}} E_u / 2\zeta_u$.

Under heuristics based allocation, the crowdsensing platform overwhelms the participating agents by maximal task load, so as to maximize the immediate utility. However, under response based allocation scheme we can allocate an optimal utility-task load that not just only maintains agent's participation in the crowdsensing framework but also considers its inconvenience metrics (energy constraints). Fig. 3 reflects the gap between optimal utility point of participating agents under energy restrictions employing historical responses and heuristics. The platform overloads participating agents with excessive tasks in normal scenario which needs to be prevented for improving agent's participation.

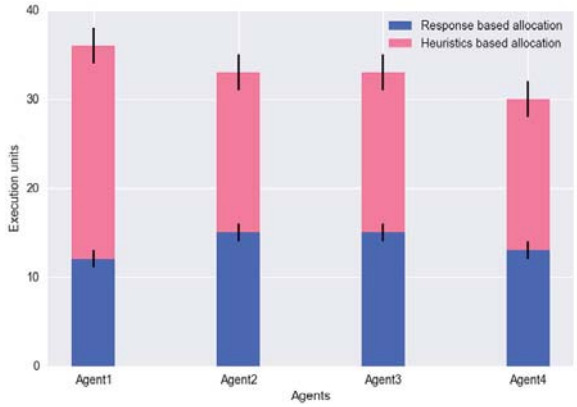

Fig. 3 Task load allocation for participating agents

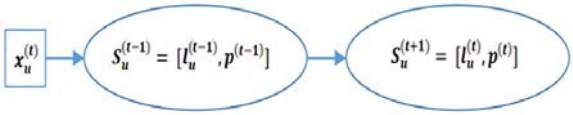

Fig. 4 State transition of agent u with Q-learning based participation strategy for utility maximization

Once the resourceful broker assists for evaluation optimal utility point for each agent, the modeled distributed Q-learning framework enables agents to achieve better cumulative reward under this dynamic interaction environment in a reliable and efficient way. Fig. 4 illustrates the state transition of agent u, where it explores the interaction environment until it reaches the state representing maximal utility point.

Algorithm 1 explains the Q-learning [13] approach implemented under this scenario. The learning parameters are initialized (line 1) with the initial state of the participating node. In each iteration, until convertgence we evaluate the reward function for the possible action sets. For this, equation (5) incorporates Q-value of current state, defined with immediate reward function R, learning α set between 0 and 1, and discount factor γ ∈ [0,1] that incorporates weights for future reward that the immediate one.

$$Q_u\left(S_u^{(t)}, x_u^{(t)}\right) \leftarrow (1-\alpha) Q_u\left(S_u^{(t)}, x_u^{(t)}\right) + \alpha\left(R + \gamma V_u\left(S_u^{(t+1)}\right)\right) \quad (5)$$

where,

$$V_u\left(S_u^{(t)}\right) = \max_{\{x_u\}} \quad Q_u\left(S_u^{(t)}, x_u\right) \quad (6)$$

The Q-values are updated in each round with the corresponding states (line 6) and are stored in the Q-table. On convergence, the best state-action association is based upon the Q-table, choosing maximum Q value. Under ε - greedy approach, the agent randomly explores the environment to obtain the terminal state.

5. Simulation Results

In simulation we've instantiated the ε - greedy Q-learning algorithm with 50 states. The optimal utility task load is facilitated by the broker to the agent. And for each successive state change the agent is rewarded or penalized with the value 10, until the terminal state where the reward value is 100. The agents training parameters are $\alpha = 0.9$, γ = 0.9, and ε = 0.9 respectively.

In Fig. 5 we can observe performance result of our algorithm. The number steps converges quickly over each episodes. Under ε - greedy approach, the agent randomly explores the environment to obtain the terminal state for maximal cumulative reward. Fig. 6 reflects the convergence of cumulative rewards for the agent under successive explorations following the Q-learning approach. The actions are based upon ε = greedy policy, which is from the Q-table value for a particular state. We can observe the quick convertgence also because of the fact that our state values are

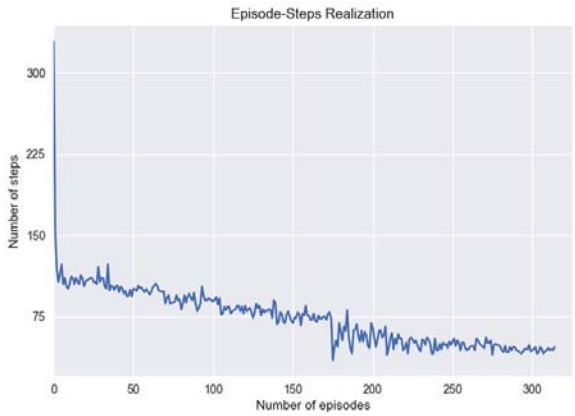

Fig. 5 Episode versus Steps

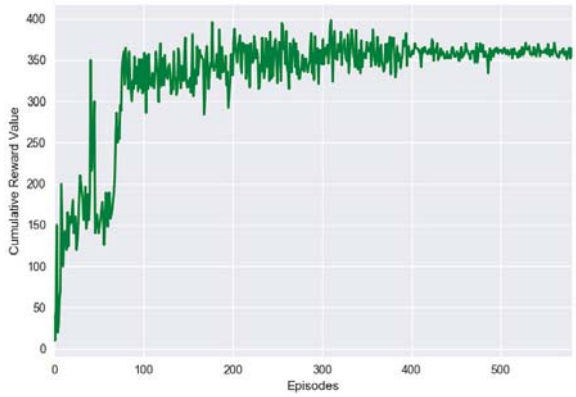

Fig. 6 Episode versus Cumulative Reward

limited. Once randomly instantiated agent will train to choose proper action set for improving his reward.

6. Conclusion

In this paper, we've proposed a distributed Q-learning framework of crowdsensing to improve cumulative reward for participating resource constrained nodes. The broker facilitates the interaction environment to improve agents 'utility under energy constraints to choose appropriate task load. The simulation results show the algorithm being efficient and converging to provide maximal cumulative reward for the participants. In future, we would like to extend this model for the dynamic crowdsensing setting and practical application environment.

References

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, Vol. 49, No. 11, pp. 32-39, Nov. 2011.

[2] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Secaucus, NJ, USA, 2006.

[3] M. Karaliopoulos, I. Koutsopoulos, M. Titsias, "First Learn then Earn: Optimizing Mobile Crowdsensing Campaigns through Data-driven User Profiling," *MobiHoc'16*, Jul. 04-08, 2016, Paderborn, Germany.

[4] Pandey, S. R., Manzoor, A., & Hong, C. S., "User Profile Based Fair Incentive Management for Participation Maximization Using Learning Mechanism," *Proc. of the KIISE Korea Computer Congress 2017*, pp. 1519-1521, 2017. (in Korean)

[5] Pandey, Shashi Raj, Sabah Suhail, and Choong Seon Hong, "Q-learning Supplemented Crowdsensing Framework for Resource Constrained Devices," *Proc. of the KIISE Korea Computer Congress 2017*, pp. 1239-1241, 2017. (in Korean)

[6] "Location based services market to reach $43.3bn by 2019, driven by context aware mobile services, accessed on aug. 2014," [Online]. Available: http://www.juniperresearch.com/pressrelease/contextandlocation-based-services-pr2.

[7] Q. Kong, J. Yu, R. Lu, and Q. Zhang, "Incentive mechanism design for crowdsourcing-based cooperative transmission," *Global Communications Conference (GLOBECOM), 2014 IEEE. IEEE*, 2014, pp. 4904-4909.

[8] T. Luo, S. S. Kanhere, and H.-P. Tan, "Optimal prizes for all-pay contests in heterogeneous crowdsourcing," *Mobile Ad Hoc and Sensor Systems (MASS), 2014 IEEE 11th International Conference on. IEEE*, 2014, pp. 136-144.

[9] Y. Zhang, C. Jiang, L. Song, M. Pan, Z. Dawy, and Z. Han, "Incentive mechanism for mobile crowdsourcing using an optimized tournament model," *IEEE Journal on Selected Areas in Communications*, Vol. 35, No. 4, pp. 880-892, 2017.

[10] Pandey, S. R. (2017), Response Driven Efficient Task Load Assignment in Mobile Crowdsourcing. Manuscript submitted for publication.

[11] Karp, Richard M., "Reducibility among combinatorial problems," *Complexity of computer computations*, Springer, Boston, MA, 1972. 85-103.

[12] Udell, Madeleine, and Stephen Boyd, "Bounding duality gap for separable problems with linear constraints," *Computational Optimization and Applications 64.2* (2016): 355-378.

[13] Reinforcement Learning: An Introduction, Richard Sutton and Andrew Barto, MIT Press, 1998.

Shashi Raj Pandey received the B.E degree in Electrical and Electronics with specialization in Communication from Kathmandu University, Nepal in 2013. After graduation, he served as a Network Engineer at Huawei Technologies Nepal Co. Pvt. Ltd, Nepal from 2013 to 2016. Since March 2016, he is working for his Ph.D in Computer Science and Engineering at Kyung Hee University, South Korea.

Sabah Suhail is a Ph.D. scholar at Kyung Hee University, Korea. She did her MS in Information Security from NUST, Pakistan. Her research interests include security and privacy in IPv6-connected IoT, secure provenance, and cloud computing.

Seung Il Moon received the B.S. and M.S. degrees in Computer Science and Engineering from Kyung Hee University, Seoul, South Korea, in 2011 and 2013, Since September 2013, he is working for his Ph.D in Computer Science and Engineering at Kyung Hee University, South Korea.

Choong Seon Hong received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Minato, Japan, in 1997. He served Korea Telecom Telecommunications Network Laboratory as a Director of Networking Team from 1988 to 1999. Since 1999, he has been Professor, Department of Computer Science and Engineering, Kyung Hee University. His research interests include future Internet, ad hoc networks, network management, and network security.